

Hwa-Young Jeong · Mohammad S. Obaidat  
Neil Y. Yen · James J. (Jong Hyuk) Park  
*Editors*

# Advanced in Computer Science and Its Applications

CSA 2013

# **Lecture Notes in Electrical Engineering**

Volume 279

For further volumes:

<http://www.springer.com/series/7818>



## *About this Series*

“Lecture Notes in Electrical Engineering (LNEE)” is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer’s high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer’s other Lecture Notes series, LNEE will be distributed through Springer’s print and electronic publishing channels.

Hwa-Young Jeong · Mohammad S. Obaidat  
Neil Y. Yen · James J. (Jong Hyuk) Park  
Editors

# Advanced in Computer Science and Its Applications

CSA 2013

 Springer

*Editors*

Hwa-Young Jeong  
Kyung Hee University  
Seoul  
Korea  
Republic of (South Korea)

Neil Y. Yen  
School of Computer Science and  
Engineering  
The University of Aizu  
Fukushima  
Japan

Mohammad S. Obaidat  
Department of Computer Science and  
Software Engineering  
Monmouth University  
New Jersey  
USA

James J. (Jong Hyuk) Park  
Department of Computer Science and  
Engineering  
Seoul University of Science and Technology  
(SeoulTech)  
Seoul  
Korea  
Republic of (South Korea)

ISSN 1876-1100

ISSN 1876-1119 (electronic)

ISBN 978-3-642-41673-6

ISBN 978-3-642-41674-3 (eBook)

Printed in 2 Volumes

DOI 10.1007/978-3-642-41674-3

Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013954014

© Springer-Verlag Berlin Heidelberg 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Message from the CSA2013 General Chairs

International Conference on Computer Science and its Applications (CSA 2013) is the FTRA 5<sup>th</sup> event of the series of international scientific conference. This conference takes place December 18–21, 2013, in Danang, Vietnam. CSA-13 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications. CSA-13 will provide an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of CSA. In addition, the conference will publish high quality papers which are closely related to the various theories and practical applications in CSA. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA-13 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA-12 (4<sup>th</sup> Edition: Jeju, November, 2012), 1CSA-11 (3<sup>rd</sup> Edition: Jeju, December, 2011), CSA-09 (2<sup>nd</sup> Edition: Jeju, December, 2009), and CSA-08 (1<sup>st</sup> Edition: Australia, October, 2008).

The papers included in the proceedings cover the following topics: Mobile and ubiquitous computing, Dependable, reliable and autonomic computing, Security and trust management, Multimedia systems and services, Networking and communications, Database and data mining, Game and software engineering, Grid and scalable computing, Embedded system and software, Artificial intelligence, Distributed and parallel algorithms, Web and internet computing and IT policy and business management.

Accepted and presented papers highlight new trends and challenges of Computer Science and its Applications. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to Steering Chairs: James J. (Jong Hyuk) Park (SeoulTech, Korea), Han-Chieh Chao (National Ilan University, Taiwan) and Mohammad S. Obaidat (Monmouth University, USA). Our special thanks go to the Program Chairs: Neil Y. Yen (The University of Aizu, Japan), Won Woo Ro (Yonsei University, Korea), Yo-Ping Huang (National Taipei

University of Science and Technology, Taiwan) and Hanmin Jung (KISTI, Korea), all Program Committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

December 2013 Young-Sik Jeong, Dongguk University, Korea (Leading Chair)  
Vincenzo Loia, University of Salerno, Italy  
Frode Eika Sandnes, Oslo University College, Norway

# Message from the CSA 2013 Program Chairs

Welcome to the 5th FTRA International Conference on Computer Science and its Applications (CSA 2013) which will be held in Danang, Vietnam, Dec. 18–21, 2013. CSA 2013 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications.

CSA 2013 provides an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of Computer Science. In addition, the conference contains high quality papers which are closely related to the various theories and practical applications in Computer Science. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA 2013 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA 2012 (4th Edition: Jeju, November, 2012), CSA 2011 (3rd Edition: Jeju, December, 2011), CSA 2009 (2nd Edition: Jeju, December, 2009), and CSA 2008 (1st Edition: Australia, October, 2008).

CSA 2013 contains high quality research papers submitted by researchers from all over the world. Each submitted paper was peer-reviewed by reviewers who are experts in the subject area of the paper. Based on the review results, the Program Committee accepted 201 papers. For organizing an International Conference, the support and help of many people is needed. First, we would like to thank all authors for submitting their papers. We also appreciate the support from program committee members and reviewers who carried out the most difficult work of carefully evaluating the submitted papers.

We would like to give my special thanks to Prof. James J. (Jong Hyuk) Park, Prof. Han-Chieh Chao, and Prof. Mohammad S. Obaidat, the Steering Committee Chairs of CSA for their strong encouragement and guidance to organize the symposium. We would like to thank CSA 2013 General Co-Chairs, Prof. Young-Sik Jeong, Prof. Vincenzo Loia, and Prof. Frode Eika Sandnes for their advices to make possible organization of CSA 2013. We also thank the Workshops Chairs, Prof. Eun Ser Lee, Prof. Xiaojun Cao, Prof. Amiya Nayak, Prof. Weiwei Fang, Prof. Ford Lumban Gaol, and Prof. Raymond Huang for organizing CSA-2013

workshops. We would like to express special thanks to FTRA members for their timely unlimited support.

CSA 2013 Program Chairs  
Neil Y. Yen, University of Aizu, Japan  
Won Woo Ro, Yonsei University, Korea  
Yo-Ping Huang, National Taipei University of Science and Technology, Taiwan  
Gangman Yi, Gangneung-Wonju National University, Korea (PC Co-chair)  
Hanmin Jung, KISTI, Korea

# Message from CSA 2013 Workshop Chair

We are very proud and honored to organize the **2013 International Conference on Computer Science and its Application (CSA 2013)**, in Danang, Vietnam, December 18–21, 2013.

Computer science is the scientific and practical approach to computation and its applications. It is the systematic study of the feasibility, structure, expression, and mechanization of the methodical processes (or algorithms) that underlie the acquisition, representation, processing, storage, communication of, and access to information, whether such information is encoded in bits and bytes in a computer memory or transcribed in genes and protein structures in a human cell. A computer scientist specializes in the theory of computation and the design of computational systems. Its subfields can be divided into a variety of theoretical and practical disciplines. Some fields, such as computational complexity theory (which explores the fundamental properties of computational problems), are highly abstract, while fields such as computer graphics emphasize real-world visual applications. Still other fields focus on the challenges in implementing computation. For example, programming language theory considers various approaches to the description of computation, whilst the study of computer programming itself investigates various aspects of the use of programming language and complex systems. Human-computer interaction considers the challenges in making computers and computations useful, usable, and universally accessible to humans. *The major issue of this workshop is to present novel technologies and ideas for Mobile and ubiquitous computing, Dependable, reliable and autonomous computing, Security and trust management, Multimedia systems and services, Networking and communications, Database and data mining, Game and software engineering, Grid and scalable computing, Embedded system and software, Artificial intelligence, Distributed and parallel algorithms, Web and internet computing, IT policy and business management.*

This year, we have received 16 submissions. All manuscripts underwent a rigorous peer-review process with three reviewers per paper. Only 7 papers were accepted for presentation and inclusion in the proceedings, comprising a 44%



acceptance rate. Thanks to all authors who contributed to the success of this workshop.

Finally, we especially like to thank the organization team of the CSA 2013 conference, for their organization of the proceedings, the invitation of valuable speakers, and the social events. The conference and CSA 2013 were not been successful without their great contributions.

CSA 2013 Workshop Chair  
Eunser Lee, Andong National University, Korea

# Message from the SECS-2013 General Chair

SECS-2013 be organized as a workshop of the 5th FTRA International Conference on Computer Science and its Applications (CSA 2013). This conference take place Dec. 18–21, 2013, in Danang, Vietnam. The aim of the SECS-2013 was to provide an international forum for scientific research in the System Engineering and Computer Simulation. It was organized by BOSI Education & Consultancy Co, Ltd in cooperation with The Future Technology Research Association International (FTRA).

The papers included in the proceedings cover the following topics: Mobile and ubiquitous computing; Dependable, reliable and autonomic computing; Security and trust management; Multimedia systems and services; Networking and communications; Database and data mining; Game and software engineering; Grid and scalable computing; Embedded system and software; Artificial intelligence; Distributed and parallel algorithms; Web and internet computing; IT policy and business management. Accepted and presented papers highlight new trends and challenges of System Engineering and Computer Simulation. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to committee Chair: Zhaocong Wu (Wuhan University, China), Zhihong\_Qian (Jilin University, China), Guijun\_Hu (Jilin University, China), Xinsheng Gu (East China University of Science and Technology, China), Shuwen Guo (Dalian Institute of Science and Technology, China), Arthur P. Ramirez (University of California–Santa Cruz, USA), Zawiyah Mohammad Yusof (Universiti Kebangsaan Malaysia, Malaysia), Madya Dr. Md. Jan Nordin (Universiti Kebangsaan Malaysia, Malaysia), Lim Chee Peng (University of Science Malaysia, Malaysia), Abdullah Mohd Zin (Universiti Kebangsaan Malaysia, Malaysia), Phalguni Gupta (Indian Institute of Technology Kanpur, India), P.S. Avadhani (College of Engineering Andhra University, India), A. Senthilrajan (Alagappa University, India), T.V. Gopal (Anna University, India), T. Meyyappan (Alagappa University, India), Rajender Singh Chhillar (Maharshi Dayanand University, India), Khurram Mustafa (Jamia Millia Islamia (Central University), India), Rajesh Ramachandran (Vinayaka

Missions University, India), Sasidhar Babu Suvanam (Sree Narayana Gurukulam College of Engineering, India), Farooq Ahmad (Information Technology University of Central Punjab, Pakistan). all Program Committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

September 2013

Hedy He  
SECS-2013 General Chair

# Message from the SECS-2013 Program Chairs

Welcome to International Symposium on System Engineering and Computer Simulation (SECS-2013), which will be held in Danang, Vietnam, Dec. 18–21, 2013. SECS-2013 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications. SECS-2013 will provide an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of System Engineering and Computer Simulation. In addition, the conference will publish high quality papers which are closely related to the various theories and practical applications in System Engineering and Computer Simulation. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject.

For SECS-2013, we received many paper submissions, after a rigorous peer review process, we accepted 79 articles with high quality for the SECS-2013 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation. We would like to sincerely thank the following invited speaker who kindly accepted our invitations. Finally, we would like to thank all of you for your participation in our conference, and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

Joy Guo, China  
Jenny Ji, China  
Tony Sun, China  
SECS-2013 Program Chairs

# Message from the CMAMBD 2013 Workshop Organizer

CMAMBD 2013 is the Second International Workshop on Creation, Management and Application of Medical and Biological Data organized by KISTI, which is held in conjunction with the 5th International Conference on Computer Science and its Applications (CSA 2013) at Danang, Vietnam, in December 18–21, 2013. Given the success of the previous edition of this workshop, which has been collocated with SERSC MMHS 2012 (Beijing, China), we were encouraged to organize the second version of this series.

This workshop aims to discuss key issues and practices of data-based medical & biological science and its fusion with information science. The medical & biological science explores the physical structures and molecular, cellular and systematic organization of the human body while the computer and information science supports computational approach to the medical & biological mechanisms through computer-based modeling and simulating with supercomputers. We need to understand the importance of the medical & biological data and fusion of medical & biological and information sciences to achieve the well-being of people. This workshop will provide a cross-disciplinary forum for researchers to share their research efforts and ideas between medical & biological and information sciences.

For CMAMBD 2013, we had 5 submissions with high quality accepted for publication. Every paper was reviewed by at least two reviewers in this workshop Program Committee. We take this opportunity to thank for their great contributions. We also wish to express our special thanks to the CSA 2013 chairs including Prof. James J. Park and Prof. Hwa-Young Jeong for allowing and helping this workshop to be successful. Finally, we heartily thank all the authors for their valuable contributions.

December 2013

Won-Kyung Sung  
Sang-Ho Lee  
Seungwoo Lee  
CMAMBD 2013 Workshop Organizer

# Organization

## Steering Chairs

James J. (Jong Hyuk) Park	SeoulTech, Korea
Han-Chieh Chao	National Ilan University, Taiwan
Mohammad S. Obaidat	Monmouth University, USA

## General Chairs

Young-Sik Jeong	Dongguk University, Korea (Leading Chair)
Vincenzo Loia	University of Salerno, Italy
Frode Eika Sandnes	Oslo University College, Norway

## Program Chairs

Neil Y. Yen	The University of Aizu, Japan
Won Woo Ro	Yonsei University, Korea
Yo-Ping Huang	National Taipei University of Science and Technology, Taiwan
Hanmin Jung	KISTI, Korea

## PC Co-Chair

Gangman Yi	Gangneung-Wonju National University, Korea
------------	--

## Workshop Chairs

Eun Ser Lee	Andong National University, Korea (Leading Chair)
Xiaojun (Matt) Cao	Georgia State University, USA
Amiya Nayak	University of Ottawa, Canada
Weiwei Fang	Beijing Jiaotong University, China
Ford Lumban Gaol	Bina Nusantara University, Indonesia

Raymond Huang National Chin-Yi University of Technology,  
Taiwan

## International Advisory Boards

Hamid R. Arabnia	The University of Georgia, USA
Doo-soon Park	SoonChunHyang University, Korea
Hsiao-Hwa Chen	Sun Yat-Sen University, Taiwan
Martin Sang-Soo Yeo	Mokwon University, Korea
Philip S. Yu	University of Illinois at Chicago, USA
Yi Pan	Georgia State University, USA
Salim Hariri	University of Arizona, USA
Leonard Barolli	Fukuoka Institute of Technology, Japan
Jiankun Hu	RMIT University, Australia
Shu-Ching Chen	Florida International University, USA
Victor Leung	University of British Columbia, Canada
Qun Jin	Waseda University, Japan
Bong-Hwa Hong	Kyung Hee Cyber University, Korea
Won-kyung Sung	KISTI, Korea

## Publicity Chairs

Namje Park	Jeju National University, Korea
Jinhua Guo	University of Michigan - Dearborn, USA
Weili Han	Fudan University, China
Waltenegus Dargie	Technical University of Dresden, Germany
Junaid Chaudhry	Universiti Teknologi Malaysia, Malaysia
Deok-Gyu Lee	ETRI, Korea
Stanislav V. Klimenko	Moscow State University, Russia
Madani Cathy	Universite Paris-Est Creteil Val de Marne, France
Jean-Luc Dugelay	Institut EURECOM, France
Daniel Lukac	University of Geneva, Switzerland
Nigel Kyseler	Microsoft Research U.S., USA
Vu Hoang Linh	Hanoi University of Technology, Vietnam
Hangbae Chang	Sangmyung University, Korea

## Publication Chair

Hwa Young Jeong Kyung Hee University, Korea

## Program Committees

Agostino Marengo	University of Study of Bari, Italy
Ahmed EL Oualkadi	Abdelmalek Essaädi University, Morocco

Alexey Rodionov	Institute of Computational Mathematics and Mathematical Geophysics, Russia
Ana Isabel Pereira	Polytechnic Institute of Bragança, Portugal
Andrzej M. Goscinski	Deakin University, Australia
Anirban Mondal	Indraprastha Institute of Information Technology, India
Aw Yoke Cheng	The Asia Pacific University of Technology and Innovation (A.P.U), Malaysia
Chan Yeob Yeun	KUSTAR, UAE
Chang Wu Yu	Chung Hua University, Taiwan
Chang-Tien Lu	Virginia Tech, USA
Chen Liu	Clarkson University, USA
Chengjiu Yin	Kyushu University, Japan
Cheng-Ming Huang	National Taipei University of Technology, Taiwan
Chia-Hung Yeh	National Sun Yat-sen University, Taiwan
Cho-Chin Lin	National Yilan University, Taiwan
Chuan-Kang Ting	National Chung Cheng University, Taiwan
Chuan-Ming Liu	National Taipei University of Technology, Taiwan
Chwan-Lu Tseng	National Taipei University of Technology, Taiwan
Claudio Ardagna	University of Milan, Italy
Clement Leung	Hong Kong Baptist University, Hong Kong
D.J. Lee	Brigham Young University, USA
Debajyoti Mukhopadhyay	Balaji Institute of Telecom & Management, India
Deok-Joo Lee	Kyung Hee University, Korea
Dion Hoe-Lian Goh	Nanyang Technological University, Singapore
DongSeong Kim	University of Canterbury, New Zealand
Don-Lin Yang	Feng Chia University, Taiwan
Duc T. Pham	Cardiff University, UK
El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Evi Syukur	University of New South Wales, Australia
Fengjun Shang	Chongqing University of Posts and Telecommunications, China
Genge Bela	University of Targu Mures, Romania
Gregorio Martinez	University of Murcia (UMU), Spain
Guandong Xu	Victoria University, Australia
Guo Bin	Institute Telecom & Management SudParis, France
Gyanendra Prasad Joshi	Yeungnam University, Korea
Haiduke Sarafian	The Pennsylvania State University, USA



Haifeng Zhao	University of California, Davis, USA
HeeSeok Kim	Korea University, Korea
Hideyuki Sotobayashi	Aoyama Gakuin University, Japan
Hiroyuki Tomiyama	Nagoya University, Japan
Hocine Cherifi	Universite de Bourgogne, France
Hoon Choi	Chungnam National University, Korea
Hung-Chi Chu	Chaoyang University of Technology, Taiwan
Igor Lemberski	Baltic International Academy, Latvia
Irene Yu-Hua Gu	Chalmers University of Technology, Sweden
Jae Joon Lee	Ajou University, Korea
Jason C. Hung	Oversea Chinese University, Taiwan
Javier Martinez Torres	Centro Universitario de la Defensa Zaragoza, Spain
Jehn-Ruey Jiang	National Central University, Taiwan
Jen Li	North Dakota State University, USA
Jen-Wei Hsieh	National Taiwan University, Taiwan
Jerzy Respondek	Silesian University of Technology, Poland
Jianwei Zhang	Tsukuba University of Technology, Japan
Jie Shen	University of Michigan, USA
Jin-Shyan Lee	National Taipei University of Technology, Taiwan
Jongwook Woo	California State University, Los Angeles, USA
Joon-Sang Park	Hongik University, Korea
Joonseok Park	Inha University, Korea
Jose' A. Cardoso e Cunha	Universidade Nova de Lisboa, Portugal
Jose Filho	University of Grenoble, France
Joseph C. Tsai	Tamkang University, Taiwan
Juan Martinez	Gran Mariscal de Ayacucho University, Venezuela
Jui-Feng Yeh	National Chia-Yi University, Taiwan
Jun Wu	National Pingtung Institute of Commerce, Taiwan
Kaikai Chi	Zhejiang University of Technology, China
Kenny Adamson	University of Ulster, United Kingdom
Kok-Leong Ong	Daekin University, Australia
Kuei-Ping Shih	Tamkang University, Taiwan
Kwangjin Park	Wonkwang University, Korea
Ladislav Burita	University of Defence, Czech Republic
Li-Jen Kao	Hwa Hsia Institute of Technology, Taiwan
Lixin Tao	Pace University, USA
M. Dominguez Morales	University of Seville, Spain
Mahasweta Sarkar	San Diego State University, USA
Maiga Chang	Athabasca University, Canada
Malinka Ivanova	Technical University, Bulgaria

Marcin Paprzycki	Polish Academy of Sciences, Poland
Marco Listanti	DIET, Italy
Margaret Tan	Nanyang Technological University, Singapore
Mario Valle	Swiss National Supercomputing Centre, Switzerland
Massimo Cafaro	University of Salento, Italy
Maurizio Lazzari	National Research Council, Italy
Maytham Safar	Kuwait University, Kuwait
Metin Basarir	Sakarya University, Turkey
Ming Li	California State University, Fresno, USA
Muhammad Javed	Dublin City University, Ireland
Muhammad Naufal Bin Mansor	University Malaysia Perlis, Malaysia
Nader F. Mir	San Jose State University, USA
Nelson Passos	Midwestern State University, USA
Nicola Masini	National Research Council, Italy
Nikolai Guschinsky	Byelorussian National Academy of Sciences, Belarus
Parimala Thulasiraman	University of Manitoba, Canada
Prabu Dorairaj	NetApp, India/USA
Prakash Veeraraghavan	La Trobe University, Australia
Qi Shi	Liverpool John Moores University, UK
Qian Yu	University of Regina, Canada
Qingguo Zhou	Lanzhou University, China
Qingyuan Bai	Fuzhou University, China
Radjef Mohammed Said	University of Bejaia, Algeria
Robert Henry	
Morelos-Zaragoza	San Jose State University, USA
Rosa Lasaponara	National Research Council, Italy
Ruay-Shiung Chang	National Dong Hwa University, Taiwan
Savino Longo	University of Bari, Italy
Schulz Frank	SAP, Germany
Seungwoo Lee	KISTI, Korea
Shang-Lin Hsieh	Tatung University, Taiwan
Shaoshan Liu	Microsoft, USA
Shengquan Wang	University of Michigan Dearborn, USA
Shu-Ching Chen	Florida International University, USA
Somchai Chatvichienchai	University of Nagasaki, Japan
Soonil Kwon	Sejong University, Korea
Stephan Chalup	Heidelberg University, Germany
Sunil Kumar Khatri	Amity University Uttar Pradesh, Noida, India
Sunyoung Han	Konkuk University, Korea
Sun-Yuan Hsieh	National Cheng Kung University, Taiwan
Sung-Pil Choi	KISTI, Korea
Takafumi Matsumaru	Waseda University, Japan
Tee Sim Hui	Multimedia University, Malaysia

Tinghuai Ma	NanJing University of Information Science and Technology, China
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Unal Ufuktepe	Izmir University of Economics, Turkey
Valentina Dagiene	Institute of Mathematics and Informatics, Lithuania
Ventzeslav Valev	Bulgarian Academy of Sciences, Bulgaria
Vidyasagar Potdar	Curtin University of Technology, Australia
Vong Chi-Man	University of Macau, Macau, China
Waralak V. Siricharoen	University of the Thai Chamber of Commerce, Thailand
Weizhong Dai	Louisiana Tech University, USA
Wenny Rahayu	La Trobe University, Australia
Wojciech Mazurczyk	Warsaw University of Technology, Poland
Wojciech Zabierowski	Technical University of Lodz, Poland
Xiaokang Zhou	Waseda University, Japan
Yan Pei	Kyushu University, Japan
Yao-Nan Lien	National Chengchi University
Yashu Chen	National Taiwan University, Taiwan
Ye Duan	University of Missouri – Columbia, USA
Yue-Shan Chang	National Taipei University, Taiwan
Yuh-Shyan Chen	National Taipei University, Taiwan
Yutaka Watanobe	University of Aizu, Japan
Zanifa Omary	Dublin Institute of Technology, Ireland
Zhenglu Yang	University of Tokyo, Japan

# International Symposium on System Engineering and Computer Simulation - Workshop Organization -

## International Program Committee Chair (China)

Zhaocong Wu	Wuhan University, China
Zhihong Qian	Jilin University, China
Guijun Hu	Jilin University, China
Xinsheng Gu	East China University of Science and Technology, China
Shuwen Guo	Dalian Institute of Science and Technology, China

## International Program Committee Chair (USA)

Arthur P. Ramirez	University of California–Santa Cruz, USA
-------------------	--

## International Program Committee Chair (Malaysia)

Zawiyah Mohammad Yusof	Universiti Kebangsaan Malaysia, Malaysia
Madya Dr. Md. Jan Nordin	Universiti Kebangsaan Malaysia, Malaysia
Lim Chee Peng	University of Science Malaysia, Malaysia
Abdullah Mohd Zin	Universiti Kebangsaan Malaysia, Malaysia

## International Program Committee Chair (India)

Phalguni Gupta	Indian Institute of Technology Kanpur, India
P.S. Avadhani	College of Engineering Andhra University, India
A. Senthilrajan	Alagappa University, India
T.V. Gopal	Anna University, India
T. Meyyappan	Alagappa University, India

Rajender Singh Chhillar  
Khurram Mustafa  
Rajesh Ramachandran  
Sasidhar Babu Suvanam

Maharshi Dayanand University, India  
Jamia Millia Islamia (Central University), India  
Vinayaka Missions University, India  
Sree Narayana Gurukulam College  
of Engineering, India

**International Program Committee Chair (Pakistan)**

Farooq Ahmad

Information Technology University of Central  
Punjab, Pakistan

# Workshop Organization Creation, Management and Application of Medical and Biological Data - Workshop Organization -

## Organizer/Chair

Won-Kyung Sung	Korea Institute of Science and Technology Information, Korea
Sang-Ho Lee	Korea Institute of Science and Technology Information, Korea
Seungwoo Lee	Korea Institute of Science and Technology Information, Korea

## Program Committee

Dai-Soon Kwak	The Catholic University of Korea, Korea
Dongmin Seo	Korea Institute of Science and Technology Information, Korea
Dukyun Nam	Korea Institute of Science and Technology Information, Korea
Hyunchul Jang	Korea Institute of Oriental Medicine, Korea
In-Su Kang	Kyung Sung University, Korea
Junglok Yu	Korea Institute of Science and Technology Information, Korea
Kum Won Cho	Korea Institute of Science and Technology Information, Korea
Kwan-Hee Yoo	Chungbuk National University, Korea
Kyung Soo Kim	Kyung Hee University, Korea
Pyung Kim	Jeonju National University of Education, Korea
Shuo Xu	Institute of Scientific and Technical Information of China, China
Sukil Kim	The Catholic University of Korea, Korea
Sun Rae Park	Korea Institute of Science and Technology Information, Korea
Wonman Park	Kyung Hee University, Korea

Yeong-Su Lee  
Yoon Hyuk Kim  
Yunliang Zhang

Cylex Inc., Germany  
Kyung Hee University, Korea  
Institute of Scientific and Technical Information  
of China, China

# Contents

## Volume 1

<b>Fast Mode Decision Algorithm Based on Adaptive Search Direction for Combined Scalability in Scalable Video Coding ...</b>	1
<i>Tae-Jung Kim, Yeon-Kyu Jeong, Byung-Gyu Kim, Gwang-Soo Hong</i>	
<b>Monitoring and Automatic Control for Heating System of the Plant .....</b>	7
<i>Roman Kuznetsov, Valeri Chipulis</i>	
<b>Distributed Sensor-Driven Web Applications through Multi-device Usage Patterns .....</b>	13
<i>Heiko Desruelle, Frank Gielen</i>	
<b>Touch Logger Resistant Mobile Authentication Scheme Using Multimodal Sensors .....</b>	19
<i>Hyunyi Yi, Yuxue Piao, Jeong Hyun Yi</i>	
<b>Commodity Recommendation Algorithm Based on Social Network .....</b>	27
<i>Zhimin Yin, Xiangzhan Yu, Hongli Zhang</i>	
<b>A Layered View Model of Social Experience Design: Beyond Single-User User Experience .....</b>	35
<i>Toshihiko Yamakami</i>	
<b>A New Schedule Strategy of Embedded Multi-core SoC .....</b>	43
<i>Tie Qiu, Fangbing Liu, Feng Xia, Ruixuan Qiao</i>	
<b>A Power-Aware Scheduler Exploiting All Slacks under EDF Scheduling .....</b>	51
<i>Ye-In Seol, Jeong-Uk Kim, Young-Kuk Kim</i>	



<b>Network-Threatening Element Extraction and Quantification Framework</b> .....	59
<i>Sungmo Jung, Donghyun Kim, Seoksoo Kim</i>	
<b>Document Clustering Based on a Weighted Exponential Measurement</b> .....	65
<i>Shahrooz Taheri, Alex Tze Hiang Sim, Seyed Hamid Ghorashi</i>	
<b>A Robust Feature Selection Method for Classification of Cognitive States with fMRI Data</b> .....	71
<i>Luu-Ngoc Do, Hyung-Jeong Yang</i>	
<b>Adaptive Networking for Continuous and Reliable Data Delivery in Wireless Sensor Networks</b> .....	77
<i>Jae-Joon Lee, Jinsuk Kang, Jaesung Lim</i>	
<b>Model Transformation for Cyber Physical Systems</b> .....	83
<i>Shuguang Feng, Lichen Zhang</i>	
<b>Load Balancing Improvement Methods in Video Conferencing Based on H.323 Standard</b> .....	89
<i>Vu Thanh Nguyen, Huynh Tuan Anh, Vu Thanh Hien</i>	
<b>A Combination of Clonal Selection Algorithm and Artificial Neural Networks for Virus Detection</b> .....	95
<i>Vu Thanh Nguyen, Nguyen Phuong Anh, Mai Trong Khang, Nguyen Hoang Ngan, Nguyen Quoc Thai, Nguyen Trong Quoc</i>	
<b>Muti-dimensional Architecture Modeling for Cyber Physical Systems</b> .....	101
<i>Bingqing Xu, Lichen Zhang</i>	
<b>An Algorithm of Group Scheduling with Void Filling in OBS Core Nodes</b> .....	107
<i>Nguyen Hong Quoc, Vo Viet Minh Nhat, Nguyen Hoang Son</i>	
<b>A Fast TRW Algorithm Using Binary Pattern</b> .....	115
<i>Jun-Young Park, Chang-Suk Cho</i>	
<b>Efficient Descriptor-Filtering Algorithm for Speeded Up Robust Features Matching</b> .....	125
<i>Minwoo Kim, Deokho Kim, Kyungah Kim, Won Woo Ro</i>	
<b>Towards Effective 3D Model Management on Hadoop</b> .....	131
<i>Hua Luan, Yachun Fan, Mingquan Zhou, Xuesong Wang</i>	
<b>Double Hot/Cold Clustering for Solid State Drives</b> .....	141
<i>Taedong Jung, Yongmyoung Lee, Junhyun Woo, Ilhoon Shin</i>	

<b>A Case Study for Cyber Physical System with Hybrid Relation Calculus</b> .....	147
<i>Shuguang Feng, Lichen Zhang</i>	
<b>Optimizing Functional Link Neural Network Learning Using Modified Bee Colony on Multi-class Classifications</b> .....	153
<i>Yana Mazwin Mohamad Hassim, Rozaida Ghazali</i>	
<b>Adult Contents Analysis and Remote Management Framework for Parental Control Based on Android Platform</b> .....	161
<i>Jae-Deok Lim, Byeong-Cheol Choi, Seung-Wan Han, Jeong-Nyeo Kim</i>	
<b>A Dual Approach of GeneralMatch in Time-Series Subsequence Matching</b> .....	167
<i>Hea-Suk Kim, Minwoo Lee, Yang-Sae Moon</i>	
<b>An Approximate Multi-step <math>k</math>-NN Search in Time-Series Databases</b> .....	173
<i>Sanghun Lee, Bum-Soo Kim, Mi-Jung Choi, Yang-Sae Moon</i>	
<b>The Role of Graph Theory in Solving Euclidean Shortest Path Problems in 2D and 3D</b> .....	179
<i>Phan Thanh An, Nguyen Ngoc Hai, Tran Van Hoai</i>	
<b>Illumination-Robust Local Pattern Descriptor for Face Recognition</b> .....	185
<i>Dong-Ju Kim, Sang-Heon Lee, Myoung-Kyu Shon, Hyunduk Kim, Nuri Ryu</i>	
<b>MAP: An Optimized Energy-Efficient Cluster Header Selection Technique for Wireless Sensor Networks</b> .....	191
<i>Kanokporn Udompongsuk, Chakchai So-In, Comdet Phaudphut, Kanokmon Rujirakul, Chitsutha Soomlek, Boonsup Waikham</i>	
<b>MRP-NEP: A Non-Equal-Probability Multicast Routing Protocol for Target Tracking in Wireless Sensor Networks</b> .....	201
<i>Boyu Diao, Peipei Li, Zhulin An, Fei Wang, Yongjun Xu</i>	
<b>Predicting Osteoporosis by Analyzing Fracture Risk Factors and Trabecular Microarchitectures of the Proximal Femur from DXA Images</b> .....	209
<i>Eun Byeol Jo, Ju Hwan Lee, Sung Yun Park, Sung Min Kim</i>	
<b>Malicious Web Page Detection: A Machine Learning Approach</b> .....	217
<i>Abubakr Sirageldin, Baharum B. Baharudin, Low Tang Jung</i>	
<b>Proposal for Virtual Web Browser by Using HTML5</b> .....	225
<i>Tomokazu Hayakawa, Teruo Hikita</i>	

<b>Design and Implementation of Task Context-Aware E-mail Platform for Collaborative Tasks</b> .....	233
<i>Masashi Katsumata</i>	
<b>Empirical Analysis on the Users' Reply Behaviors of Online Forums</b> .....	241
<i>Guirong Chen, Wandong Cai, Aiwang Chen, Huijie Xu, Rong Wang</i>	
<b>On Companding Transform Techniques for OFDM Visible Light Communication over Indoor Dispersive Channels</b> .....	249
<i>Kasun Bandara, Pararajasingam Niroopan, Y.H. Chung</i>	
<b>Head Pose Estimation Based on Random Forests with Binary Pattern Run Length Matrix</b> .....	255
<i>Hyunduk Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Dong-Ju Kim, Nuri Ryu</i>	
<b>Triggers That Increase Co-Creation Risks: A Consumer Perspective</b> .....	261
<i>Freida Palma, Soon Goo Hong</i>	
<b>A New Lightweight Protection Method against Impersonation Attack on SIP</b> .....	273
<i>Yoon Mi Koh, Kyung Hee Kwon</i>	
<b>Measuring Syntactic Sugar Usage in Programming Languages: An Empirical Study of C# and Java Projects</b> .....	279
<i>Donghoon Kim, Gangman Yi</i>	
<b>Log Pre-processor for Security Visualization</b> .....	285
<i>Jinwon Seo, Jin Kwak</i>	
<b>Fast Coding Unit (CU) Depth Decision Algorithm for High Efficiency Video Coding (HEVC)</b> .....	293
<i>Chan-seob Park, Byung-Gyu Kim, Gwang-Soo Hong, Sung-Ki Kim</i>	
<b>A Study on Using OWC with Battery Storage for Providing Power to Sensor Nodes in a Fish Farm</b> .....	301
<i>Olly Roy Chowdhury, Hong-Geun Kim, Dae-Heon Park, Chang-Sun Shin, Yong-Yun Cho, Jang-Woo Park</i>	
<b>Location-Based Intelligent Robot Management Service Model Using RGPSi with AoA for Vertical Farm</b> .....	309
<i>Hong-Geun Kim, Dae-Heon Park, Olly Roy Chowdhury, Chang-Sun Shin, Yong-Yun Cho, Jang-Woo Park</i>	

<b>Design and Implementation of Intelligent Video Surveillance System Using Dual Cameras</b> .....	315
<i>Seung-Hyeok Yoo, Mi-Jeong Park, Chang-Gyun Lim, Heon Jeong, Eung-Kon Kim</i>	
<b>Context-Aware Control Service Model Based on Ontology for Greenhouse Environment</b> .....	321
<i>Nam-Jin Bae, Kyung-Hun Kwak, Saraswathi Sivamani, Chang-Sun Shin, Jang-Woo Park, Kyungryong Cho, Yong-Yun Cho</i>	
<b>An OWL-Based Ontology Model for Intelligent Service in Vertical Farm</b> .....	327
<i>Saraswathi Sivamani, Nam-Jin Bae, Chang-Sun Shin, Jang-Woo Park, Yong-Yun Cho</i>	
<b>Parallel Genetic Algorithm for Solving the Multilayer Survivable Optical Network Design Problem</b> .....	333
<i>Huynh Thi Thanh Binh, Nguyen Xuan Tung</i>	
<b>A Context-Aware Collaborative Filtering Algorithm through Identifying Similar Preference Trends in Different Contextual Information</b> .....	339
<i>Phung Do, Hiep Le, Vu Thanh Nguyen, Tran Nam Dung</i>	
<b>Survivable Flows Routing in Large Scale Network Design Using Genetic Algorithm</b> .....	345
<i>Huynh Thi Thanh Binh, Son Hong Ngo</i>	
<b>Belief Propagation in Bayesian Network</b> .....	353
<i>Ji Ryang Chung, Gangman Yi</i>	
<b>A Method of Search Scope Compaction for Image Indexes</b> .....	363
<i>Samuel Sangkon Lee</i>	
<b>Efficient Locking Scheme with OPOF on Smart Devices</b> .....	369
<i>Hyun-Woo Kim, Anna Kang, Leonard Barolli, Young-Sik Jeong</i>	
<b>A Requirement Analysis of Awareness-Based Vessel Traffic Service System for Maritime Safety</b> .....	379
<i>Byunggil Lee, Namje Park, Juyoung Kim</i>	
<b>Analysis of Open Interface Function in USN Service Middleware System</b> .....	385
<i>Yilip Kim, Dong-Hwan Park, Hyo-Chan Bang, Namje Park</i>	

<b>Security Controls Based on K-ISMS in Cloud Computing Service</b> .....	391
<i>Jongho Mun, Youngman Jung, Jeeyeon Kim, Youngsook Lee, Kidam Park, Dongho Won</i>	
<b>Implementation of Smart Grid Educational Application</b> .....	405
<i>Yeonghae Ko, Namje Park</i>	
<b>Development of Open Service Interface’s Instructional Design Model in USN Middleware Platform Environment</b> .....	411
<i>Jeongyeun Kim, Dong-Hwan Park, Hyo-Chan Bang, Namje Park</i>	
<b>The Analysis of Case Result and Satisfaction of Digital Textbooks for Elementary School Students</b> .....	417
<i>Jeongyeun Kim, Namje Park</i>	
<b>Cryptanalysis of Encrypted Remote User Authentication Scheme by Using Smart Card</b> .....	423
<i>Jongho Mun, Jiye Kim, Woongryul Jeon, Youngsook Lee, Dongho Won</i>	
<b>The Development of Convergent STEAM Program Focused on Rube Goldberg for Improvement of Engineer Career Awareness of Elementary School Students</b> .....	429
<i>Yilip Kim, Namje Park</i>	
<b>Key Management Scheme Using Dynamic Identity-Based Broadcast Encryption for Social Network Services</b> .....	435
<i>Youngman Jung, Yoonho Nam, Jiye Kim, Woongryul Jeon, Hanwook Lee, Dongho Won</i>	
<b>Protecting Mobile Devices from Adversarial User by Fine-Grained Analysis of User Behavior</b> .....	445
<i>Yonggon Kim, Ohmin Kwon, Sunwoo Kim, Byungjin Jeong, Hyunsoo Yoon</i>	
<b>Design of an RDFizer for Online Social Network Services</b> .....	453
<i>Junsik Hwang, Hyosook Jung, Sujin Yoo, Seongbin Park</i>	
<b>A High Performance and Bandwidth Efficient IDMA Scheme with Large Receiver MIMO Technologies</b> .....	459
<i>Pararajasingam Niroopan, Kasun Bandara, Yeon-ho Chung</i>	
<b>Security Considerations for Smart Phone Smishing Attacks</b> .....	467
<i>Anna Kang, Jae Dong Lee, Won Min Kang, Leonard Barolli, Jong Hyuk Park</i>	
<b>Technology Venture Startup Invigoration Strategy for Building Infrastructures for the Business Startup Ecosystem</b> ...	475
<i>Hye-Sun Kim, Yunho Lee, Hyoung-Ro Kim</i>	

<b>Smart-Contents Visualization of Vehicle Big Data Using Vehicle Navigation Status Information</b> .....	483
<i>Hae-Jong Joo, Suck-Joo Hong, Dong-Su Park</i>	
<b>A Model for Analyzing the Effectiveness of Smart Mobile Communication Quality Measurement</b> .....	489
<i>Bong-Hwa Hong, Hae-Jong Joo, Sang-Soo Kim</i>	
<b>Characteristics Analysis and Library Development for Common Lamps by Using PSPICE Modeling</b> .....	497
<i>Young-Choon Kim, Moon-Taek Cho, Ho-Bin Song, Ok-Hwan Kim</i>	
<b>A Study on Brake Torque for Traction Motors by Using the Electric Brake</b> .....	503
<i>Young-Choon Kim, Moon-Taek Cho, Ho-Bin Song</i>	
<b>A Virtual Cluster Scheme Technology for Efficient Wireless Sensor Networks</b> .....	509
<i>Seongsoo Cho, Bhanu Shrestha, Young-gi Kim, Bong-Hwa Hong</i>	
<b>Implementation of Wireless Electronic Acupuncture System</b> ....	515
<i>You-Sik Hong, Bong-Hwa Hong, Baek-Ki Kim</i>	
<b>Network Based Intelligent Agent for Ubiquitous Environments</b> .....	523
<i>Dong W. Kim, Ho-Dong Lee, Sung-Wook Park, Jong-Wook Park</i>	
<b>A Study on the Clustering Scheme for Node Mobility in Mobile Ad-hoc Network</b> .....	531
<i>Hyun-Jong Cha, Jin-Mook Kim, Hwnag-Bin Ryou</i>	
<b>Design of User Access Authentication and Authorization System for VoIP Service</b> .....	537
<i>Ho-Kyung Yang, Jin-Mook Kim, Hwang-Bin Ryou</i>	
<b>Approach of Secure Authentication System for Hybrid Cloud Service</b> .....	543
<i>Jin-Mook Kim, Jeong-Kyung Moon</i>	
<b>Intelligent Inference System for Smart Electronic Acupuncture</b> .....	551
<i>You-Sik Hong, Chang-Hoon Choi, Baek-Ki Kim</i>	
<b>Electric Braking Control System to Secure Braking Force in the Wide Speed Range of Traction Motor</b> .....	559
<i>Young-Choon Kim, Moon-Taek Cho, Ok-Hwan Kim</i>	

<b>Optimized Design of Charger for Electric Vehicles with Enabled Efficient CCCV Mode Movement</b> .....	565
<i>Ji-Yong Chun, Young-Choon Kim, Moon-Taek Cho</i>	
<b>Ubiquitous Mobile Computing on Cloud Infrastructure</b> .....	571
<i>DongBum Seo, In-Yong Jung, Jong Hyuk Park, Chang-Sung Jeong</i>	
<b>Identifying Invalid Data within Operating System for Higher Flash Utilization</b> .....	577
<i>June Young Kim, Seung-Ho Lim</i>	
<b>Moving Human Detection Using Motion Depth in Depth Image Sequences</b> .....	583
<i>Bong-Hwa Hong, Kiseo Park</i>	
<b>Microbial Fuel Cell for STEAM Teaching Tools and Method</b> ....	591
<i>Yeonghae Ko, Namje Park</i>	
<b>Research on Educational Use of Smart-phone Applications with Smart Clicker Technique</b> .....	597
<i>Ji-Hye Bae, Sung-Ki Kim</i>	
<b>Hierarchical Customization Method for Ubiquitous Web Applications</b> .....	603
<i>Wonjae Lee, Hyun-Woo Lee, Min Choi, Jong Hyuk Park, Young-Sik Jeong</i>	
<b>PV System for Medical Devices in the Hospital</b> .....	609
<i>Young-Choon Kim, Moon-Taek Cho, Ok-Hwan Kim</i>	
<b>Color Image Retrieval Using Fuzzy Measure Hamming and S-Tree</b> .....	615
<i>Thanh The Van, Thanh Manh Le</i>	
<b>Pipeline Verification via Closed-Loop Feedback</b> .....	621
<i>Danghui Wang, Yongda Li, Xiaoping Huang</i>	
<b>Fuzzy Rule-Based to Predict the Minimum Surface Roughness in the Laser Assisted Machining (LAM)</b> .....	627
<i>M.R.H. Mohd Adnan, Azlan Mohd Zain, Habibollah Haron</i>	
<b>GRT-Multigraphs for Communication Networks: A Fuzzy Theoretical Model</b> .....	633
<i>Siddhartha Sankar Biswas, Bashir Alam, M.N. Doja</i>	

<b>Exploring the Key Determinants of Successful ICT Innovation Adoption: A Case Study of a Fishing Community in Thailand</b> .....	643
<i>Heng Wei Lee, Maziani Sabudin, Rosnah Idrus, Mohd Azam Osman, Thongchai Kruahong, Nattayanee Leanjay</i>	
<b>Dynamic Bandwidth Allocation Algorithm for Ethernet Passive Optical Networks Based on Traffic Prediction</b> .....	649
<i>Chengwei Xiao, Hong Fan, Dongjing Liu</i>	
<b>Game Based Learning for Teaching Electrical and Electronic Engineering</b> .....	655
<i>Michael Callaghan, Mark Cullen</i>	
<b>Mobile WiMAX Resource Allocation Design Goals: Key Features/Factors/Issues</b> .....	661
<i>Zaid G. Ali, R.B. Ahmad, Abid Yahya, L.A. Hassnawi</i>	
<b>Hybrid Trust Framework for Loss of Control in Cloud Computing</b> .....	669
<i>Tang Jia Li, Manmeet Mahinderjit Singh</i>	
<b>Secure Authorized Proxy Signature Scheme for Value-Added Service in Vehicular Ad Hoc Networks</b> .....	677
<i>Iuon-Chang Lin, Chen-Hsiang Chen, Pham Thuy Linh</i>	
<b>The Computer-Aided System for Promoting Judgments of Umpires</b> .....	685
<i>Chin-Fa Chen, Chia-Chun Wu, Duc-Tai Dang, Iuon-Chang Lin</i>	
<b>A Distributed Model to Analyzed QoS Parameters Performance Improvement for Fixed WiMAX Networks</b> .....	695
<i>A.L. Ibrahim, A. Md. Said, K. Nisar, Peer Azmat Shah, Abubakar Aminu Mu'azu</i>	
<b>A Hybrid of Fixed-Size and Dynamic-Size Tile Algorithm for Panoramic View on Mobile Devices</b> .....	703
<i>Chen Kim Lim, Kian Lam Tan, Abdullah Zawawi Talib</i>	
<b>A Relationship Strength-Aware Topic Model for Communities Discovery in Online Social Networks</b> .....	709
<i>Juan Bi, Jia Huang, Zhiguang Qin</i>	
<b>Replication in Data Grid: Determining Important Resources</b> ...	717
<i>Yuhanis Yusof</i>	
<b>Backpropagation Neural Network for Sex Determination from Patella in Forensic Anthropology</b> .....	723
<i>Iis Afrianty, Dewi Nasien, Mohammed R.A. Kadir, Habibollah Haron</i>	



<b>Learning History Using Role-Playing Game (RPG) on Mobile Platform</b> .....	729
<i>Guo Haur Lee, Abdullah Zawawi Talib, Wan Mohd Nazmee Wan Zainon, Chen Kim Lim</i>	
<b>Experiment on Modified Artificial Bee Colony for Better Global Optimisation</b> .....	735
<i>Muhammad Shahrizan Shahrudin, M. Mahmuddin</i>	
<b>Ontology-Supported Development for Drug Analysis Laboratory Corresponding to the ISO/IEC 17025 Standard</b> .....	743
<i>Orapan Apirakkan, Wanna Sirisangtragul, Pusadee Seresangtakul</i>	
<b>Volume 2</b>	
<b>Modeling the Natural Capital Investment on Tourism Industry Using a Predator-Prey Model</b> .....	751
<i>Zhonghua Cai, Qing Wang, Guangqing Liu</i>	
<b>Study on Tracking Derivative Based Method for DC System Grounding Fault Detection</b> .....	757
<i>Wang Peng, Li Huashun, Wang Ning, Qu Linkui, Yu Yin</i>	
<b>Towards a VLIW Architecture for the 32-Bit Digital Signal Processor Core</b> .....	763
<i>Khoi-Nguyen Le-Huu, Thanh T. Vu, Diem N. Ho, Anh-Vu Dinh-Duc</i>	
<b>Application and Effectiveness of Cooperative Teaching Strategies in Entrepreneurship Education</b> .....	769
<i>Chien-Hua Shen, Chun-Mei Chou, Hsi-Chi Hsiao</i>	
<b>Isarn Dharma Alphabets to Thai Language Translation by ATNs</b> .....	775
<i>Phoemporn Lakkhawannakun, Pusadee Seresangtakul</i>	
<b>Simulation and Analysis of Space Station Redocking</b> .....	783
<i>Rui Chen, Shengjing Tang, Guojiang Sun</i>	
<b>Campus Mobile Navigation System Based on Shortest-Path Algorithm and Users Collaborations</b> .....	791
<i>Shir Ni Ler, Wan Mohd Nazmee Wan Zainon</i>	
<b>Tracking Multiple Fish in a Single Tank Using an Improved Particle Filter</b> .....	799
<i>Wong Poh Lee, Mohd Azam Osman, Abdullah Zawawi Talib, Jean-Marc Ogier, Khairun Yahya</i>	

<b>Efficient 3D Model Visualization System of Design Drawing Based on Mobile Augmented Reality</b> .....	805
<i>Yeon-Jae Oh, Eung-Kon Kim</i>	
<b>A New Co-evolutionary Immune Algorithm for Flow Shop with Zero Wait</b> .....	811
<i>Zhenhao Xu, Shun Zhang, Xingsheng Gu</i>	
<b>Design and Implementation of Safety Verification for Civil Aviation Processing System Based on BPMN</b> .....	819
<i>Lisong Wang, Qian Zhang</i>	
<b>A Survey on Ontology Mapping Techniques</b> .....	829
<i>Yew Kwang Hooi, M. Fadzil Hassan, Azmi M. Shariff</i>	
<b>The Parametric Design and Automatic Assembly of Hydrostatic Rotary Table Based on Pro/Engineer</b> .....	837
<i>Qiang Cheng, Can Wu, Zhifeng Liu, Jianhua Wang, Peihua Gu</i>	
<b>Modeling and Simulation of China's Competition Strategy for Highly Educated Talents Based on System Dynamic</b> .....	843
<i>Wen Guo, Feng Dai</i>	
<b>Kernelized-QEMU: A Study of System-Level Virtual Layer in Linux Kernel</b> .....	853
<i>Yongxuan Xu, Xianglan Chen, Hua-Ping Chen, Huang Wang</i>	
<b>A Multi-pattern Matching Algorithm Based on Double Array Trie</b> .....	863
<i>Miao Hou, YingHui Song, Dongliang Xu, Hongli Zhang</i>	
<b>Lower Bound Estimation for Required Number of Nodes in the Opportunistic Communication – Based Wireless Sensor Network</b> .....	869
<i>Yubo Deng, Shizhong Wu, Shoupeng Li, Yongping Xiong, Tao Zhang</i>	
<b>Potential Attacks against k-Anonymity on LBS and Solutions for Defending the Attacks</b> .....	877
<i>Pan Juncheng, Deng Huimin, Song Yinghui, Li Dong</i>	
<b>To Practice Management Innovation and Reinforce the Foundation of Intellectual Property Strategy</b> .....	885
<i>Wen-Liang Sun, Yong-Fei Ma, Li-Wen Chen, Shang Yu</i>	
<b>Research of Intranet Security Audit in E-government Management Website Group Based on Multi-agents</b> .....	893
<i>Han Bing, Wang Bo</i>	

<b>Simulation of Three-Phase Voltage-Source PWM Rectifier with LCL Filter</b> .....	901
<i>Liu Hui, Xu Chao, Chen Chen, Wu Yibing</i>	
<b>The Research on the Computer Control Technology of Tobacco Production</b> .....	911
<i>Li Zhekun, Song Yanhong, Zou Yusheng, Wu Mingyi, Liu Geyi</i>	
<b>A Study on the Application of Financial Engineering in Supply Chain Risk Management</b> .....	919
<i>Hao Jianhua, Sun Wenliang, Chen Liwen, Yang Zhanchang</i>	
<b>Logical Symmetry Based K-means Algorithm with Self-adaptive Distance Metric</b> .....	929
<i>Wu Zu-Feng, Mu Xiao-Fan, Liu Qiao, Qin Zhi-guang</i>	
<b>Modeling and Simulation of Network-Based C2 System Based on AOA</b> .....	937
<i>Bei Huo, Xin-rong Yang</i>	
<b>Application Study of Spatiotemporal Chaotic Sequence in Code Division Multiple Access Communication System</b> .....	943
<i>Li Nan, Tang Quan-bin</i>	
<b>OPRF Secure Computation for Safety Neighbor Verification Protocols of Wireless Sensor Networks</b> .....	951
<i>Mei Meng, Xu Zhongwei, Zhu Yujun</i>	
<b>Texture Terrain Based on Land Cover Distribution</b> .....	957
<i>Du Jinlian, Xing Lifei, Wang Song</i>	
<b>Text Mining for Information Screen in Risk Assessment of Environmental Endocrine Disruptive Chemicals</b> .....	967
<i>Yong Chen, Yu Wang, Xiaolian Duan, Huatang Zhang, Yuyue Jia</i>	
<b>The Choices of Issued Inventory Valuation Methods Based on Different Accounting Standards at Home and Abroad</b> .....	973
<i>Gu Shusheng</i>	
<b>Formal Constraints Research of Military Conceptual Model Elements</b> .....	981
<i>Ding Xiao-Jian, Xie Bin</i>	
<b>Leaf Lesion Detection Method Using Artificial Bee Colony Algorithm</b> .....	989
<i>Faudziah Ahmad, Ahmad Airuddin</i>	

<b>The Accurate Positioning Method of Data Matrix Code Image Marked in Cylinder Glass</b> .....	997
<i>Wei-ping He, Gai-fang Guo, Wei Wang, Xi-zheng Cao</i>	
<b>Analysis of Urban Traffic Based on Taxi GPS Data</b> .....	1007
<i>Li Meng, Li Ru-tong, Xia Yong, Qin Zhi-guang</i>	
<b>Research and Application of Embedded System Development Based on Petri Nets</b> .....	1017
<i>Chen Yu</i>	
<b>CloStor: A Cloud Storage System for Fast Large-Scale Data I/O</b> .....	1023
<i>Wang Ying, Yang Dongri, Liu Peng</i>	
<b>Nonnegative Variable Weight Based Auto Parts Combination Demand Forecasting Model Research</b> .....	1031
<i>Jian Gong, Chun Lu, Xin Liu</i>	
<b>A Data-Hiding Method Based on TCP/IP Checksum</b> .....	1039
<i>Zhen Liu, Yuyu Jiang, Ping Qian</i>	
<b>A Parallel Full-System Emulator for Risc Architecture Host</b> .....	1045
<i>Xiao-Wu Jiang, Xiang-Lan Chen, Huang Wang, Hua-Ping Chen</i>	
<b>Density Based Active Self-training for Cross-Lingual Sentiment Classification</b> .....	1053
<i>Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, Ali Selamat</i>	
<b>Constructing 3D Model Based on Panoramic Images</b> .....	1061
<i>Jia Li, Yehua Sheng, Ping Duan, Siyang Zhang, Haiyang Lv</i>	
<b>QoE Evaluation Model of Anti-Spam Allergy Problem Based on Fuzzy Entropy Method and Hidden Markov Model</b> .....	1067
<i>Haizhuo Lin, Jilong Wang</i>	
<b>Topic Identification Strategy for English Academic Resources</b> ...	1073
<i>Huang Li, Hu Qing, Xiong Xin, Hua Lijun</i>	
<b>Performance Comparison between MLP Neural Network and Exponential Curve Fitting on Airwaves Data</b> .....	1079
<i>Muhammad Abdulkarim, Afza Shafie, Wan Fatimah Wan Ahmad, Radzuan Razali</i>	
<b>A Private Cloud Storage System Based on Multithread Optimization</b> .....	1089
<i>Zhang Wei, Xu Tao, Li Kun</i>	

<b>Cloud Storage-Based Medical Data Integration Technology</b> . . . . .	1097
<i>HongJun Zhan, Wei Zhang</i>	
<b>The Operation and Maintenance Management System of the Cloud Computing Data Center Based on ITIL</b> . . . . .	1103
<i>HongJun Zhan, Wei Zhang</i>	
<b>Underwater Object Detection by Combining the Spectral Residual and Three-Frame Algorithm</b> . . . . .	1109
<i>Zhe Chen, Huibin Wang, Jie Shen, Xin Dong</i>	
<b>The Data Mining of Breast Cancer Based-on K-Means</b> . . . . .	1115
<i>Zhang Ruilan, Feng Zhixin</i>	
<b>Compositional Modeling for Underwater Warfare Simulation Based on HLA and Flames</b> . . . . .	1121
<i>Hongtao Liang, Fengju Kang, Yanyang Zeng</i>	
<b>Spot Electricity Price Dynamics of Indian Electricity Market</b> . . .	1129
<i>G.P. Girish, S. Vijayalakshmi</i>	
<b>Guaranteed QoS for UDP and TCP Flows to Measure Throughput in VANETs</b> . . . . .	1137
<i>Abubakar Aminu Mu'azu, Low Tang Jung, Halabi Hasbullah, Ibrahim A. Lawal, Peer Azmat Shah</i>	
<b>Detecting People Using Histogram of Oriented Gradients: A Step towards Abnormal Human Activity Detection</b> . . . . .	1145
<i>Abdul-Lateef Yussiff, Suet-Peng Yong, Baharum B. Baharudin</i>	
<b>Research of Touchscreen Terminals Gesture Operation Error Based on Kansei Engineering</b> . . . . .	1151
<i>Rong Qin, Dongxiang Chen</i>	
<b>Heart Sound Feature Extraction Based on Wavelet Singular Entropy</b> . . . . .	1159
<i>Zhang Lu</i>	
<b>Research on MTMP Structure Chlorine Dosing Decoupling Control</b> . . . . .	1165
<i>Xie Peizhang, Zhou Xingpeng</i>	
<b>Development of Ontology for the Diseases of Spine</b> . . . . .	1171
<i>Geun-Hye Kim, Min-Jeoung Kang, Chai Young Jung, Joon-Yong Jung, Seung Eun Jung, Jin-Sung Kim, Ji-Seong Jeong, Do-Hyeong Kim, Kwan-Hee Yoo, Dongmin Seo, Seungwoo Lee, Seungbock Lee, Sangho Lee, Sukil Kim</i>	

<b>Implementation of Information Retrieval Service for Korean Spine Database with Degenerative Spinal Disease</b> .....	1179
<i>Dongmin Seo, Seungwoo Lee, Seungbock Lee, Sangho Lee, Hanmin Jung, Won-Kyung Sung</i>	
<b>Automatic Surface Mesh Intersection Algorithm of Spine and Implant FEM Models</b> .....	1185
<i>Dukyun Nam, Ikhwang Choi, Younho Kim, Wanho Jeon, Kum Won Cho</i>	
<b>Simulation Program Binding Technology Based on Supercomputing</b> .....	1195
<i>Sun-Rae Park, DuSeok Jin, Kyu-Chul Lee, Kum Won Cho</i>	
<b>Data Security in Cloud for Health Care Applications</b> .....	1201
<i>R. Anitha, Saswati Mukherjee</i>	
<b>Web Service Selection Using Decision Tree Analysis in a Risky Environment</b> .....	1211
<i>P. Sandhya, M. Lakshmi</i>	
<b>An Enhanced e-Voting System in Cloud Using Fingerprint Authentication</b> .....	1219
<i>Gandhi Usha Devi, Kannan Anusha, G.V. Rajyalakshmi</i>	
<b>Local Minima Jump PSO for Workflow Scheduling in Cloud Computing Environments</b> .....	1225
<i>S. Chitra, B. Madhusudhanan, G.R. Sakthidharan, P. Saravanan</i>	
<b>Respiratory Rate Estimation by Extracted PPG Signals from Embedded Smart Attire of Operation Strategists</b> .....	1235
<i>Kathir Deivanai</i>	
<b>The Assessment of Information and Communication Technology (ICT) Policy in South Korea</b> .....	1241
<i>Young-Hyun Yeo, Sung-Ki Kim, Ji-Hye Bae, Byung-Gyu Kim</i>	
<b>Computer Graphical Score and Music Education: Application to Music Animation Machine MIDI Player</b> .....	1251
<i>Soon-Hak Hwang, Se-Hak Chun</i>	
<b>A Newton's Universal Gravitation Inspired Firefly Algorithm for Document Clustering</b> .....	1259
<i>Athraa Jasim Mohammed, Yuhanis Yusof, Husniza Husni</i>	
<b>A Study on the Correlation between the Customers' Perceived Risks and Online Shopping Tendencies</b> .....	1265
<i>Chiung En Huang, Ping Kuo Chen, Chih Chung Chen</i>	

<b>Model Checking Probabilistic Timed Systems against Timed Automata Specification</b> .....	1273
<i>Junhua Zhang</i>	
<b>Particle Filter Parallel of Improved Algorithm Based on OpenMp</b> .....	1279
<i>Yang Zhang</i>	
<b>An E-commerce-System-Based Research on Trust Risk Assessment Models</b> .....	1287
<i>Mei Ting, Zhang Yi, Dai Qun</i>	
<b>An Analysis about the Security of the Operating System Trusted Platform</b> .....	1295
<i>Zhang Yi</i>	
<b>Technology Venture Startup Invigoration Strategy for Building Infrastructures for the Business Startup Ecosystem</b> ...	1303
<i>Hye-Sun Kim, Yunho Lee, Hyoung-Ro Kim</i>	
<b>Smart-Contents Visualization of Vehicle Big Data Using Vehicle Navigation Status Information</b> .....	1311
<i>Hae-Jong Joo, Suck-Joo Hong, Dong-Su Park</i>	
<b>A Model for Analyzing the Effectiveness of Smart Mobile Communication Quality Measurement</b> .....	1317
<i>Bong-Hwa Hong, Hae-Jong Joo, Sang-Soo Kim</i>	
<b>Characteristics Analysis and Library Development for Common Lamps by Using PSPICE Modeling</b> .....	1325
<i>Young-Choon Kim, Moon-Taek Cho, Ho-Bin Song, Ok-Hwan Kim</i>	
<b>A Study on Brake Torque for Traction Motors by Using the Electric Brake</b> .....	1331
<i>Young-Choon Kim, Moon-Taek Cho, Ho-Bin Song</i>	
<b>FOLI Technique Algorithm for Real-Time Efficient Image Processing</b> .....	1337
<i>Seongsoo Cho, Kwang Chul Son, Jongsup Lee, Seung Hyun Lee</i>	
<b>Stabilization Inverse Optimal Control of Nonlinear Systems with Structural Uncertainty</b> .....	1343
<i>Jong-Yong Lee, Seongsoo Cho</i>	
<b>Implementation of Wireless Electronic Acupuncture System</b> ....	1349
<i>You-Sik Hong, Bong-Hwa Hong, Baek-Ki Kim</i>	

<b>Network Based Intelligent Agent for Ubiquitous Environments</b> .....	1357
<i>Dong W. Kim, Ho-Dong Lee, Sung-Wook Park, Jong-Wook Park</i>	
<b>A Study on the Clustering Scheme for Node Mobility in Mobile Ad-hoc Network</b> .....	1365
<i>Hyun-Jong Cha, Jin-Mook Kim, Hwnag-Bin Ryou</i>	
<b>Design of User Access Authentication and Authorization System for VoIP Service</b> .....	1371
<i>Ho-Kyung Yang, Jin-Mook Kim, Hwang-Bin Ryou</i>	
<b>Approach of Secure Authentication System for Hybrid Cloud Service</b> .....	1377
<i>Jin-Mook Kim, Jeong-Kyung Moon</i>	
<b>Intelligent Inference System for Smart Electronic Acupuncture</b> .....	1385
<i>You-Sik Hong, Chang-Hoon Choi, Baek-Ki Kim</i>	
<b>Electric Braking Control System to Secure Braking Force in the Wide Speed Range of Traction Motor</b> .....	1393
<i>Young-Choon Kim, Moon-Taek Cho, Ok-Hwan Kim</i>	
<b>Optimized Design of Charger for Electric Vehicles with Enabled Efficient CCCV Mode Movement</b> .....	1399
<i>Ji-Yong Chun, Young-Choon Kim, Moon-Taek Cho</i>	
<b>A Virtual Cluster Scheme Technology for Efficient Wireless Sensor Networks</b> .....	1405
<i>Seongsoo Cho, Bhanu Shrestha, Young-gi Kim, Bong-Hwa Hong</i>	
<b>An Intelligent e-Services Composition Platform for Ubiquitous Baby Care: The Case Study of Life and Commercial Support Services for Property Management</b> .....	1411
<i>Chih-Kun Ke, Yi-Jen Yeh, Chiao-Min Chang</i>	
<b>An Authoring System of Creating Graphic Map for Item Search Based on Library OPACs</b> .....	1417
<i>Hsuan-Pu Chang, Wei-Ting Chang, Shi-Pin Fong</i>	
<b>Cloud-Based Traveling Video Editing System</b> .....	1423
<i>Joseph C. Tsai, Neil Y. Yen</i>	
<b>Linked Data-Based Service Publication for Service Clustering</b> .....	1429
<i>Incheon Paik, Wuhui Chen, Banage T.G.S. Kumara, Takazumi Tanaka, Zhenni Li, Yuichi Yaguchi</i>	



**The Practical Quality Model for Cloud Learning System** . . . . . 1437  
*Anna Kang, Leonard Barolli, HwaYoung Jeong, JongHyuk Park,  
Hae-Gill Choi*

**Realizing the Right to Be Forgotten in an SNS Environment** . . . 1443  
*Cheol Ho Sin, Nam A. Kim, Byeong Woo Go, Kim Seong Min,  
Jae Dong Lee, Jong Hyuk Park*

**Author Index** . . . . . 1451

# Fast Mode Decision Algorithm Based on Adaptive Search Direction for Combined Scalability in Scalable Video Coding

Tae-Jung Kim<sup>1</sup>, Yeon-Kyu Jeong<sup>2</sup>, Byung-Gyu Kim<sup>2</sup>, and Gwang-Soo Hong<sup>2</sup>

<sup>1</sup> Smart TV Reseach Department, ERTI,  
Daejeon, Rep. of Korea

<sup>2</sup>Dept. of Computer Science and Engineering, SunMoon University,  
A-san city, Rep. of Korea  
taejungkim@etri.re.kr,  
{JeongYK,bg.kim,hongzolv}@mpcl.sunmoon.ac.kr

**Abstract.** Scalable Video Coding (SVC) is an extension of H.264/AVC for providing several scalabilities. In the SVC, there are three types of scalabilities such as temporal, spatial, and quality scalabilities. These types of scalabilities can be combined together. As an extension of H.264/AVC video standard, the SVC coding structure is based on a mode decision using 7 variable block types for rate distortion (RD) optimization, hierarchical bi-directional motion prediction, and inter-layer prediction. These techniques achieve a coding efficiency, but they increase the computational complexity of the SVC encoding system. This paper proposes a fast motion estimation method using an adaptive search direction selection for combined scalability. We verify that the proposed algorithm can achieve up to 51% decrease in the encoding time with a negligible loss of quality.

**Keywords:** Fast mode decision, Scalable video coding, Combined scalability, Hierarchical bi-motion prediction.

## 1 Introduction

Scalable video coding (SVC) has been finalized as an extension to the H.264/AVC video standard [1]. Compared to previous video coding standards, SVC is intended to encode the signal once while enabling decoding from partial streams depending on the specific rate and resolution required by certain applications. In the SVC standard, a coded bitstream is composed of a base layer with several enhancement layers [1]. The base layer contains a reduced resolution or reduced quality version of each coded frame for mobile device, such as portable phones and smart phones with low computing power. The upper enhancement layers are used to provide a higher quality service for a peak signal-to-noise ratio (PSNR), frame rate, and image resolution.

SVC has been designed as an extension of H.264/AVC, and most H.264/AVC techniques are used for coding as mode decision in variable block sizes, bi-directional motion search, integer transform, et al. In addition, in order to improve the coding

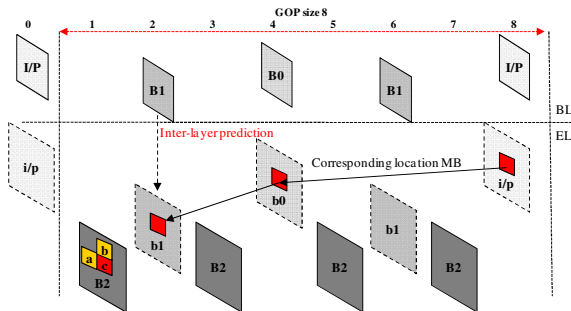
efficiency, inter-layer prediction mechanisms are incorporated [1]. These techniques achieve a coding efficiency, but they increase the computational complexity. Among them, the bi-directional motion search for variable block sizes have to calculate rate-distortion cost between forward, backward, and bi-reference frames and current frame. Therefore, we propose a fast mode search algorithm based on adaptive search direction for inter frame coding in combined scalability.

Li et al. proposed a fast mode decision algorithm for spatial scalability in SVC. In this scheme, they used the mode distribution relationship between base layer and enhancement layers [3]. Also, an efficient algorithm based on macroblock (MB) tracking scheme has been reported by Kim [4]. In this algorithm, he used the most correlated MB of the reference picture to design an early termination rule for H.264/AVC encoding system.

A layer-adaptive mode decision algorithm and a motion search scheme have been suggested for coarse grain scalability (CGS) and temporal scalability by Lin et al. [5]. To reduce the mode search, they skipped modes with limited contributions to the coding efficiency based on statistical analysis. To speed-up the motion search, they reused the reference frame indices of the base layer and determined the initial search points using the motion vector at the base layer. This scheme is not suitable for spatial scalability.

## 2 Overview of the Combined Scalability in SVC

The SVC standard includes spatial, temporal, and quality scalability. These types of scalabilities can be combined together. Spatial, temporal and quality scalability, all existing scalability may be provided at the bitstream level. The structure of the combined scalability has features of both spatio-temporal and quality scalability. If the encoding structure supports the combined scalability with two layers, then the base layer is encoded at a lower resolution, a slower temporal rate, and reduced quality.



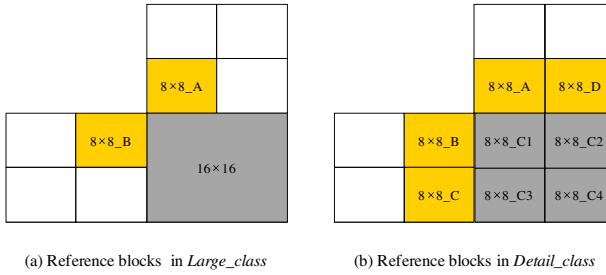
**Fig. 1.** The proposed class decision using reference MBs in the combined scalability

The enhancement layer is encoded at a higher resolution, a faster temporal rate, and a higher quality. The enhancement layer is also divided into an inter layer prediction picture (i/p, and b0,1) and a non-inter layer prediction picture (B2), as shown in Fig. 1. The inter layer prediction picture employs the base layer information, such as intra texture, motion vector, and residual coefficients to make the best prediction data.

### 3 Proposed Algorithm

We propose a fast mode prediction using correlation of search direction between neighboring MBs and the current MB. The proposed search direction prediction different calculates according to block types. We divide the available block types into two classes: *Large\_class* ( $16 \times 16$ ,  $8 \times 16$ ,  $16 \times 8$ ) and *Detail\_class* ( $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$ ). As shown in Fig. 1, the class in inter-layer prediction picture (b0,1) is determine by mode of corresponding location MB in base layer and previous temporal level.

The class in non-inter layer prediction pictures (B2) is determine by mode of neighboring MBs (a,b). If mode of reference MBs (corresponding and base layer MB for inter layer prediction picture or neighboring MB for non-inter layer prediction picture) was coded detailed blocks size ( $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$ ) or Intra, then the class of current MB is *Detail\_class*. Otherwise, the current MB can be considered as the class of *Large\_class*.



**Fig. 2.** The proposed reference blocks according to the designed classes.

According to class, reference blocks for search direction prediction are determined as shown in Fig. 2. For example, if current MB is *Detail\_class* and coding block is  $8 \times 8_{C2}$ , reference blocks are  $8 \times 8_D$  and  $8 \times 8_{C1}$ . The search direction prediction (SDP) method using search directions of reference MBs follows in Table 1. The SDP is defined for selecting the best search direction prediction (BSDP).

The BSDP is computed using the direction of the representative mode ( $16 \times 16$  type in *Large\_class* and  $8 \times 8$  type in *Detail\_class*) and the SDP information. As shown in Table 2, the bi-predictive mode is selected as the best search direction when the SDP and the direction of the representative mode are inverse relationship. Otherwise, the direction of the representative mode is selected as the best search direction for the current block type.

**Table 1.** Search Direction Prediction

Left Block	Above Block	SDP
$X$	$X$	$X$
$X$	$X'$	$Bi$
$Bi$	<i>Whichever</i>	$Bi$

**Table 2.** Best Search Direction Prediction

SDP	$16 \times 16$ or $8 \times 8$	BSDP
$X$	$X$	$X$
$X$	$X'$	$Bi$
$Bi$	$X$	$X$

$X$  is forward or Backward direction,  $X'$  is inverse direction of  $X$ .

## 4 Experimental Results

To verify the performance of the proposed fast mode determination for combined scalability in SVC, simulations were performed on various test sequences using JSVM (joint scalable video model) 9.17 reference software. Table 3 shows the simulation conditions.

The measures for evaluating the performance of the proposed algorithm were BDPSNR(dB), BDBR (%) [2], and  $\Delta$ Time (%).  $\Delta$ Time represents a comparison factor indicating the average for the amount of saved encoding time at each QP, defined as:

**Table 3.** Simulation Conditions

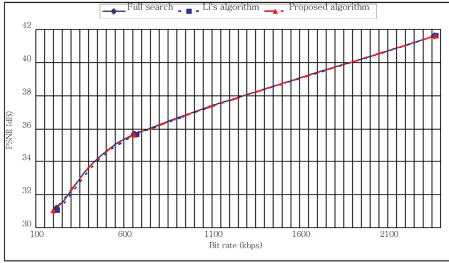
		Conditions	
QP	Base	40	
	Enhancement	24, 28, 32, 36	
Resolution	Base	QCIF (7)	CIF (2)
	Enhancement	CIF (7)	4CIF (2)
Frame Rate	Base	15Hz (7)	30Hz(2)
	Enhancement	30Hz(7)	60Hz(2)
Coding option		MV search range: 32, MV resolution: 1/4 pel, Reference frame: 1, GOP size: 8 Total encoding frame: 97 CAVLC, Loop Filter off	

$$\Delta Time = \frac{Time[reference] - Time[proposed]}{Time[reference]} \times 100. \quad (1)$$

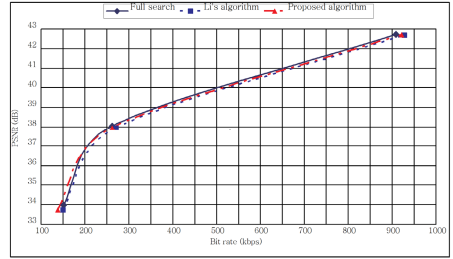
We used Li's method [3] which is well known fast mode decision technique in the SVC encoding system, for an objective comparison of the encoding performance.

**Table 4.** Simulation Results for Combined Scalability

Sequence	Algorithm	BDPSNR	BDBR	$\Delta$ Time
FOREMAN	Li's	-0.67	14.79	39.25
	Proposed	<b>-0.05</b>	<b>1.2</b>	<b>48.71</b>
MOBILE	Li's	-0.31	6.90	38.94
	Proposed	<b>-0.04</b>	<b>0.98</b>	<b>42.28</b>
CITY	Li's	-0.71	13.51	38.32
	Proposed	<b>-0.04</b>	<b>0.92</b>	<b>50.65</b>
BUS	Li's	-0.52	9.53	39.13
	Proposed	<b>-0.1</b>	<b>1.5</b>	<b>42.01</b>
SOCCER	Li's	-0.61	11.07	38.85
	Proposed	<b>-0.05</b>	<b>0.96</b>	<b>50.01</b>
FOOTBALL	Li's	-0.56	9.48	36.91
	Proposed	<b>0.08</b>	<b>1.44</b>	<b>36.83</b>
ICE(4CIF)	Li's	-0.70	17.20	37.89
	Proposed	<b>-0.05</b>	<b>1.34</b>	<b>50.74</b>
HARBOUR(4CIF)	Li's	-0.27	7.60	40.52
	Proposed	<b>-0.03</b>	<b>0.83</b>	<b>41.31</b>
AVERAGE	Li's	-0.54	11.26	38.72
	Proposed	<b>-0.04</b>	<b>1.15</b>	<b>45.32</b>



(a)



(b)

**Fig. 3.** Rate-distortion (RD) curves: (a) Harbour and (b) Foreman sequences

As shown in Table 4, the proposed algorithm increases the speed of the SVC encoding system up to 50.74% in the ICE sequence, compared to the full mode search. Compared to Li's method, the proposed algorithm achieved speed-up gain of up to 13% with a smaller bit increment for the ICE sequence. By using the proposed algorithm, we can see that the average speed-up gain of over 45% was obtained comparing to the full mode search while suffering less quality loss and a smaller bit rate increment.

With the same experimental condition, Figure 3 illustrates the rate-distortion (RD) curves for four sequences. The Foreman sequence has a medium motion property and the Harbour sequence has a little slow motion. The proposed mode decision algorithm exhibited an RDO performance similar to the JSVM original encoder with the full intra mode search. When Li's algorithm employed, a large loss of quality of approximately 0.1~ 0.24 (dB) for the Foreman and Harbour sequences was occurred

comparing to the original JSVM encoder. This is undesirable from the viewpoints of image quality and network bandwidth. However, the proposed algorithm produced almost the same performance in overall bit rate. Moreover, in very low bitrate area, the proposed algorithm yielded better RDO performance because of a lot of bit saving effect of the proposed method.

## 5 Conclusions

We have proposed fast mode prediction using correlation of search direction between neighboring MBs and the current MB for combined scalability in SVC encoding system. In our algorithm, the direction of motion estimation is determined using the direction information of neighboring MBs and block types. The proposed algorithm yield good performance because of local directional information of the current MB. Based on comparative analysis, a speed-up factor of 51% was verified with a negligible bitrate increment or large bitrate saving and a minimal loss of image quality.

**Acknowledgement.** This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (MEST), under Grant NRF-2010-0024786.

## References

1. Schwarz, H., Marpe, D., Wiegand, T.: Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. *IEEE Trans. Circuit Syst. Video Technol.* 17(9), 1103–1120 (2007)
2. Bjontegaard, G.: Calculation of average PSNR differences between RD-curves. Presented at the 13th VCEG-M33 Meeting, Austin, TX (2001)
3. Li, H., Li, Z.G., Wen, C.: Fast Mode Decision Algorithm for Inter-Frame Coding in Fully Scalable Video Coding. *IEEE Trans. Circuit Syst. Video Technol.* 16(7), 889–895 (2006)
4. Kim, B.G.: Novel Inter-Mode Decision Algorithm Based on Macroblock (MB) Tracking for the P-Slice in H.264/AVC Video Coding. *IEEE Trans. Circuit Syst. Video Technol.* 18(2), 273–279 (2008)
5. Lin, H.C., Peng, H.W., Hang, M.H., Ho, W.J.: Layer-adaptive Mode Decision and Motion Search for Scalable Video Coding with Combined Coarse Granular Scalability (CGS) and Temporal Scalability. In: *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 289–292 (September 2007)

# Monitoring and Automatic Control for Heating System of the Plant

Roman Kuznetsov and Valeri Chipulis

Institute of Automation and Control Processes FEB RAS,  
Vladivostok, Russian Federation  
{kuznetsov, chipulis}@dvo.ru

**Abstract.** The experience of the development, implementation and follow-up operation of information-analytical system on the plant is discussed. Particular attention is paid to the analysis of economic efficiency of heat energy accounting and automatic control of the heating. The possibilities of regression analysis for effective heating control are shown.

**Keywords:** energy management system, heat meter, information-analytical system, district heating.

## 1 Introduction

The main activities of the Institute of Automation and Control Processes is development, implementation and support of information and analytical systems in the heat-power engineering [1-2]. In the last decade an information-measuring systems are widely used in heat-power engineering of Russia. It became possible due to the advent of modern measuring devices. These devices allow both to perform the measurement with high accuracy and to organize data acquisition by using wired or wireless communication. In addition, Federal law “On saving energy and increasing energy efficiency and on amendments to certain legislative acts of the Russian Federation” came into force. The primary data source for information-analytical systems is heat meters. Heat meter performs the following basic functions: measuring of the heat-transfer medium (flow, temperature and pressure), the calculation of the heat consumption based on the measurement data, accumulation and storage measured values in the archives, providing access to the historical data on request via the communication interface. We will make accentuate on retrospective analysis of data collected and accumulated in the database.

The development of the system for monitoring and automatic control of the heating was carried out for the Radiopribor plant. There are various objects at the plant territory: storehouses, offices, production buildings and so on. The local heating network is connected to the district heating network by the main thermal point. In addition to traditional accounting tasks for the individual thermal points the customer required the solution of more specific ones for this complex industrial object of automation.



There are tasks of:

- monitoring in control room;
- analysis of the balance of mass flows and energy consumption;
- energy savings;
- evaluation of the effectiveness.

Tasks have been solved by stages.

## 2 Energy Accounting

At the first stage, the energy analysis of the plant, the designing, installation of heat meters and new equipments for thermal points are performed. Measurements are collected from the heat meters by an information network, which was created for the data acquisition. The supervisory point is organized and the server hardware with the information-analytical system of thermo-power engineering (IAS) is installed. The IAS is used for the monitoring of the whole heating system and separate buildings, for the accumulation measurements in database, for the analytical processing and presenting the results in the form of graphs, tables and reports.

Monitoring of heat supply is shown in Fig.1. The main parameters of the heat consumption: flow, temperature and pressure of the heat-transfer medium are on scheme of the plant. In case of emergency the relevant parameters are displayed with warning coloration on the screen. The lack of balance is verified on monitoring as the difference between all internal consumers and total consumption of the plant. The dispatcher has real-time information to control the heating modes and for emergency detection.

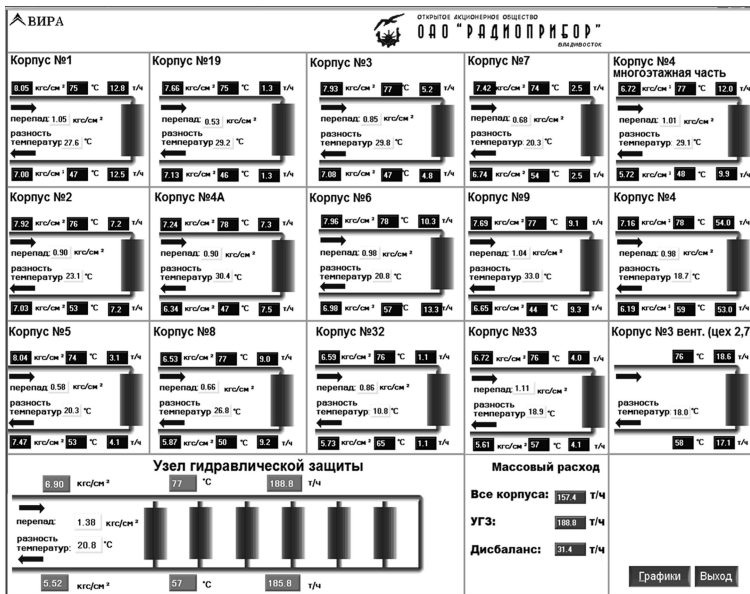


Fig. 1. Monitoring of the heating system

### 3 Automatic Control of the Heating

On the second stage the automatic controllers of heating were set up on thermal points of the plant. It's making possible to regulate the inlet temperature of heating system according to outdoor temperature by heating schedule. The principle of the control by outside air temperature consists in the following. Mass flow  $M_I$  and heat energy  $Q$  is changed according to weather conditions  $T_{air}$  (Fig. 2). At night, mass flow grows up while outside air is colder. In the daytime, the mass flow (thus the heat energy consumption) is reduced while the outdoor temperature rises. In the evening hours, the valve is triggered to open and flow rate is increased according to required heat load while the outdoor temperature falls down. Additional energy savings has been got as result of settings of regulators taking into account the operation mode of heating buildings. The lowest heating schedule has been specified for warehouses and office buildings with low heat losses.

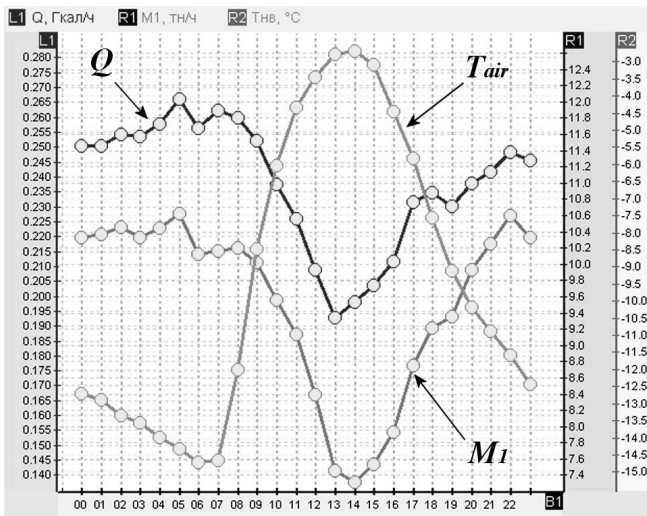


Fig. 2. The regulation by heating schedule

### 4 Effectiveness Analysis

Assessment of the economic effects of automatic control in heating system has been performed by regression analysis. The measurements of heating season before regulators installation are used for the linear regression approximations. In result the coefficients of regression equations were obtained. The Table 1 presents the formulas for calculating of the energy that corresponds to the heat load in case of the absence of automatic control.

**Table 1.** The formulas for calculating heat energy

Building	Formula	R <sup>2</sup>
N1	$M_1^p * (-1.27 * M_1^p + 0.683 * T_1 - 4.76)$	0.98
N2	$M_1^p * (-2.12 * M_1^p + 0.525 * T_1 + 0.16)$	0.97
N3	$M_1^p * (-0.32 * M_1^p + 0.334 * T_1 - 2.58)$	0.98
N4	$M_1^p * (-0.24 * M_1^p + 0.383 * T_1 - 3.23)$	0.76
N4-A	$M_1^p * (-0.28 * M_1^p + 0.154 * T_1 + 5.23)$	0.89
N4-M	$M_1^p * (-0.47 * M_1^p + 0.366 * T_1 + 1.95)$	0.93
N6	$M_1^p * (-0.87 * M_1^p + 0.42 * T_1 + 0.38)$	0.97
N9	$M_1^p * (-1.17 * M_1^p + 0.61 * T_1 - 9.47)$	0.93
N19	$M_1^p * (-0.82 * M_1^p + 0.226 * T_1 - 0.57)$	0.94
N33	$M_1^p * (-2.23 * M_1^p + 0.14 * T_1 + 17.45)$	0.87

- $M_1^p$  - the mass flow of inlet pipe in the heating season without regulators (previous season).
- $T_1$  - the temperature of inlet pipe in the heating season with regulators.
- $R^2$  - coefficient of the approximation accuracy.

The effectiveness of the automation is determined by comparing the measured heat energy consumption with the estimated one. The estimated heat energy consumption is a calculated value that corresponds to the hydraulic regime (mass flow rate) of previous heating season (before the installation of control system) as well as according to the weather and inlet temperature from district heating network of the season after installation.

**Table 2.** Energy savings by regulating (Q, Gcal)

Building	Q <sub>m</sub>	Q <sub>e</sub>	Q <sub>e</sub> - Q <sub>m</sub>
N1	233.4	319.6	86.2
N2	433	587.5	154.5
N3	423.3	724.3	301
N4	1586.7	2175.3	588.6
N4-A	575.6	778.3	202.7
N4-M	908	1500	592
N6	722.1	908.4	186.3
N9	482.8	784.5	301.7
N19	119	232	113
N33	172.3	275.3	103
Q total	5656.3	8286	2629.7

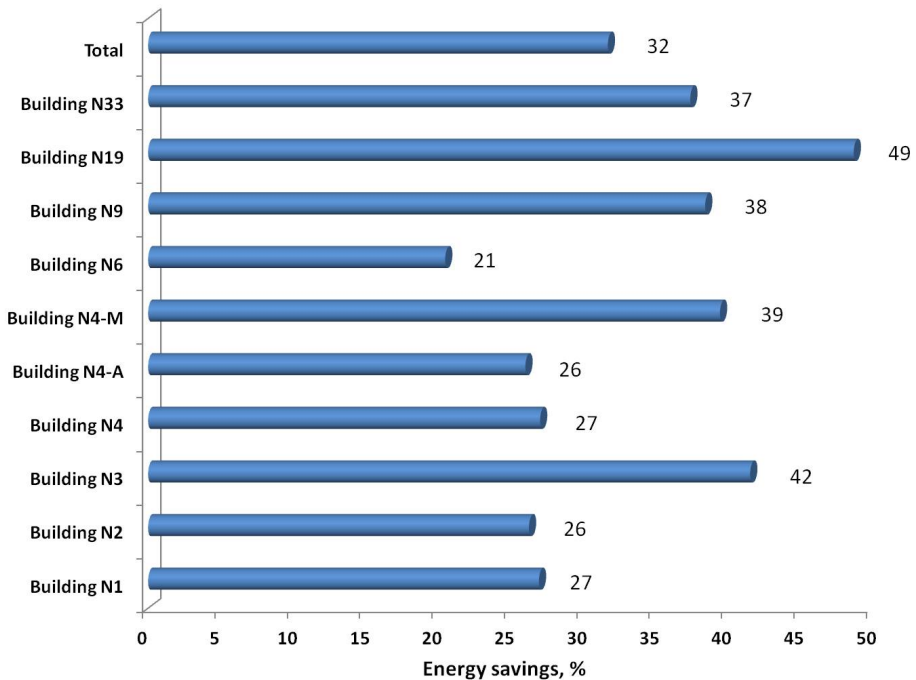
- Q<sub>m</sub> - summary energy is calculated by heat meter according with mass flow and temperatures are measured in heating system;
- Q<sub>e</sub> - the energy is calculated by regression formula (Table 1) that obtained using approximating the measurements (inlet flow and inlet temperature) in heating season without regulation.

Energy savings due to regulation in the percentage is calculated by the formula:

$$Q_s = \frac{(Q_e - Q_m)}{Q_e} \times 100\%. \quad (1)$$

The relative values of the energy savings calculated by equation (1) are shown on the diagram for each building (Fig. 3). The lowest observed values of savings (21%) have been got for the building N6. The reason of low savings is repair work that had been carried out inside the local heating system of building and thus the automatic control system was turned off. The greatest savings has been achieved for N3, N4-M, N9 and N19 buildings. The seventy percent of the total savings is obtained on the large buildings (N3, N4, N4-M, N9).

In conclusion, the total energy savings obtained due to installation of the automatic heating control system is around 32% or 2630 Gcal (in absolute units) per heating season. Hence the development of information-analytical system and the plant-wide automation were cost effective. There are perspectives to automate the ventilation system in the industrial buildings.



**Fig. 3.** Energy savings after installation of the automatic heating control system

## 5 Remote Control of the Valve

The problem of quantitative heating regulation has been solved around the plant as a whole. The main thermal point of the plant was modernized and open-close valve (part-turn actuator SG) has been installed on it. The actuator is combined with the AUMATIC AC controls. The settings of actuator controls is configured to provide the best quality heating regulation and to eliminate the risk of the water hammer. The software has been developed for the remote flow control from the dispatch room (Fig. 4). The software allows monitoring of the control system and set the current position in percent to open/close valve. HMI-interface displays indicators, the control mode and the current valve position as well elements for remote control.

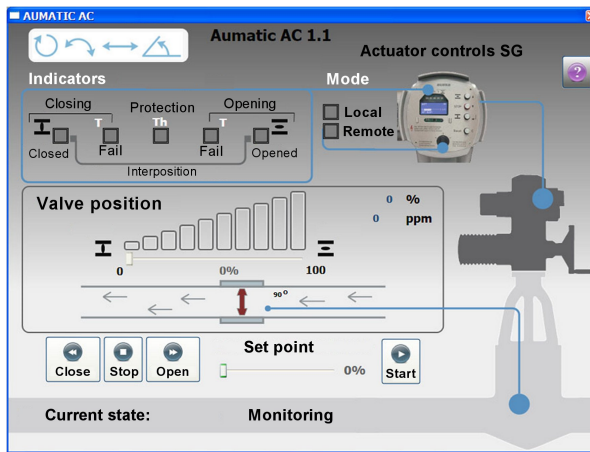


Fig. 4. The software for valve remote control

## 6 Conclusion

The information-analytical systems provide the management of the heating on a new level by using a next generation of measuring equipment and information technologies. The system's features are aimed to provide uninterrupted heating, quality maintenance, energy efficient operation modes as well as getting an economic benefit at the expense of automatic control.

## References

1. Vinogradov, A.N., Kuznetsov, R.S., Chipulis, V.P.: Monitoring and Analysis of the Operating Conditions of Heat Sources. *Measurement Techniques* 51(11), 1218–1223
2. Bogdanov, Y., Chipulis, V.: Information-Analytical Systems of Thermo-Power Engineering. In: Sénac, P., Ott, M., Seneviratne, A. (eds.) *ICWCA 2011*. LNICST, vol. 72, pp. 116–124. Springer, Heidelberg (2012)

# Distributed Sensor-Driven Web Applications through Multi-device Usage Patterns

Heiko Desruelle and Frank Gielen

Ghent University – iMinds,  
Dept. of Information Technology – IBCN, Ghent, Belgium  
{heiko.desruelle, frank.gielen}@intec.ugent.be

**Abstract.** To access their computer applications and services, people tend to use an increasing variety of consumer electronic devices. Devices range from laptops and netbooks, to smartphones and tablets, and even interactive television sets. In the context of mobile applications, this ubiquitous revolution allows for various multi-device use cases and scenarios that are based on a user's dynamic usage patterns. In this paper we discuss how people can access an application using multiple devices, both in sequence as well as in parallel. Moreover, we elaborate on the technological opportunities and challenges for such multi-device enabled applications.

**Keywords:** Multi-device applications, dynamic usage patterns, ubiquitous web, HTML5.

## 1 Introduction

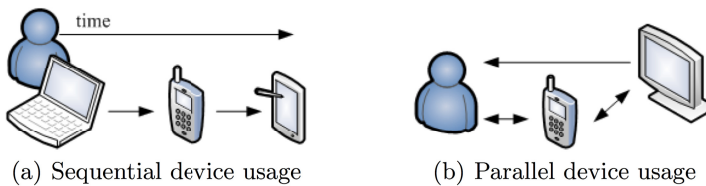
The increasing popularity of internet-enabled devices and technology is allowing people to access online content virtually anywhere, at anytime, and on any device. The available devices range from smartphones and tablets, to laptops and interactive television sets, etc. With the help of web technology, mobile applications can be built that are accessible by most of these device types (e.g., using web applications and widgets, or PhoneGap). Nevertheless, existing application solutions only partly succeed in providing end-users a convincing user experience. This issue is mainly due to the fact that most mobile applications are still tightly bound to the physical device on which they are being executed [1]. Existing application platforms barely take advantage of the diversity of devices owned by its users. The intended immersive and blended interaction aspect of such ubiquitous applications is thus mainly lost.

In this paper we introduce a web-based platform that aims to be a generic enabler for such multi-device applications. The proposed platform does so by relying on standardized technology in order to maximize its value and impact, both towards application developers as well as consumers. The remainder of the paper is structured as follows. In Section 2, we describe opportunities and related work for dynamic usage patterns that arise in environments with ubiquitous consumer electronic devices. Section 3 presents the developed application platform and discusses the main technological challenges for multi-device applications, which it aims to resolve. Section 4

presents the prototype implementation of the proposed platform for realizing an e-learning application use case. Finally, the conclusion and future work are presented in Section 5.

## 2 Background and Related Work

Online content can take various forms. It can range from documents, to presentations, webpages, videos, etc. Depending on the user's contextual setting, the typical usage patterns for accessing these resources may vary considerably. The user's context is dynamic over time. Its characterizing parameters include user preferences, as well as the available devices, the user's current location, etc. From this perspective, we identify two generic multi-device usage patterns for personal ubiquitous applications: sequential device usage, and parallel device usage (see Fig. 1).



**Fig. 1.** Multi-device usage patterns in a ubiquitous computing environment

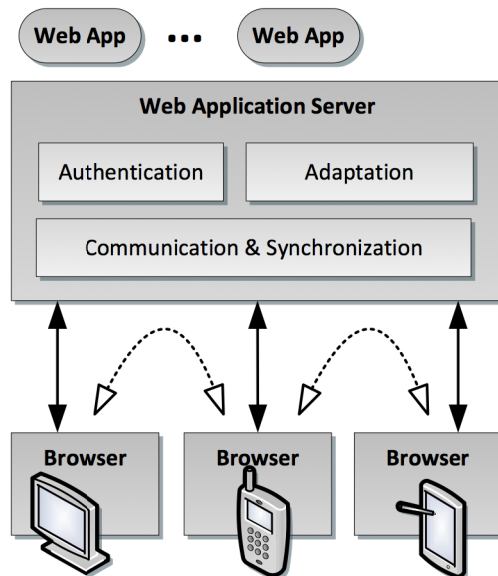
- *Sequential device usage.* Sequential usage patterns aim to smoothen the state transfer from one device to another and make it as seamless as possible. Based on this usage pattern, a user should, e.g., be able to start a session on a desktop PC and yet still be able to seamlessly pick up and continue this session with his mobile device when being away from home.
- *Parallel device usage.* This usage pattern aims to combine the interaction and presentation capabilities of multiple devices simultaneously. A real immersive experience should include the ability to distribute an application's user interface over multiple devices. Accessing content such as a video stream can, e.g., result in a rendering component being displayed on a television screen, whilst the stream's playback controls are shown on the user's mobile device.

Despite the prevalence of mobile application frameworks and operating systems, aiming to simultaneously combine features of multiple devices, solutions are often tight to a very limited set of devices, or a specific set of usage contexts. Existing work generally focuses on proprietary protocols and only supports specifically targeted platforms and vendors (e.g., the emerging interactive and connected TV platforms, which enable second screen applications via smartphone devices). As a counter, the Munin toolkit and Gibraltar framework aim to broaden this scope with a more flexible peer-to-peer design for distributed mobile applications over the Internet [2] [3].

### 3 Enabling Multi-device Usage Patterns

The platform described in this section aims to generically enable multi-device usage patterns. As depicted in Fig. 2, the proposed platform consists of a web application framework. Web technology has been selected as primary delivery channel based on our platform's goal to cover a broad range of devices (i.e., PC, mobile, tablet, TV, etc.). The application platform does so by leveraging standardized and widely adopted web technology such as HTML, CSS, and JavaScript. In result, applications can be accessed from virtually any web-enabled device's browser via their uniform resource identifier (URI).

In comparison to traditional web application frameworks, the proposed platform is able to automatically adapt its served applications' user interfaces (UI) based on the number and types of devices operated by the end-user. For this purpose, the application platform must be capable of dynamically enrolling requesting devices to a particular user session. The platform does so by generating QR codes (Quick Response code). This two-dimensional barcode is encoded with the active application's URI and a session identifier. Any device with a camera can in turn scan the code to automate its enrollment.



**Fig. 2.** High-level architecture for a web application framework, enabling multi-device usage patterns

The subsequent adaptation of an application's user interface is supported via two mechanisms, i.e., server-side and client-side adaptation. Server-side adaptation allows for the optimization of an application's user interface before it is sent to the requesting device. This type of adaptation enables developers to perform server-side



UI adaptations based on the user's contextual setting. To do so, the proposed platform provides access to detailed device feature and capability information, which are detected via user agent matching. Moreover, this step aims to minimize resource usage on the client's device (network, CPU, memory, etc.).

However, the usage patterns described in Section 2 are primarily characterized by dynamic session handovers. As devices are allowed to randomly join and leave active sessions, support is needed for on-the-fly UI adaptation as well. Hereto, the client-side adaptation mechanism aims to enable the adaptation of a UI that is already being rendered by one or more particular devices. Within the proposed platform, client-side adaptation relies on the at runtime manipulation of the application's DOM (Document Object Model) via JavaScript instructions.

By default, client devices communicate with the application server over standard HTTP (Hypertext Transfer Protocol). Additionally, a WebSocket communication channel is set up for efficient bi-directional communication once the initial HTTP request is closed. This way, application state changes can easily be propagated to all devices within the same user session. Moreover, server-initiated adaptation instructions can be pushed to a client after another device has joined or left the session.

## 4 Proof of Concept Implementation

A prototype of the proposed platform has been implemented as part of the webinos open source project [4]. The project consortium involves over 30 partner companies, including device manufacturers, service providers, universities, and research organizations. The prototype's server-side components are implemented on top of Node.js, a flexible and event-driven runtime for Google's V8 JavaScript engine [5]. In order for the prototype to cover a broad range of devices, the client-side requirements have been kept to a minimum and encompass all devices with at least an Internet connection and a browser supporting HTML5 WebSockets.

Moreover, a proof of concept e-learning application was implemented to showcase the multi-device capabilities of the proposed platform. The implemented application focuses on providing students with a blended learning experience when accessing educational content. The application's intended end-user experience is based on the two multi-device usage patterns presented in Section 2. The application provides traditional e-learning functionality by enabling users to navigate through various types of learning content. This content ranges from static text, to presentations, videos, etc. The added value of the proposed platform, however, is the built-in support for enrolling additional devices. When a secondary device joins the user's active session, the presentation and interaction components of the application are automatically distributed between the active devices.

Fig. 3 depicts the use case of a mobile device joining a session started on a television set. Scanning the displayed QR code starts the enrollment procedure. With the obtained application URI and session id, the mobile device opens a browser window and requests the application platform access. This request initiates the first adaptation phase, i.e., server-side adaptation. As elaborated in Section 3, the application platform

aims to optimize the returned user interface based on the a-priori knowledge of the client device's characteristics and capabilities. For the proof of concept implementation, this data is gathered by matching the browser's user agent string with the WURFL device description repository [6].

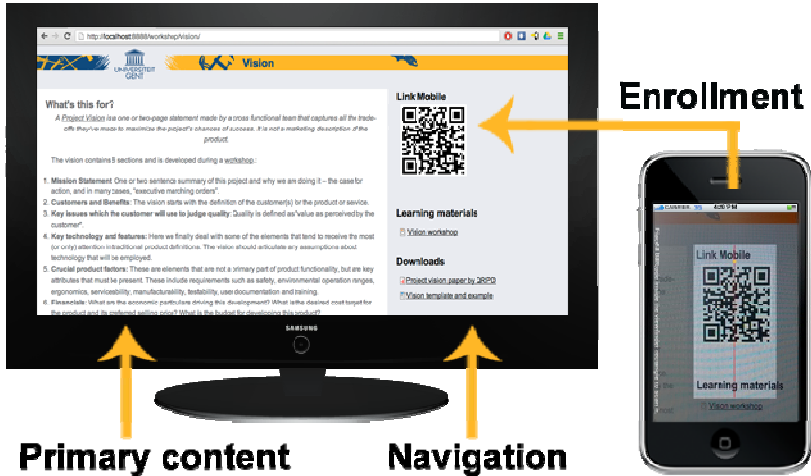


Fig. 3. Device enrollment and session synchronization via QR code scanning



Fig. 4. Distribution of the application's presentation and navigation modules between enrolled devices

Right after the enrollment, the television set is notified about the newly connected device. The application platform pushes client-side adaptation instructions to that device via their shared WebSocket communication channel. For the user's convenience, the television set is instructed to focus on primary content rendering and to remove the navigation bar (see Fig. 4). The client-side adaptation mechanism is implemented via jQuery's DOM manipulation API (Application Programming Interface). The API allows for the insertion and removal of specific DOM elements, as well as the modification of their contents and styling properties.

## 5 Conclusion and Future Work

In this paper, we've elaborated on the evolution towards multi-device usage patterns for accessing mobile and ubiquitous applications. We've presented the design for a web-based application platform, capable of automatically coping with the enrollment and synchronization of multiple devices owned by a particular user. Moreover, the platform supports the on-the-fly adaptation of its served application user interfaces. A platform prototype is implemented, as well as a proof of concept application for blended e-learning using multiple devices.

Future work includes a thorough quantitative study on the proposed platform's performance and scalability, as well as its contextual adaptability. Moreover, a qualitative user study is planned, which will be based on the prototype application presented in this paper. This evaluation data will serve to further validate and refine the assumptions made with regards to multi-device usage patterns.

**Acknowledgments.** The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement number 257103 (webinos project).

## References

1. Desruelle, H., Blomme, D., Gielen, F.: Adaptive mobile web applications: a quantitative evaluation approach. In: Auer, S., Díaz, O., Papadopoulos, G.A. (eds.) ICWE 2011. LNCS, vol. 6757, pp. 375–378. Springer, Heidelberg (2011)
2. Elmquist, N.: Munin: a peer-to-peer middleware for ubiquitous visualization spaces. In: Proc. of the 1st Workshop on Distributed User Interfaces (DUI 2011), pp. 17–20. University of Castilla-La-Mancha (2011)
3. Lin, K., Chu, D., Mickens, J., Zhuang, L., Zhao, F., Qui, J.: Gibraltar: exposing hardware devices to web pages using AJAX. In: Proc. of the 3rd USENIX Conference on Web Application Development (WebApps 2012). USENIX Association, Berkeley (2012)
4. Desruelle, H., Isenberg, S., Lyle, J., Gielen, F.: Multi-device application middleware: leveraging the ubiquity of the Web with webinos. *Journal of Supercomputing* (2013)
5. Node.js, <http://www.nodejs.org>
6. WURFL – Mobile Device Database, <http://wurfl.sourceforge.net>

# Touch Logger Resistant Mobile Authentication Scheme Using Multimodal Sensors

Hyunyi Yi, Yuxue Piao, and Jeong Hyun Yi\*

Department of Computer Science and Engineering  
Soongsil University, Seoul, Korea  
{hyunyiyi, qianbin127, jhyi}@ssu.ac.kr

**Abstract.** The PIN that is widely used for various services in mobile devices is highly vulnerable to attacks such as shoulder surfing. Various schemes have been proposed to solve this vulnerability of the PIN. However, despite the enhanced security of existing schemes, usability such as authentication time and error rate has decreased. In this paper, we propose a new scheme called PassWindow that allows enter a PIN securely through a window moving on the virtual keypad. PassWindow provides improved usability in the mobile devices and prevents shoulder-surfing attacks at the same time. We also propose an input method using multimodal sensors. This method strengthens the security against recording attacks and touch logger attacks.

**Keywords:** usable security, password-based authentication, shoulder-surfing attack, touch logger attack.

## 1 Introduction

Mobile devices store not only private information, such as pictures and contacts, but also store important information that allows the user to use various services conveniently such as e-mail and financial transactions. Therefore, the access to critical information and services are restricted through the password-based authentication. However, the user authentication for mobile devices is frequently performed in public places. As a result, there is a risk of password exposure by attacker observing the authentication interface over the user's shoulder [1].

Typical password authentication schemes well known to users are PIN(Personal Identification Numbers) and alphanumeric password. The PIN, which uses four digits numeric (0 to 9) as a password, is commonly used in financial services because it is easy to remember and input. However, there are only 10,000 available password combinations. Thus, the security against brute force attack is low. The alphanumeric password, which uses 6 to 12 digits including numbers and letters, is secure than the PIN against brute force attack. However, the alphanumeric password increases the burden on the user's memorization capabilities by requiring the user to set a strong password with restrictions such as text length, composition, etc. As a result, users tend

---

\* Corresponding author.

to write down their passwords or apply the same password in several systems [2, 3], which reduces the security of the alphanumeric password. Therefore, the existing schemes [4-7] do not satisfy the requirements of security and usability at the same time. In this paper, we propose a user authentication scheme that allows for the secure input of PIN. The proposed scheme makes the user to enter different values each time through a window that moves freely in the virtual keypad, thereby preventing shoulder-surfing attacks. Also, an additional method to input the password is proposed that using the sensors in mobile devices to defend against touch logger attacks.

The rest of the paper consists of the following sections. In section 2, we review related studies. Section 3 introduces the proposed scheme in detail and section 4 analyzes the security of proposed scheme. Section 5 presents the experimental results. Finally, Section 6 presents the conclusion.

## **2 Related Works**

### **2.1 Shoulder-Surfing Attacks in Mobile Devices**

Shoulder-surfing attacks are attacks that extort personal information deliberately by observing the user's behavior [8]. A shoulder-surfing attack can be categorized into two types depending on the attacker's capabilities: cognitive shoulder surfing and recording-based shoulder surfing [9].

### **2.2 Touch Logger Attacks in Mobile Devices**

A touch logger is a type of spyware that applies the keylogger to the touch interface. The touch logger, by logging touch location, can determine which characters are selected on the virtual keypad, and can be utilized as a means to capturing a state of screen when a touch event generates. The touch logger attacks that have been studied to date can be categorized into two types depending on the method for detecting the touch location. Cai and Chen proposed a new key logging scheme, TouchLogger [10], that utilizes the motion sensors in mobile devices. This scheme measures the variability of the values of motion data depending on the screen location that is being touched. This information is then used to extract the input. Damopoulos et al. proposed a touch logger that can record all of the touch events that occur on the screen [11]. This scheme makes it possible to collect the touch events in the background by hooking the touch event API of iPhone.

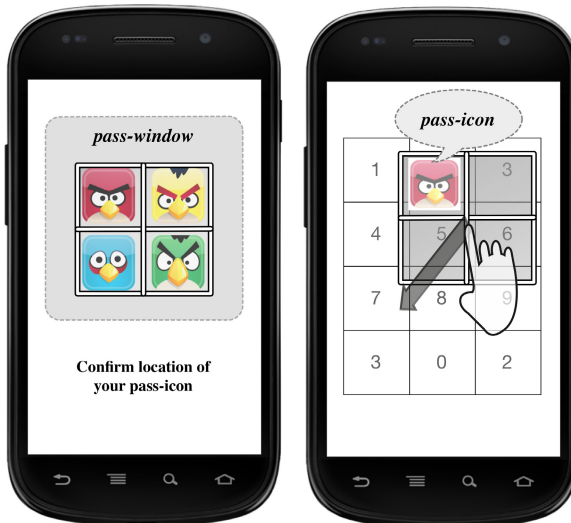
## **3 Proposed Scheme**

We propose a new PIN-based authentication scheme that considers the security issues that arise when authentication in mobile devices and the usability of small touch screens on these devices. This scheme is called "PassWindow [12]." PassWindow blocks direct exposure of password by enabling the input of the password through a grid-configured window, thereby preventing shoulder-surfing attacks. Users can move

the window freely over the virtual keypad using a simple touch event. This provides an additional input method that using the sensors in mobile devices. As a result, it strengthens security against recording and touch logger attacks.

PassWindow uses a PIN and an additional secret value as the password. In this section, the proposed scheme using an image as the additional value will be explained. The corresponding image is called a “pass-icon” and is used by the user to identify the location where the PIN is entered. The user selects the number  $N$  digits and the pass-icon during the password setup step. The authentication system randomly selects decoy icons, which are then stored along with the pass-icon. The secret value that should be memorized with the PIN can influence the ease of memorization. Therefore, it should be up to the users to determine which kind of elements are easier for them to memorize as secret values among the elements of image, text, color, etc.

The user identifies the location that matches with the PIN using the image that is displayed on the  $x \times y$  sized grid at the time of user authentication. This grid, which is called the “pass-window,” consists of the pass-icon and other decoy icons. Each cell in the pass-window is an image location and is represented as  $(i, j)$ . The user recalls the pass-icon within the pass-window and memorizes the location. The location of pass-icon within the pass-window is different whenever authenticated. After the pass-location is confirmed, then the virtual keypad consisted as numbers and the pass-window with its images disappeared, show up in the center. The user moves the pass-window, which is floating over the virtual keypad, so that the pass-icon moves over each digit of the PIN on the pass-location for authentication. Fig. 1 shows the PassWindow authentication process.



**Fig. 1.** PassWindow Authentication Process

The process of authentication is explained by assuming that the user selects 1234 as the PIN and by taking Fig. 1 as an example. First, the user, after confirming that the pass-icon is located at (1, 1) in the 2 × 2 pass-window, memorizes this as the pass-location. After this, we let the virtual keypad numbers “1,” “2,” “3,” and “4” overlap in the proper order with the pass-location and then make a selection. The pass-window is an important element that has a direct influence on both the security and usability of the proposed scheme. Therefore, some methods are provided below for pass-window composition and operation. There are two methods for operating the pass-window: the Touch Enter and the Hidden Enter method. These control methods are specialized for mobile devices, are convenient, and improve security against interface attacks. Fig. 2 shows the operation method for the pass-window.

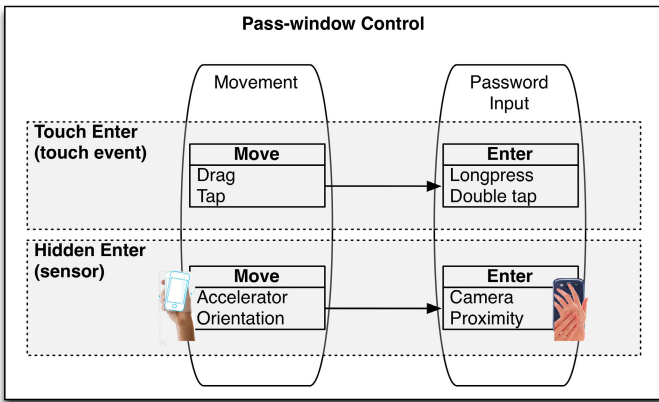


Fig. 2. Methods of Pass-window Control

**Pass-Window with Touch Enter.** This method utilizes the touch event in mobile devices. For example, after matching the number on the pass-location by moving the pass-window using a drag event, the input of password can be performed by touching the screen. As the user touches the screen directly, it may be possible for an attacker to identify the numbers that are being entered through the pass-window, but it is still not easy to acquire the PIN because the pass-icon is not known.

**Pass-Window with Hidden Enter.** This method operates the pass-window using the embedded sensors in mobile devices [13]. As the user moves the mobile device, the sensors detect the movement and use the information from this movement to move the pass-window. The entry of the password can be performed by covering the rear camera lens with a finger. Operating the pass-window via sensors helps to prevent touch logger attacks because it does not generate touch events, and entering the password through the rear camera sensor enables the users to hide the password input, thereby improving the security against shoulder-surfing and recording attacks.

## 4 Security Analysis

In this section, the security of the PassWindow is analyzed in relation to shoulder-surfing attacks and touch logger attacks.

### 4.1 Resistance to Shoulder-Surfing Attacks

Table 1 shows a comparison of the resistance of existing authentication schemes and the proposed scheme to shoulder-surfing attacks where  $N$  denotes the length of the password and  $L$  does the number of cells in the pass-window.

**Table 1.** Resistance to Shoulder-surfing Attacks

	Resistant to cognitive shoulder-surfing attack	One-time recording attack probability
Passfaces [4]	NO	1
DAS [5]	NO	1
PIN-Entry [6]	YES	1
ColorPIN [7]	YES	$1/3^N$
PassWindow	YES	$1/L$

Authentication schemes where the original password is input, such as DAS and Passfaces, are vulnerable to cognitive shoulder-surfing attacks. On the other hand, PIN-Entry and ColorPIN perform the authentication using varied response values and thereby prevent direct exposure of the password. Further, the PassWindow allows the indirect entry of the password through the pass-window. Therefore, an attacker who attempts a shoulder-surfing attack must memorize each of the  $L \times N$  characters that are entered through the pass-window in order to obtain the password. With regards to recording attacks, the security of the PassWindow is influenced by the size and composition of the pass-window and the input methods. The number of response values that an attacker can acquire with a one-time recording attack is equal to the number of the pass-window cells. Because each one of the cell values, respectively, becomes the actual password, the one-time recording attack success probability is  $1/L$ . The probability of acquiring the password with a one-time recording attack when ColorPIN is used is  $1/3^N$ , which is more secure than PassWindow. As the pass-icon is not known to the attacker, it is not easy to find out the PIN from a single observation. However, if the authentication process is recorded several times for analyses, the PIN can be deduced. Nevertheless, the security of PassWindow can be improved against recording attacks through the use of the Hidden Enter. The attacker may identify the pass-window movement location throughout the authentication process recording, but it is not easy to detect the actions that cover the camera lens on the rear side of the device. This makes it possible to hide the selections of the response values as the pass-window is moving, thereby preventing the attacker from acquiring the input response values.



## 4.2 Resistance to Touch Logger Attacks

Table 2 shows a comparison of the probabilities of success of the proposed and existing schemes for touch logger attacks.

**Table 2.** Resistance to Touch Logger Attacks

	Motion sensor value-based attack probability	Touch event-based attack probability
Passfaces [4]	1	1
DAS [5]	1	1
PIN-Entry [6]	1	1
ColorPIN [7]	$1/3^N$	$1/3^N$
PassWindow (Touch Enter)	$1/L$	$1/L$
PassWindow (Hidden Enter)	$1/L$	N/A

As the authentication for Passfaces, DAS, PIN-Entry, and ColorPIN schemes is performed by touching the screen, it is possible to detect the generation of the touch event and to obtain the entire authentication screen and touch locations as well. Therefore, the probabilities of success for touch logger attacks on the existing schemes are equal to those for recording attacks. The security of PassWindow against touch loggers attack varies depending on the operation method that is used for the pass-window. When Touch Enter is used, the resistances to touch logger attacks and to recording attacks are equal to the resistance of the existing schemes. The Hidden Enter, on the other hand, allows the authentication to be performed using acceleration sensors and the camera lens. However, as this does not generate touch events, touch loggers cannot acquire any useful information.

## 5 Experimental Results

This section compares and analyzes the usability of existing authentication schemes and the proposed scheme using two experiments. One experiment is based on a user interface evaluation tool and the other experiment is based on a real user experiment. Table 3 shows a comparison of the experimental results for two experiments.

**Table 3.** Experimental Results

	CogTool Authentication time [s]	User Test Authentication time [s]	User Test Error rate [%]
Passfaces [4]	8.65	14.55	14.00
DAS [5]	11.23	9.87	12.00
PIN-Entry (Immediate) [6]	20.05	19.55	28.00
PIN-Entry (Delayed) [6]	16.99	26.60	34.00
ColorPIN [7]	19.58	20.10	16.00
PassWindow	18.12	17.86	4.00

## 5.1 CogTool Test

We describe experiments that predict the authentication times for existing authentication schemes and the proposed scheme using CogTool [14].

The authentication time predicted by CogTool for PassWindow is measured at 18.12 s, which is longer than those for the simpler Passfaces and DAS authentication methods. In the meanwhile, it is expected that the authentication for PassWindow will be faster than the authentications for the PIN-Entry and ColorPIN schemes that require complicated password input processes.

## 5.2 User Test

For the user experiments, the existing and proposed schemes were implemented using an Android-based application program. The implementation used Eclipse Helios, Android SDK 2.3, and JAVA 1.6.0. Five units of the Samsung Smartphone SHW-M250S (1.2 GHz) were used as test equipment. The tests were performed by a total of ten persons. The testers requested to set up password and repeated the authentication process 5 times. Table 3 summarizes the average authentication times and error rates that were calculated using the authentication logs.

The test results show that PassWindow has more improved authentication time than the existing schemes that have similar or higher security do. The results also demonstrate its superiority for the ease of password memorization with relatively lower error rates than the schemes with speedier authentication speeds.

## 6 Conclusion

In this paper, we proposed a new PIN authentication scheme that prevents shoulder-surfing attacks. The proposed scheme uses a secret value in addition to the PIN. The authentication method works by matching the PIN with the particular location of the window on which the virtual keypad moves freely. A particular location of the window varies each time based on the secret values, thereby preventing the exposure of the real PIN. The Hidden Enter that utilizes multimodal sensors was also proposed. This input method blocks the generation of touch events and is hard to catch a timing of the password input.

The security analysis results showed that the proposed scheme prevents shoulder-surfing attacks. Further, the input method that uses sensors showed that it strengthens the security against recording and touch logger attacks. From the analysis of the usability test results, it was observed that the proposed scheme enabled speedier authentication than existing schemes that are secure against shoulder-surfing and recording attacks. Furthermore, an ease of password memorization through significantly reduced error rates was demonstrated.

**Acknowledgment.** This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A1A2013041).

## References

1. Schaub, F., Deyhle, R., Weber, M.: Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms. In: 11th International Conference on Mobile and Ubiquitous Multimedia, pp. 1–10 (2012)
2. Sasse, M.A., Brostoff, S., Weirich, D.: Transforming the ‘weakest link’— a Human/Computer Interaction Approach to Usable and Effective Security. *BT Technology Journal* 19(3), 122–131 (2001)
3. Payne, B.D., Edwards, W.K.: A Brief Introduction to Usable Security. *IEEE Internet Computing* 12(3), 13–21 (2008)
4. Passfaces, <http://www.realuser.com>
5. Park, S.B.: A Method for Preventing Input Information from Exposing to Observers. 10-2004-0039209, Korea (2004)
6. Roth, V., Richter, K., Freidinger, R.: A PIN-entry Method Resilient against Shoulder Surfing. In: 11th ACM Conference on Computer and Communications Security, pp. 236–245 (2004)
7. Luca, A.D., Hertzschuch, K., Hussmann, H.: ColorPIN: Securing PIN Entry through Indirect Input. In: 28th International Conference on Human Factors in Computing Systems, pp. 1103–1106 (2010)
8. Tari, F., Ozok, A.A., Holden, S.H.: A Comparison of Perceived and Real Shoulder-surfing Risks between Alphanumeric and Graphical Passwords. In: 2nd Symposium on Usable Privacy and Security, pp. 56–66 (2006)
9. Shi, P., Zhu, B., Youssef, A.: A PIN Entry Scheme Resistant to Recording-based Shoulder-surfing. In: 3th International Conference on Emerging Security Information, Systems and Technologies, pp. 237–241 (2009)
10. Cai, L., Chen, H.: TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion. In: 6th USENIX Workshop on Hot Topics in Security, p. 9 (2011)
11. Damopoulos, D., Kambourakis, G., Gritzalis, S.: From Keyloggers to Touchloggers: Take the Rough with the Smooth. *Computers & Security* 32, 102–114 (2013)
12. Yi, J.H., Ma, G., Yi, H., Kim, S.: Method and Apparatus for Authenticating Password of User Device. 10-1175042, Korea (2012)
13. Yi, J.H., Yi, H., Piao, Y., Kim, T.: Method and Apparatus for Authenticating Password using Sensing Information. 10-2012-0103897, Korea (2012)
14. CogTool, <http://cogtool.hcii.cs.cmu.edu/>

# Commodity Recommendation Algorithm Based on Social Network

Zhimin Yin, Xiangzhan Yu, and Hongli Zhang

Department of Computer Science and Technology,  
Harbin Institute of Technology, Harbin, China  
yzm0621@163.com, {yxz, zhanghongli}@hit.edu.cn

**Abstract.** According to the records of the user's past shopping and the properties of merchandise, recommended related products is a core function of online shopping, the feature also has an important value in the search of the Internet page, is emerging data mining research field. Its core is that recommending maybe interested information to the user. This subject according to user feedback records of the shopping system or retrieval system, using the cluster analysis techniques to identify the relationship among objects, and analyzing cluster characteristics of the object semantics to support the efficient product recommendations and sort of related information. In order to effectively improve commodity recommendation and the sorting effect of related information, this paper suggested that with the user access records and feedback record of shopping system or retrieval system analysis of relationship between objects independently, to support the efficient product recommendations and sort of related information.

**Keywords:** shopping, data mining, clustering, recommended.

## 1 Introduction

With the rapid development of Internet, personalized information recommendation service has become one of the hotspots in e-Commerce. However, the e-commerce system for its user more choice, while commodity information overload is more and more severe, and its structure becomes more complicated, it is difficult to the space quickly find our looking for the real goods among a lot of product information[1,2]. Then, how to use e-commerce to organize information effectively, how to understand the customer's hobbies and interests as much as possible, thus optimizing website design to facilitate consumer shopping, becoming an e-commerce development urgent problem to be solved[3]. As a result, a lot of teams have started to study the information recommended, and achieved some good results. Some knowledge of social networks can also be a good recommendation algorithm applied to information.

In this paper, we introduce the basic theory of e-commerce recommendation system and collaborative filtering techniques, content-based recommendation, as well as mixed recommend recommendation based on association rules. Then we presents a way of mining useful and valuable information using social networking content, and

provide recommendations and information through the result. Through social network analysis techniques to get users interested in the same cluster, clustering in the user based on the user access and interest combined with relevant content recommendation algorithm items. According to shopping system or retrieval system users to access records and relevance feedback records, we propose a user feedback product recommendation model based on social network recommendation algorithm.

## **2 Related Recommendation Algorithm**

### **2.1 User-Based Collaborative Filtering Algorithm**

The basic idea of user-based collaborative filtering algorithm is based on an assumption that "people with similar preferences with your favorite things you probably like", so user-based collaborative filtering main task is to find the user's K most similar neighbors, which according to nearest neighbor preferences to make the score predicted unknowns [4]. This algorithm has a performance bottleneck, when the number of users increases, the complexity of the nearest neighbor will be substantial growth. So this method can't meet the requirements of the recommended time.

### **2.2 Content-Based Filtering Algorithm**

Content-based filtering algorithm is the idea that a user would prefer those similar products which he had to buy. Content-based recommendation is based on the characteristics of learning history information, its advantage performance in improving the recommended scalability, and can make good explanation based on the results of recommended recommend can help users find interested in the content of the project, but can't find new content for the user[5].

### **2.3 Association Rule-Based Mining Recommendation Algorithm**

The basic purpose of association rule mining is to look for correlation in the merchandise sales records, to better guide the marketing strategy formulation [6]. A typical rule is: "43% of the purchase instant coffee Nescafe, Nestle Coffee Mate will be bought." Goods association rules can be divided into spatial correlation and time correlation. In the general study of association rules is just the space, also in the same time (the same time buy), analysis of consumers often buy goods together, this is also known as basket analysis of main support technology.

## **3 Social Network Analysis**

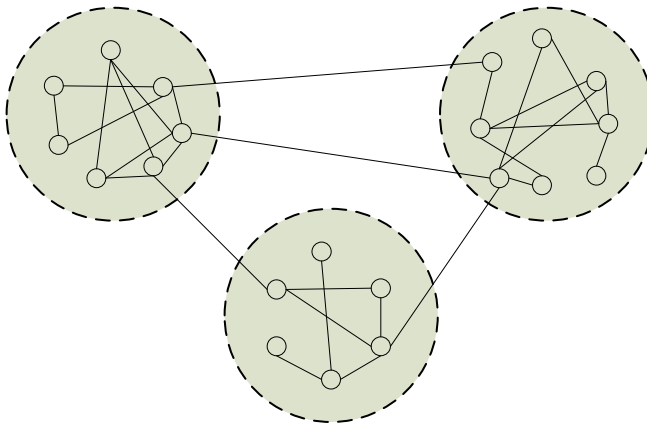
### **3.1 Social Network Analysis**

Because of the interaction among social networks, the social individual members influence to form a relatively stable relationship system. Social network focused on

the interaction and relationship between people and social interaction will affect people's social behavior [7]. Social network analysis (SNA) is used to measure individual actors and their complicated relationship between social network members. The method of SNA allows researchers perspective to see the whole community interaction between network users to see their interconnected relationship diagram. SNA consists of three units of analysis: 1) Actor, node in the network, that is, each user or events in the social network etc.2) Relationship, that is, the connections between various nodes in the network, reflecting the formation of a variety of actors, social relations and interactions.3) Connection, a set of relations. SNA relationship study consists of three parts: the content, direction and intensity [8]. As one of SNA research relationships is the content, the use of SNA technology Mining same interests or similar products User Clustering.

### 3.2 Community Structure in the Network

As the physical meaning and mathematical characteristics of the network properties of thorough research, people found that many real networks all have a common nature, namely the community structure. That is to say, this network is composed of several groups or clusters. Inside each cluster, node has a relatively very close connection, but the connection between the various groups is relatively sparse. Figure 1 shows a small community structural qualitative network diagram. The figure includes three community network, corresponding to the three circle surrounded with dotted line [9].

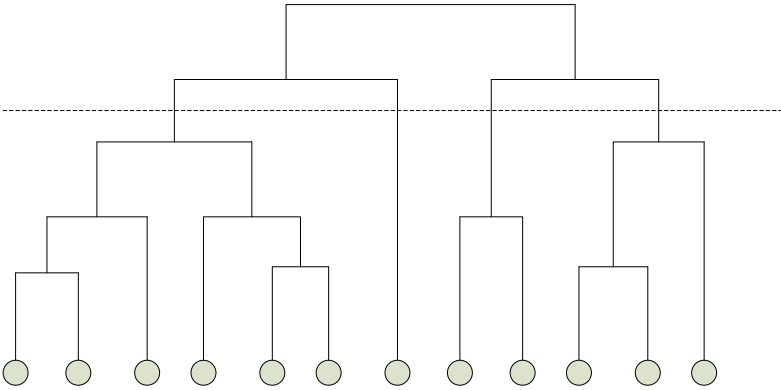


**Fig. 1.** A small community structural qualitative network diagram

For example, WWW may be regarded as consisting of a lot of web site community network, including all of the sites within the same community are discussing some topics of mutual interest.

### 3.3 Community Detection

In general, to find the exact solution of a partition problem is a NP difficult problem. Therefore, there is no effective accurate solution when the data size is large. In most cases, however, there are a lot of exploratory algorithm can get a satisfied solution. Hierarchical clustering is a kind of traditional algorithm which looking for social network community structure. It is based on the similarity or connection strength between each node, the nature of the whole network is divided into several sub groups. The algorithm can be divided into two categories according to a side changing method which removed from or added to the network: splitting method and condensation method [10]. The basic idea of condensation is to calculate on the similarity of each node in a network, then from the highest similarity of nodes added to the network which has  $n$  nodes and 0 edges. This process can be terminated at any point, and then the network is composed of several communities. The formation process can be expressed in Figure 2.



**Fig. 2.** Record the results of the algorithm using the clustering tree

At the bottom of the tree, each circle represents a node in the network. When horizontal dotted line gradually moves up from the very bottom of the tree, each node has gradually aggregated into some of the larger society. When the dotted line moved to the top, it indicates that the entire network becomes a general community. Any location in the tree with a dotted line disconnect, it corresponds to produce a kind of community structure. Aggregation method based on a series of similarity metrics has been applied to many different networks. For example, in the user's shopping network, if two users purchased the same two products, then there is an edge between them. So, we can use the number of two users buying the same items to measure the similarity of user nodes. Also we can use the two goods being purchased the number of times at the same time to measure similarity of goods node.

## 4 Recommendation Algorithm Based on Newman

### 4.1 Fast Algorithm Based on Newman Thought

This fast algorithm based on greedy algorithm is actually a condensation algorithm.

*Algorithm:*

- a) Initialize the network consists of  $n$  Societies, each node represents an independent community,  $e_{ij}$  and  $a_i$  meet:

$$e_{ij} = \begin{cases} 1/2m, & \text{If there is an edge between the nodes } i \text{ and } j \\ 0, & \text{else} \end{cases} \quad (1)$$

$$a_i = k_i / 2m \quad (2)$$

$k_i$  is the degree of node  $i$ ,  $m$  represents the total number of edges in the network.

- b) If the two societies are adjacent, then we merge them together in turn and calculate the combined module degree increments:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (3)$$

According to the principle of the greedy algorithm, the merger should be along the  $Q$  of the increase or decrease the biggest or smallest direction every time. This step of the algorithm complexity is  $O(m)$ . After the merger, updates the value  $e_{ij}$ . This step of the algorithm complexity is  $O(n)$ . Therefore, the second step for the general algorithm complexity is  $O(m+n)$ .

- c) Repeat the step b, and constantly consolidated associations until the entire network have been merged to a community. Here,  $n-1$  times merger execution at most.

Therefore, the time complexity of the algorithm is  $O(n(m+n))$ , After completion of the whole algorithm we can get a community structure decomposition tree. You can select different network community structure in different position disconnected. In the community structure, the choice of corresponds to a local maximum  $Q$  value, is to find the best online community structure.

### 4.2 Recommendation Algorithm Based on Newman Thought

First, initialize the network consists of  $n$  Societies, each node represents an independent community, in which each node has only the corresponding goods,  $e_{ij}$  and  $a_i$  meet:

$$e_{ij} = \begin{cases} c_{ij} / 2m, & \text{If there is an edge between the nodes } i \text{ and } j \\ 0, & \text{else} \end{cases} \quad (4)$$

$$a_i = k_i / 2m \quad (5)$$



$k_i$  is the degree of node  $i$ ,  $m$  represents the total number of edges in the network,  $c_{ij}$  represents the number of edges between node  $i$  and node  $j$ .

Second, if the two societies are adjacent, then we merge them together in turn and calculate the combined module degree increments:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (6)$$

According to the principle of the greedy algorithm, the merger should be along the  $Q$  of the increase or decrease the biggest or smallest direction every time.

Then, repeat the previous step, and the merger of the commodity associations until the entire community members exceeds a certain critical value. The algorithm as follows:

- a) According to the existing trading records, to build a diagram of the goods and commodities.
- b) According to the merchandise diagram, initialize the network entry  $e_{ij}$  and  $a_i$ .
- c) After merging associations, calculate  $\Delta Q$ , take  $\Delta Q$  maximum of two societies increased to merge.
- d) Determine whether the members of the new society is greater than the critical value, if it is greater than the critical value, the newly generated independent society to stop merger.
- e) If it is less than the critical value, the new generation of commodity society continued participation in the merger.

## 5 Conclusion

In this paper, we studied the commodity recommendation algorithm. According to the user's purchasing power, in order to let users find what they are interested in as fast as possible, we try our best to make bold speculation. We recommend users might be interested in something to the user. The paper in general can be divided into two main parts. The first part is a brief introduction to some common ideological recommendation algorithm. We put forward the commodity recommendation algorithm based on the content of social networks in the second part. Recommendation algorithm mainly based on social network has realized the social network in the classic partitioning clustering algorithm to the merchandise and clustering division made by the user. The user makes the same cluster with high interest similarity and clustering in the same commodity with high similarity. Thus, we realize personalized recommendation. This paper mainly makes a study of recommended algorithm and we will be focus on the algorithm timeliness in future work.

**Acknowledgement.** This research was partially supported by the National High Technology Research and Development Program of China (863 Program) under grant No. 2012AA012506, the National Science Foundation of China (NSF) under grant No. 61173145.

## References

1. Agichtein, E., et al.: Improving Web Search Ranking by Incorporating User Behavior Information. In: Proceeding SIGIR 2006 Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, USA (2006)
2. Deli, C.: Design of Personalization E-commerce Model Based on Data Mining. *Journal of Information* 8 (2006)
3. Hillard, D., et al.: Improving Ad Relevance in Sponsored Search. In: Proceeding WSDM 2010 Proceedings of the Third ACM International Conference on Web Search and Data Mining, New York, USA (2010)
4. Sun, T., Wang, L., Guo, Q.: A Collaborative Filtering Recommendation Algorithm Based on Item Similarity of User Preference. In: *Knowledge Discovery and Data Mining* (2009)
5. Peng, Y., Cheng, X.: Item-based collaborative filtering algorithm using attribute similarity. *Computer Engineering and Applications* 43(14), 144–147 (2007)
6. Zhang, X.-Z.: Building Personalized Recommendation System in E-Commerce using Association Rule-Based Mining and Classification. *Machine Learning and Cybernetics* (2007)
7. Streeter, C.L., Gillespie, D.F.: Social Network Analysis. *Journal of Social Service Research* 16 (1993)
8. Giupponi, C., Camera, R., et al.: Network Analysis, Creative System Modelling and Decision Support: The NetSyMoD Approach (March 2006)
9. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *The American Physical Society* (2004)
10. Fortunato, S.: Community detection in graphs. *Physics Reports* (2010)

# A Layered View Model of Social Experience Design: Beyond Single-User User Experience

Toshihiko Yamakami

ACCESS, Software Solution  
1-10-2 Nakase, Mihama-ku, Chiba-shi, Japan 261-0023  
<http://www.access-company.com/>

**Abstract.** As the time on site of Internet social services increases, it is important to develop a systematic framework that facilitates good social experience design. As an analogy of user experience design, it is natural to lead to “social experience design.” In this paper, the author discusses research questions involving social experience design. Then, the author provides a three-dimensional view model that facilitates systematic approaches toward better social experience design. Finally, the author provides a layered view model of social experience design for one of the above-proposed dimension of approaches.

## 1 Introduction

As the role of the Internet infrastructure changes from data access to social interactions, it is crucial to develop a systematic methodology for creating better social services. Social experience design is one candidate for such a methodology. Social experience poses an interesting set of research questions beyond user experience design for single users.

Better social experience is one of the important goals in the era of social services. Social experience design is different from user experience design on several crucial points. Social experience is intangible, difficult to visualize, evolves over a span of time, and emerges with emotion and engagement, which are difficult to measure. It will be a challenging topic of research for the next decade, considering the profound level of acceptance of social services worldwide.

First, the author provides a dimensional model for bird's-eye views of social experience design. Then, the author proposes a layered view of social experience design for one of the proposed dimensions.

## 2 Background

### 2.1 Purpose of Research

The aim of this research is to identify a framework that can cope with the demands for social experience design.

## 2.2 Related Works

The term User Experience Design was coined by Donald Norman. He also discussed emotional design and mentioned how emotion is a necessary part of life, affecting how we feel, how we behave, and how we think [4]. Buxton is another advocate of user experience design. He discussed detailed design methodologies for creating great user experiences [1].

The author presented the concept of social experience design [5]. He presented an evolutionary-path-based approach for social experience design [6]. The details of social experience design have not been explored in past literature.

The research of social experience design is still in its early stages. Social experience design needs to address the same wide range of complicated issues that organization management and collaboration studies have faced over the decades [2]. The originality of this paper lies in its examination of a systematic approach toward social experience design methodologies.

## 3 Definitions and Methods

The definitions for user interface, user experience, social interface, and social experience, are depicted in Table 1 [6].

**Table 1.** Definitions

Term	Description
User interface	Design of human-machine interaction where interaction between humans and machines takes place. Its aim is the effective operation and control of the machine with usable feedback from the machine.
User Experience	Design of how a person feels about using a product, system, or service. It highlights valuable aspects of human-computer interaction and product ownership.
Social interface	User-computer interface that deals with human-human interactions. User-computer interface that deals with Multi-user interactions. (This is the definition used in this paper. Social interface may represent human-like computer interface in other contexts)
Social experience	Design of the way a person feels about other humans through a computer-user interface.

### 3.1 Research Questions

This field is still immature. There are many unsolved research questions, as depicted in Table 2.

**Table 2.** Research questions

Item	Description
Comparison to user experience design	What is unique and different in social experience design compared to user experience design? User experience design is a design paradigm of architectures and interaction models that impact the user's perspective of a system/product/service. User experience is based on user-centered design. What is the main paradigm for social experience design?
Frameworks for social experience design	What are the basic logical building blocks for social experience design? Are they structural view models, taxonomies, or an encyclopedia of successful social experience design rules?
Design steps	What are the design steps in social experience design?
Evaluation methodologies	What are the evaluation methodologies for social experience design? What are its standard scenarios? What are the sets of evaluation methods and environments?
Metrics	What are the metrics of good social experience? Is there any single state of ideal social experience to pursue?

### 3.2 Methods

Social experience design covers the multi-faceted approaches of user and social interaction. Social aspects include a wide range of unwritten features that are difficult to evaluate in a quantitative or consistent manner. The author performs the following:

- Examines examples of good social experience design
- Examines metrics of good social experience design
- Identifies dimensional models for bird's-eye views of social experience design
- Identifies a systematic framework of social experience design in one of the above-identified dimensions

## 4 Frameworks

### 4.1 Framework Requirements

The requirements of the framework for social experience design are as follows:

- Abstractness and applicability to a wide range of social activities
- Consistency
- Constructability
- Learn-ability
- Enabling social emotions
- Ability to facilitate social activities

Examples of deliverables of good social experience design are as follows:

- Easy to construct
- Fits to social interactions

- High engagement
- High aggregated user satisfaction
- High performance
- Easy to learn

One challenge to social experience design is the fact that there is no intuitive final step for social experience, which is different from user experience design. Social experience is partially open-ended and may span over a long duration of time.

### 4.2 Metrics

It is a challenge to coin metrics for evaluating social experience design. A dimensional view model of the metrics of good social experience design is depicted in Fig. 1.

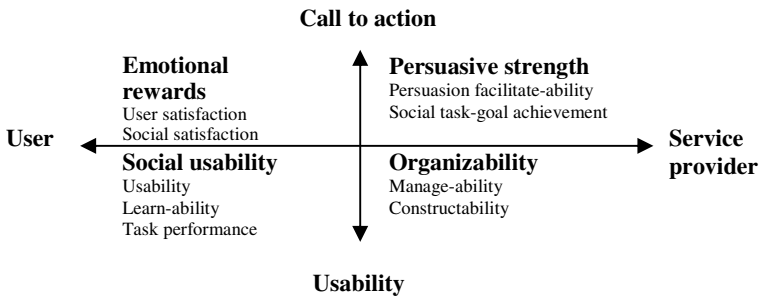


Fig. 1. Dimensional view model of the metrics of good social experience design

The multi-user aspect, asynchronicity, variety of emotion, heterogeneity, culture, and long-term aspect add further complexity to the existing user experience design.

### 4.3 Dimensional Model

User-centered design is a central paradigm for user experience. The author assumes that socially-empowered engagement as the first step of social experience design.

It is important to identify dimensions of social experience design. A systematic methodology requires a logical model of social experience design. The candidates for entities in a logical model of social experience design are as follows: (a) components of social experience design, (b) relationships among components, and (c) Laws and best practices for each component and/or relationships.

The components of user experience design are as follows: (a) social cognition, (b) social interaction, (c) social relationship,(d) emotion,(f) social meanings.

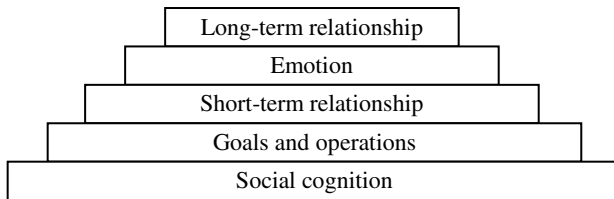
Assuming socially-empowered engagement, the dimensions of social experience design are depicted in Table 3.

**Table 3.** Dimensions of social experience design

Dimension	Description
Hierarchical dimension	Structural relationships.
Taxonomy dimension	Components and best practices are grouped with aspects, domains, emotion deliverables and functions.
Design Time-dimension	Step-wise, stage-wise design methodologies. Methods are listed in chronological order.

## 5 Layered View of Social Experience Design

There are a wide variety of relationships between social experience design components. The author attempts to create a simple structure in order to provide a solid ground for later design methodologies. Examining the best practices of mobile social game design for constructing good social experience has led the author to a hierarchical structure. The author proposes a layered view of social experience design depicted in Fig. 2.



**Fig. 2.** A layered view model of social experience design

The layer of social cognition serves as a function for awareness that facilitates the need for cognition. The layer of emotion serves as a function for cognitive aspects beyond each operation. The layer of short-term relationship serves as a social framework for each operation. The layer of goals and operation serves as a function of targeting and attention. The layer of long-term relationship serves as a function of building roles and trust over a long span of time.

**Table 4.** Examples of techniques in each layer from best practices of mobile social games

Layer	Example in mobile social games
Long-term relationship	High engagement and commitment through team-play in weekly events
Emotion	Joy of achievement in team-play through individual, incremental achievement of high levels.
Short-term relationship	Support and counter-support as social rewards.
Goals, operations	Visual feedback from team-goal achievement. Interaction around team-goal achievement. Real-time human feedback to support actions.
Social cognition	Timeline of team member activities.

The metrics for measuring the success of social experience design are not yet clear. The author has chosen one of the candidate goals: high social engagement. Examples of the techniques in each layer, from the best practices of mobile social games are depicted in Table 4.

## **6 Discussion**

### **6.1 The Advantages of the Proposed Approach**

The proposed layered view model provides a building block for the logical relationships between the components of social experience design. The model serves as a basis for deconstructing the high-level goals of social experience design such as high social engagement into detailed design components and techniques. Social dynamics are complicated and difficult to design in advance. However, the proposed layered deconstruction provides clues for mapping high-level goals into concrete design techniques. The proposed model facilitates the guiding process of social experience design using the layered deconstruction.

The proposed model is different from the Gamenics theory by Saito [3](Section 23.1 What is “Gamenics” from the viewpoint of explicitly defining the relationship among design elements. The proposed model is different from user experience design from the viewpoint of explicitly handling long-term and short-term social relationships. The proposed model is different from the CSCW layered model such as Awareness, coordination, cooperation, and collaboration from the viewpoint of focusing on relationships rather than social tasks.

The proposed model provides a simple structure: a hierarchy of design elements, which is a good starting point for exploring social experience design. The proposed model has good fits with the best practices of mobile social game design in each layer. The proposed model has general applicability to social experience design because it separates any specific tasks.

### **6.2 Limitations**

This research is a qualitative study. Quantitative measures for verifying the multiple aspects of social experience design discussed in this paper remain for further study.

The two other dimensions of the proposed dimensional model of social experience design are not covered in this paper. Detailed study of each layer of the proposed view models requires future research.

Acceptance of social experience in a real world environment is beyond the scope of this paper. Quantitative analysis of usability and the satisfaction of social experience design requires future research.

The concrete design methodology of social experience is beyond the scope of this paper.



## 7 Conclusion

An evolutionary step from the concept of user experience design brought user interaction design into a new stage where design is focused on good user experience rather than simple consistency, integrity, learn-ability and good performance. A similar shift is likely happening in social experience design. It is still unclear how the goal of good social experience design can be achieved. No intuitive approach to good social experience is feasible because social experience can be formed in the collective mind of users through long-term engagement with a group of users.

The author examines the concept of social experience design. The author discusses a three-dimensional model of social experience design to guide the formation of systematic methodologies in social experience design. In order to explore one of the dimensions, the structural dimension, the author proposed a layered view model of social experience design. The proposed model provides a hierarchical view of social experience design, which helps form a systematic approach to social experience design. The author discusses the comparison to other approaches and highlights the applicability of the proposed model. This is a forward-moving approach to the many unanswered research questions of social experience design.

The increasing importance of social services on the Internet has leveraged the needs of social experience design. The proposed model is a stepping stone toward a systematic approach in social experience design.

## References

1. Buxton, B.: *Sketching User Experiences: Getting the Design Right and the Right Design* (Interactive Technologies). Morgan Kaufmann, New York (2007)
2. Grudin, J.: Groupware and social dynamics: eight challenges for developers. *CACM* 37(1), 92–105 (1994)
3. Isbister, K., Schaffer, N.: *Game Usability: Advancing the Player Experience*, illustrated edition. Morgan Kaufmann (2008)
4. Norman, D.A.: Emotional design. *Ubiquity* 2004, 1 (January 2004)
5. Yamakami, T.: From user experience to social experience: A new perspective for mobile social game design. In: *UFIRST 2012*, pp. 792–796. IEEE CS (2012)
6. Yamakami, T.: An evolutionary path-based analysis of social experience design. In: Park, J.J.(J.H.), Ng, J.K.-Y., Jeong, H.Y., Waluyo, B. (eds.) *MUE 2013*. LNEE, vol. 240, pp. 69–76. Springer, Heidelberg (2013)

# A New Schedule Strategy of Embedded Multi-core SoC

Tie Qiu, Fangbing Liu, Feng Xia, and Ruixuan Qiao

School of Software, Dalian University of Technology, Dalian 116620, China  
{qiutie, f.xia}@ieee.org

**Abstract.** In this paper, we propose the weighted least-connection (WLC) task schedule strategy for isomerism architecture SoC. Each core communicates with other cores through mailbox. The master core has two main threads: communication thread and task allocation thread. The slave core has two threads: communication thread and data deal thread. The dynamic weighted least-connection task schedule strategy takes each core's computing performance and each core's real-time processing information into consideration to make each core's load balanced and to reduce the congestion. We do the experiment on Xilinx V5 platform and compare dynamic weighted least-connection task schedule strategy with round-robin task schedule strategy.

**Keywords:** Embedded Multicore SoC, Schedule Strategy with Least-connection, Inter-core communication.

## 1 Introduction

Embedded Multi-core System on Chip (EMSoC) is a kind of high-performance parallel processor. Multiple executing cores of EMSoC can compute synchronously [1], which is important for embedded devices to execute large amounts of data processing and complex computing [2, 3]. In EMSoC, multiple computing cores are integrated into a processor chip, and each core is an independent processing unit. The relationship between these cores is tightly coupled, and these cores are often interconnected by shared-bus or the high-speed data channel [4].

EMSoC can execute process parallel and improve performance greatly than single core processor [5]. Multi-core processor can be divided into two main types according to its architecture: isomorphism and isomerism [6]. The isomerism can be expanded easily and has richer functions than isomorphism, so it's widely used in a variety of applications [7], for example large amount of data collection and processing in embedded area. The enormous performance enhancement in multi-core platform injects lot of challenges into task allocation. Altogether task schedule is a crucial part in the view of the operating system [8].

In this paper, a high-performance processor EMSoC has been designed. We use Xilinx V5 as the hardware platform, and MicroBlaze soft core as processors on the platform. Every MicroBlaze core can run a simple operating system, and each soft core can work independently. Among the cores, there is one master core and the left cores are slave cores. The master core is responsible for allocation coming tasks. We

design shared bus communication architecture on the platform and the cores can communicate with others through mailbox. We propose Weighted-Least Connection (WLC) strategy for task allocation. In D-WLC, the master core collects each slave core’s real-time load situation and adjusts the task allocation in order to achieve load balancing and to avoid congestion.

## 2 EMSoC Architecture

The architecture is shown in Figure 1. The architecture has eight executing cores: *MB\_0* to *MB\_7*. Each core is realized by MicroBlaze core. *MB\_0* is the master core, which is responsible for handling task allocation. The other seven cores, *MB\_1* to *MB\_7*, are slave cores. The tasks are running on slave cores. We must allocate memory for each core to store program code. We use *Bram\_block* IP-core as the memory of each core, *Bram0* to *Bram7*. Each core use *plb\_v46* IP-core as the local bus module, *plb\_0* to *plb\_7*. The local bus is used to transfer data and instruction. In order to reduce the inter-core communication overhead, we use the shared bus to realize the inter-core communication. Each salve core communicates with master core through mailbox, *mbox0* to *mbox6*. The mailbox is realized by mailbox IP-core. In order to prevent several cores visiting the same memory address at the same time, conflict

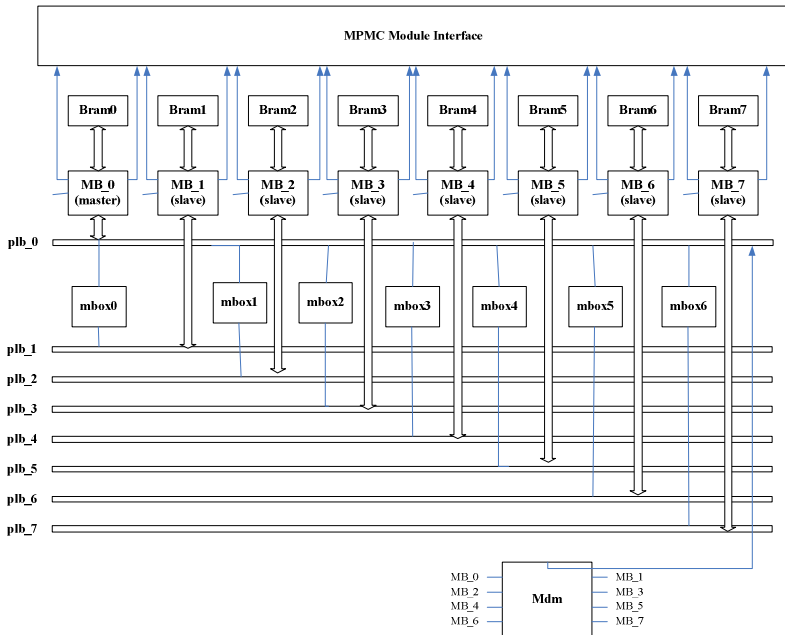


Fig. 1. Architecture of EMSoC

control module is needed. We use mpmc IP-core as the conflict control module, *MPMC*. *MPMC* has eight ports to control memory visits from cores. It has an independent bram to store register data. At last each core must have a clock to work normally. We use xps\_timer IP-core as the timer module. *Mdm* is used for debugging.

### 3 EMSoC Scheduling Strategy

#### 3.1 Inter-core Communication Strategy

Each core communicates with other cores through communication module. There are three inter-core communication strategies: mailbox mechanism, message queue based on memory sharing, and Fixed Access Unit (FAU) which supports atomic reading and writing. Considering the characteristics of hardware platform, we select mailbox mechanism to implement the inter-core communication. Every core can send messages to other cores through corresponding mailbox. The read-write operation is unblocked.

#### 3.2 Multi-core Task Allocation Strategy

Suppose that the multi-core processor has  $N_{core}$  cores,  $C_0, C_1, \dots, C_{N_{core}-1}$ , and  $M_{thread}$  threads are need to be allocated,  $T_0, T_1, \dots, T_{M_{thread}-1}$ .  $N_{thread}$  is the total number of threads that have been finished.  $taskgiven[N_{cores}]$  is the number of tasks that each core is allocated,  $taskfinish[N_{cores}]$  is the number of tasks that every core has finished,  $load[N_{cores}]$  is the load of each core,  $con[N_{cores}]$  is the congestion situation of each core. So the number of each core's left tasks can be computed by Formula (1).

$$taskcurrent[i] = taskgiven[i] - taskfinish[i] \quad (1)$$

When a new task comes, the master core updates  $M_{thread}$  plus one, then it collects data form *mailbox1-7* and store the data, in  $recvmsg[N_{cores}]$ , which is the finished tasks number of each slave core. We update  $taskfinish[i] = recvmsg[i]$ . Then we can compute  $taskcurrent[N_{cores}]$  by Formula (1).

Next we can allocate the coming task to a proper core. We consider the core's computing performance and the number of each core's unfinished tasks. We get a weight for each slave core. The smaller is the weight, the greater performance that the core has, for example, *MB\_1*' clock frequency is 100Mhz, while the *MB\_2*' clock frequency is 50Mhz, then we assume that *MB\_1*'s performance is twice of *MB\_2*, and store them in  $SW[N_{cores}]$ , thus  $SW[1]=1$ ,  $SW[2]=2$ .  $SW[N_{core}]$  is the weight value of each core. So the final weighted value for each core can be computed by  $order[i] = SW[i] * taskcurrent[i]$ . Then we get least  $order[i]$ , and  $i$  is the index of the core that the task should be allocated to.

## 4 Schedule Algorithm Design

### 4.1 Timed Interrupt Module

Each core has its own timer. When the interrupt happens, each slave core sends its own load situation to the master core. The master core reads data from corresponding mailbox and allocates task to adjust slave cores' loads. Interrupt module is implied as Algorithm 1.

<b>Algorithm 1: Timed interrupt algorithm</b>
<p><b>Begin</b>            Step 1. initialize the timer and interrupt controller;            Step 2. set the timer operating mode, load the initial value automatically;            Step 3. set the initial value;            Step 4. thread switch, resource allocation;            Step 5. check if timer's value equal 0, if so, go to step6, otherwise go to step4;            Step 6. slave cores calculate the load and send to the master core through mailbox; master core reads the mailbox and makes corresponding changes;            Step 7. the timer load value automatically , go to setp4;  <b>End</b></p>

### 4.2 Internal-Core Thread Switch Module

Slave cores have following threads: data dealing thread and inter-core communication thread. Data dealing thread is one of the most important parts of slave core threads. It is used for computing data and updating global variables.

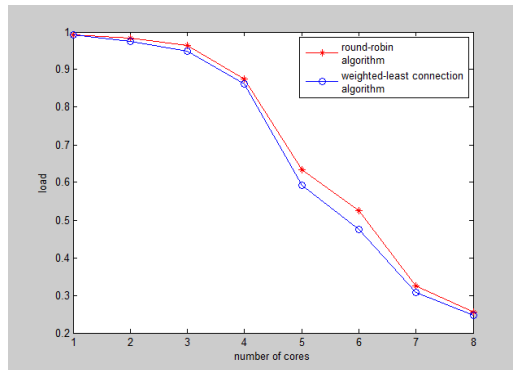
Data dealing thread is realized in Algorithm 2.

<b>Algorithm 2: Data dealing algorithm</b>
<p><b>Begin</b>            Step 1. read data from memory;            Step 2. get the data type label;            Step 3. if the label is 1, go to step4; otherwise, go to step5;            Step 4. execute the computing-intensive data-dealing thread;            Step 5. execute the computing-general data-dealing thread;            Step 6. update the global variable, <i>taskfinish</i>++;            Step 7. check if there is data in the memory, if so, go to step1; otherwise go to step8;            Step 8. update global variable <i>idle</i>=1;  <b>End</b></p>

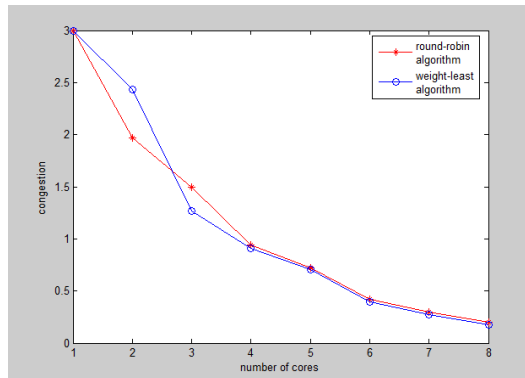
In the data-dealing thread, global variables, *taskfinish*, *taskcurrent*, *idle*, *end* are updated. All these variables' values should be sent to the master core in order to make proper task allocation.

## 5 Experiment Results and Performance Evaluation

The result of using ticks is shown in Figure 2. In the figure, horizontal axis is the number of cores that the processor has; vertical axis is the total ticks used to finish all tasks. The total number of simulated tasks is 300. From the figure, we can see that scheduling strategy is useless when the number of core is one. When the number of core is larger than one, the master core is responsible for the task allocation and the slave cores deal with data. The processing time decrease with the increase of core number. When the number of cores is two, there is no difference between the two strategies. We can find that the weight-least connection strategy takes fewer ticks to finish all tasks than the round-robin strategy when the number of core is more than two.



**Fig. 2.** The load situation when dealing with 300 tasks using two strategies



**Fig. 3.** The congestion situation dealing with 300 tasks using two strategies

Figure 3 is the congestion situation of two strategies. The total number of tasks is also 300. When the number of cores is less than two, the round-robin algorithm is better than weight-least connection algorithm. Because the feedback from slave cores increases the communication time consumption. However, weight-least connection strategy performs better than round-robin when the total number of cores is larger than two.

## 6 Conclusion

In this paper, we propose the WLC task schedule strategy for isomerism SoC. The strategy allocates the coming tasks according to the slave core's processing speed and the slave core's real-time processing situation. We build an eight-core EMSoC on Xilinx V5 platform. We do the experiment on Xilinx V5 platform and compare weighted least-connection task schedule strategy with round-robin task schedule strategy. Through simulation, we find that when the core number changes from one to two, because of the master-slave communication cost, processing speed decreases, but the load and congestion situation is improved. If the number of cores exceeds two, processing time becomes less with the increasing number of cores. And the load and congestion situation are also better.

**Acknowledgment.** This work was supported in part by Natural Science Foundation of P.R. China (Grant No. 61202443), and the Fundamental Research Funds for the Central Universities (No. DUT13JS07).

## References

- [1] Baklouti, M., Ammar, M., Marquet, P.: A model-driven based framework for rapid parallel SoC FPGA prototyping. In: Proceedings -2nd IEEE International Symposium on Rapid System Prototyping 2011, pp. 149–155 (May 2011)
- [2] Chen, J.M., Liu, C.N., Yang, J.K.: Parallel architecture core (PAC)-the first multicore application processor SoC in Taiwan part II: Application programming. *Journal of Signal Processing Systems* 2011 62(3), 383–402 (2011)
- [3] Flor, J.P.P., Muck, T.R., Frohlich, A.A.: High-level design and synthesis of a resource scheduler. In: 2011 18th IEEE International Conference on Electronics, Circuits, and Systems 2011, pp. 736–739 (December 2011)
- [4] Ravi, V.T., Agrawal, G.: Performance issues in parallelizing data-intensive applications on a multi-core cluster. In: 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid, pp. 308–315 (May 2009)
- [5] Yun, Y.N., Kim, J.B., Kim, N.D., Min, B.: Beyond UVM for practical SoC verification. In: 2011 International SoC Design Conference, pp. 158–162 (2011)
- [6] Huang, X., Liu, L., Li, Y.: FPGA verification methodology for SiSoC based SoC design. In: 2011 IEEE International Conference of Electron Devices and Solid-State Circuits (2011)

- [7] Zhang, E.Z., Jiang, Y., Shen, X.: The significance of CMP cache sharing on contemporary multithreaded applications. *IEEE Transactions on Parallel and Distributed Systems* 2012 23(8), 367–374 (2012)
- [8] Bechara, C., Berhault, A., Ventroux, N.: A small footprint interleaved multithreaded processor for embedded systems. In: 18th IEEE International Conference on Electronics, Circuits, and Systems 2011, pp. 685–690 (2011)



# A Power-Aware Scheduler Exploiting All Slacks under EDF Scheduling

Ye-In Seol<sup>1</sup>, Jeong-Uk Kim<sup>1</sup>, and Young-Kuk Kim<sup>2,\*</sup>

<sup>1</sup>Green Energy Institute, Sangmyung University, Seoul, South Korea  
sauntert1@gmail.com, jukim@smu.ac.kr

<sup>2</sup>Dept. of Computer Science & Engineering,  
Chungnam Nat'l University, Daejeon, South Korea  
ykim@cnu.ac.kr

**Abstract.** Power-aware scheduling reduces CPU energy consumption in hard real-time systems through dynamic voltage scaling(DVS). The basic idea of power-aware scheduling is to find slacks available to tasks and reduce CPU's frequency or lower its voltage using the found slacks. In this paper, we introduce a novel power-aware scheduling algorithm which exploits all slacks under preemptive early-deadline first scheduling. The simulation results show that proposed algorithm with the algorithmic complexity of  $O(n)$  reduces the energy consumption by 10-70% over the existing algorithms.

**Keywords:** real-time scheduling, power-aware scheduling, embedded systems.

## 1 Introduction

Energy consumption issues are becoming more important for mobile or battery-operated systems. Since the energy consumption of CMOS circuits, used in various microprocessors, has a quadratic dependency on the operating voltage( $E \propto V^2$ )[1], it is a very useful method for reducing energy consumption to lower the operating voltage of circuits. But, lowering the operating voltage also decreases its clock speed, so the execution times of tasks are prolonged. This makes problem more complex for hard real-time systems where timing constraints of tasks should be met. There has been significant research effort on Dynamic Voltage Scaling(DVS) for real-time systems to reduce energy consumption while satisfying the timing constraints[2,4-8].

DVS algorithms depend on the scheduling policy, task model, and processor architecture. In this paper, we adopt Early-Deadline First(EDF)[3] scheduling policy, periodic or sporadic task model and uniprocessor system. We present an algorithm which adopts the results of CC-EDF[5] and improves them. The simulation results show that the proposed algorithm reduces more energy consumption than previous work.

The rest of the paper is organized as follows. In section 2, we present the system model and notations adopted in this paper and introduce previous work which

---

\* Corresponding author.

motivates the work done in this paper. In section 3, we present a power-aware scheduling algorithm. In section 4, simulation results will be provided and section 5 will conclude and discuss the future directions of this paper.

## 2 Motivation

In this section we present the system model and introduce the result of the related work.

### 2.1 System Model

We consider preemptive hard real-time system in which all tasks are periodic or sporadic and mutually independent. The target processor is DVS enabled uniprocessor and its supply voltage and frequency are varied continuously between  $[v_{min}, v_{max}]$  and  $[f_{min}, f_{max}]$ , respectively. Let  $T=\{T_1, T_2, \dots, T_n\}$  be a set of periodic or sporadic tasks. Each task is represented as  $T_i = (P_i, C_i, D_i)$ , where

- $P_i$  is period for periodic task or minimum inter-arrival time for sporadic task;
- $C_i$  is work-case computation time for task  $T_i$  at the maximum frequency;
- $D_i$  is relative deadline of a task  $T_i$ .

If a instance or job of task  $T_i$  released at  $R_i$ , then its absolute deadline( $d_i$ ) is  $R_i + D_i$ . We will consider tasks only with  $D_i = P_i$ , so task  $T_i$  could be represented as  $(P_i, C_i)$ . Also the following notations will be used.

- $TU_i$  : the worst-case utilization of task  $T_i$  at the maximum frequency, i.e.,  $U_i=C_i/P_i$ .
- $TU$  : total utilization of all tasks in the system, i.e.,  $TU=\sum_i TU_i$ .
- $CC_i$  : task's actual computation time which should be less than  $C_i$ .
- $CU_i$  : actual utilization of task, i.e.,  $CU_i=CC_i /P_i$ .
- $CU$  : actual total utilization of system, i.e.,  $CU=\sum_i CU_i$
- $RC_i$  : task's remaining computation time.
- $\alpha$  : current frequency ratio, i.e.,  $f_{cur}/f_{max}$ .

### 2.2 Related Work

EDF scheduling has been extensively investigated in the area of real-time and power-aware scheduling[2-7]. While devising a new power-aware scheduling algorithm, we especially considered the results presented by Pillai and Shin[5]. They introduced a cycle-conserving method to real-time DVS. This method reduces the operating frequency on each task completion and increases on each task release. When a task completes its current invocation after using  $CC_i$  computation time, they treat the task as if its worst-case execution time were  $CC_i$ . So processor speed could be set as the actual total utilization  $CU$  which is always less than or equals to worst-case total utilization  $TU$ .

Mei et al.[4] integrated the above cycle-conserving method and the result of Qadi et al.[6] for sporadic task set. But, these method doesn't fully utilize the slack

generated. Let's see the following figure. If a task completed at  $t_c$ , then during the time interval  $[R_i, t_c]$  the system operated at higher frequency than required. This observation provides a clue to more slow down the processor when a task completes. We will show later that the amount of slack which could be used for lowering processor frequency is related with temporal idleness of the completed task.

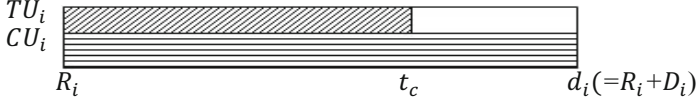


Fig. 1. CC-EDF Schedule and Its Available Slack

### 3 Power-Aware Scheduling Algorithm

As we stated at section 2, previous algorithms such as CC-EDF[5], DVSST[6] and CC-DVSST[4] don't fully utilize the slacks generated by early completed tasks. Before presenting more discussion, let's introduce new definition.

**Definition 1:** Temporal idleness  $TI_i(t)$  of a task  $T_i$  at time  $t$  is defined as following.

$$\left\{ \begin{array}{l} 0 \text{ until its completion and after its deadline.} \\ \frac{RC_i}{D_i - t_c} \text{ if it was completed at time } t_c \text{ and } t = t_c. \\ TI_i(t_c) \pm \gamma \text{ if } t > t_c. \end{array} \right.$$

The real value of  $\gamma$  depends on the status of system and how to calculate will be presented later. Following lemma computes the available slack at the fig.1 of CC-EDF and CC-DVSST algorithm.

**Lemma 1:** At CC-EDF or CC-DVSST scheduling, the amount of computation time exceeding its actual pace until its completion time(= $t_c$ ) is the same as  $[TI_i(t_c) - (TU_i - CU_i)] \times (D_i - t_c)$ .

$$\begin{aligned} \text{Proof: } [TI_i(t_c) - (TU_i - CU_i)] \times (D_i - t_c) &= [RC_i / (D_i - t_c) - C_i / D_i + CC_i / D_i] \times (D_i - t_c) \\ &= [RC_i D_i - C_i D_i + C_i t_c + CC_i D_i - CC_i t_c] / (D_i (D_i - t_c)) \times (D_i - t_c) \\ &= [(RC_i + CC_i) D_i - C_i D_i + C_i t_c - CC_i t_c] / D_i = [C_i t_c - CC_i t_c] / D_i \quad \because RC_i + CC_i = C_i \\ &= (C_i / D_i - CC_i / D_i) \times t_c = (TU_i - CU_i) \times t_c \quad \blacksquare \end{aligned}$$

Using Lemma 1, if a task  $T_i$  completed at time  $t$ , then we can slow down processing speed by amount of  $TI_i$  when executing lower priority tasks than  $T_i$  until its deadline, because CC-EDF or CC-DVSST slows down processing speed by the amount of  $(TU_i - CU_i)$ . The proposed algorithm tries to use slacks of already completed higher priority tasks which are necessarily generated by assuming worst-case execution scenarios and follows the result of CC-EDF when running task's priority is higher than those of already completed tasks. Also, if we cannot utilize those slacks by some reasons, i.e., when executing higher priority tasks or when slacks are too much to fully utilize, then we evenly distribute those unused slacks until corresponding deadlines.

One more consideration occurs when there is idle period. Let's consider periodic task set  $A = \{T_i = (P_i, C_i) \mid T_1 = (3, 1), T_2 = (3, 1), T_3 = (6, 2)\}$ . If  $T_{1,1}$  and  $T_{2,1}$  completes at

$t=1$  and  $t=2$ , respectively, and  $T_{3,1}$  completes early at  $t=2.5$ , then  $TI_3(2.5)=1.5/4.5=3/7$ . If we lower the processing speed as much as  $TI_3(2.5)$ , then actual processing capacity during  $t=[3,6)$  is  $(1-3/7)*3=12/7$  which is less than sum of WCET of  $T_{1,2}$  and  $T_{2,2}$ .

Deadline miss occurs because there is idle period  $t=[2.5,3)$ . During the idle period, the total processing capacity which should be processed under the actual execution scenarios is larger than sum of slacks used. So, we should reduce future slacks to compensate those mismatch. Following figure shows it.

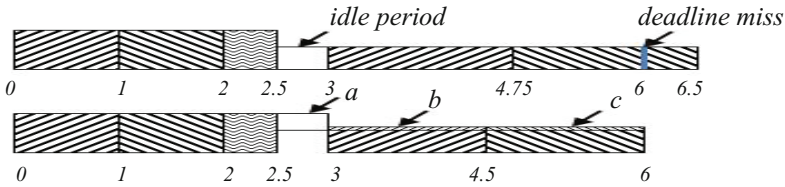


Fig. 2. Slack Reduction Example

```

when task  $T_i$  arrived,  $T_i$  completed and at deadline
if ( ex_flag) increase_temporal_idleness()
else if ( cpu was idle) decrease_temporal_idleness()
insert  $T_i$  into TC(completed) or delete(at deadline)
recompute  $TU = TU \pm C_i/P_i$ (when arrived or at deadline)
last_cpu_speed = compute_cpu_speed( $T_{cur}$ )
if there is no task to execute then set cpu as idle
else set cpu speed as last_cpu_speed
compute_cpu_speed( $T_j$ ) // $T_j$  is highest priority ready
task or null
cpu_speed = CU
for all tasks  $T_i$  at TC
  if  $d_i \leq d_j$  or  $T_j$  is null
    cpu_speed -= ( $T_{ii} - (T_{ui} - C_{ui})$ )
  if cpu_speed < 0 then ex_flag = 1; break
return (cpu_speed) or (0 if cpu_speed < 0)
decrease_temporal_idleness()
idle_work = idle_period * last_cpu_speed
for all tasks  $T_i$  at TC
  if  $T_{ii} > idle\_work$ 
    decrease  $T_{ii}$  as much as idle_work; idle_work = 0
  else
     $T_{ii} = 0$ ; idle_work -=  $T_{ii}$ 
increase_temporal_idleness()
for each unused slack of tasks
  distribute unused slack evenly until deadline

```

Fig. 3. Power-Aware Scheduling Algorithm

The amount of slack which should be reduced is  $TU - TI$ , and is the same as  $(CU - (TI - (TU - CU)))$ . At the above figure, the area 'a' is the same as 'b'+c', and if  $TI$  exceeds  $TU$ , we could save some slacks because only  $TU * (\text{length of idle period})$  should be processed by CPU.

## 4 Experimental Results

We evaluated our proposed algorithm using RTSIM[9] which is a real-time simulator. RTSIM can simulate the behaviors of dynamic voltage scaling algorithms as well as traditional real-time scheduling algorithms. In this simulation, it is assumed that a constant amount of energy is required for each cycle of operation at a given voltage. This quantum is scaled by the square of the operating voltage, consistent with energy dissipation in CMOS circuits ( $E \propto V^2$ )[1,2]. Only the energy consumed by CPU was computed and any other source of energy consumption was ignored. Also we do not consider preemption overheads, task switch overheads, and operating frequency change overheads. It is also assumed that the CPU consumes no energy during idle period and its operating frequency range is continuous at  $[f_{\min}=0, f_{\max}=1]$ . We compared our proposed algorithm with DVSST, and CC-DVSST. CC-EDF assumes periodic task model and CC-DVSST is direct result of CC-EDF. So we skipped for CC-EDF. DVSST and CC-DVSST assume sporadic task model, so we compared them at sporadic task system.

To evaluate the effect of number of tasks in the system, we generated 10 or 20 tasks for each comparison. Their periods or minimum inter arrival times are chose randomly in the interval [1-1000]ms. We divided task set into three groups to reflect more real environments. One group of tasks have short period in the interval [1-10]ms, another group of tasks have medium period in the interval [10-100]ms, and the last group of tasks have long period in the interval [100-1000]ms. The simulation also performed by varying the load ratio of tasks, i.e., the ratio of the actual computation time to the worst case computation time. For all simulations, the worst-case total utilization of system is always 1, i.e.,  $TU = 1$ . Following figure shows the simulation results.

Figure 4 shows the result for sporadic task system. In this case, our proposed algorithm outperforms both DVSST and CC-DVSST. For DVSST, the ratio of energy saving is up to 70% and for CC-DVSST, up to 10 %. The effect of number of tasks in the system could be also neglected, but the number of CPU frequency change of our algorithm was larger than that of DVSST and almost the same as that of CC-DVSST. So, the ratio of energy saving to DVSST could be shrinked. But our algorithm has huge performance gain to DVSST, so in spite of frequency change overheads, it is expected that our algorithm still outperforms to DVSST at real environments.

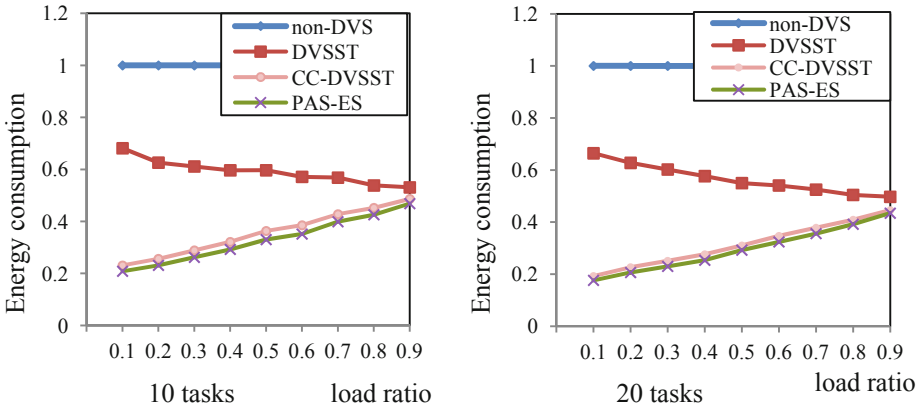


Fig. 4. Simulation Result for Sporadic Tasks

## 5 Conclusion

In this paper, we presented a power-aware scheduling algorithm for periodic and sporadic tasks. The proposed algorithm adopts the results of cycle conserving method(CC-EDF) and sporadic task scheduling(DVSST) and improves them. The simulation results show that this algorithm outperforms existing algorithms up to 10-70 % with respect to CPU energy saving.

In the future we would like to improve the proposed algorithm. This could be done if we assign all slacks generated by early completed higher priority tasks into the task of highest priority among uncompleted ready tasks instead of evenly distributing them until the ends of deadlines. This method may lower processor frequency much more than the proposed algorithm.

**Acknowledgement.** This work was supported by the Industrial Strategic Technology Development Program(10041740) funded by the Ministry of Trade, Industry and Energy(MOTIE, Korea).

## References

1. Burd, T.D., Brodersen, R.W.: Energy efficient CMOS microprocessor design. In: Proc. of Twenty-Eighth Hawaii Int'l Conf. on System Sciences, vol. 1 (1995)
2. Lee, C.H., Shin, K.G.: On-line dynamic voltage scaling for hard real-time systems using the EDF algorithm. In: Proc. of IEEE Int'l Real-Time Systems Symposium, pp. 319–327 (2004)
3. Liu, C.L., Layland, J.W.: Scheduling algorithms for multiprogramming in a hard real-time environment. J. ACM 20(1), 46–61 (1973)
4. Mei, J., Li, K., Hu, J., Yin, S., Sha, E.H.-M.: Energy-aware preemptive scheduling algorithm for sporadic tasks on DVS platform. Microprocessors & Microsystems 37, 99–112 (2013)

5. Pillai, P., Shin, K.G.: Real-time dynamic voltage scaling for low-power embedded operating systems. *ACM SIGOPS Operating System Review* 35(5), 89–102 (2001)
6. Qadi, A., Goddard, S., Farritor, S.: A dynamic voltage scaling algorithm for sporadic tasks. In: *Proc. of IEEE Int'l Real-Time Systems Symposium*, pp. 52–62 (2003)
7. Shin, D., Kim, J.: Dynamic voltage scaling of periodic and aperiodic tasks in priority-driven systems. In: *Proc. of the Asia and South Pacific Design Automation Conference*, pp. 653–658 (2004)
8. Yao, F., Demers, A., Shenker, S.: A scheduling model for reduced CPU energy. In: *Proc. of the IEEE Foundations of Computer Science*, pp. 374–382 (1995)
9. RTSIM:Real-Time system SIMulator, <http://rtsim.sssup.it>

# Network-Threatening Element Extraction and Quantification Framework

Sungmo Jung, Donghyun Kim, and Seoksoo Kim\*

Department of Multimedia, Hannam University, Daejeon 306-791, Korea  
{sungmoj, donghyunk1986}@gmail.com, sskim0123@naver.com

**Abstract.** APT attacks' forecasts and warning technologies are the most effective strategy to detect and analyze pre-attack indicators. Here, a study on a quantification framework for threatening conditions is necessary to being able to extract various network threatening elements. In this study, we collected data on security threatening elements to analyze network threatening elements. We analyzed the limitations of pre-technologies to analyze the quantification technologies of pre-threatening conditions. We also categorized their degrees of risks by grading threatening elements to deduce a method of graded degrees of risks.

**Keywords:** Quantification, Framework, Network, Security, Risk.

## 1 Introduction

With the advancements in information technology (IT), cyber threats have become a major factor in national security. Internal and external security agencies are cautioning users to take care in their self-evaluations when noticing symptoms or forecasting cyber-attacks [1].

In the event of a security breach, businesses face greater survival risks owing to their high level of dependence on information systems in their operational activities. Therefore, many companies have expressed an interest in the construction of a framework for minimizing the extent of damage caused by cyber-attacks. The use of forecasting and warning technologies is the most effective strategy to detect and analyze pre-attack indicators.

However, existing risk degrees are about to become enacted quantifications though the computation formula of risk degree. Therefore, security agents generally understand the necessity of a forecasting computation formula.

Therefore, a study on a quantification framework for threatening conditions is necessary to allow network-threatening elements to be extracted. We therefore carried out this pre-basic survey and study to investigate both network-threatening element extraction and a quantification framework.

---

\* Corresponding author.



## 2 Network-Threatening Element Extraction

Generally, network-threatening elements are categorized into malware such as worms and viruses, security vulnerabilities, and hacking techniques. Malware, in particular, is one of the most dangerous cyber-attacks factors [2].

Incipient malware is mostly in the form of viruses created to destroy specific systems. However, daily malware is usually in a social-engineering hacking form arousing the curiosity of the users, such as an APT attack [3], and is based on a continuous evolution

An advanced persistent threat (APT) attack applying a social engineering method has a high level of malleability and strategy compared with the distribution method of existing malware, and has a low probability of detection of less than ten% according to the attack features [4].

The phases of an APT attack [5] are as follows:

- ① Intelligence Gathering: Attack preparation with the target selection and requirement
- ① Point of Entry: Network penetration using a social engineering method
- ① Command & Control (C&C) Communication: Host observation and control
- ① Lateral Movement: Other host penetration, off-the-record information extortion, and authority securement
- ① Asset/Data Discovery: Primary server and service confirmation and interesting data correction using port scanning
- ① Data Exfiltration: Data sent to the penetration system and external system using encryption after correcting the primary data

The core of an APT attack is authorization acquisition through the extortion of a user’s account from a malware infection. We analyze the network-threatening elements by defining the first two phases of a six-phase APT attack.

For the quantification method, a risk computation formula for an enacted quantification owing to a confined period of technical security is used while

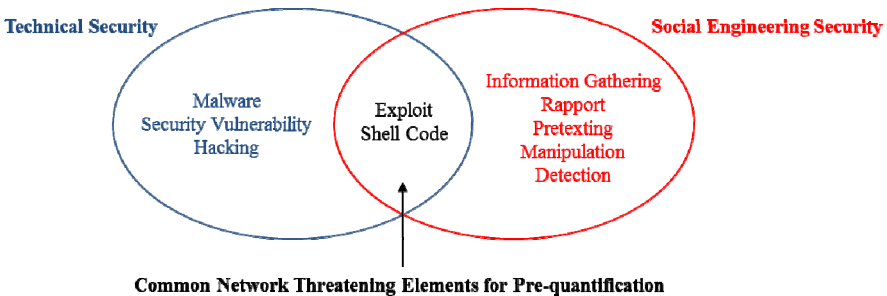


Fig. 1. Network-Threatening Elements

quantification from a forecasting perspective is available, only the number of attacks ( $G$ ) is known. This is because existing methods do not recognize different viewpoints regarding pre- and enacted quantifications.

In this study, we deal with the following network-threatening elements.

### 3 Categorization of Risk Degrees

The risk degree for a pre-quantification must consider the risk of threatening elements over time. We therefore have to consider the frequency of elements based on temporal characteristics focusing on the unique characteristics of the threatening and technical elements of social engineering.

$$\text{Pre-risks degree } (R) = \text{Risk possibility } (A) + \text{Situation frequency } (B) \tag{1}$$

The reason for classifying the frequency and possibility of risk for a pre-quantification is to coordinate the overall risk degree. Although risk possibility is lower than risk frequency, the overall risk degree will be relatively lower for formulating the hypothesis used in this study.

**Table 1.** Risks degree though risk possibility and frequency

(B) \ (A)	High	Average	Low
High	Upper	Upper	Middle
Average	Middle	Middle	Lower
Low	Lower	Lower	Lower

The risk frequency enumerates the threatening elements with a possibility of risk over time. The quantitative score is also arbitrarily calculated though a qualitative analysis of the risk possibility and frequency.

In this paper, we allocate each score as follows.

**Table 2.** Score categorization

Categorization	Elements	Detail	Risk Possibility	Situation Frequency	
Threatening Social Engineering Elements	Information Gathering (4)	Specific Information	2	4	
		System Information	1		
		Etc.	1		
	Rapport (10)	Cellular Phone	2	10	
		E-mail	2		
		P2P	2		
		Web-hard	2		
			Etc.	2	

**Table 2.** (continued)

	Pretexting (20)	Cellular Phone		2	20
		E-mail		3	
		P2P		3	
		Web-hard		3	
		USB		5	
		Etc.		4	
	Detection (30)	Cellular Phone		5	30
		E-mail		5	
		P2P		5	
		Web-hard		5	
		USB		5	
Etc.		5			
Threatening Technical Elements	Manipulation (36)	Exploit	White	1	15
			Black	4	
			Etc.	10	
	Shell Code	White	1	15	
		Black	4		
		Etc.	10		
	Etc. (Code)	Dormant	2	6	
		Self-remove	2		
		Driver	2		
<b>Total</b>				<b>Score A</b>	<b>Score B</b>

#### 4 Quantification Framework

The risk possibility and frequency are 25% and 75%, respectively, of the total pre-risk degree. The following formula can therefore be established.

$$\text{Risk possibility (A)} = \text{Score A} \cdot 0.25 \tag{2}$$

$$\text{Situation frequency (B)} = \text{Score B} \cdot 0.75$$

From this, the following formula for the pre-risk degree ( $R$ ) is used to consider the frequency of elements with temporal characteristics.

$$R = \sum_{i=1}^k A_i + \sum_{j=1}^n B_j \tag{3}$$

*A: Risk possibility*

*B: Frequency of occurrence*

*k: Number of detail threatening elements*

*n: Number of threatening elements*

## 5 Simulation

Based on the cyber crisis warning levels [6] obtained from a national cyber security center, a simulation of the evaluation results from stuxnet [7] was performed.

**Table 3.** Comparison of the evaluation results with the common vulnerability scoring system (CVSS) [8] and pre-risk degree

Standards	by CVSS	by the Proposed Method
Risk possibility (25)	11.50	24.25
Frequency of occurrence (75)	34.50	72.75
Pre-quantification risk degree	46.00	97.00
Expected warning level	Normal	Serious

## 6 Conclusion

This paper proposed a method to detect network-threatening elements and a quantification framework.

In the event of a security breach, businesses face increasing risk owing to a high level of dependence on information systems in their activities. However, many companies have to overcome this technical obstacle and spend considerable amounts of money to construct and operate a pre-warning framework. This study can provide the basic data on how to respond most effectively to an attack and prevent its proliferation by detecting and analyzing pre-attack symptoms.

In addition, since many companies require such a solution to detect network-threatening elements and a quantification framework, these results will be available for future research through technology transfers and patent registrations.

**Acknowledgement.** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2006026).

## References

1. Korea National Intelligence Service. White book (2012)
2. Robert Lemos, Counting the Cost of Slammer, CNet News.Com (2003)
3. Bodmer, Kilger, Carpenter, Jones: Reverse Deception: Organized Cyber Threat Counter-Exploitation. McGraw-Hill Osborne Media, New York (2012)
4. Tankard, C.: Advanced Persistent threats and how to monitor and deter them 2011(8), 16–19 (2011)
5. Trendmicro, Recent APT Attacks' Style and Solution, Cloudsec (2012)

6. National Cyber-Alert System, Vulnerability Summary for CVE-2010-2743,  
<http://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2010-2743>
7. Wikipedia, <http://en.wikipedia.org/wiki/Stuxnet>
8. A Complete Guide to the Common Vulnerability Scoring System Version 2.0,  
<http://www.first.org/cvss/cvss-guide.html>

# Document Clustering Based on a Weighted Exponential Measurement

Shahrooz Taheri, Alex Tze Hiang Sim, and Seyed Hamid Ghorashi

Department of Information Systems,  
Faculty of Computer Science and Information Systems,  
Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia  
{shahrooz64,h.ghorashi}@gmail.com, alex@utm.my

**Abstract.** Frequent terms sets clustering method has been proposed to overcome hardship of high dimensionality, and finding meaningful labels for clusters. Although this method provides meaningful labels for clusters, it has low accuracy. In this research, candidate clusters are extracted by mining frequent terms set within documents dataset. Each document is assigned to these clusters with considering the value of supports. A new similarity measurement function for clusters is designed based on similarity and weight of clusters and is proposed to remove unwanted clusters in a noise reduction step. The proposed method operates based on the concept of terms sets, value of support and weight of each cluster. Experimental results show that our proposed method provides more accurate clusters in comparison with previous efforts done on “Re0” and “Hitech” datasets.

**Keywords:** Clustering, Frequent Terms Set, Noise Reduction.

## 1 Introduction

Document clustering is an effective unsupervised classification approach in information retrieval systems for organizing and managing documents. [1], [2]. The main challenge in document clustering is to group documents into clusters that are either derived from data or predetermined by users. Although documents clustering can be considered as a general form of data clustering, it is much more complicated. It involves analyzing very high dimensionality data. Typically, text documents are also unstructured and involve conjugation [3]. Ironically, having clustered the documents, it is fundamentally challenging to assign meaningful labels to it.

To date, a number of clustering methods have been proposed taking the advantages of frequent terms set mining methods. The method was first proposed in [4] who labeled clusters using frequent terms found within documents. Experimental results suggested that this method is effective in labeling a cluster. A natural extension in this breed of research is to further improve the accuracy of clustering. In this paper, we proposed a novel function to remove unwanted candidate cluster. This function is based on weight and frequent patterns of each cluster. The subsequent sections of this paper are organized as follows: In Section 2, we review related literature. This was

followed by Section 3 that has been dedicated to present our approach. Experimental results are presented and analyzed in Section 4. Finally, we conclude and propose some future directions in Section 5.

## 2 Frequent Terms Sets Clustering Methods

Documents clustering based on the frequent term sets can be defined as a class of clustering methods that use frequent terms set miner and clustering module. These classes of documents clustering methods are designed to provide meaningful label for clusters and to reduce the dimension of document model. Concept of “similar documents share similar frequent terms sets” is considered as a basis for these methods. Typically, a frequent terms set miner is to extract candidate clusters by mining each frequent terms set within documents and a clustering module is to assign documents to the extracted candidate.

In term of reducing dimension and providing label, Frequent Terms Clustering (FTC) and Hierarchical Frequent Terms Clustering (HFTC) are basic methods in frequent term-set clustering which were proposed in [4]. They were designed to cluster documents based on reducing overlapping among clusters. Standard Overlap (SO) and Entropy Overlap (EO) were used as metrics to reduce overlaps.

In 2002, Hierarchical Frequent Term-based Clustering (FIHC) was proposed by [5] to provide scalability and increase accuracy. It also proves that using frequent terms set for clustering can reduce size of documents model. In this approach, document was assigned to candidate clusters based on the number of occurrences of terms. Moreover, this approach generated a cluster tree with respect to structures in frequent terms set in order to provide hierarchical clusters.

F<sup>2</sup>IHC is a novel method for clustering documents based on the frequent fuzzy term set, which was developed by [6]. In this method, fuzzy Apriori was used to mine fuzzy frequent terms sets and extract candidate clusters. In their approach, two matrixes were proposed in order to obtain similarity of document for candidate clusters, namely Document-Term Matrix (DTM) and Term-Cluster Matrix (TCM). They were designed to assign documents to candidate clusters. The former provides a model of dependency of documents to terms and the latter presents dependency of terms to extracted candidate clusters. In this method, hierarchical tree with regard to hierarchical structure of frequent terms set was generated in order to generate hierarchical clustering solution. This method takes the advantage of sibling merging clusters by considering the inner similarity in order to generate  $x$  number of clusters.

According to literature, many frequent terms set clustering methods have been proposed. They relatively tried to increase the accuracy. In order to increase accuracy, we proposed a novel method based on the concept of terms sets and new clusters measurement function, which was designed to increase accuracy based on the intuition. Results show that using support value and weight of each candidate clusters can provide clusters that are more accurate.

### 3 Proposed Method

In this study, required procedure for clustering documents is conducted in 5 general steps of (i) Preprocessing, (ii) Term set extraction, (iii) Document assignment, (iv) Pruning and (v) Noise Reduction. This section explains how the proposed method can be applied in order to generate clusters that are more accurate.

**(i) Preprocessing.** This method contains some general preprocessing which is involved with word extraction, stop words elimination, words stemming and words selection. In the proposed method, we use the matrix which comprises of term frequency and inverse document frequency [7] for selecting the terms. This metric can be defined as follows:

$$TFIDF(t_j, d_i) = \frac{f_{t_j, d_i}}{\sum_{j=1}^n f_{t_j, d_i}} * \log\left(\frac{|D|}{|\{d_i | t_j \in d_i, d_i \in D\}|}\right) \quad (1)$$

Where,  $TFIDF(t_j, d_i)$  presents importance of  $j$ th term ( $t_j$ ) in  $i$ -th document ( $d_i$ ). Where  $f_{t_j, d_i}$  is the frequency of term  $t_j$  in document  $d_i$ .  $|D|$  is the total number of documents in the document set  $D$ , and  $|\{d_i | t_j \in d_i, d_i \in D\}|$  is the number of documents that contains term  $t_j$ .

For elaboration proposes, the highest value of  $TFIDF$  is achieved when  $t_j$  occurs many times within a small number of documents. To the contrary, the lowest values are achieved when a term occurs in most of the documents. Theoretically, this function ignores terms that are too frequent and terms that are too few. These are less significant for the above given reasons.

**(ii) Term Set Extraction.** In the frequent terms set clustering method, using an appropriate frequent items sets miner is an important issue. One of the critical drawbacks of algorithms that generate candidate item set such as Apriori is related to execution time. Any generate-and-test algorithm wastes time in generating none relevant frequent items set that finally was removed. FP-Growth is a more suitable algorithm which omits the procedure to generating candidate items sets. Extracting frequent items sets by FP-Growth follows two basic procedures to generate and mine an FP-tree [8]. The later is a structure to represent a transactional database in a tree structure. We could omit the need to generate exponentially large number of candidate frequent items sets by traversing the tree for all frequent items sets.

A group of these frequent item sets form a candidate cluster. Each candidate cluster is later assigned with a document. Uniquely, we consider the maximum distance among clusters in measuring their similarity. This is explained again below.

**(iii) Assign Documents to Candidate Clusters.** We first find the similar candidate clusters and each document is assigned to these clusters. Consequently, empty candidate cluster suggests there are more than one candidate clusters that are similar to a specific document. In this situation, among the similar candidate clusters, the one with higher value of support will be selected as the most similar candidate cluster following our policy, and a document will be assigned to it.



Using larger support values as a policy for selecting a candidate cluster can be described as: “every cluster’s center with bigger value of support is more valid than the other centers” because it is frequent and has been shared among many documents.

**(iv) Pruning.** Candidate clusters pruning can be defined as, removing vacant clusters. By assigning the membership value of documents to candidate clusters, based on the pervious section, some candidates will be vacant; it means that there is not any documents exist for assigning to these clusters. Removing these vacant clusters reduces the total number of output clusters and provides more appropriate input data for the next noise reduction step.

**(v) Noise Reduction.** In frequent terms sets clustering methods, mining terms sets with a low threshold value would generate too many candidate clusters. We employ a novel noise reduction procedure based on complete linkage algorithm in order to merge similar candidate clusters with respect to their weight. In our approach, clusters with smaller number documents are merged. On the contrary clusters with larger number of documents are kept. In order to implement this policy, we modified a similarity function [9] as shown in Equation (2) as follows:

$$\text{Similarity}(C_i, C_j) = \exp \frac{-\text{Distance}(C_i, C_j)}{\beta} \quad (2)$$

We define,

$$\text{Distance}(C_i, C_j) = |C_i - C_j| * \exp^{(C_i.\text{weight} + C_j.\text{weight})} \quad (3)$$

Where,

$$C_i.\text{weight} = \text{number of documents assign to cluster } i \quad (4)$$

$$\beta = \sum_{i=0}^n |C_i - \bar{C}| / n \quad (5)$$

$$\bar{C} = \sum_{i=0}^n C_i / n \quad (6)$$

In this study, the weight of each cluster is defined by the number of documents that it contains, and  $C_i$  is the  $i_{\text{th}}$  candidate cluster.  $\beta$  is used to describe a sample riance.  $\bar{C}$  is an average of all candidate cluster which in this research is presented using vector space.

Based on this formula, cluster that is weighted more is an autonomous cluster compare to lighter clusters. With this similarity function, a complete linkage algorithm selects two most similar clusters and merges them together as a new cluster where the distance of heavier candidate are now further compare to distance of lighter cluster. This procedure continues on new clusters until termination condition is met.

## 4 Experimental Results

This section is presented to provide results and to compare the proposed method with pervious works. Sets of experimental have been applied in order to evaluate the

effectiveness of the proposed method. All testes have been done on Fedora 15 on Core2Dual CPU with 4G RAM. For comparing with other work, two standard datasets of ‘Re0’ and ‘Hitech’ have been used as primitive dataset. The dataset can be downloaded from “<http://glaros.dtc.umn.edu/gkhome/fetch/sw/cluto/datasets.tar.gz>”. An Overall F-measure was proposed in [10] and is used as metric for evaluating the proposed method. It measured the quality of a clustering result, C using the weighted sum of maximum F-measures for all natural classes according to [5].

In this section, we compare experimental results from the proposed method with others work as reported in [6]. Based on the same settings of minimum support, which is 0.04, 0.05 and 0.06, we tested our method and selected average of overall F-measure in different threshold as the accuracy of our proposed method.

**Table 1.** Comapre Results Based on Overall F-measure

Dataset	# of clusters	Proposed method	F <sup>2</sup> IHC	FIHC	UPGMA	Bi. <i>k</i> -means
<i>Hitech</i>	3	0.54	0.47	0.48	0.33	0.54
	15	0.50	0.47	0.45	0.33	0.44
	30	0.50	0.48	0.46	0.47	0.29
	60	0.47	0.45	0.42	0.40	0.21
Average		0.50	0.47	0.45	0.38	0.37
<i>Re0</i>	3	0.59	0.55	0.40	0.36	0.34
	15	0.50	0.54	0.41	0.47	0.38
	30	0.38	0.54	0.38	0.42	0.38
	60	0.36	0.54	0.40	0.34	0.28
Average		0.46	0.54	0.40	0.40	0.34

Table 1 illustrates the overall F-measure values for the proposed method and others work which was reported by [6] for four different cluster numbers: 3, 15, 30 and 60.

In ‘Hitech’ dataset and for all four clusters, our method scores 0.54, 0.50, 0.50 and 0.46 and provides 14%, 6%, 4.1% and 4.5% improvement in comparison with outputs of F<sup>2</sup>IHC. For ‘Re0’ dataset, the most improvement was achieved on three clusters with 7% improvement compare to the current best method F<sup>2</sup>IHC. The proposed method does not provide further improvements in 15, 30 and 60 number of clusters but performs better compare to FIHC, UPGMA and Bi. K-means in general.

These improvements can be justified via concepts of centers selection and centers merging. In this research, centers that provide bigger value of support have been selected as candidate clusters. In addition, the proposed method uses weight of candidate clusters as shown in Equation (2) and Equation (3) to merge most appropriate pair of candidates in the noise reduction step as read in Section 3.5. Through these procedures, lighter clusters are noise and are dissolved by heavier candidate clusters.

## 5 Conclusion

Although numerous studies have been done on documents clustering, accuracy of document clustering methods is still a challenge that needs to be improved. This study proposed a novel clustering approach which integrates a new similarity function to remove unwanted candidate clusters thus the increase in accuracy. We are experimenting a different measure that further enhances our experimental results.

**Acknowledgment.** The author A.T.H. Sim would like to acknowledge in received of Research University Grant (vot. 07J79).

## References

1. Luo, C., Li, Y., Chung, S.M.: Text document clustering based on neighbors. *Data & Knowledge Engineering* 68(11), 1271–1288 (2009)
2. Jain, A.K.: Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31(8), 651–666 (2010)
3. Tan, A.H.: Text Mining: The state of the art and the challenges. In: *Pacific Asia Conf. on Knowledge Discovery and Data Mining* (1999)
4. Beil, F., Ester, M., Xu, X.: Frequent term-based text clustering. *ACM* (2002)
5. Fung, B.C.M.: Hierarchical document clustering using frequent itemsets. Simon Fraser University (2003)
6. Chen, C.L., Tseng, F.S.C., Liang, T.: Mining fuzzy frequent itemsets for hierarchical document clustering. *Information Processing & Management* 46(2), 193–211 (2010)
7. Salton, G.: *The SMART Retrieval System—Experiments in Automatic Document Retrieval*. Prentice Hall Inc., Englewood Cliffs (1971)
8. Han, J., et al.: Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery* 8(1), 53–87 (2004)
9. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences* 8(3), 199–249 (1975)
10. Zhao, Y., Karypis, G.: Empirical and theoretical comparisons of selected criterion functions for document clustering. *Machine Learning* 55(3), 311–331 (2004)
11. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. *Communications of the ACM* 18(11), 613–620 (1975)

# A Robust Feature Selection Method for Classification of Cognitive States with fMRI Data

Luu-Ngoc Do and Hyung-Jeong Yang\*

Department of Electronics and Computer Engineering  
Chonnam National University  
Gwangju, South Korea  
doluungoc@gmail.com, hjyang@jnu.ac.kr

**Abstract.** The functional magnetic resonance imaging (fMRI) technique is a powerful imaging tool for analyzing the brain activity by localizing patterns of activity related to specific mental processes. Recently, researchers have started to solve the inverse problem of detecting the cognitive states at a particular point of time by applying the multi-voxel pattern classification approach. Since fMRI data are high dimensional, extremely sparse and noisy, feature selection is a key challenge in this kind of approach. In this paper, we propose a new method for selecting the most informative features from fMRI data. By computing Fisher Discriminant Ratio, we can identify the most active voxels from several Regions of Interest. These active voxels are considered as the most powerful discriminative features. We investigated the performance of this method by classifying the human's cognitive states of "observing a picture" versus "reading a sentence". The experimental results showed that our method achieved the highest accuracy compared to other feature selection methods with the Gaussian Naïve Bayes (GNB) classifier. The average accuracy of six human subjects is approximately 96.45%.

**Keywords:** fMRI, Regions of Interest, feature selection, Fisher Discriminant Ratio, active voxel.

## 1 Introduction

Neuroimaging is a key process to access the human's cognitive states through brain activation. In this domain, the functional magnetic resonance imaging (fMRI) has the promise of achieving good performance for studying human cognitive processes. fMRI technique is commonly performed using blood oxygenation level-dependent (BOLD) contrast to local changes in deoxyhemoglobin concentration in a brain [1]. Based on the diamagnetic property of oxygenated hemoglobin and paramagnetic property of deoxygenated hemoglobin, BOLD signals can construct the measurements of the activation of brain for generating three-dimensional images. Each image consists of a number of uniformly spaced volume elements, called voxels. If some parts of the brain are activated by mental states, the voxel's intensity will be changed.

---

\* Corresponding author.

Tracking the variation of this value across time can get the knowledge of the brain's activation.

There are two main approaches for analyzing fMRI data: localization and classification. In the localization approach, researchers have to identify which regions of the brain are activated when a human performs a particular cognitive function. In contrast, many other researchers are interested in mapping from fMRI data to the human subject's cognitive states [2-7] which are known as classification problem. In this approach, a machine learning classifier is necessary to automatically detect the subject's cognitive state at a single time interval. This classifier needs support from a powerful feature selection method to deal with the issue of extremely high dimensional, sparse and noisy data. For example, in our case, we encountered problems where the examples are described by more than 80,000 features and we have only several dozens of examples per class. Selecting the most informative features from fMRI data will not only improve the accuracy of classifier but also reducing the processing time effectively.

In this paper, we propose a new method for selecting the most informative features from fMRI data. The main idea of our proposed method is to extract the most active voxels from the most active Regions of Interest (ROIs). These active voxels are considered as the most powerful discriminative features for the cognitive state classification. The activity of voxels was measured by the Fisher Discriminant Ratio. We performed classification of the human's cognitive states of "observing a picture" versus "reading a sentence". By using Gaussian Naïve Bayes classifier, our system achieved the average accuracy approximately 96.45%.

The remainder of this paper is organized as follows. Section 2 describes approaches of others researchers and their achievements. Section 3 discusses about the proposed method. The data description and experimental results are detailed in Section 4. Section 5 is our conclusion.

## 2 Related Work

Since fMRI data is high dimensional, dimensionality reduction is typically performed before classification to improve the performance of system. The Principle Component Analysis (PCA) and the Regions of Interest (ROIs) are two of the most well-known feature selection methods in the fMRI data analysis system. PCA is a conventional dimensionality reduction method that has been proved to be a powerful approach for many kinds of data mining analysis. However, when the dimension of original data is much higher than the number of observations which is the common case of fMRI data, PCA cannot achieve a good performance. Therefore, many researchers tried to apply the enhanced version of PCA or the combination of PCA and other techniques. T. Hoang et al. used incremental PCA (iPCA) to develop an incremental subspace tracking for reducing computation and storage requirements [3]. Y. Fan et al. applied PCA after extracting regional features which are formed by statistical information [4].

Brain regions that are relevant to the problem under study must first be selected from a background of brain activity [4]. Since only some specific regions of the brain are activated when a mental state is performed, selecting the features from restricted

Regions of Interest (ROIs) is a good approach. Etzel et al. tried to solve classification problem of fMRI by using anatomical ROIs [5]. They selected all voxels containing in some specific ROIs to evaluate the performance of anatomical ROIs-based fMRI classification approach. Mitchell et al. and S. Bapi et al. selected the most active voxels per ROIs [6,8]. They applied t-test for each voxel of each target class to measure the power of active voxels.

Instead of using standard classifiers such as Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), Bernard et al. proposed a new group of classifiers, called Generalized Sparse Classifiers (GSC) to alleviate the over-fitting problem of standard classifiers [7]. They constructed a Spatial-Smooth Sparse Linear Discriminative Analysis (SSLDA) classifier to demonstrate the power of GSC. SSLDA has good performance for the Picture versus Sentence study provided by Mitchell [8].

### 3 Proposed Method

#### 3.1 Regions of Interest

In this paper, we followed Mitchell et al. [8] to mark up the brain with 25 anatomical Regions of Interest (ROIs). They included: calcarine sulcus (CALC), dorsolateral prefrontal cortex – left & right (LDLPFC, RDLPFC), frontal eye fields – left & right (LFEF, RFEF), left inferior frontal gyrus (LIFG), inferior parietal lobule – left & right (LIPL, RIPL), intraparietal sulcus – left & right (LIPS, RIPS), opercularis – left & right (LOPER, ROPER), posterior precentral sulcus – left & right (LPPREC, RPPREC), supramarginal gyrus – left & right (LSGA, RSGA), superior parietal lobule – left & right (LSPL, RSPL), temporal lobe –left & right (LT, RT), triangularis – left & right (LTRIA, RTRIA), supplementary motor areas (SMA), inferior temporal lobule – left & right (LIT, RIT). A structural image which captures the static physical brain structure at high resolution was used to identify the anatomical regions of interest, using the parcellation scheme of Caviness and Rademacher [9].

#### 3.2 Voxel Activity

The most common approach for feature selection when training classifiers is select the features that best distinguish the target classes. In fMRI data, not only the data that describes the mental states can be considered but also the data of non-state can be considered too. This kind of data is corresponding to the fixation or rest condition which contains the data observed when the subject is generally at rest. In this paper, we follow Mitchell's definition [8] for voxel discriminability and voxel activity:

- Voxel discriminability indicates how well the feature distinguishes class 1 from class 2.
- Voxel activity indicates how well the feature distinguishes class 1 or class 2 from the zero signal class.

By using the voxel activity, instead of selecting the features that best describe the correlation between-classes, we can choose features that best describe the mental state. Figure 1 illustrates the definition of voxel activity and voxel discriminability. In this paper, we measure the voxel activity by computing the Fisher Discriminant Ratio for every feature.

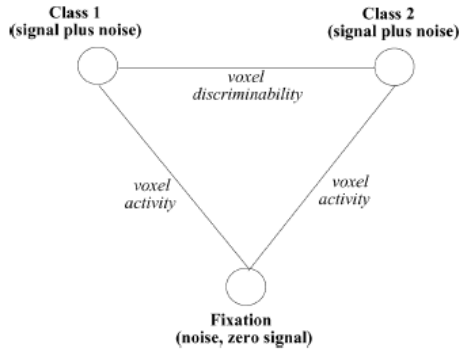


Fig. 1. Voxel discriminability and voxel activity [8]

## 4 Experimental Result

### 4.1 Data Description

We used the StarPlus data collected by Mitchell et al. for validation [8]. This data was conducted from six human subjects who were shown a sequence of sentences and a simple picture. We have to train a classifier to distinguish whether the subject is seeing a picture or reading a sentence. The pictures simply were geometric arrangements of symbols such as +, \*, @, \$, --. For each subject, we trained a classifier of the form:

$$f : \text{fMRI-sequence}(t, t+8) \rightarrow \{\text{Class1}, \text{Class2}\}$$

where  $t$  is the starting time of stimulus. Signals in the interval of 8 seconds are considered to avoid lacking the brain activity.

Each subject includes 80 samples, 40 samples for each label. Table 1 shows the number of all features of each subject. Generally, each sample includes approximately 80,000 features.

Table 1. Data Description

Subject ID	04799	04820	04847	05675	05680	05710
Number of features	79184	80240	75168	82160	80992	74144

## 4.2 Evaluation

We firstly choose the most active ROIs by considering the performance of each region in our study. In our experiments, a set of ROIs: {CALC, LIPL, LT, LTRIA, LOPER, LIPS, LDLPFC} produced the best accuracy for our studies. We then extract the most active voxels from these ROIs to build the classifier.

For evaluating the classification performance, we applied k-fold cross validation with  $k = 10$ . The average accuracy was computed and compared to other methods such as all features, iPCA [3], and ROIs. Our proposed method will be described as ROIs+FDR with 250 features selected and performed by using GNB classifier. Table 2 shows the classification accuracy of our method for each human subject and the comparison with other methods. For all subjects, our proposed method had classification accuracy much higher than the others. Table 3 shows the comparison between the proposed method and the method of using voxel discriminability with the same measurement and classifier which are FDR and GNB. The voxel activity provides high accuracy for all of subjects while the voxel discriminability cannot achieve a good performance for subject 04820 and 05680.

**Table 2.** Classification result of single subject

<i>Feature Selection</i>	04799	04820	04847	05675	05680	05710
All features	56.75%	57.5%	75%	58.75%	67.5%	70%
ROIs	61.5%	70%	97.5%	75%	80%	80%
iPCA	80%	80%	90%	88.75%	78.75%	85%
ROIs+FDR	92.5%	97.5%	100%	98.75%	95%	95%

**Table 3.** Voxel Discriminability & Voxel activity

	04799	04820	04847	05675	05680	05710
Voxel Discriminability	91.25%	77.5%	98.75%	90%	85%	92.5%
Voxel Activity	92.5%	97.5%	100%	98.75%	95%	95%

## 5 Conclusion

In this paper, we have presented a new approach for classifying specific cognitive states of a single subject from fMRI data. By selecting the most active features with highest FDR values from several ROIs, we achieved a set of the most powerful discriminative features. The experimental results showed that our proposed method had a good performance with the highest accuracy compared to the other ones. In the



future, we will apply this method to another kind of study and dataset to solve the problem of classifying human cognitive states. We will also extend to the problem of detecting multiple-subject cognitive states.

**Acknowledgements.** This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA (National IT Industry Promotion Agency). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2012-0007810).

## References

1. Lindquist, M.A.: The Statistical Analysis of fMRI Data. *Statistical Science* 28, 439–464 (2008)
2. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fMRI) “brain reading”: Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 19, 261–270 (2003)
3. Hoang, M.T.T., Won, Y.G., Yang, H.J.: Cognitive States Detection in fMRI Data Analysis using incremental PCA. In: ICCSA, pp. 335–341 (2007)
4. Yong, F., Shen, D., Davatzikos, C.: Detecting Cognitive States from fMRI Images by Machine Learning and Multivariate Classification. In: *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop* (2006)
5. Etzel, J.A., Gazzola, V., Keysers, C.: An introduction to anatomical ROI-based fMRI classification analysis. *Brain Research* 1282, 114–125 (2009)
6. Bapi, R.S., Singh, V., Miyapuram, K.P.: Detection of Cognitive States from fMRI data using Machine Learning Techniques. In: *IJCAI*, pp. 587–592 (2007)
7. Ng, B., Vahdat, A., Hamarneh, G., Abugharbieh, R.: Generalized Sparse Classifiers for Decoding Cognitive States in fMRI. In: Wang, F., Yan, P., Suzuki, K., Shen, D. (eds.) *MLMI 2010. LNCS*, vol. 6357, pp. 108–115. Springer, Heidelberg (2010)
8. Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S.: Learning to decode Cognitive States from Brain Images. *Machine Learning* 57, 145–175 (2004)
9. Rademacher, J., Galaburda, A.M., Kennedy, D.N., Filipek, P.A., Caviness, V.S.: Human cerebral cortex: Localization, parcellation, and morphometry with magnetic resonance imaging. *Journal of Cognitive Neuroscience* 4, 352–374 (1992)

# Adaptive Networking for Continuous and Reliable Data Delivery in Wireless Sensor Networks

Jae-Joon Lee, Jinsuk Kang, and Jaesung Lim

Ajou University, Suwon, Korea  
jjnlee, @ajou.ac.kr

**Abstract.** A rapid increase in personal mobile communication devices and wireless communication capability of electronic devices in home automations and security systems will in turn increase interference among communication devices. Especially, interference source with mobility can significantly affect data delivery among wireless communication devices, which deteriorates service in the system. In this paper, we propose an Adaptive Networking for Continuous and Reliable data Delivery (ANCORD) scheme that guarantees lossless and reliable data delivery by employing the adaptive networking method.

**Keywords:** Wireless sensor network, mobile, interference, adaptive networking.

## 1 Introduction

Home automation systems and home security use multiple wireless sensor devices that communicate with each other autonomously to exchange information on detection and control [1-3]. In addition, use of mobile communication devices and machine-to-machine (M2M) or Internet of Things (IoT) devices is rapidly increasing. Consequently, communication among hand-held devices, consumer electronic devices, home automation and security monitoring systems, M2M/IoT devices will be widely required for various services including surveillance or automation in home, office, factory, or battlefield. Users carry smart communication devices to access data from the network at any time and place. Especially in security monitoring systems that consist of multiple sensor devices, reliable data delivery in required time is important to provide the desired service. However, communication interference can significantly deteriorate data delivery, which results in degradation of service quality.

Diverse communication devices use the same frequency channels including Bluetooth, 802.15.4 devices, or 802.11, which causes interference [4-6]. As the number of smart communication devices and M2M/IoT devices for security monitoring or other services increases, interference among these communication devices becomes a significant factor. Interference and jamming result in data loss and interruptions on service [7-9]. In the case of security monitoring systems equipped with sensor communication devices, unreliable data delivery and interruptions significantly degrade service. When interference sources are mobile, the effect on the networks becomes much significant because data delivery path can continuously experience interference due to mobility of the interference source.

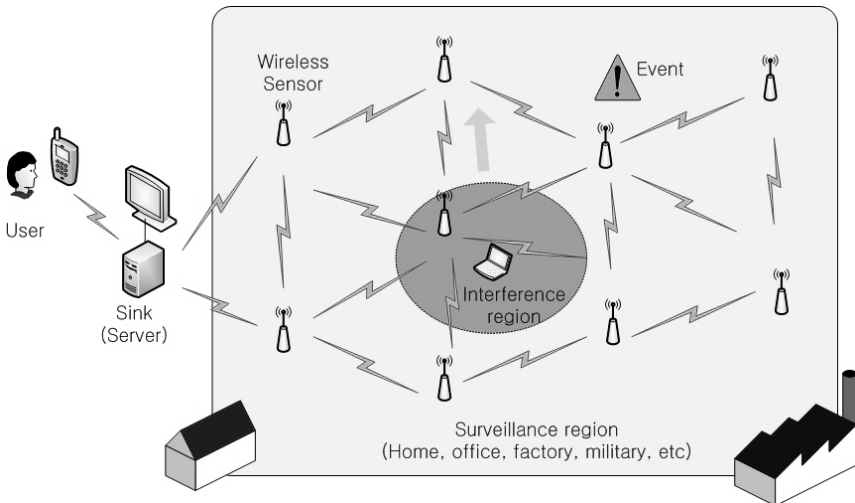
In this paper, to solve the problem of mobile interference in wireless networks that consist of multiple sensor nodes for security monitoring or military applications, we present an Adaptive Networking scheme for Continuous and Reliable data Delivery (ANCORD). ANCORD provides the reliable data delivery by avoiding the nodes that are experiencing interference or likely to be interfered. In addition, ANCORD reduces the route reconstruction processes, which is another important aspect in wireless sensor communication devices including IoT/M2M devices that use limited energy resources.

The rest of the paper is organized as follows. Section 2 presents the proposed algorithm and the procedure of the adaptive scheme. Section 3 presents the simulation results and Section 4 concludes the paper.

## 2 Adaptive Networking Scheme against Mobile Interference

### 2.1 Wireless Sensor Network and Mobile Interference

Fig. 1 illustrates the surveillance system that utilizes a wireless sensor network and the effect of interference with mobility. The purpose of this system is to provide security of a target area and notify any event to a user through wireless sensor networking system. The event can be transmitted through multi-hop wireless links to a sink that collects data and informs to the user. When the network constructs network topology and delivers events to the sink through wireless links, a routing protocol should be able to recover reliable data delivery paths if problem occurs in the existing paths by reconstructing a detour that avoids links or nodes experiencing any problem such as



**Fig. 1.** Illustration of the surveillance system with wireless sensor network and mobile interference

interference. But, when the interference source is mobile, repeated problem of data loss and route destruction becomes more severe. If the interference source moves toward the reconstructed route, it results in data loss and new route should be established again. This kind of data loss and route construction process also consume energy, which results in faster depletion of energy especially in wireless sensor networks that consist of communication devices with limited battery energy.

## 2.2 Adaptive Networking Scheme for Continuous and Reliable Data Delivery

ANCORD uses two mobility characteristics of interference source to achieve continuous and reliable data communication, avoiding any interruption to application service in mobile environment. The overall procedure of the adaptive networking scheme is as follows. First, when a node detects interference, its neighbors are notified through messages that notify the occurrence of interference similar to the method in [9]. Second, when a node receives a notification message about interference, it obtains the distance from the interference region and its velocity. Third, the risk level of the node is calculated; this is the main part of ANCORD. Finally, the cost of the node is obtained from the risk level, and the path that provides the minimum cost is selected for routing.

The proximity of node  $i$  is obtained as  $d_i = \min_{j \in J} dist_{ij}$ , where  $J$  is the set of nodes that are affected by the interference and  $dist_{ij}$  is the distance between nodes  $i$  and  $j$ . Then, each node calculates the risk level based on the proximity at time step  $t$  as follows:

$$r_{i,i}(t) = d_i(t)^{-\alpha}, \quad (1)$$

where  $r_{i,i}(t)$  is the risk level of node  $i$  at time step  $t$  based on the proximity, and  $\alpha$  is greater than equal to 1. As  $\alpha$  increases, the effect of proximity increases. Thus, as the distance between the interference region and the node decreases, the risk level based on the proximity increases.

If a node is located in the path of an interference region or a node is moving toward the interference region, then the chance of being affected by the interference source increases. On the other hand, a node in the direction opposite to the interference region has a lower chance of being interfered with. Each node obtains the direction of the interference region by checking the change in the proximity. When the distance from the interference region increases, risk will decrease.

Each node examines the proximity at each time step and evaluates the velocity as  $v_i(t) = \frac{d_i(t-1) - d_i(t)}{\Delta t}$ . This information contains both speed and direction. The absolute value indicates the speed. The plus and minus sign indicates whether the interference source is approaching or moving away. We note that the risk level estimation of direction needs to identify only whether the interference source is inbound or outbound. For obtaining the risk level based on direction, we use the Exponentially Weighted Moving Average, which is widely used for predicting a trend and smooth short-term fluctuation.

$$r_{2,i}(t) = \beta v_i(t)^\gamma - (1 - \beta)r_{2,i}(t-1) \quad (2)$$

$\beta$  is a positive value less than 1 and  $\gamma$  is greater than or equal to 1.

The risk levels are used to obtain the cost of the node. This cost value is applied to the cost of the inbound links of the node. The cost of the inbound links of a node  $i$  is as follows.

$$c_i(t) = \sum_k w_k r_{k,i}(t) \quad (3)$$

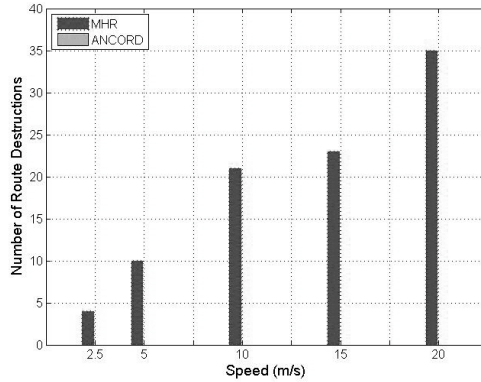
$w_k$  is a weight of the risk level  $r_k$  and determines the relative effect of each risk level. For example, if an interference source tends to move in a certain direction, the weight of the risk level based on direction is set to high. On the other hand, when a movement is unpredictable, the weight of the proximity-based risk level should increase. Basically, the proximity-based risk level has much higher weight than the direction-based risk level. The obtained inbound link costs of each node are used to select the reliable path for data delivery.

### 3 Simulation Results

We evaluate our proposed scheme, ANCORD, and compare it with a minimum hop routing (MHR) scheme. The main objective of the simulation is to identify the significant impact of interference on data delivery paths determined by routing selection schemes in the face of mobile interference source and to verify the resilience of the paths selected by ANCORD.

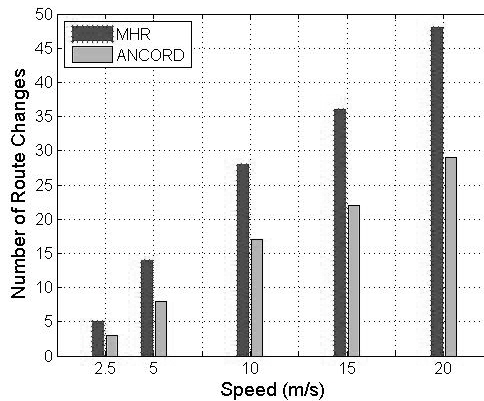
In this simulation, we deploy 25 nodes in a grid topology. A sender and a receiver are located at each side of the network. The distance between adjacent nodes is set as 30 m, and the radio radius is set to allow a node to communicate directly with 4 adjacent nodes. An interference source moves between the sender and the receiver via the middle of the network. The interference radius is set as 40m.

Fig. 2 shows the number of route destructions when the interference source moves with constant speed and direction. The results indicate the number of nodes in the data delivery paths that experience interference owing to the mobile jammer during simulations. We examine five different speeds and each simulation runtime is 60 s. As the speed of the mobile interference increases, the number of route destructions increases in the case of the minimum hop routing scheme. However, in the case of ANCORD, no nodes experience interference during data delivery throughout simulation regardless of the speed of the interference source. Since the recovered route can be repeatedly affected by the mobile interference in the case of the MHR scheme, significant route destructions are observed. However, ANCORD can provide reliable routes that are far from the interference source, thereby preventing any interruptions and losses due to the interference.



**Fig. 2.** Comparison of route destructions

The number of route changes is presented in Fig. 3. Route re-establishments occur when nodes in a routing path are affected by interference or when a more optimal path based on a routing algorithm is found. The MHR scheme results in a greater number of route changes, which is almost double that of ANCORD owing to route destructions. In ANCORD, as the interference source moves, the data delivery path is changed to provide a more reliable route by examining the distance from the interference region and the direction of its movement. However, route changes are much less when using ANCORD than when using the other scheme, because no destruction occurs and the data delivery path is much more stable. In addition, more route changes require more network resources, an important factor in mobile communication devices with limited resources.



**Fig. 3.** Comparison of route changes

## 4 Conclusion

With the increased use of personal mobile communication devices and the enhanced communication capability of sensor devices, as well as home automation and security systems, wireless networking systems needs to provide a dynamically adaptive networking mechanism when interference occurs. The impact of interference on routing is significant particularly in mobile environments, which can result in continual service interruptions and data losses. To solve this problem, we have presented ANCORD, a method for providing reliable data delivery paths adaptive to environments. By selecting more reliable paths, ANCORD prevents route destructions as well as significantly reduces route reconstructions when interference occurs in mobile environments.

**Acknowledgements.** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2009569) and (2011-0015329), and also by MSIP under the ITRC supervised by the NIPA (NIPA-2013-(H0301-13-2003)).

## References

1. Gomez, C., Paradells, J.: Wireless home automation networks: A survey of architectures and technologies. *IEEE Communications Magazine* 48(6), 92–101 (2010)
2. Chiang, S., Huang, C., Chang, K.: A Minimum Hop Routing Protocol for Home Security Systems Using Wireless Sensor Networks. *IEEE Transactions on Consumer Electronics* 53(4), 1483–1489 (2007)
3. Song, G., Wei, Z., Zhang, W., Song, A.: A Hybrid Sensor Network System for Home Monitoring Applications. *IEEE Transactions on Consumer Electronics* 53(4), 1434–1439 (2007)
4. Shin, S.Y., Park, H.S., Choi, S., Kwon, W.H.: Packet Error Rate Analysis of ZigBee Under WLAN and Bluetooth Interferences. *IEEE Transactions on Wireless Communications* 6(8), 2825–2830 (2007)
5. Srinivasan, K., Dutta, P., Tavakoli, A., Levis, P.: An empirical study of low-power wireless. *ACM Transactions on Sensor Networks (TOSN)* 6(2), 1–49 (2010)
6. Kim, J., Kwon, Y.: Interference-aware topology control for low rate wireless personal area networks. *IEEE Transactions on Consumer Electronics* 55(1), 97–104 (2009)
7. Bayraktaroglu, E., King, C., Liu, X., Noubir, G., Rajaraman, R., Thapa, B.: On the performance of IEEE 802.11 under jamming. In: *Proc. IEEE Infocom*, pp. 1265–1273 (April 2008)
8. Xu, W., Ma, K., Trappe, W., Zhang, Y.: Jamming sensor networks: attack and defense strategies. *IEEE Network* 20, 41–47 (2006)
9. Wood, A.D., Stankovic, J.A.: Denial of service in sensor networks. *IEEE Computer* 35(10), 54–62 (2002)

# Model Transformation for Cyber Physical Systems

Shuguang Feng and Lichen Zhang

Shanghai Key Laboratory of Trustworthy Computing, East China Normal University  
Shanghai, China 200062  
zhanglichen1962@163.com

**Abstract.** Cyber Physical Systems(CPS) are composed with discrete and continuous dynamics. Traditional modeling techniques can't implement the requirement of modeling CPS. One way of solving this is to model CPS parts with different techniques and translate them into a uniform model. The approach of integrating Modelica with AADL is a suitable choice. AADL is a modeling language aimed at modeling the system architecture and check the consistency in discrete time. Modelica has emerged as a standard for modeling the dynamics of cyber physical system and verify the discrete events with differential algebraic equation systems. Combining the descriptive feature of AADL models and the equation-based power of Modelica forms a more powerful way of Modeling CPS. The model transformation from AADL to Modelica provides an efficient way to unify the CPS model that helps to verify the properties of the whole model.

**Keywords:** model transformation, AADL, Modelica, Cyber Physical Systems.

## 1 Introduction

CPS is becoming more popular and important in modern days. It has both discrete and dynamic behavior. CPS is not a new concept and has been used in many domains for example: automotive embedded systems. With the development of the techniques and requirements, new functionalities has been proposed and extend the scope. [1, 2] The computing and communication requirements are major issues to be solved. This requires a comprehensive integrated modeling framework for modeling of architecture and tracing their relationships. The current technique can't modeling both hardware part and software part or can't check the consistency and continuous dynamics simultaneously.[3]

In order to meet the challenge of cyber-physical system design, a new modeling technique is needed. The new modeling technique can both model the discrete physical part and the continuous dynamics.[4] Among the existing modeling techniques, AADL[5] can model the architecture and verify the consistency and Modelica[6] is a multi-domain modeling language that is suitable for physical systems and versifying the continuous dynamics. The integrate Modelica with AADL is a suitable choice to build cyber physical systems. AADL is designed for modeling system architecture. Society of Automotive Engineers (SAE) released the AADL in November 2004. AADL can build the system architecture and analysis the time property, reliability, efficiency and some other properties.[7] Modelica is a multi-domain modeling



language. This language is proposed in 1997 with a group of international efforts. Modelica is object-oriented, acasual-modeling and equation-based language. With object-oriented property, model reusability can be improved.[8] It can describe the physical world in a direct way. In the period of model checking, Modelica can analyze the properties of this model in continuous time.[9] In comparison with AADL, Modelica can check model properties in continuous time, while AADL can't model the system and check the properties in this way. AADL is mainly aimed at modeling the architecture of the system. In a system architecture, there are many components interacting. The consistency among components can be checked and verified to meet the requirements. We can implement the visualization of a component model with Modelica. The transformation from AADL into Modelica is needed and developed that will support implementations to transfer efficiently the modeling information between AADL and Modelica models without ambiguity.

In this paper, we propose an approach to transform the models of AADL into the models of Modelica to model Cyber Physical System. The precise transformation principles are stated and clarified. Based on this transformation mechanism.

## 2 AADL and Modelica Features

AADL is aimed at modeling system architecture. Modelica is aimed at modeling the physical model. Many Cyber Physical Systems applications are system-of-systems, integrating various mechanical, electronic, and information technology system. Cyber Physical Systems should be multi-domain. The Model-Driven Engineering has been put into use in many domains. Cyber Physical Systems can also use the principle of Model-Driven Engineering to build models. By integrating AADL and Modelica, we combine the formal language for differential algebraic equations and discrete events of Modelica with AADL constructs for requirements, structural decomposition, logical behavior and corresponding cross-cutting constructs.

### 2.1 AADL Features

AADL is used to build the system architecture. We can use AADL to model the Cyber Physical Systems architecture. The main element of AADL is component. These components are mainly divided into three layers: application software, execution platform, composite.

### 2.2 Modelica Feature

Modelica is designed to model the physical world. It is object-oriented modeling, acasual-modeling and equation-based modeling. Modelica can check the model properties in continuous time with visualization.

The basic element of Modelica is class. There are eight classes in Modelica: class, model, record, block, function, connector, type, package. The physical world can be modeled with Modelica in a direct way.[10]

### 3 Cyber Physical Systems Modeled with AADL and Modelica

Cyber Physical Systems should have both discrete and continuous behaviors. The discrete part can be modeled with AADL, and the continuous behavior can be modeled with Modelica.

#### 3.1 Model Transformation from AADL into Modelica

We design this transformation from AADL into Modelica to model and check a complete Cyber Physical System. The following are the principles of transforming AADL into Modelica.

##### 3.1.1 Inheritance Transformation

AADL has a mechanism of object-oriented modeling technique. Modelica is object-oriented. So, there is the inheritance transformation from AADL components into Modelica classes. In table 3, the inheritance relationship transformation is listed.

**Table 1.** Inheritance transformation from AADL to Modelica

Label	Modelica class relationship	AADL components relationship	Description
1			In AADL, the <i>Type</i> component define features and <i>Implement</i> component implements <i>Type</i> component
2			The relationship of component extension can be represented with Modelica class extension.
3			AADL component instantiation can be represented by Modelica class instantiation.

##### 3.1.2 Transform AADL Components into Modelica Classes

Components can be projected into Modelica classes. After the comparing AADL components and Modelica classes. The projection table is listed. In table 4, AADL components can be transformed into Modelica classes.

**Table 2.** AADL components to Modelica classes

Label	AADL components	Modelica classes
1	thread	model
2	thread group	class
3	process	model
4	data	record & type
5	subprogram	model
6	processor	model
7	memory	model
8	device	model
9	bus	connector
10	system	model

### 3.1.3 AADL Keywords Transformation

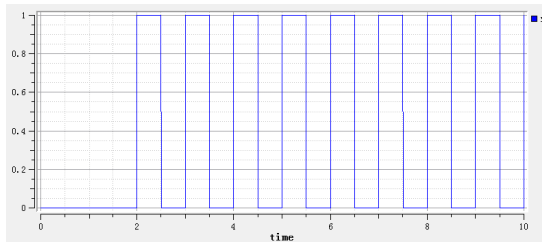
The keywords in a modeling language is unique. The AADL model can transfer into Modelica in order to meet the transformation requirement. The transformation table is in Table 5.

**Table 3.** AADL keywords to Modelica keywords

Label	AADL keywords	Modelica keywords
1	extends	extends
2	feature	parameter
3	flows	connect
4	properties	equation
5	packages	packages
6	implementation	extends
7	port	connector
8	connections	connect
9	subcomponents	Declared as variables
10	modes	Transformed as comments in Modelica

## 3.2 Transformation Case Study

A control\_process is modeled with AADL. To analyze the working state of this process, we transform control\_process into Modelica. Figure 1 is the result of analyzing control\_process with Modelica.



**Fig. 1.** The result of transforming AADL into Modelica

## 4 Conclusion

Cyber Physical Systems are becoming more popular and important. Cyber Physical Systems are composed with discrete and continuous behaviors of a system. AADL can be used to construct Cyber Physical System architecture. Modelica can be used to model the physical components of Cyber Physical System and verify the continuous dynamic behavior. By integrating AADL and Modelica, the properties of AADL components can be checked with Modelica tool, which validates the security and consistency of an architecture.

**Acknowledgment.** This work is supported by national high technology research and development program of China (No.2011AA010101), national basic research

program of China (No.2011CB302904), the national science foundation of China under grant No.61173046, No.61021004, No.61061130541), doctoral program foundation of institutions of higher education of China (No. 200802690018), national science foundation of Guangdong province under grant No.S2011010004905.

## References

1. Rajkumar, R.R., et al.: Cyber-physical systems: the next computing revolution. In: Proceedings of the 47th Design Automation Conference. ACM (2010)
2. Baheti, R., Gill, H.: Cyber-physical systems. *The Impact of Control Technology*, 161–166 (2011)
3. Broy, M.: Challenges in modeling cyber-physical systems. In: IPSN 2013, pp. 5–6 (2013)
4. Eidson, J.C., et al.: Distributed Real-Time Software for Cyber-Physical Systems. *Proceedings of the IEEE* 100(1), 45–59 (2012)
5. Feiler, P.H., Gluch, D.P., Hudak, J.J.: The architecture analysis & design language (AADL): An introduction. No.CMU/SEI-2006-TN-011, Carnegie-Mellon Univ. Pittsburgh Pa Software Engineering Inst. (2006)
6. Fritzson, P., Engelson, V.: Modelica—A unified object-oriented language for system modeling and simulation. In: Jul, E. (ed.) ECOOP 1998. LNCS, vol. 1445, pp. 67–90. Springer, Heidelberg (1998)
7. Feiler, P.H., Lewis, B.A., Vestal, S.: The SAE Architecture Analysis & Design Language (AADL) a standard for engineering performance critical systems. In: *Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*. IEEE (2006)
8. Modelica Association. Modelica: A Unified Object- Oriented Language for Physical Systems Modeling: Language Specification Version 3.0 (September 2007), <http://www.modelica.org>
9. Modelica - a unified object-oriented language for physical systems modelling. Language specification. Technical report, Modelica Association (2002)
10. Fritzson, P.: Principles of object-oriented modeling and simulation with Modelica 2.1. Wiley. com (2010)

# Load Balancing Improvement Methods in Video Conferencing Based on H.323 Standard

Vu Thanh Nguyen<sup>1</sup>, Huynh Tuan Anh<sup>1</sup>, and Vu Thanh Hien<sup>2</sup>

<sup>1</sup> University of Information Technology, Vietnam National University, HCM City, Vietnam  
{nguyenvt, anhht}@uit.edu.vn

<sup>2</sup> University of Foreign Languages – Information Technology, HCM City, Vietnam  
hienvt2000@yahoo.com

**Abstract.** H.323 standard provides a foundation for audio communications, video, and other data through IP-based networks, including the Internet. By following H.323, products and multimedia applications from multiple manufacturers can interoperate with each other, allowing users to communicate with each other without worrying about interoperability. While the H.323 is continuously revised, however, still many important issues need to be further discussed. In particular, bandwidth management, address translation and intelligent routing, which are the main functions of an H.323 gatekeeper. In fact, the call routing methods used in the current version of H.323 is very basic and will cause an imbalance to keep the gatekeeper. This article will propose a method named Multiple-Registration with Single Database (MRSD). It also describes the design principles and discuss the effectiveness of this method is based on the performance of the system is implemented.

**Keywords:** load balancing, H.323 gatekeeper, Multiple-Registration with Single Database (MRSD).

## 1 Introduction

Video conferencing is one of the most complex communication systems. If a gatekeeper is only used to process all the requirements for your system, then certainly that the load of the gatekeeper can't afford to meet all requirements of extremely large. The increasing of traffic and then upgrading hardware upgrade are no longer economically and more efficient. With the aim to create a dynamic gatekeeper, more gatekeepers will need to be added to split the load between the groups of gatekeepers, still known as the gatekeeper cluster. The distribution of load between the gatekeepers is called load balancing.

H.323 is a recommendation of the International Telecommunications Union (International Telecommunication Union, ITU for short). While the H.323 is continuously revised, however, still many important issues need to be further discussed. In particular, bandwidth management, address translation and intelligent routing, which are the main functions of an H.323 gatekeeper. To solve the problem of load balancing between the gatekeepers, the two methods improved call routing proposed in [1] which

is Multiple-Registration with Multicast Location Request (MRML) and Multiple-Registration with Hierarchical Database (MRHD), and introduce methods Multiple-Registration with Single Database (MRSD) proposed by the research team. This paper will describe the design principles and discuss the effectiveness of these two approaches based on the performance of the system is implemented, and is based on two metrics call blocking rate and effectiveness of channels to assess the load balancing of the system, the results indicated that both metrics have improved on the system is implemented.

## 2 Methods for Improving Load Balancing

### 2.1 The Problem of Load Balancing

The primitive call routing method of H.323 will cause load imbalance between the gatekeepers. One of the reasons is that H.323 only allows endpoints register with a gatekeeper, but does not provide any mechanism to load balance. Therefore, if the endpoints are not shared evenly between the gatekeepers, the gatekeeper will load imbalance.

Second, the problem of load balancing will occur even though the endpoint is shared evenly between the gatekeepers. For example, suppose that both the GK11 and GK12 had two endpoints registered on MAN1. However, if the endpoint has been registered by GK11 making calls more than the endpoint registered GK12, the load of the GK11 will be heavier than the GK12. Moreover, the gatekeepers may vary in computing power, storage capacity and bandwidth connection. If the endpoint is not shared suitable with the gatekeeper capabilities, load balancing problems also exist.

### 2.2 MRSD

The improved routing method for load balancing based on the MRSD. MRSD method registers the endpoints as MRHD method. This method uses a single database and proposes an architecture extension for the purpose of address translation and load

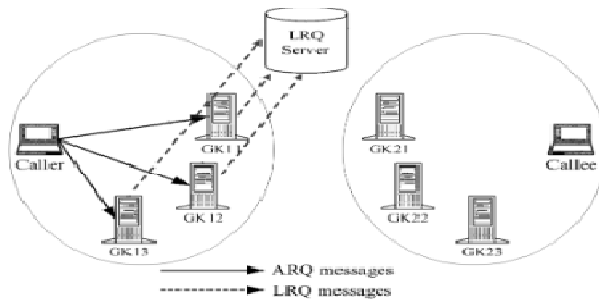
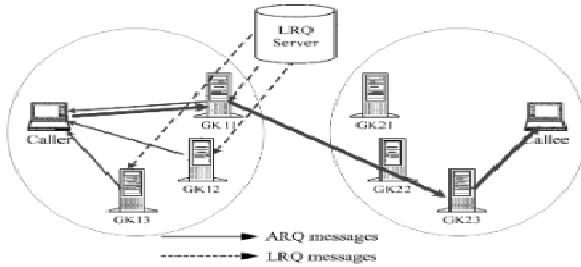


Fig. 1. Scenario of admission request and LRQ processes in MRSD method

balancing. This database includes all of the registration information of the gatekeepers registered in the area. When one of the endpoints is trying to register with the gatekeeper, the gatekeeper will keep a version of the registration information in this database. Similarly, when the endpoint is trying to unregister to the gatekeeper, the gatekeeper will not only delete the registration information in the database. Moreover, to provide load balancing mechanism, LRQ server also monitors the load condition of each gatekeeper. Whenever a message is sent to the LRQ server, load condition of the gatekeepers will be updated.



**Fig. 2.** Scenarios of LCF and admission confirm processes in the MRSD method

The MRSD method is implemented as scenarios illustrated in Figure 2 and Figure 3. At the start of a call, the caller will send the ARQ message to all registered gatekeeper (eg, GK11, GK12, GK13). To get the address of the callee the gatekeeper and load conditions is called, the calling gatekeeper GK11, GK12, and GK13 sends LRQ messages for LRQ Server. Server LRQ will get the address of the callee and load conditions of the called gatekeeper. LRQ Server sends back the LCF message with the load condition of the called gatekeeper to GK11, GK12, GK13. So the calling gatekeeper will know the lightest load gatekeeper and route calls to that gatekeeper later. Finally, GK11, GK12, GK13 will send a ACF message with load conditions to the caller, the callee will identify the gatekeeper of the caller side which is the lightest load gatekeeper then set the call through this gatekeeper. Assuming that is the gatekeeper GK11 and GK23 lightest corresponding side call and be called. AAs a result, the call setup message Q.931 is sent from the caller, routed through GK11 and GK23, and then forwarded to the callee, as shown in Figure 3. The MRSD algorithm is summarized as follows:

- MRSD algorithm:
  - Start calling, caller send ARQ message to the gatekeepers GK11, GK12, GK13.
  - The gatekeepers GK11, GK12, GK13 send back LRQ messages for LRQ Server.
  - LRQ Server send queries retrieve the address of the called and load conditions gatekeeper GK21, GK22, GK23.
  - After receiving feedback, LRQ Server will return LCF messages with load conditions of the called gatekeepers GK21, GK22, GK23 to the gatekeepers GK11, GK12, GK13.
  - GK11, GK12, GK13 identify the called gatekeeper (assuming the GK23) has lightest load conditions and will route the call to the GK23.

- GK11, GK12, GK13 send ACF message to the caller with load condition of the lightest gatekeeper (assuming GK11).
- Finally, Q.931 call setup message is sent from the caller, routing to GK11 and GK23, then will be transferred to the callee.

### 3 Implementation and Experimental Results

For testing, our team had implemented a video conferencing system based on the H.323 standard open source software. we installed the three algorithms MRML, MRHD, MRSD and integrated it into the system for improving load balancing. The goal of this testing is showing the influence of three load balancing methods on call blocking ratio and channel utilization.

#### 3.1 Testing Model

Currently there are many open source projects to build conference systems following H.323 standard. Each system based on the H.323 standard but have different scale, along with the additional features that make a difference. However, for the experimental subject, the research team has selected and implemented a simple system, basic and versatile. This system ensures the functions of the H.323 system, at the same time it's easy to install, set up specifications, and as a development easily platform to develop scale of system at large and small levels different.

H323plus Project, before is OpenH323, with the goal is developing a fully functional installation, open source (MPL license) of the H.323 standard. My Phone is a IP phone software, also known as softphone. GNU Gatekeeper is an open source project implementing an H.323 gatekeeper. GNU Gatekeeper is also developing on H323Plus library. Talkez is free software supporting Internet telephony, conferencing with both H.323 and SIP. On last time, Talkez is open source software, but currently it is only a free software. System components for testing:

- One or two PCs play role as a server to control system that are installed GNU Gatekeeper and MCU Talkez. This can be installed on the same a PC or individual.
- Two more PC play role as a client installed MyPhone.
- The practical test model consists of three PC that is installed Windows 7 operating system, are connected together via LAN:
  - One PC is also a server and a client, is installed these software: GNU Gatekeeper, MCU Talkez, My Phone.
  - Two PC is two client will be installed My Phone.

#### 3.2 Experimentation and Evaluation Results

To compare the call blocking rate in the cases exist and not exist the load balancing methods and check the call blocking rate when the value of time average of two setup time is as increasing as. Call blocking rate is calculated as follows



$$\text{Call blocking ratio} = \frac{\text{number of call setup rejects}}{\text{number of call setup attempts}}$$

Call blocking rate in cases that using of load balancing methods are generally smaller than different values of the average time of two consecutive calls. Specifically, the average improvement of call blocking rate was 34.2% . The reason is that there is no load balancing method, the probability of overloading on some gatekeeper will rise and will increase the call blocking rate. However, with the load balancing method, the overall load will be shared equally to each gatekeeper, so traffic will be smooth. Thus, the call blocking rate can be reduced.

To evaluate the effectiveness of the using the channel in cases with the load balancing methods, we will check the using of the channel when the value of the average time of two consecutive setup the call. The using channel efficiency is calculated as the meaning channel using in the number of channels available accounts.

$$\text{Channel utilization} = \frac{\text{mean number of channels occupied}}{\text{total number of available channels}}$$

It can be seen that the using channel is implemented better than with load balancing methods, especially when traffic is high (for example, the average time of two consecutive call setup is short). Specifically, the average improvement of using channel is 13.1%. The reason is that the call blocking rate is reduced by load balancing method, so the using of channels increases respectively.

Besides, the research team implemented the experiments processing use in video-conferencing version with two members, do not use the audio capture device to test the quality of the image is transmitted when using the system. Each meeting cost 2 minutes is tested with load balancing experimental above. With:

- Part A: Use of video source from webcam with contents fewer changes.
- Part B: use of fake video source was created from the MyPhone software with content frequently changing.

The system is experimented with a resolution of CIF and QCIF. In each conference, the implementing team took the bit rate data that is transmitted over the average of the conference of each participant using WVMT codec [2].Each of testing is implemented three times and the final result count by the average of three test results.

The following table is the experimental results:

Resolution	Bit rate A (kb/s)	Bit rate B (kb/s)	Updating image speed
QCIF	66.3	282.4	moderate
CIF	114.8	597.2	moderate

From the results table we have the following comments:

- First, the bit rate of part A is always lower than part B, that is show that the system is more effective with the video not more changing from webcam.
- Second, we are easy to see the quality of images obtained from the system was decreasing, and a few places the phenomenon of interference. However, with this level of quality is still acceptable.

- Third, the bit rate of the conference sessions to collect data from the webcam (side A) from 66.3 kb/ s to 114.8 kb/ s of the LAN connection is 10Mb/ s and the ADSL upload speed from 256 kb/ s to 512 kb/ s, download speed can reach over 1MB/ s.
- Result Table of Call blocking rate and effectiveness of channels

Method	Call blocking rate	Channel utilization
MRML	34.3%	13.1%
MRHD	40.2%	15.2%
MRSD	42.7%	18.4%

In conclusion, after the experiment, the system shows effective to meet the demand for video conferencing applications on the LAN and ADSL internet connection. Good image quality, call blocking rate and effectiveness of channels has been improved by those methods.

## 4 Conclusion

Based on the research facility systems and multimedia communications technology, the article has been counted the basic knowledge important to deploy successfully on open source H.323 video conferencing system. We also introduced three methods to improve load balancing in the video conferencing base on H.323 standard. For quality control, these algorithms of MRML, MRHD and MRSD were implemented in open source H.323 system. The result shows the efficiency from three methods. We can be seen as a first step towards the foundation for a fascinating researching following to applications that is building the application system for conferencing with good image quality and low cost.

**Acknowledgment.** This research is funded by Vietnam National University HoChi-Minh City (VNU-HCM) under grant number B2012-26-04.

## References

1. Chang, C.-Y., Chen, M.-S., Huang, P.-H.: An H.323 Gatekeeper Prototype: Design, Implementation, and Performance Analysis. *IEEE Transactions on Multimedia* 6(6), 936–946 (2004)
2. Vu, T.N., Mai, H.H.T.: Motion-Based Video Transferring Algorithm. In: *International Conference on Computer Technology and Development*, 647–651 (2011)
3. *Introduction To Multimedia Conferencing*, Computer Science Department. University College London, London (1998)
4. ITU Radiocommunication Sector. Parameter values for the HDTV standards for production and international exchange Programme (2008)
5. Tri, N.M., Tuan, N.T.: Research and build software systems to support online teaching on the internet / intranet. Thesis, Faculty of Information Technology, University of Natural Sciences, Ho Chi Minh City (2007)

# A Combination of Clonal Selection Algorithm and Artificial Neural Networks for Virus Detection

Vu Thanh Nguyen, Nguyen Phuong Anh, Mai Trong Khang,  
Nguyen Hoang Ngan, Nguyen Quoc Thai, and Nguyen Trong Quoc

University of Information Technology, Vietnam National University, HCM City, Vietnam

**Abstract.** In this paper, we proposed a new approach using bio-inspired algorithms such as Clonal Selection Algorithm (CLONALG) and Artificial Neural Networks (ANNs) which aims to handle virus detection problem. The point of difference is using ANNs as the detectors and CLONALG as the algorithm for finding the best ANN's structure and weights. According to experimental results, the proposed model has an acceptable detection rate and false positive rate.

**Keywords:** Artificial Immune System (AIS), Artificial Neural Network (ANN), Clonal Selection Algorithm (CLONALG), Virus Detection System (VDS).

## 1 Introduction

In recent years, virus recognition and elimination become critical and compulsory problems. However, the development of anti-virus systems is complicated because of the continuous change in virus signatures and attacking methods. Corresponding to the increase in the number of virus, traditional data-based detecting methods reduce efficiency while behavior-based methods produce high false positive rates. Recently, there are many attempts using AIS and ANNs in constructing new VDS.

AIS is a field of study devoted to the development of computational models based on the principles of the biological immune system (BIS). It is an emerging area explores and employs different immunological mechanisms to solve computational problems.

Clonal selection algorithms are developed based on the clonal selection theory (Burnet, 1959) [1] proposed nearly 50 years ago, which are mainly inspired by B cells' response to antigens. Clonal selection algorithms [2] [3] however, are very similar to a kind of evolutionary algorithm; namely, evolutionary strategies. Clonal selection algorithms are also population-based search and optimization algorithms generating a memory pool of suitable antibodies for solving a particular problem.

In 1943, Warren McCulloch and Walter Pitts proposed the ANNs [4], which is an information processing paradigm-mimicking nervous system. It has been used in many fields, such as robotics, control, recognizing and making predictions.

In this paper, we proposed a new approach in resolving Virus Detection problem by combining Clonal Selection Algorithms and Artificial Neural Networks.

## 2 Related Work

In 1995, Anastasia Dumas et al used ANNs for recognition and classification computer viruses [5]. They implemented ANNs to recognize a set of attributes of system activity which implies the presence of specific viruses.

In 2009, Rui Chao and Ying Tan proposed a virus detection system based on AIS [6]. In their model, they used NSA and Clonal Selection in detector generation.

In the same year, Essam Al Daoud [7] introduced an AIS-based Metamorphic Viruses Detection model. In this model, they construct a multilayer immune system; each layer uses an AIS-based algorithm for self/nonself discrimination.

In 2010, Vladimir Golovko et al [8] propose an integration of ANN and AIS. They described an attack detection system design based on artificial immune network.

Recently, Suha Afaneh and Raed Abu Zita [9] developed an algorithm based on AIS for detecting viruses, called Virus Detection Clonal algorithm.

Despite still having some limits, these researches brought us to a new kind of approach that opens a new prospect to further researches.

## 3 Proposed Approach

### 3.1 Ideas

Assume that we have a set of strings, each string represents a virus data fragment and we need to discriminate between these strings and other benign strings. With CLONALG, antibody and antigen are represented in strings. In our approach, each antibody is an ANN and characterized by 2 strings, the first represents the ANN's structure and the second is a list of weights. By this way, we can use a smaller number of antibodies as detectors than using string representation while getting a better result in coverage rate. In addition, each antigen is a cluster of similar virus fragments.

### 3.2 Objects

Beside ordinary objects in CLONALG-based system, we introduce some new concepts that make the algorithm easy to understand and model.

- Antibody: an ANN encoded in strings and has an antigenic affinity value.
- Antigen: a cluster of similar virus fragments.
- Memory set: population of memory antibodies.
- Population: others antibodies.
- Total set: contains all the antibodies in the memory set and population.
- Clonal set: clonal of n highest affinity antibodies selected from total set.
- Matured set: all the antibodies in clonal set will be suffered affinity maturation.
- Environment: a population of antigens and benign set.
- Epoch: a process that converts the current stage of system into new stage.
- Fitness function: used for calculate the antigenic affinity of an antibody.

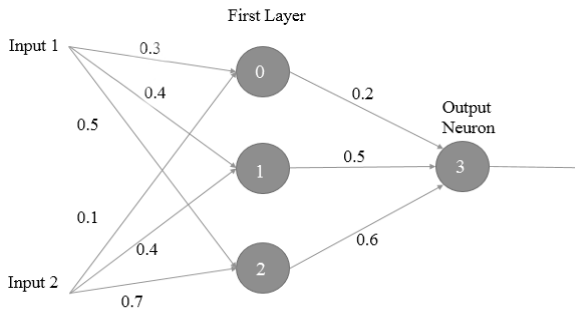
### 3.3 Artificial Neural Networks

ANNs are multilayer feed-forward neural networks with the bipolar sigmoid activation function.

A string describes ANN's weights list. Each ANN's neuron has a set of weights  $W: w_0 w_1 \dots w_m$  where  $w_j$  denotes the weight of a connection from neurons in previous layer to this neuron. In some cases, we add  $w_{m+1}$  to denote the neuron's threshold. Therefore, the string is constructed by:  $W_0 W_1 W_2 \dots W_n$  where  $n$  is the number of neurons.

We define a structural string to describe the ANN's structure: the number of neurons each layer, the number of inputs and the number of outputs. Particularly, this structure string consist 4 components A, B, C, D where:

- A denotes the number of inputs (equals the length of antigenic strings).
- B, C denotes the number of neurons in the first and second layer of hidden layer.
- D denotes the number of outputs (here  $D = 1$ ).



**Fig. 1.** An ANN's structure example

Structure string: 2 3 0 1

Chromosome string: 0.3 0.1 0.0 0.4 0.4 0.0 0.5 0.7 0.0 0.2 0.5 0.6 0.0

### 3.4 Clonal Selection Algorithm

The CLONALG is employed in training the detectors. The main operators are described as follows:

1) Antigenic Affinity Evaluation: for every antibody in the total Set, we calculate its fitness value (antigenic value) with respect to current presented antigen by using fitness function object. We input the strings into the Neural Networks constructed by antibody's information to get the output as well as the error value. Because each antigen is a cluster of virus fragments, so we will calculate the average error of the antibody with respect to all the strings in the antigen and normalize it. In addition, we calculate the False Positive Rate of the current antibody with respect to all benign fragments. By this way, we prevent our detectors from detecting benign strings. Therefore, the fitness value will be calculated as follows:

$$f=2/(\alpha.\text{error}+\beta.\text{fpr})$$

Where  $f$  is the fitness value or antigenic affinity value of the antibody,  $\alpha, \beta$  are important factor whose default value is 1, error is the average error of the antibody with respect to all strings in the antigen, and fpr is the false positive rate.

2) Proliferation: Select the  $n$  highest affinity antibodies in the total Set and clone them independently and proportionally to their antigenic affinities. The higher the antigenic affinity, the higher the number of clones will be generated.

3) Affinity maturation: All antibodies in clonal set will be suffered the affinity maturation, where the mutation rate is inversely proportional to the antigenic affinity.

4) Metadynamics: Update the memory Set by inserting the highest affinity antibodies chosen from the matured Set if its fitness value is larger than its respective memory antibody. Update the population by regenerating new antibodies to replace the  $d$  lowest affinity antibodies.

### 3.5 Detectors Generation Algorithm

The detectors generation algorithm is described as follows:

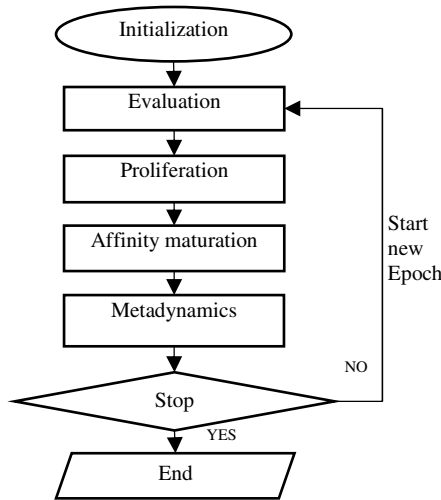


Fig. 2. Detectors generation algorithms

Step 1: Initialization: Generate a population with  $m$  individuals randomly.

Step 2: For every antigen in the environment:

- Antigenic Affinity Evaluation
- Proliferation
- Affinity Maturation
- Metadynamics

Step 3: If we reach our maximal number of iterations or meet our stop condition, stop our generation process. Otherwise, go to step 2.

## 4 Experimental Result

### 4.1 Training Data

We extract data fragments from virus files and benign files. The string length is 4 bytes and the step-size is 2 bytes. NSA with R-continuous distance is used to discard any string in virus files that matches strings in benign files. Afterwards, K-Means clustering algorithm is applied to get clusters of virus fragments. The number of clusters is 50 and 50% of them will be chosen for training. The antigenic set contains virus fragments in all clusters, the benign set denotes benign fragments extract from benign files.

**Table 1.** The number of strings in the antigenic set and benign set used

Antigenic Set	Benign Set
1200	200

### 4.2 Experiments

In our experiment, the number of antibodies in memory set is about 10% to 20% the number of antibodies in the population. As the learning algorithm trains ANN's weights and structure, the smaller number of detectors we defined, the more complicated the ANN's structure is. If the number of detectors is too small, the training algorithm will force detectors to extract the most important feature and the false positive will be high, on the contrary, too many detectors will leads to inefficient training process because of the overlap between them.

**Table 2.** The number of antigenic strings and benign strings used in training and testing

Training Set		Testing Set	
Antigenic Set	Benign Set	Antigenic Set	Benign Set
1000	200	1200	200

**Table 3.** Results when changing number of detectors (the number of antigen equals 1000)

Number of Detectors	Detection Rate	False Positive Rate
5%	85.58%	22.5%
10%	87.41%	18.5%
25%	88.42%	16.0%

**Table 4.** Results when changing number of antigenic strings used for training (when number of detectors equals 10%)

Number of antigenic strings	Detection Rate	False Positive Rate
800	84.58%	27.5%
900	88.33%	19.0%
1000	87.41%	18.5%

When the number of antigenic strings presented for training increases, the result gets better in both detection rate and false positive rate.

**Table 5.** Detection rate and false positive rate when changing number of epochs

Number of Epochs	Detection Rate	False Positive Rate
50	77.91%	34.0%
100	87.41%	18.5%
150	88.75%	15.5%

The result gets better when the number of epochs increases. However, the number of epochs should not be too large to avoid over-training.

## 5 Conclusions

In this approach, we evolve ANN's structure and weights using CLONALG. We introduce some special concepts, which are available in biology to make the model understandable. This research indicates that the new approach using biology-inspired algorithms opens a new prospect for dealing with the virus detection problem.

**Acknowledgment.** This research is funded by University of Information Technology – VNU, HCM under grant number C2011CTTT-02.

## References

1. Dasgupta, D., Niño, L.F.: Immunological Computation - Theory and Applications, pp. 27–29. CRC Press (2009)
2. de Castro, L.N., Von Zuben, F.J.: Learning and Optimization Using the Clonal Selection Principle. *IEEE Transactions on Evolutionary Computation*, Special Issue on Artificial Immune Systems 6(3), 239–251 (2002)
3. Cutello, V., Narzisi, G., Nicosia, G., Pavone, M.: Clonal Selection Algorithms: A Comparative Case Study Using Effective Mutation Potentials. In: Jacob, C., Pilat, M.L., Bentley, P.J., Timmis, J.I. (eds.) ICARIS 2005. LNCS, vol. 3627, pp. 13–28. Springer, Heidelberg (2005)
4. McCulloch, W.S., Pitts, W.H.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 115–133 (1943)
5. Doulas, A., Mavrouidakis, K., Gritzalis, D., Katsikas, S.: Design of a Neural Network for Recognition and Classification of Computer Viruses. *Computers & Security* 14(5) (1995)
6. Chao, R., Tan, Y.: A Virus Detection System Based on AIS. In: *Proceedings of the 2009 International Conference on Computational Intelligence & Security*, vol. 1, pp. 6–10 (2009)
7. Daoud, E.A.: Metamorphic Viruses Detection Using Artificial Immune System. In: *International Conference on Communication Software and Networks*, pp. 168–172 (2009)
8. Golovko, V., Komar, M., Sachenko, A.: Principles of Neural Network Artificial Immune System Design to Detect Attacks on Computers. *Modern Problems of Radio Engineering, Telecommunications and Computer Science* (2010)
9. Afaneh, S., Zita, R.A.: Virus Detection Using Clonal Selection Algorithm with Genetic Algorithm. *Applied Soft Computing* 13(1), 239–246 (2013)



# Muti-dimensional Architecture Modeling for Cyber Physical Systems

Bingqing Xu and Lichen Zhang

Shanghai Key Laboratory of Trustworthy Computing  
East China Normal University  
Shanghai 200062, China  
zhanglichen1962@163.com

**Abstract.** Cyber physical systems (CPS) are based on the computation, communication and control technology, and the use of cyber physical systems is so extensive. Cyber physical systems contain three parts: physical parts, communication parts and computation parts. Though many existing methods have their own merits in describing single aspect of the cyber physical system, single method is not complete enough to cover all the aspects in the cyber physical systems. In this paper, we proposed a joint approach to model cyber physical systems with multi-dimension architecture. This joint method integrates Modelica, UML, AADL and aspect-oriented modeling method based on model-driven architecture.

**Keywords:** Cyber Physical Systems, Multi-dimension, AOP, Architecture, AADL.

## 1 Introduction

Nowadays, applications of cyber physical systems appear in various fields as aerospace system, public transportation control system, mechatronic system, automotive, robotics and so on. Such systems based on network demand accurate and real-time convey of spatial parameters, physical parameters and so on, thus any unnoticed mistake will lead to unimaginable result. Therefore, cyber physical systems not only request a complete functional system, but also demands non-functional properties a lot, such as reliability, safety, security, fault-tolerance, automation as well. As the integration of computation process and physical process, CPS request accurate and synchronous interaction among human beings and other nodes on the whole network [1].

Although many researchers have done much related work with CPS, many technical problems and challenges still remain as follows: how can physical parameters transfer in discrete time slots; how can dimension, distance, topology of objects be traced in real-time; how can non-functional properties be well defined and described; how to describe the architecture of an typical CPS etc. This paper mainly focuses on the last question.

Architecture Analysis & Design Language (AADL) can analyze and prove the main property of system architecture repeatedly such as reliability etc through tool extension, tool framework and accurate semantics. It supports the analysis of real-time embedded

system, mapping from software to hardware, especially the modeling of real-time embedded system. AADL takes UML, MetaH as reference, and integrates the advantages together to provide a standard for designing and analyzing the software architecture, hardware architecture and main property [2].

Modelica is a object-oriented and equation based language which is good at modeling of complicated physical system. It is used in thermo system, control system etc.

## 2 Related Work

From the end of 2006, America government has funded in CPS summit and research. MIT developed the Distributed Robot Garden program. It aims to equip each tomato plant with sensor, in order to let robots take care of these plants within the help of real-time data gathered by the sensors [3].

Researchers have got some solutions for CPS architecture which is the high level of abstraction, based on various concepts. Li et al also designed a real-time CPS architecture based on service-oriented-architecture; it is composed of sensor layer, communication layer, computation layer, control layer and service layer. Koubaa et al considered the the character of Internet and embedded system, and proposed Cyber-physical Internet [4-6].

## 3 Proposed solution

To model a multi-dimensional architecture for cyber physical system, we should first work out the architecture framework. In the paper, we proposed the architecture of CPS in Figure 1. Based on the information flow, CPS can be divided into sensor part, communication network, computation system, data center, controller and strategy, actuator part. CPS communicates with terminals physical environment outside through network.

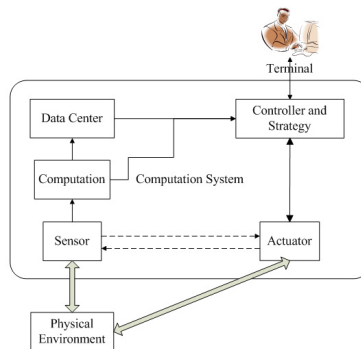


Fig. 1. Architecture Framework of CPS

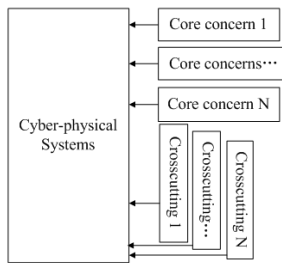
We first divide CPS into three dimensions: physical world dimension, communication dimension and computation dimension. The three dimensions rely on various domains such as machinery, control, hydraulic power etc, and some major languages and tools such as Modelica, hybrid automation, simulink etc are specialized in the design and analysis of domains.

In physical world dimension, general equations are used for modeling of the physical phenomena, the architectural components and connectors correspond to intuitive notions of physical dynamics in the same way that cyber components and connectors correspond to elements of computational systems.

Communication layer handles the data communication between devices or layers. Between the computation and the physical worlds, there exists the communication channel which connects the two worlds together. We provide two directed connector: physical-to-cyber (P2C) connector and cyber-to-physical (C2P) connector.

In computation dimension, we provide the architectures for describing a system in terms of its components, interfaces, and the connectors between the interfaces. We depicts component behaviors in describing the inputs, outputs, sequences, and conditions for coordinating various system behaviors , specifying the flow of control between the components. All these can be partly achieved in the AADL tool, but still need extension. Because the object-oriented modeling is not suitable for non-functional properties description, and the traditional MDA is based on the object-oriented thinking.

Aspect-oriented modeling can separate core concerns and crosscutting concerns, and encapsulate the non-functional part as Figure 2 shows. Certificated annexes such as error model, behavior annex of AADL can model the dynamic part of the system [7], and OSATE, the open source AADL tool environment provided a tool set solution [8]. In addition, the property set in AADL enables us to introduce the Modelica and UML description to AADL [9].

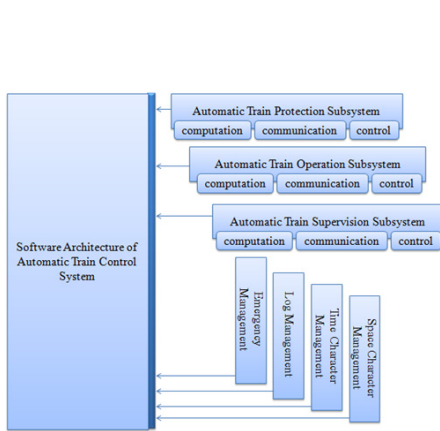


**Fig. 2.** AOP Architecture for CPS

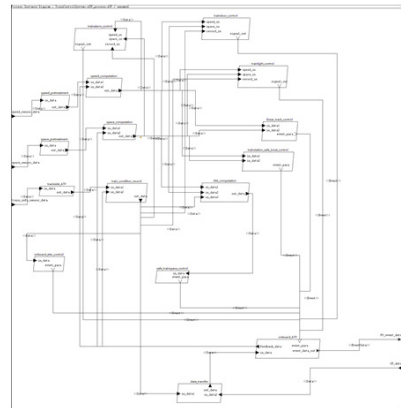
In addition, to be concrete, we propose a multi-view to illustrate the multi-dimensional architecture as follows. Views includes physical environment view, characteristic view, logical view, physical view, data view, behavior view.

## 4 Case Study

In this paper, we introduce an Automatic Train Control (ATC) system as CPS example. It should be emphasized that traditional train control system uses many signal control, but many operations still rely on the train driver. This operation mode is full of hidden dangers, thus we need a fully automated ATC system badly. According to the proposed solution, we first analyze the requirement of the system. ATC consists of three subsystems: Automatic Train Protection (ATP) system, Automatic Train Supervisor (ATS) system, Automatic Train Operation (ATO) system. The architecture for ATC is shown in Figure 3.



**Fig. 3.** Architecture for Case Study ATC



**Fig. 4.** ATP Logical View

Figure 4 shows the logical view of ATP subsystem emphasizing on the software part. The entire core concerns are described by components such as process, thread; the collaboration among components such as port, link.

## 5 Conclusion

This paper mainly puts forward a joint approach to model cyber physical systems architecture. According to the research and understanding of CPS, we integrate AADL, Modelica, and UML as a joint method and describe the architecture in multi-view. We use AOP for separating core business and crosscutting concerns. In addition, to make up the shortage of AADL description in multi-view, we put forward characteristic view, behavior view, data view, physical environment view, and expand AADL in description of physical world. Then we take ATC system as an example, apply the method to the description of ATC.

In the future, we tend to work harder to make up the shortage of AADL description in multi-view. We want to introduce more mature languages into the AADL tool to

make a more complete and integrated tool for CPS, especially in description in physical, cyber layer and hardware part. In order to guarantee a safe tool, we are still working on the verification method on such integrated approach by AADL annex.

**Acknowledgment.** This work is supported by Shanghai Knowledge Service Platform Project (No.ZF1213), national high technology research and development program of China (No.2011AA010101), national basic research program of China (No.2011CB302904), the national science foundation of China under grant (No.61173046, No.61021004, No.61061130541, No.91118008), doctoral program foundation of institutions of higher education of China (No.20120076130003), national science foundation of Guangdong province under grant (No.S2011010004905).

## References

1. Lee, E.A.: Cyber Physical Systems: Design Challenges. In: Proceedings of the 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing, pp. 363–369 (2008)
2. Yang, Z., et al.: AADL: An Architecture Design and Analysis Language for Complex Embedded Real-Time Systems. *Journal of Software* 21(5), 899–915 (2010)
3. Cyber-physical system system, <http://en.wikipedia.org/wiki/Cyber-physical>
4. Hoang, D.D., Paik, H.-Y., Kim, C.-K.: Service-Oriented Middleware Architectures for Cyber-Physical Systems. *International Journal of Computer Science and Network Security* 12(1), 79–87 (2012)
5. Li, Y., Sun, D., Liu, W., Zhang, X.: A Service-Oriented Architecture for the Transportation Cyber-Physical Systems. In: Proceeding of the 31st Chinese Control Conference, Hefei, China, July 25-27 (2012)
6. Tabuada, P.: Cyber-physical systems: Position paper. In: NSF Workshop on Cyber-Physical Systems, Austin, Texas (2006)
7. Feiler, P., Hansson, J.: Flow Latency Analysis with the Architecture Analysis and Design Language (AADL) (2007)
8. Feiler, P.: SAE AADL V2: An Overview (2010)
9. Kandula, S., et al.: Cyber-Physical Systems(CPS) Annex for AADL (2010)

# An Algorithm of Group Scheduling with Void Filling in OBS Core Nodes

Nguyen Hong Quoc, Vo Viet Minh Nhat, and Nguyen Hoang Son

Hue University, Hue, Vietnam  
{nhquoc, vvmnhat, nhson}@hueuni.edu.vn

**Abstract.** Scheduling is one of the activities which have a great impact on the communication performance of optical burst switching networks. There have been many scheduling algorithms proposed but most of them are online. Recently, several algorithms of group scheduling have been published, but they all have a high complexity (NP-complete) and have not exploited the voids created among the previous scheduled bursts yet. This article proposes an algorithm of group scheduling with void filling. The analyses and evaluations of the effectiveness based on the simulation results will confirm the advantages of our algorithm.

**Keywords:** Optical Burst Switching, Linear and Adaptive Group Scheduling with Void Filling, NS2-obs.

## 1 Introduction

Optical communications, since its inception in the 90s so far, has developed through several generations. From the initial models of wavelength routing, with the end-to-end lightpaths, to the models of optical packet switching proposed recently, which comes from the inspiration of mature electronic packet switching. However, due to some limitations of optical technology, for example: optical buffers (similar to RAM in electronic domain) or the optical packet switches of nanosecond rate cannot be produced; the optical packet switching cannot have become the real thing in a near future. A compromise solution of the optical circuit switching and the optical packet switching is the optical burst switching, which has opened a new research and is considered as a promising technology for the next generation of all-optical Internet.

A typical characteristic of optical burst switching (OBS) networks is the burst header packet (BHP) transmitted separately from its data burst in terms of space and time. This means that the BHP will be sent on a control channel separated from the data channels transporting its burst and the BHP will be sent before its data burst an offset time. This offset time must be predetermined sufficiently so that the BHP can reserve needed resources and configure switches just before its burst arriving at each node on the path from source to destination.

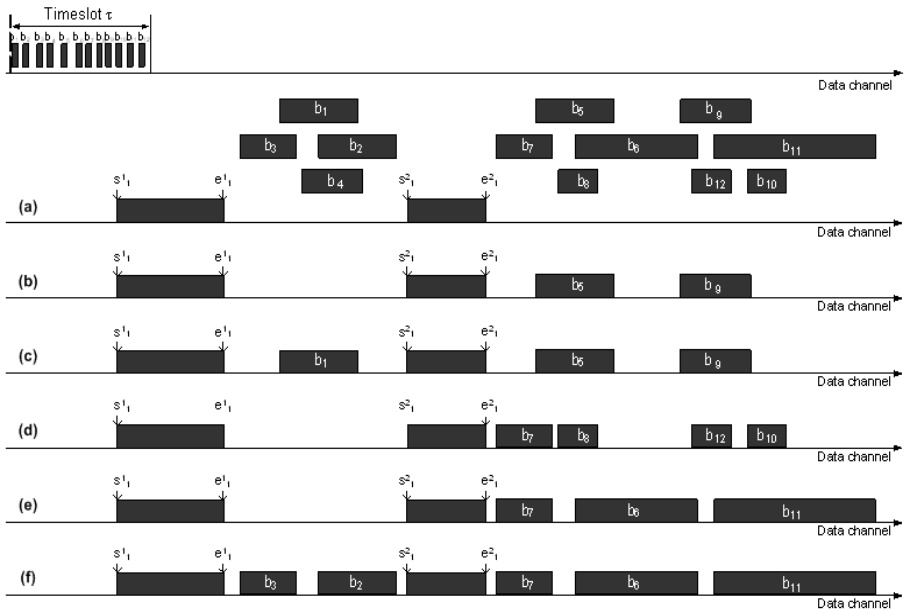
The activity of resource reservation of a BHP at an OBS core node is a part of scheduling operation. There are two approaches for scheduling: (1) the online scheduling as LAUC and LAUC-VF in [1,2], in which each arriving BHP schedules

immediately for its following burst, and (2) the group scheduling as OBS-GS and MWIS-OS in [3,4], in which the BHPs arriving in each group scheduling timeslot schedule together for their following bursts. The simulation results in [4] have proved that the group scheduling is better than the online one, basing on the rate of lost bytes. However, the complexity of group scheduling algorithms in [3,4] is large (NP-complete) [5]; its group scheduling timeslot is established fixedly without considering the impact of the arriving data rate; and the algorithms have still not exploited the bandwidth in the voids created among the previous scheduled bursts. This article proposes a new algorithm of group scheduling which overcomes the above problems.

The next structure of our article is as follows: Section 2 presents some related works; Section 3 describes our algorithm of group scheduling with void filling; Section 4 adds an algorithm of adjusting the group scheduling timeslot which changes adaptively with the rate of arriving data; Section 5 shows our scenario, simulation results and analyses basing on NS2-obs; and the conclusion is in Section 6.

## 2 Related Works

Considering a case of 12 BHPs (from  $b_1$  to  $b_{12}$ ) arriving at an OBS core node in timeslot  $\tau$  on a control channel and their following bursts as shown in Fig. 1a.



**Fig. 1.** A description of BHPs arriving in timeslot  $\tau$  and the possible cases of online and group scheduling

If an online scheduling algorithm without void filling as LAUC [1] is used, only  $b_1$  and  $b_5$  are scheduled (as shown in Fig. 1b), while the others are dropped. With the case of online scheduling algorithm with void filling as LAUC-VF [2], there are also only  $b_1$ ,  $b_5$  and  $b_9$  scheduled (as shown in Fig. 1c).

To increase the number of scheduled bursts (or the total length of scheduled bursts), some group scheduling algorithms have been proposed in which the BHPs arriving in timeslot  $\tau$  schedule together for their bursts. The algorithm of OBS-GS in [3] maximizes the number of bursts scheduled (e.g.  $b_3$ ,  $b_4$ ,  $b_6$ , and  $b_8$  are scheduled in Fig. 1d); while the algorithm of MWIS-OS in [4] maximizes the total length of scheduled bursts (e.g.  $b_3$ ,  $b_2$ ,  $b_7$  are scheduled in Fig. 1d). Basing on the simulation results in [4], MWIS-OS is better than OBS-GS in terms of the bandwidth utilization, but both of them have an algorithm complexity of NP-complete [5]. Moreover, its group scheduling timeslot is established fixedly without considering the impact of the arriving data; and the algorithm has still not exploited the voids created among the previous scheduled bursts. The new group scheduling algorithms we propose below overcome all the above problems (e.g.  $b_3$ ,  $b_2$ ,  $b_6$ ,  $b_7$  and  $b_{11}$  are scheduled in Fig. 1d).

### 3 The Model of Group Scheduling without and with Void Filling

Assuming that OBS core nodes have no use of the fiber delay link, no wavelength converter and the initiated offset time is large enough to perform the group scheduling at each intermediate node along the path from source to destination. Arriving BHPs in each timeslot  $\tau$  at an OBS core node will schedule together for their bursts. Our new model of group scheduling can be described as a process of two phases:

**Phase 1:** Each arriving BHPs will be classified into an appropriate queue depending on the output data channel of its burst. The information on the  $i^{\text{th}}$  arriving BHP includes the arriving time ( $s_i^k$ ), the length ( $l_i^k$ ) and the wavelength ( $w_i^k$ ) of its burst on the  $k^{\text{th}}$  data channel, which is described as a triple  $b_i^k(s_i^k, l_i^k, w_i^k)$ . Depending on without or with void filling, each  $k^{\text{th}}$  output data channel keeps different information and therefore, the scheduling conditions are also different.

(A) The case of without void filling: Each output data channel keeps a threshold of latest available unscheduled time (LAUT), in which an arriving burst is only scheduled if its arriving time is after this threshold,  $s_i^k > \text{LAUT}$ .

(B) The case of with void filling: Each output data channel keeps the start time ( $s_j^k$ ) and the end time ( $e_j^k$ ) of scheduled bursts ( $j=[1..m]$ , where the  $m^{\text{th}}$  burst is the last one scheduled in the  $k^{\text{th}}$  output data channel). An arriving burst is only scheduled if it matches one of the two following conditions: (1)  $s_i^k \geq s_m^k$  (where  $s_m^k$  is equivalent to LAUT in the case of without void filling) and (2)  $s_i^k \geq e_j^k$  và  $s_{j+1}^k \geq s_i^k + l_i^k$  (in the case of with void filling).

**Phase 2:** After each timeslot  $\tau$ , the BHPs in each queue are taken out to schedule together for their bursts. Being similar to MWIS-OS, our algorithm of group scheduling also maximizes the total length of scheduled bursts.



### 3.1 The Algorithms of Group Scheduling without and with Void Filling

Considering a set of arriving BHPs  $I = \{b_1, b_2, \dots, b_n\}$  in timeslot  $\tau$ , where  $n$  is the total of arriving BHPs. Two cases of group scheduling without and with void filling are only different from the keeping information and the scheduling conditions, so a common description of our group scheduling algorithms without and with void filling is as follows:

#### The algorithm of group scheduling without/with void filling

**Input:**  $n, I = \{b_1^k, b_2^k, \dots, b_n^k\}$  where  $b_i^k(s_i^k, l_i^k, w_i^k), i=1,2,\dots,n$ .

**Output:** A set of scheduled bursts ( $I'$ ), the total length of which is maximum.

**Process:**

**Step 1:** (Initializing)

The arriving bursts are arranged in ascending of their end time.

**Step 2:** (Determining  $\text{index}(j)$ , where  $j=1,2,\dots,n$ )

Setting  $\text{index}(j)=k$  if the burst  $k$  does not overlap the precedent burst  $j$ .

**Step 3:** (Determining  $C(j)$ , that is the total length when the burst  $j$  is scheduled, where  $j=1,2,\dots,n$ )

$C(j)$  is calculated by the following equation:

$$C(j) = \begin{cases} 0 & \text{if } j = 0 \\ \text{Max}\{C(j-1), l_j^k + C(\text{index}(j))\} & \text{if } j > 0 \end{cases}$$

**Step 4:** (Determining the set of scheduled bursts by tracing)

**Step 4.1:** Setting  $j=n$  and  $\text{cost}=C(n)$ .

**Step 4.2:** while  $j>0$ :

- if  $\text{cost} = C(j-1)$  then  $j=j-1$ .

- if not, scheduling the burst  $j$  and setting  $j=\text{index}(j)$ .

### 3.2 The Complexity of Group Scheduling Algorithm

The complexity of our algorithm is determined as follows:

- Step 1: A quick sort algorithm can be used to sort bursts in ascending of their end time, so the complexity is  $O(n \log_2(n))$ .

- Step 2: To determine  $\text{index}(j)$ , we need to make  $k$  steps ( $k=1,\dots,j-1$ ) to check if the burst  $j$  overlaps the burst  $k$ . Because the bursts have been arranged in Step 1, we can use a binary search algorithm and therefore the complexity is  $O(\log_2(n))$ .

- Step 3: The complexity is  $O(n)$ .

- Step 4: In the step of determining the set of scheduled bursts by tracing, the worst case is to repeat  $n$  Step 4.2, so the complexity is  $O(n)$ .

Since the steps of our LGS algorithm are independent, so the final complexity is  $O(n \log_2(n))$ . It means that the complexity of our algorithm is only linear, so it is called Linear Group Scheduling (LGS) (in the case of without void filling) or Linear Group Scheduling with Void Filling (LGS-VF) (in the case of with void filling).

Another improvement to reduce the running time and the complexity of LGS and LGS-VF algorithms is to sort each arriving burst and to calculate  $\text{index}(j)$  immediately before the timeout of timeslot  $\tau$ ; the group scheduling time will be reduced and the complexity of our LGS and LGS-VF algorithms is now only  $O(n)$ .

### 3.3 Adjusting the Group Scheduling Timeslot

In the precedent model of group scheduling, the timeslot  $\tau$  is predetermined and unchangeable. However, if the rate of arriving BHPs is slow and the timeslot  $\tau$  is small, there may not have many, or even just a burst arriving in timeslot  $\tau$ ; the group scheduling is evidently inefficient and the timeslot  $\tau$  needs increasing to have more BHPs for group scheduling. Conversely, if the rate of arriving BHPs is fast, there will be many bursts overlapped and then dropped; the large timeslot  $\tau$  is insignificant and it is necessary to decrease it in order to reduce the source-destination delay. In summary, it is necessary to change the size of timeslot  $\tau$  negatively to the rate of arriving data.

We keep in mind that the timeslot  $\tau$  is limited by 2 thresholds in above and below:

- The timeslot  $\tau$  should be large enough to have at least 2 consecutive BHPs for scheduling their bursts. Assuming that  $\tau_{\min}$  is the allowable minimum distance for 2 consecutive BHPs on a control channel, the minimum threshold of timeslot  $\tau$  is then  $2*\tau_{\min}$ , in which there are at least 2 arriving BHPs.

- The timeslot  $\tau$  should not be larger than the allowable maximum time ( $\tau_{\max}$ ) for a BHP which transits at an OBS core node. The threshold  $\tau_{\max}$  obviously depends on the original offset time at each edge node and  $\tau_{\max}$  must be greater than  $2*\tau_{\min}$ . In this article, we assume that the offset time is large enough so that at each intermediate node on its path, a BHP can wait a maximum delay  $\tau_{\max}$  before scheduling for its burst.

Some other notations are the following:

- +  $\tau_{\text{init}}=(2*\tau_{\min}+\tau_{\max})/2$ : the initiated value of timeslot  $\tau$ .
- +  $\tau_{\text{step}}=(\tau_{\max}-2*\tau_{\min})/M$ : the adjusting step of timeslot  $\tau$  and  $M$  is the number of adjusting steps.  $M$  must be chosen so that  $\tau_{\text{step}} \geq \tau_{\min}$ .
- +  $V_a$ : the rate of arriving BHPs in timeslot  $\tau$ .
- +  $V_{\text{avg}}$ : the average rate of previous arriving BHPs.

The algorithm which adjusts the timeslot  $\tau$  depending on the rate of arriving BHPs is described as follows:

#### The algorithm of adjusting the group scheduling timeslot

**Input:**  $\tau_{\min}, \tau_{\max}, V_{\text{avg}}$

**Output:**  $\tau, V_{\text{avg}}$

**Process:**

**Step 1:** (Determining the current rate of arriving BHPs basing on the number of arriving BHPs in the timeslot  $\tau$ )

Setting  $V_a = N_a/\tau$ .

**Step 2:** (Determining the correlation between the current rate of arriving BHPs and the average rate of previous arriving BHPs)

- if  $V_a > V_{\text{avg}}$  then  $\Delta = -V_a/V_{\text{avg}}$

- if  $V_a = V_{\text{avg}}$  then  $\Delta = 0$

- if  $V_a < V_{\text{avg}}$  then  $\Delta = V_{\text{avg}}/V_a$

**Step 3:** (Adjusting the timeslot  $\tau$ )

Setting  $\tau = \tau + \Delta*\tau_{\text{step}}$

**Step 4:** (checking if the timeslot  $\tau$  is valid)

- if  $\tau < 2 * \tau_{\min}$  then  $\tau = 2 * \tau_{\min}$
- if  $\tau > \tau_{\max}$  then  $\tau = \tau_{\max}$

**Step 5:** (Updating the average rate of previous arriving BHPs)

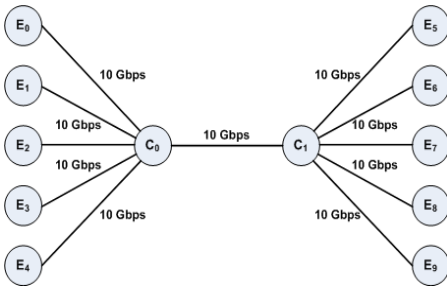
Setting  $V_{\text{avg}} = (V_a + V_{\text{avg}}) / 2$

The complexity of this algorithm is  $O(1)$ .

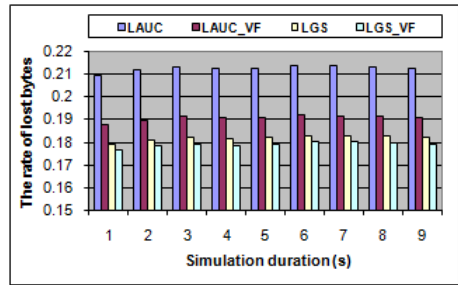
Noting that if we add this algorithm of adjusting the timeslot  $\tau$  to Step 5 of the precedent group scheduling (LGS and LGS-VF) algorithm, we have an algorithm of linear adaptive group scheduling (LAGS and LAGS-VF), because the timeslot  $\tau$  now is adjusted adaptively depending on the rate of the arriving data.

### 4 Simulation and Analysis

We install our algorithms of LGS, LAGS, LGS-VF, LAGS-VF and compare them (basing on the rate of lost bytes and the delay) with the online scheduling algorithm of LAUC, LAUC-VF [1,2] and also with two group scheduling algorithms of OBS-GS, MWIS-OS, [3,4]. The simulation environment is NS-obs0.9a, on a PC of Intel 2 Core CPU 2.4 GHz, 2G RAM.



**Fig. 2.** The simulation network



**Fig. 3.** A comparison between LAUC, LAUC-VF, LGS and LGS-VF

The simulation network includes two core nodes ( $C_0$  and  $C_1$ ), in which each connects with 5 edge nodes ( $E_i, i=0, \dots, 9$ ) as shown in Fig. 2. Assuming that the data arriving at a core node have the Poisson distribution and the arriving bursts have varied lengths. Each link has 8 data channels and 2 control channels. The bandwidth of a data channel is 10GB/s. The simulation time is from 1 to 9 seconds. Other parameters are established as follows:  $\tau_{\min}=500\mu s$ ,  $\tau_{\max}=1500\mu s$ ,  $V_{\text{avg}}=1$ .

As described in Fig. 3, the simulation results show that LGS and LGS-VF are better than LAUC and LAUC-VF, basing on the rate of lost bytes (approximately 13% and 11%). It is evident because, as shown in Fig. 1b and Fig. 1c, the online scheduling algorithms as LAUC, LAUC-VF do not consider the impact of a current scheduled burst on the scheduling of the later others; and, as shown in Fig. 1b and Fig. 1e,

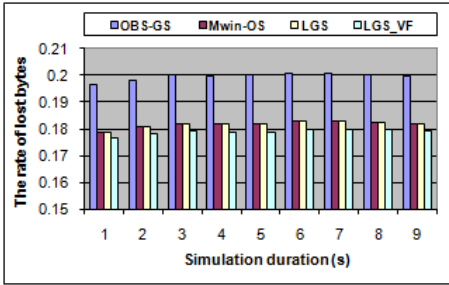


Fig. 4. A comparison of OBS-GS, MWIS-OS, LGS and LGS-VF.

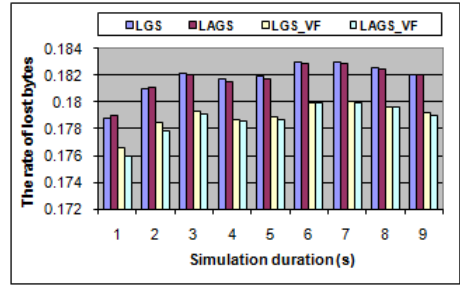


Fig. 5. A comparison between LGS, LGS-VF, LAGS and LAGS-VF.

LAUC, LGS (being equivalent to MWIS-OS) do not take care for the fragmentation of the data channel (caused by the previous scheduled bursts). LGS-VF overcomes the both above problems, so it has the least rate of lost bytes.

When comparing with OBS-GS and MWIS-OS, the simulation results in Fig. 4 show that the rate of lost bytes of LGS is lower than OBS-GS and equal to that of MWIS-OS (because MWIS-OS and LGS both reach the maximum total length of scheduled bursts). However, LGS-VF has the least rate of lost bytes, because it exploits the idle bandwidths among the scheduled bursts.

To compare LAGS with LGS and LAGS-VF with LGS-VF, we install the algorithm of adjusting the timeslot  $\tau$  and the received result is LAGS-VF that has least rate of lost bursts when the timeslot  $\tau$  is adjusted adaptively (Fig. 5).

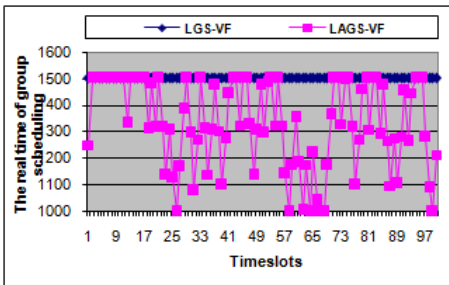


Fig. 6. A comparison between LGS-VF and LAGS-VF about the real time of group scheduling for the first 300 simulated timeslots

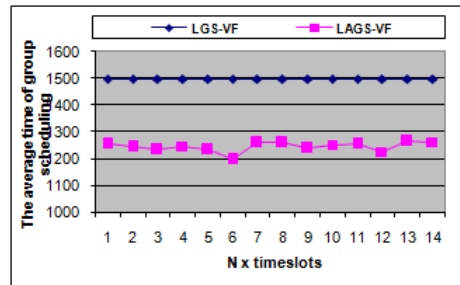


Fig. 7. A comparison between LGS-VF and LAGS-VF about the real time of each 500 consecutive timeslots averaged of 7500 simulation timeslots

However, the most important advantage of LAGS, LAGS-VF is to have the real time of group scheduling (shown by the size of the adjusted timeslot) which is much less than that of LGS, LGS-VF. As shown in Fig. 6 (in case of the first 300 simulated timeslots), the timeslots of LGS, LGS-VF are always fixed and equal to the allowable maximum timeslot ( $\tau_{max}$ ) at each core node, while the real timeslots of LAGS, LAGS-VF are changeable and less than this maximum value. Another clearer

presentation is described in Fig. 7, in which the real time of each 500 consecutive timeslots is averaged of 7500 simulation timeslots.

Reducing the real scheduling time is very significant in decreasing the offset time and the communication delay. Moreover, reducing the scheduling time at each node will provide an opportunity for increasing the scheduling time at the next nodes, thus the ability of scheduling can be increased and the congestion may be reduced. Therefore, basing on the simulation results, in comparison with LGS-VF, LAGS-VF does not improve much the rate of lost bytes (Fig. 6), but it brings a lasting effect on the scheduling activities at the other nodes and other communication activities (such as avoiding congestion, reducing the end-to-end delay ...) in OBS networks.

## 5 Conclusion

Scheduling is one of the activities which have a significant effect on the performance of OBS network. In fact, there are many proposed scheduling algorithms but most of them are the online scheduling ones. Recently, several algorithms of group scheduling have been published, but they have the high complexity (NP-complete). This article proposes a model of adaptive group scheduling with void filling with its linear complexity. The simulation results show that the LGS-VF and LAGS-VF are more effective (based on the rate of lost bursts and the algorithm complexity) than the online scheduling algorithms of LAUC, LAUC-VF and 2 group scheduling algorithms of OBS-GS and MWIS-OS. Especially, LAGS and LAGS-VF improve significantly the scheduling time, therefore, it provides the long-term effectiveness of the scheduling activities at the other nodes and other communication activities in OBS networks.

## References

1. Turner, J.: Terabit burst switching. *J. High Speed Networks* 8, 3–16 (1999)
2. Xiong, Y., Vandenhoute, M., Cankaya, H.C.: Control architecture in optical burst-switched WDM networks. *IEEE Journal on Selected Areas in Communications* 18, 1838–1851 (2000)
3. Charcranon, S., El-Bawab, T.S., Cankaya, H.C., Shin, J.: Group-Scheduling for Optical Burst Switched (OBS) Networks. In: *Proceedings of Globecom 2003*, pp. 2745–2749 (2003)
4. Zheng, H., Chen, C.: Performance Analysis of Scheduling Algorithms in Optical Burst Switching (OBS) Networks. In: *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, p. 557 (2007)
5. Östergård, R.J.: A fast algorithm for the maximum clique problem. *Discrete Applied Mathematics* (2002)
6. Network simulator (NS), <http://www.isi.edu/nsnam/ns/>

# A Fast TRW Algorithm Using Binary Pattern

Jun-Young Park<sup>1</sup> and Chang-Suk Cho<sup>2</sup>

<sup>1</sup>Eyenix Co., Ltd.,

980-3, Yeongtong-Dong, Yeongtong-Gu, Suwon-Si, Korea

<sup>2</sup>Div. of Information & Telecommunication, Han-Shin University,

137 Hanshinda-Gil, Osan-Si, Korea

cscho@hs.ac.kr

**Abstract.** A fast TRW algorithm that calculates disparity map for stereo vision is proposed. The fast TRW algorithm using binary pattern could reduce processing time by skipping invalid area. In order to skipping the invalid area we designed a binary pattern image to classify valid and invalid area of pattern from the object. To verify the effectiveness of our method we investigated reliability of disparity map between the results of TRW and our algorithm. As a result of our investigation for the reliability, the result by our method shows slightly larger error rate than TRW but the difference is not so large, whereas our method can generally reduce 11.14% of processing time than TRW. If calibration work need not the case that very small error rate is requested, our method will be a very effective one.

**Keywords:** Stereo vision, Stereo matching, Belief propagation, TRW.

## 1 Introduction

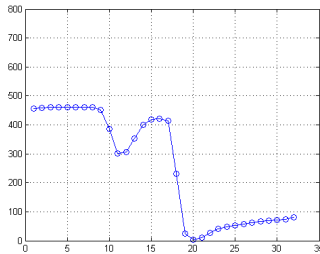
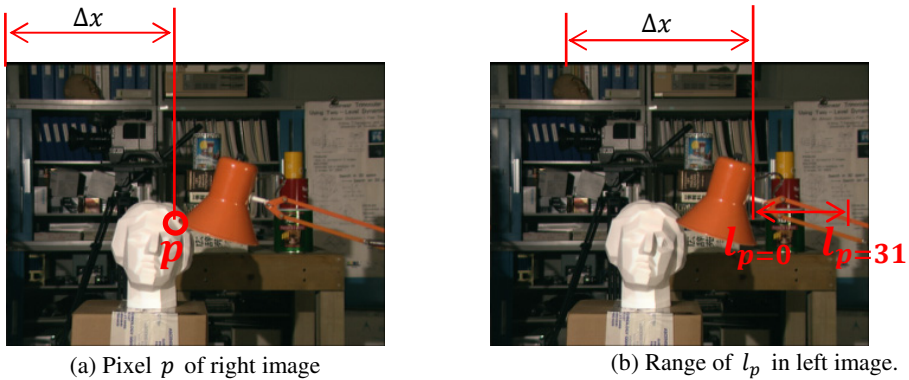
Stereo vision method infers 3D scene geometry from two images with different viewpoints. The fundamental basis for stereo vision is the fact that a single three-dimensional physical location is projected to a unique pair of scene locations in two observing cameras. Therefore it is the key problem to find the corresponding points between the two images obtained by the two cameras. In order to find the corresponding points the two approaches such as local stereo matching algorithm and global stereo matching algorithm were reported. Local stereo matching algorithm [1-4] is efficient method to get disparity map easily, but sensitive to locally ambiguous regions in scenes. The block matching algorithms [1-2] and the feature matching algorithms [3-4] are a kind of local stereo matching algorithm. Global stereo matching algorithm [5-7] uses nonlocal constraints in order to reduce sensitivity to local regions in the scene which has areas fail to match due to occlusion or uniform texture, etc. However the use of these constraints makes the computational complexity of global matching significantly greater than that of local matching. The common approaches to the global matching are Graph-Cut [5-6], Belief Propagation (BP) [7] and TRW (Tree-Reweighted) methods [7]. Over the last few years, energy minimization approaches such as Graph-Cut, BP and TRW have had a renaissance, primarily due to

powerful new optimization algorithms. In Middlebury university’s benchmarks [8], TRW produced results whose energy is lower than the energy of the ground-truth solution. Although it is generally accepted that algorithms such as TRW are a huge improvement over older algorithms, less is known about the efficiency versus accuracy trade-off among more recently developed algorithms. In this paper, we propose an enhanced algorithm of TRW as a fast TRW, which improves computational processing time using binary pattern.

## 2 TRW Algorithm

TRW (tree-reweighted) message passing algorithm is a message-passing algorithm which is similar to BP [9] on the surface. The message of TRW update rule is equation 1 [10].

$$M_{p \rightarrow q}^t(l_q) = \min \left( C_{pq} \left( d_p(l_p) + \sum_{s \in N(p)} M_{s \rightarrow p}^{t-1}(l_p) \right) - M_{q \rightarrow p}^{t-1}(l_p) + V_{pq}(l_p, l_q) \right) \quad (1)$$



(c) Graph of  $d_p$  function ( $x$ -axis :  $l_p$ ,  $y$ -axis :  $d_p(l_p)$ )

**Fig. 1.** Function graph of  $d_p$  for pixel  $p$

TRW algorithm works by passing messages around the graph defined by the four connected image grids. The method has iterative steps with messages from all nodes being passed in parallel. Each message is a vector of dimension given by the number of possible labels. Let  $M_{p \rightarrow q}^t$  be the message that node  $p$  sends to a neighboring node  $q$  at iteration  $t$ . Note that if  $c_{pq}$  was set to 1, the message passing formula would be identical to that of the standard BP. Here,  $l$  represents disparity having criteria  $l \in [0, disparity\ max - 1]$ . If disparity maximum is set to 32,  $l$  has 0 ~ 31. The data term  $d_p$  measures how well the disparity function  $d$  agrees with input image pair to use SAD (Sum of Absolute Differences) or SSD (Sum of Squared Differences) to measure. In Fig. 1, a pixel having the smallest value of  $d_p$  is determined to the similar pixel with pixel  $p$ . This similar pixel position is position of pixel  $p$  + position of  $l_p$

The smoothness term  $V_{pq}(l_p, l_q)$  in equation 2 encodes the smoothness assumptions made by the algorithm [8-10]. It does not find similarities among pixels but compares continuities among pixel  $p$  and neighbor pixels, since a pixel and its neighbors have similar disparities generally. To make the optimization computationally tractable, the smoothness term is often restricted to only measuring the differences between neighboring pixel's disparities.

$$V_{pq}(l_p, l_q) = V(|l_p - l_q|) = \min(|l_p - l_q|^k, V_{max}) \quad k \in \{1, 2\} \quad (2)$$

If the disparity  $l_p$  of a pixel  $p$  and  $l_q$  of  $q$  are the same, continuity penalty of smoothness term is 0. If a difference of disparities becomes bigger, continuity penalty becomes also bigger.  $\sum_{s \in N(p)} M_{s \rightarrow p}^{t-1}(l_p)$  in equation 1 is a function for pixel  $p$  and neighbor node  $s$ . Neighbor node  $s$  includes left node, right node, up node, down node of pixel  $p$ . So,  $\sum_{s \in N(p)} M_{s \rightarrow p}^{t-1}(l_p)$  represents sum of messages from neighbors.

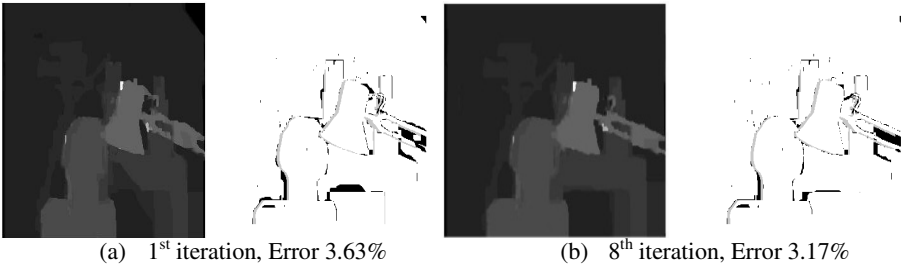


Fig. 2. Tsukuba disparity map in each iteration

When the stereo matching algorithm is estimated, the reliability of disparity map and the processing time are the most important factors. In order to reduce the processing time, many suggestions using PC and FPGA (Field-Programmable Gate Array) were reported [13]. Generally global stereo matching algorithm has more reliability than local stereo matching algorithm. However it needs a lot of processing time, so that TRW algorithm also needs so much time. Especially, TRW algorithm needs a lot of iterations in order that disparity map becomes reliable. In Fig. 2, the left



scenes show disparity maps obtained by iteration process. Right scenes show errors of matching. The errors are taken by Middlebury University, which introduced it in stereo vision web page [8], at occlusion area and discontinuous area. This web page provides objective results of estimation to each algorithm.

Fig. 3 shows processing time in graph at 1, 2, 4 and 8 iterations. Processing time is increasing while iteration time is increasing largely. The processing time was measured under Intel core 2 duo 2.0GHz CPU with 2GB RAM and the Tsukuba image has  $384 \times 288$  as size and 16 step disparity, Teddy size is  $450 \times 375$  with 60 step disparities, Venus size is  $434 \times 383$  with 20 step disparity, Wood size is  $300 \times 242$  with 32 step disparities.

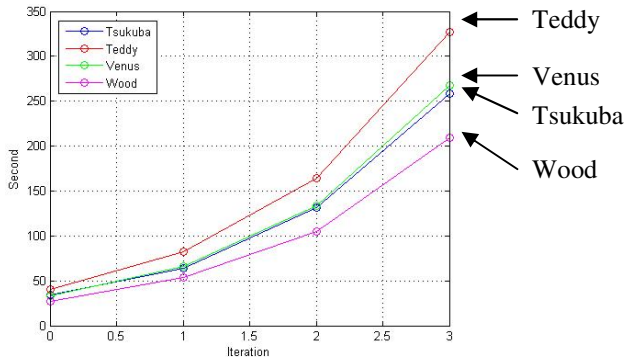
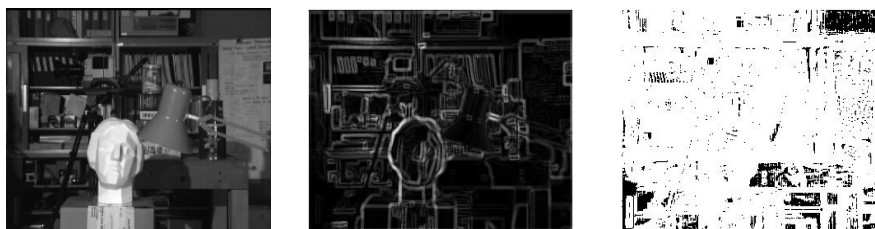


Fig. 3. Processing time of TRW algorithm by iteration

### 3 A Fast TRW Algorithm Using Binary Pattern

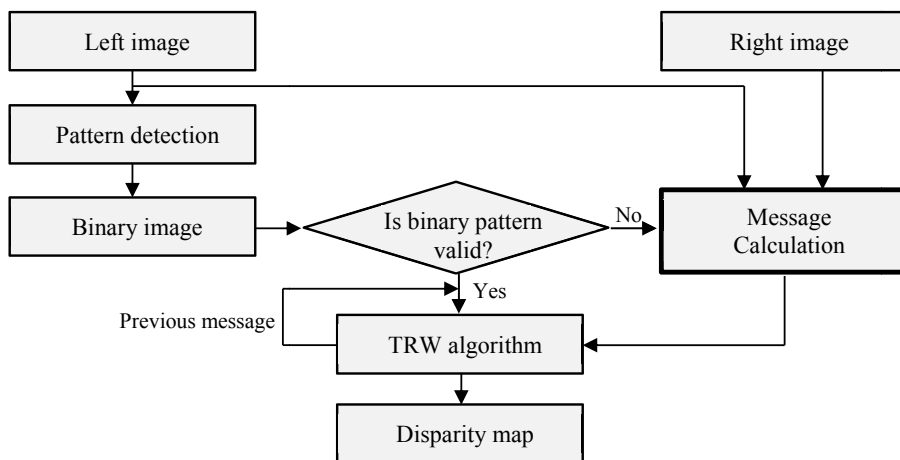
TRW algorithm has a feature to make pixels form a group affected by neighbor if there are no high frequency areas. Therefore, it is robust at occluded area and noise. There is a distinguished difference between message values of pixel and its neighbor in high frequency area (which has many edges and patterns), but there is little difference between message values in simple or no pattern area. Thus if we find areas having simple or no pattern in image, processing time for the area can be reduced to use previous message value instead of calculating new message. There is much pattern information at the area having many variations in pattern or at the edge of an object. Hence the computation of message value is carried out using equation 1 only at the edge or the patterned area, whereas at the simple pattern area the previous message value is used instead of computing a new value. For detecting pattern information, Laplacian mask was used to detect the edge since Laplacian mask can detect edges from all directions and its processing time is fast. Fig. 4 shows binarization result. Bright area has complex pattern in image but dark area shows a simple pattern in image. The right image of Fig. 4 is the binarized image. In the image the white area has valid pattern but the black has no valid pattern.



**Fig. 4.** Left : Tsukuba test image, *Center* : Edge detection result(Amount of pattern) *Right* : Binarization image (*Threshold* = 10)



**Fig. 5.** Left : Venus test image, *Center*: Edge detection result(Amount of pattern) *Right* : Binarization image (*Threshold* = 10)



**Fig. 6.** Block Diagram for the Fast TRW Algorithm using Binary Pattern

After classifying the areas into the two groups of valid and invalid area, the fast TRW algorithm applied to computation of message and propagation. If the area is invalid, our TRW algorithm uses previous message and propagation instead of computing. Block diagram of this algorithm is shown in Fig. 6.

## 4 Experimental Result

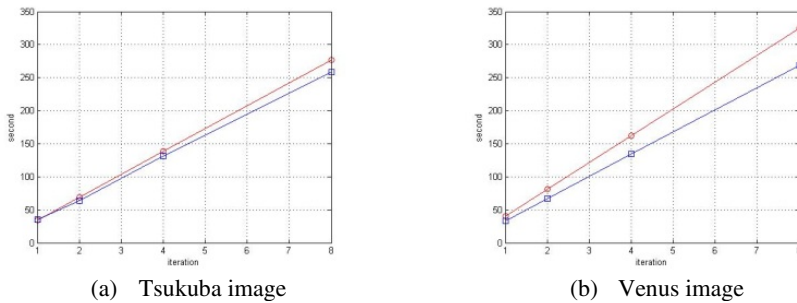
Using the fast TRW using pattern information, we confirmed it could reduce the processing time.. Table 1 and Table 2 show comparison of processing time between our fast TRW and the TRW. The differences in time between the two method are small in Tsukuba image because Tsukuba image has a lot of complex patterns. But, the differences are large in Venus image that has not so many complex patterns.

**Table 1.** Comparison of processing time in Tsukuba with TRW and the fast TRW using binary pattern

Tsukuba Image	TRW Not Using Binary Pattern	The fast TRW Using Binary Pattern
1 <sup>st</sup> iteration	34.46 sec	34.99 sec
2 <sup>nd</sup> iterations	68.85 sec	64.25 sec
4 <sup>th</sup> iterations	138.26 sec	131.87 sec
8 <sup>th</sup> iterations	276.44 sec	258.69 sec

**Table 2.** Comparison of processing time in Venus with TRW and the fast TRW using binary pattern.

Venus Image	TRW Not Using Binary Pattern	The fast TRW Using Binary Pattern
1 <sup>st</sup> iteration	40.59 sec	33.07 sec
2 <sup>nd</sup> iterations	81.27 sec	66.99 sec
4 <sup>th</sup> iterations	162.16 sec	134.43 sec
8 <sup>th</sup> iterations	324.63 sec	268.31 sec



**Fig. 7.** Processing time of comparative algorithms. (○ : Not using binary pattern, □ : Using binary pattern)

In Fig. 7, graph ○ represents processing time of TRW algorithm. Graph of □ does processing time of a fast TRW algorithm using binary pattern. Processing time of Tsukuba image shows small difference between the two algorithms. However, processing time of Venus image shows large difference between them.

Table 3 shows the error rates of mismatching for the two methods. The differences in error rate show not a big difference between the two methods. The error rate is obtained from Middlebury University's stereo web site as in the case of the figure 2. In Tsukuba image the error rates between the two images show no difference, but in Venus image TRW shows slightly smaller error rate than that of the fast TRW. However the difference of the error rates is slightly small and the benefit obtained from the reduced time is larger than loss of the error rate, so that the proposed method can be used effectively.

**Table 3.** Error comparison of disparity map

	Iterations	TRW	A Fast TRW algorithm using Binary Pattern
Tsukuba image	1	5.24%	4.79%
	2	4.89%	5.10%
	4	4.71%	4.64%
	8	4.50%	4.51%
Venus image	1	3.63%	5.46%
	2	3.60%	5.41%
	4	3.32%	5.27%
	8	3.17%	5.22%

Benefit from reducing time is shown in Table 4. Tsukuba image can get 5.91% reducing rate for processing time using pur method. Venus image get 17.64% reducing rate for processing time. Venus image can get the reducing effect largely in processing time than Tsukuba since Venus is not complex then Tsukuba image.

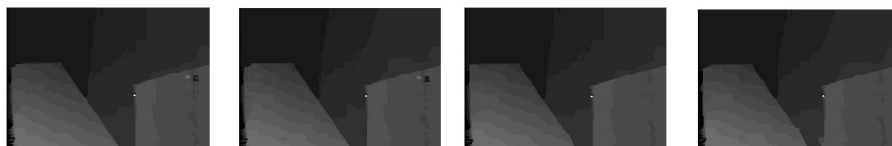
**Table 4.** Benefit of the reducing effect for processing time using our method

	Tsukuba	Venus
Rate	5.91%	17.64%



(a) 1<sup>st</sup> iteration, Tsukuba image.

(b) 8<sup>th</sup> iteration, Tsukuba image.



(c) 1<sup>st</sup> iteration, Venus image.

(d) 8<sup>th</sup> iteration, Venus image.

**Fig. 8.** Processing time of comparative algorithms. (○ : Not use binary pattern, □ : Use binary)

## 5 Conclusion and Future Work

In this paper, we suggest a new algorithm that calculates disparity map. The fast TRW algorithm using binary pattern could reduce processing time by skipping invalid area. In order to skipping the invalid area we designed a binary pattern image to classify valid and invalid area of pattern from the object. To verify the effectiveness of our method we investigated reliability of disparity map between the results of TRW and our algorithm. As a result of our investigation for the reliability, the result by our method shows slightly larger error rate but the difference is not so large, whereas our method can generally reduce 11.14% of processing time. If calibration work need not the case that very small error rate is requested, our method will be a very effective one.

**Acknowledgements.** This work was supported by Hanshin research grant.

## References

1. Aschwanen, P., Guggenbuhl, W.: Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms. In: Forstner, Ruwiedel (eds.) *Robust Computer Vision*, pp. 268–289. Wickmann (1993)
2. Bhat, D.N., Nayar, S.K.: Ordinal Measures for Image Correspondence. *IEEE Trans. Pattern Analysis and Machine Intelligence* 20, 415–423 (1998)
3. Bigone, F., Henricsson, O., Fua, P., Stricker, M.: Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery. In: Buxton, B.F., Cipolla, R. (eds.) *ECCV 1996*. LNCS, vol. 1064, pp. 85–96. Springer, Heidelberg (1996)
4. Birchfield, S., Tomasi, C.: Multiway Cut for Stereo and Motion with Slanted Surfaces. In: *Proc. Int'l Conf. Computer Vision*, vol. 1, pp. 489–495 (1999)
5. Belhumeur, P.N.: A Bayesian Approach to Binocular Stereopsis. *Int'l J. Computer Vision* 19(3), 237–260 (1996)
6. Birchfield, S., Tomasi, C.: Depth Discontinuities by Pixel-to-Pixel Stereo. Technical Report STAN-CS-TR-96-1573, Stanford Univ. (1996)
7. Boykov, Y., Veksler, O., Zabih, R.: Fast Approximate Energy Minimization via Graph Cuts. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23(11), 1222–1239 (2001)
8. <http://vision.middlebury.edu/stereo/eval/>
9. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision. In: Figueiredo, M., Zerubia, J., Jain, A.K. (eds.) *EMMCVPR 2001*. LNCS, vol. 2134, pp. 359–374. Springer, Heidelberg (2001)
10. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(6), 4 (2008)
11. Wainwright, M., Jaakkola, T., Willsky, A.: Map Estimation via Agreement on Trees: Message-Passing and Linear Programming. *IEEE Trans. Information Theory* 51(11), 3697–3717 (2005)

12. Felzenszwalb, P., Huttenlocher, D.: Efficient Belief Propagation for Early Vision. *Int'l J. Computer Vision* 70(1), 41–54 (2006)
13. Yu, T., Lin, R., Super, B., Tang, B.: Efficient Message Representations for Belief Propagation. In: *ICCV*, pp. 1–2 (October 2007)

# Efficient Descriptor-Filtering Algorithm for Speeded Up Robust Features Matching

Minwoo Kim, Deokho Kim, Kyungah Kim, and Won Woo Ro

School of Electrical and Electronic Engineering, Yonsei University,  
262 Seongsanno, Seodaemun-gu, Seoul, Korea  
{kenstars, nautes87, kyungah.kim, wro}@yonsei.ac.kr

**Abstract.** This paper presents an efficient descriptor filtering algorithm for the feature matching process of SURF. The matching algorithm used in OpenSURF compares each and every feature descriptors by calculating the root-mean-square error of the descriptor vectors. The proposed instant-termination and Bloom filtering algorithm pre-compares the feature descriptors and decides whether the compared descriptor pairs should be further inspected. The proposed pre-comparison process compares the most significant bits of the descriptor for early decision. Also, the descriptor bits are interleaved to adapt to the Bloom filter, increasing the reliability of the filtering process. Our proposed filtering algorithm effectively reduces the number of root-mean-square error calculations.

**Keywords:** Speeded Up Robust Features, Feature detection and description, Feature matching, Bloom filter.

## 1 Introduction

Finding correspondences between different images has always been an important issue in the computer vision field. In order to do so, features from an image are first detected by extracting the interest points. Then, a set of vectors is calculated to describe the local features of the detected interest points. This process is called *feature description*. The detected and described features from different images are then matched by comparing the feature descriptors.

Many works have been conducted in order to detect, describe, and match the visual features. Among them, Scale Invariant Feature Transform (SIFT) has shown significant progress by providing invariance in scale changes, rotation, *etc.* [1]. Later, Speeded Up Robust Features (SURF), which also provides robustness in scale changes and rotation, has accelerated the feature detection and description process by using integral images and approximated scale space analysis [2].

Despite the accelerated detection and description process, however, feature matching process of the open source SURF (OpenSURF) exhaustively compares each and every feature one by one by calculating the root-mean-square error (RMSE) of the 64 local descriptor vectors [3]. That is, the descriptor comparison process gives  $O(N^2)$  complexity which results in a heavy computation overhead.

Thus in this research work we propose an instant-termination and filtering algorithm for SURF matching process in order to reduce the amount of RMSE calculation. There have been several previous researches that use hashing method in order to match the features [4, 5]. However in order to adapt to the SURF matching algorithm, direct use of hashing method is not appropriate since the natural hashing method provides only exact matches, whereas the descriptor vectors of SURF used for matching show differences in values, forcing it to use RMSE for comparison. In order to efficiently perform exact match comparisons between the SURF descriptors, the proposed algorithm investigates the format of the descriptors and compares the Most Significant Bits (MSB) which we will call Maximum Likelihood Bounding (MLB) bits to precede the RMSE calculation. We additionally propose a filtering method using a Bloom filter [6], which is widely used for network applications [7, 8]. To adequately utilize the advantages of Bloom filter for SURF matching, we propose a *descriptor-interleaving* process in addition to the MLB method. Thus, bits in one position are collected, concatenated, and independently compared to provide efficiency and reliability to the filter. Direct comparison of the MLB bits enables the use of hashing method as well as reducing the unnecessary calculation of RMSE of mismatching features.

## 2 SURF Algorithm

### 2.1 Feature Detection and Description of SURF

Detecting interest points using SURF basically begins by computing the integral image. Then box filters of multiple sizes are applied to the integral image according to the different scales. Box filters allow approximation of the Gaussian second order derivative that is used for scale space analysis, thus lowering the computational overhead compared to SIFT. The scale space is used to extract the interest points by comparison with  $3 \times 3 \times 3$  neighboring pixels:  $3 \times 3$  neighbors each from current, upper, and lower scales, and deciding as an interest point if the current pixel is the extremum.

The extracted interest points are described by computing Haar wavelet responses on each pixel around the interest point within 6 times of scale. This enables to assign the orientation of the interest point, which provides invariance to image rotation. Then,  $4 \times 4$  grids each of  $5 \times 5$  samples are constructed over the interest point along the orientation. Haar wavelet responses are computed for each sample point, one in horizontal direction ( $d_x$ ) and the other in the vertical direction ( $d_y$ ). The computed Haar responses of each grid are summed up to form the descriptor vectors ( $\sum d_x$ ,  $\sum d_y$ ,  $\sum |d_x|$ , and  $\sum |d_y|$ ). Thus, each feature contains 64 descriptor vectors.

### 2.2 Feature Matching Algorithm of OpenSURF

The matching algorithm used in OpenSURF utilizes the computed descriptor vectors. First, the RMSE is calculated as



$$(\sum(\mathbf{d}_{curr} - \mathbf{d}_{ref})^2)^{0.5}. \quad (1)$$

for 64 local descriptor vectors, where  $\mathbf{d}_{curr}$  and  $\mathbf{d}_{ref}$  denotes the descriptor vector of the feature from the current image and the reference image. All the error values are calculated for each feature in the current image and all the features from the reference image, respectively. Then, two reference features that show the least RMSEs are stored as candidates. If the RMSE ratio between the candidates is less than 0.65, the best candidate is set as the match for the feature in the current image. As explained thus far, each and every feature from both current and reference images are compared by calculating the RMSE. This results in a heavy computation overhead since  $O(MN) \doteq O(N^2)$  comparisons should be performed, whereas the number of matches is limited to the value  $\min(M, N)$ . Therefore, in the following section, we propose an efficient instant-termination and filtering algorithm that reduce the number of error value computations.

### 3 Maximum Likelihood Bounding (MLB)

#### 3.1 Instant-Termination Using MLB

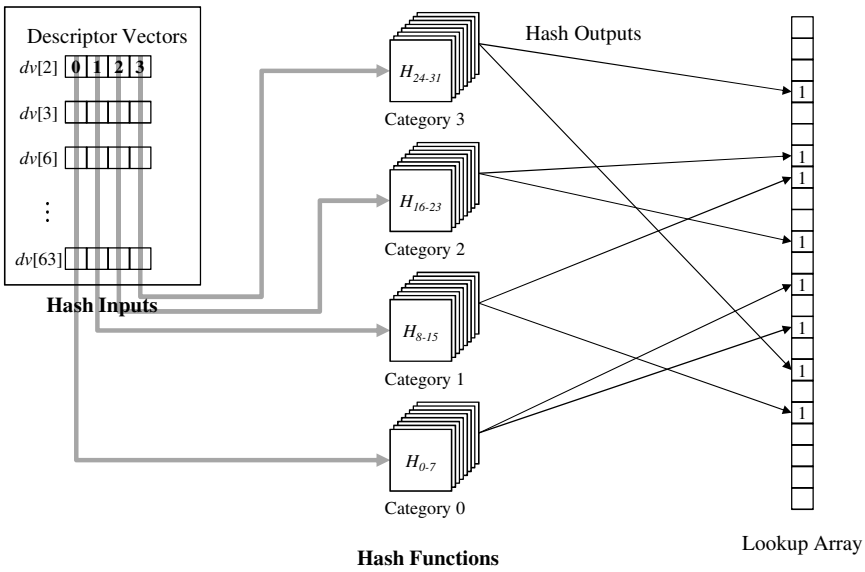
The proposed instant-termination algorithm compares the first  $n$  (eight in this research) descriptors among 64. If the number of identical comparisons fails to exceed a specific threshold (1 in this research), the features are considered mismatching and are excluded from the comparison process using RMSE. Since the descriptor vectors compared rarely show an exactly identical value, we first scale the descriptor vector value to integer space by multiplying by a large value, specifically  $2^{20}$  in this research work. Then, in order to compare the exact equality between the feature descriptor vectors, the bits from the higher positions, which show high probability of similarity, are extracted from the scaled integer space. Specifically, the values of the four least significant bits (LSBs) from the five most significant bits (MSBs) are compared. It is designed this way for the descriptor vector value does not exceed 0.5 for almost all cases, allowing four LSBs from five MSBs sufficient for the comparison. We call the extracted bits *MLB bits* throughout this paper. The proposed scaling algorithm enables to examine the equality of the integers from the current and the reference features, which is much simpler and computationally friendlier than the RMSE calculation.

#### 3.2 Bloom Filter Adaptation

This section presents the utilization of the Bloom filter in order to compare the MLB bits from the scaled descriptor vector illustrated in the previous section. Since the direct usage of the instant-termination algorithm only filters out the most obvious mismatches, we utilize the Bloom filter in order to compare all the descriptor vectors efficiently. In order to do so, the MLB bits of the scaled descriptor vectors of the features from the *reference* image first go through a set of hash functions, which are categorized into four groups as shown in Fig. 1. Then the descriptor vectors of the

features from the *current* image are hashed in the same way, and the number of matches in the lookup array is calculated. Basically, the number of matches in the lookup array that exceeds a certain threshold value is considered as possible matching features, and the original matching process is performed using RMSE, as described in Section 2.2.

The proposed method using Bloom filter interleaves the input descriptor vectors such that each hash function category  $n$  collects and concatenates all the  $n^{\text{th}}$  bits of the original input descriptor vectors and hashes the interleaved values, also illustrated in Fig. 1. The inputs are handled this way so that the descriptor vectors could be rearranged from higher probability of similarity to lower probability. This design enables to implicitly utilize the instant-termination algorithm because the bits with higher probability of similarity could be compared in advance. That is, the number of matches of the eight hash outputs of  $H_{0-7}$  is calculated in advance and if the number of matches does not exceed the local threshold value, the rest of the hash function operations are skipped and the next feature of the current image is handled. In other words, mismatching higher order bits are considered as having higher probability of mismatching features and are filtered out by the proposed Bloom filter.



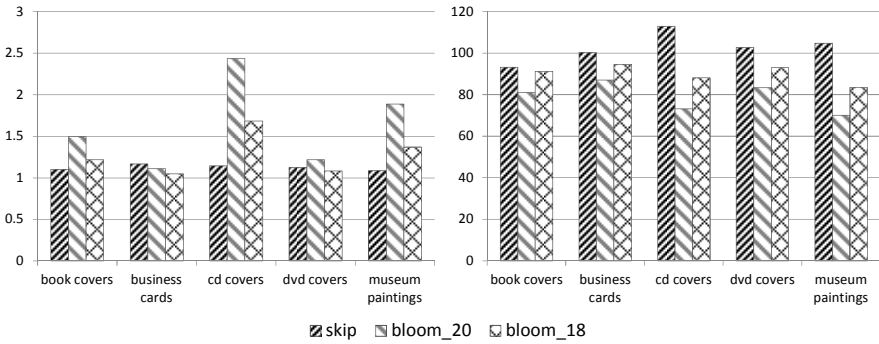
**Fig. 1.** The hash inputs consist of 32 local descriptor vectors ( $\sum |d_{x,l}|$  and  $\sum |d_{y,l}|$ ). The inputs are interleaved according to the bit position.

## 4 Experimental Results

This section provides the experimental results of the proposed filtering algorithm. A set of various types of images are used for the experiment [9]. First, the features of the reference and the current images are extracted and described. Then the features are

matched using the proposed method. The execution time of the matching algorithms is measured and the performance of the proposed schemes is evaluated. The experiment is performed on a real machine equipped with AMD Opteron 6176SE processors. The legends of the result graph denote the following: the *skip* field represents the proposed algorithm that uses only the instant-termination method introduced in Section 3.1. The *bloom<sub>x</sub>* field denotes the adoption of Bloom filter as well as the instant-termination method, where *x* represents the threshold value. A heuristic method was used to determine the appropriate threshold values for effective matching rates.

The left of Fig. 2 shows the speedup results of the proposed algorithm. As shown in the figure, the proposed method efficiently filters out the unnecessary comparisons, achieving maximum speedup of 2.44 and average of 1.35. Especially when the number of feature is way larger than the number of matches, the proposed methods achieve extremely speeded up matching process. In other words, mismatching features are effectively removed from being compared by calculating RMSE. Also, the speed up of *bloom<sub>20</sub>* outperforms the other algorithms in almost all cases since its large threshold value filters out lots of matching tries. However, this results in a low value of matching rate, as will be discussed in the next paragraph.



**Fig. 2.** Speed up (left) and matching rate (right, %) result of the proposed method

The matching rate of the proposed algorithms is shown in the right of Fig. 2. The matching rate is calculated by the ratio of the number of matched feature of the proposed algorithm and that of original OpenSURF matcher. The matching rate of the proposed algorithm shows the maximum value of 112.88% and average of 90.57%. It can be seen from the results that *skip* provides larger rate of matching compared to using Bloom filter since its low value of threshold skips only the *obvious* mismatches. On the contrary however, the speedup of *bloom<sub>x</sub>* significantly exceeds that of *skip* in most cases. Because *bloom<sub>20</sub>* has a high value of threshold, it filters out lots of positive matches compared to *bloom<sub>18</sub>*. This means that carefully adjusting the threshold value is critical for the performance of the Bloom filter. In overall, we can conclude that the proposed method using Bloom filter provides both speeded up and relatively high rate of matching. In addition, because the original OpenSURF matcher compares the relative similarities of the descriptor vectors by

calculating RMSE, there naturally exists potential probability of false match caused by relative comparison using candidates. On the contrary, our proposed algorithm reproduces the descriptor vector values in order to directly compare the absolute value, potentially increasing the rate of exact matches, respectively.

## 5 Conclusion

This research proposed an efficient instant-termination and filtering method in order to accelerate the matching process of OpenSURF. So as to compare the descriptor vectors of the features of reference and current image, we scaled the original descriptor vector values to integer space and extracted four LSBs from five MSBs which we called MLB bits. This approach enables to compare the absolute values of the various descriptor values, simplifying the matching process. Also, the interleaving of the MLB bits made it possible to perform instant-termination for the Bloom filter since mismatching higher bits could be considered as mismatching features. Through experiments, our approach has shown significant improvements in execution time of the matching process, still maintaining high probability of matching rate.

As future work, we would investigate on more efficient comparison method in order to further increase the matching rate of the instant-termination and filtering algorithm, while maintaining the fast speed of matching.

**Acknowledgements.** This work was funded by grants from the Digital Media & Communication R&D Team, Samsung Electronics Co., Ltd.

## References

1. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* 60, 92–110 (2004)
2. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* 110, 346–359 (2008)
3. Evans, C.: Notes on the OpenSURF Library. Technical report, University of Bristol (2009)
4. Dong, W., Wang, Z., Charikar, M., Li, K.: Efficiently Matching Sets of Features with Random Histograms. In: 16th ACM International Conference on Multimedia, pp. 179–188. ACM, New York (2008)
5. Bayram, S., Sencar, H.T., Memon, N.: An Efficient and Robust Method for Detecting Copy-Move Forgery. In: 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 1053–1056. IEEE Press, New York (2009)
6. Bloom, B.H.: Space/Time Trade-Offs in Hash Coding with Allowable Errors. *Commun. ACM* 13, 422–426 (1970)
7. Broder, A., Mitzenmacher, M.: Network Applications of Bloom Filters: A Survey. *Internet Mathematics* 1, 485–509 (2004)
8. Jain, N., Dahlin, M., Tewari, R.: Using Bloom Filters to Refine Web Search Results. In: 8th International Workshop on the Web and Databases, pp. 25–30 (2005)
9. Chandrasekhar, V., Chen, D., Tsai, S., Cheung, N.-M., Chen, H., Takacs, G., Reznik, Y., Vedantham, R., Grzeszczuk, R., Bach, J., Girod, B.: The Stanford Mobile Visual Search Data Set. In: ACM Multimedia Systems Conference, pp. 117–122. ACM, New York (2011)

# Towards Effective 3D Model Management on Hadoop\*

Hua Luan\*\*, Yachun Fan, Mingquan Zhou, and Xuesong Wang

College of Information Science and Technology, Beijing Normal University  
No. 19 Xijiekou Wai Street, Beijing, China

{luanhua, fanyachun, mqzhou, wangxs}@bnu.edu.cn

**Abstract.** In the age of big data, the volume of unstructured data, such as photo, image, video and 3D model, is rapidly growing. This paper proposes a 3D model management strategy based on Hadoop distributed file system to tackle the big data problem that 3D model management is facing. The file access characteristics are analyzed and two kinds of storage models are designed to store and index different data types in 3D models. Efficient storing and accessing algorithms are presented to load data and meet user requests. An experimental study is conducted to show the effectiveness and efficiency of the proposed methods.

**Keywords:** 3D Model; Hadoop; Storage Model; Batch Reading.

## 1 Introduction

IDC study [1] predicts that world's data will reach 40 zettabytes (ZB) by 2020, resulting in a 50-fold growth from the beginning of 2010 and unstructured data will account for 90% of all data. Both industry and research community are facing challenges to efficiently meet exploding unstructured data storage demands and fast information access requirements. Hadoop [2] is an open-source framework that allows for the distributed processing of large data sets across clusters of computers. Hadoop Distributed File System (HDFS), derived from Google File System (GFS) [3], stores massive data sets reliably and can be used to support many applications as file system.

3D model retrieval is an effective approach to manage, access and share 3D information and is also a fundamental task for further 3D applications like modeling virtual scene. Many useful research work and systems [4], [5], [6] have been developed to tackle various challenges of 3D retrieval. Generally, retrieval performance for example precision and recall is the main concern associated with the existing research work. With the expansion of 3D data, how to provide good scalable data storage and request processing is a challenging task that 3D models management is facing.

In this paper, we present an efficient approach to manage unstructured data in 3D by using Hadoop distributed file system. Original files in 3D models are organized into large files in HDFS. By analyzing the data components and access patterns, two

---

\* Supported by “the Fundamental Research Funds for the Central Universities” (No. 2012LYB47).

\*\* Corresponding author.

types of storage models are utilized, one for 2D thumbnail images and the other for specific 3D information files. Data in the same minor category is treated as a unit which is archived in the same large file. In each storage model, a diverse index is built to help access the original files. The corresponding data loading and file accessing algorithms are developed. OBJ, MTL and texture files belonging to one model are fetched simultaneously in a single read operation to increase performance. The proposed management methods are evaluated on a cluster of commodity servers.

## 2 Preliminaries

Google File System (GFS) [3] and MapReduce [7] are two key technologies for Google. Hadoop provides an open-source implementation for institutions and companies with very large data processing challenges, including Yahoo! and Facebook. Hadoop distributed file system is called HDFS, where file metadata and application data are separately stored. There exist two kinds of nodes namely NameNode and DataNode. The NameNode maintains all the metadata of files and useful information such as data block distribution records. File blocks are placed on DataNodes.

When file requests are sent to NameNode through HDFS clients, NameNode responds the requests by returning a list of relevant DataNode servers where the actual file data live. Then file data transfer happens between clients and DataNodes. NameNode is the central manager of HDFS. If files stored in HDFS are small in size and the number of files is huge, a great deal of metadata need be maintained in NameNode which leads to two issues. First, the overhead of managing metadata is increased and too many metadata operations become the performance bottleneck. Second, high memory usage is introduced since the information need be kept in memory of NameNode. Therefore, Hadoop itself is not optimized for a massive number of small files.

## 3 3D Models on Hadoop

### 3.1 Data and Behavior Analysis

A 3D model in our data includes at least three kinds of files. A JPG image file contains a thumbnail view of 3D object. The second type of file in a 3D model is an OBJ file, which is a kind of data-format that represents 3D geometry (although 3D geometry is represented by OBJ format in this paper, the proposed methods are not just suitable for OBJ files but easily adapted to be applicable to other types of files such as OFF files). An OBJ file usually supports materials by referring to an or more external MTL material files, which are the third type of file in 3D models (there is one MTL file for each model in our current data). In addition, one or more external image or texture files may be referenced from within some MTL files via texture map statements. In our models, these files have the .jpg extension.

For user queries, our retrieval system first returns the matching 3D models from data repository in the form of thumbnail images, which provides users a quick glance

at the available results. So JPG file accesses of these 3D models show a kind of spatial locality. When the user clicks the interested thumbnail image, the system displays the associated 3D model, which means further accesses to the OBJ file, MTL file and texture files. Generally 3D model data are bulk-loaded, and then accessing becomes the primary operation.

Before adopting HDFS as underlying storage layer, our 3D models are grouped in categories and stored in the directory hierarchies of file system. Fig. 1 shows the storage structure. All the data are kept in a top level directory which is named “Data” or anything else. The 3D models are first organized by major categories with category names such as “Animal”, “Clothing” as folder names and each category is further subdivided into minor categories. In every small class, for each 3D model there is a low level directory named with numbers which contains the thumbnail image file, OBJ file, MTL file and texture files. During a search, files in the same category have high probability to be accessed in a consecutive manner due to similarity.

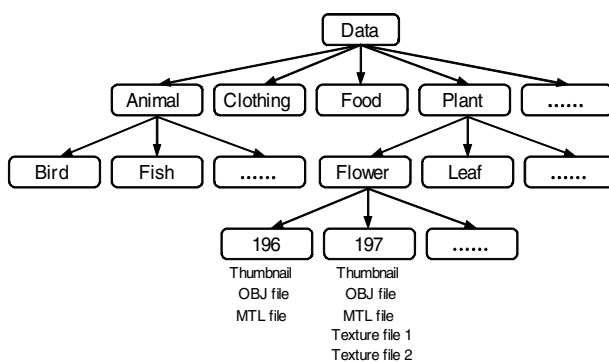


Fig. 1. The original 3D data storage structure

## 3.2 The Details of Our Approach

### 3.2.1 File Structure

A hybrid of horizontal and vertical integration techniques is used to merge original files. Two kinds of storage models are introduced in the process. Every minor category produces two HDFS large files no matter how many models it owns. The thumbnail files are selected to form one HDFS file, and the remaining files are merged together into the other big file. For ease of presentation, the former large file is called TF (Thumbnail File), and the latter is named OF (3D Object related File). Models in a TF are arranged in the same order as in an OF. For one 3D model, files in the OF in turn are OBJ file, MTL file and texture files according to the order in which they appear in MTL file if exist. Fig. 2 shows the file layouts in two different storage models.

The large files do not have length limitation. A file is broken into block-sized chunks, which are stored as independent units in HDFS. The block layout is shown in Fig. 3. Each file has a metadata footer to store information such as the number of original files in it. Thus the last block (HDFS can figure out where it is) ends with

this file footer, following multiple original files. The other blocks do not have any metadata in our current implementation. But if it is required in future, each block can have block header in the beginning. For example, this could facilitate parallel block processing in MapReduce programs. In addition, the content of file footer can also be extended if needed.

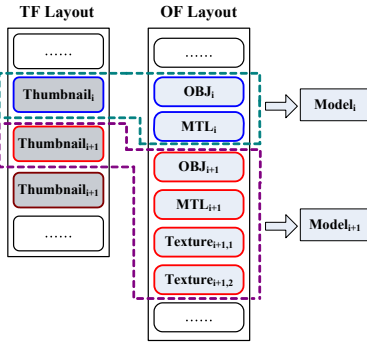


Fig. 2. The file layout

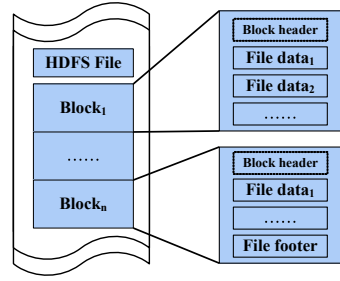


Fig. 3. The block layout

### 3.2.2 Index Structure

In order to find the original file in a large one, useful mapping information should be maintained. Given an original file, we should know where it resides, that is the name of large file, which block with offset it starts from, as well as its length. To retrieve files quickly, an in-memory hash index is utilized to map key to the corresponding large file, block, offset and file size. Path names of small files are treated as hash keys. The path name can be constructed by major category, minor category and file name since data are stored under category directories and classification information is available when reading files. So files with duplicate name are allowed in different categories. We make use of category names to name the large files. For example, for the “Flower” class in Fig. 1, the TF is called “Plant-Flower-Thumbnail”, and the OF is named “Plant-Flower-Object”. By means of classification information, the target merged file can be identified. Thus large file names are removed from hash values.

Since OBJ, MTL and related texture files in one model are stored contiguously in an OF, they can be fetched as a whole in a read operation. In order to use this optimization strategy, the hash value of OBJ file is designed differently, which contains the offset of OBJ in OF, sum of sizes of OBJ, MTL and texture files, and the OBJ file length. The concrete methods are introduced in the following subsection. In addition, index files are created and persisted in HDFS in data loading phase. At system start, the files are scanned to rebuild in-memory indexes.

### 3.2.3 Storing and Accessing Algorithms

Data loading refers to the process of merging the original files displayed in Fig. 1 into large files in HDFS using two kinds of storage models. The detailed steps are shown in Algorithm 1. First the algorithm creates the corresponding TF and OF for each minor category (line 3). Then it selects the thumbnail file in the current 3D model and



appends the JPG image to the TF file (line 6). A  $\langle key, value \rangle$  pair is formed (line 7) and is recorded in the index structure of TF (line 8). The *blockSize* is typically 64MB. For OBJ file in the above 3D model, line 10 performs the appending operation. Then MTL file is analyzed to get the texture files it references and appended to OF (line 11). The information for indexing MTL file is generated just like for thumbnail image (line 12). Lines 13 to 14 conduct the merging process for each texture file. Then, the  $\langle key, value \rangle$  pair for OBJ file is produced (line 15) and recorded in the index structure (line 16). The OBJ index information is delayed until the end of the algorithm because it needs to collect all the file lengths.

---

**Algorithm 1.** Data Loading (*data*)
 

---

<b>Input:</b> <i>data</i> : where the model files are	<b>Output:</b> files and indexes in HDFS
---	--

```

1: for each major category in data do
2:   for each minor category in major category do
3:     create TF and OF;
4:     for each model in minor category do
5:       size = length of TF;
6:       append the thumbnail JPG file to TF;
7:       set  $\langle key, value \rangle$  with pathname of JPG file and size, blockSize, length of JPG file;
8:       record  $\langle key, value \rangle$  in index structure for TF;
9:       size = length of OF;
10:      write OBJ file to OF;
11:      analyze the MTL file and append it to OF;
12:      set  $\langle key, value \rangle$  of MTL and put it in index of OF;
13:      merge each texture file;
14:      set  $\langle key, value \rangle$  of each texture file and put it in index of OF;
15:      set  $\langle key, value \rangle$  with pathname of OBJ and size, sum of file lengths, length of OBJ;
16:      record  $\langle key, value \rangle$  of OBJ in index for OF;

```

---

There are two types of file accesses, one for a single thumbnail file and the other is to get OBJ, MTL and texture files of a model. The method is shown in Algorithm 2. Based on supplied pathname of the requested file (thumbnail file or OBJ file), the corresponding TF or OF as well as offset in large file and file length are identified (lines 1 to 4). For a thumbnail image, the algorithm reads the file data from the large file according to index information (line 5). Then, the read request is met (lines 6 to 7). For an OBJ file, it reads a chunk of data from OF which also contains MTL file and texture files (line 5). The metadata of MTL file in the in-memory index is acquired and the MTL data is determined from the data chunk, then the algorithm gets the related texture file names by analyzing MTL file and their metadata, and separates them from the whole data (line 9).

In order to show the contribution of batch reading to the performance gains, we will also test the variant of Algorithm 2 which reads OBJ, MTL and texture files one by one instead of acquiring a data chunk in the experimental study. In addition, since some small files are in the same large file, they can share the same file handler. Files do not need to be opened and closed frequently, which can reduce the system cost. This intuitive and side benefit will also be verified in the performance study.

---

**Algorithm 2.** File Access (*file*)

---

**Input:** *file*: the requested file      **Output:** the found files

- 1: *key* = pathname of *file*;
- 2: identify the TF or OF based on pathname of *file*;
- 3: get *value* from index based on *key*;
- 4: parse *value* to calculate *offset* and *length*;
- 5: read *data* from TF or OF according to *offset* and *length*;
- 6: **if** *file* is a thumbnail image JPG file **then**
- 7:     return;
- 8: **else**
- 9:     split *data*;

---

## 4 Experimental Study

Our test platform contains 17 server nodes. One node acts as NameNode, and the other 16 nodes are DataNodes. Each server runs Linux operating system with 2.6.32 kernel and is equipped with two 2.93GHz Intel CPUs and 48GB memory. The Hadoop version used is 2.0.3 and the number of replicas is set to 3. There are 25 major categories and 409 minor categories in our current real 3D models. Synthetic data sets are generated based on the real data to verify the performance of algorithms. The sizes of files range from a few kilobytes to tens of megabytes. The detailed distributions of sizes and numbers of files are omitted here due to space limitation.

Our strategy is compared with a baseline approach. The baseline storing method uses Hadoop simply, that is it stores original files in HDFS directly and original files are kept as separate small files and not merged into large ones, which will lead to problems introduced in Section 2. When accessing files, the baseline approach requests files from where the data are loaded using the baseline storing approach.

### 4.1 Data Loading

Fig. 4 shows the execution time for the proposed Algorithm 1 and the baseline approach when the size of data loaded into HDFS is varied from 50GB to 500GB, which means the maximum size of data actually loaded and stored in HDFS is 1.5TB since the number of replicas is 3. As the data size is increased, both of them spend more time on the batch file writing, while our method achieves a speedup between 4.7 and 5.9 on the baseline method. For example, the proposed approach takes 862 seconds to load 100GB data (actually 300GB) while the baseline one needs 4,823 seconds. The memory usage of NameNode is monitored and the baseline method consumes more space than ours, which is 1.6GB and 574MB respectively when loading 100GB data. This is due to extra file metadata maintained by the baseline method.

### 4.2 File Access

The performance of reading thumbnail files is displayed in Fig. 5 and Fig. 6. Fig. 5 shows the time of three algorithms when accessing 1,000 to 10,000 files randomly on

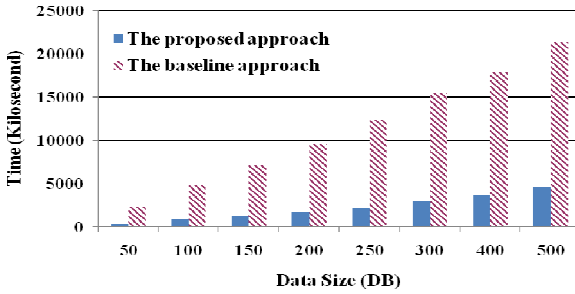


Fig. 4. Data loading time w.r.t data size

50GB data, and Fig. 6 illustrates the results of random accesses of 12,000 to 20,000 thumbnail images on 100GB data (the relative performance on more data is similar thus omitted due to space limitation). “BA” represents the baseline file accessing approach. “PA1” refers to Algorithm 2 with each request leading to an open file descriptor, while “PA2” is Algorithm 2 with shared file descriptors. The three algorithms show similar performance trends as the number of reads is increased in two figures. PA2 provides a speedup between 4.2 and 5.6 over BA. The side benefit can be validated by comparing PA1 with PA2. It is obvious that PA2 spends less time than PA1 and it affords an improvement up to 47% compared to PA1.

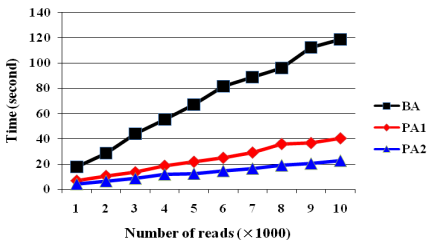


Fig. 5. Access time w.r.t images on 50G

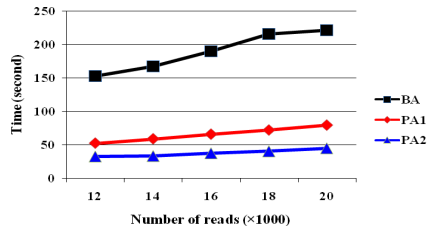


Fig. 6. Access time w.r.t images on 100G

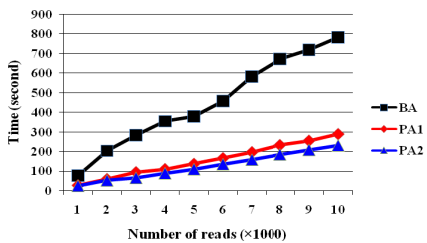


Fig. 7. Access time w.r.t models on 50G

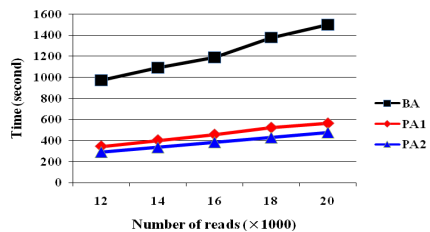


Fig. 8. Access time w.r.t models on 100G

The execution time of accessing OBJ, MTL and texture files of 1,000 to 10,000 models randomly is shown in Fig. 7. Fig. 8 is the performance with more model accesses. The variant of Algorithm 2, which reads OBJ, MTL and texture files one by one from large files, is represented by “PA1” in figures. Algorithm 2 is denoted by “PA2”. Both PA1 and PA2 reduce the chance of opening files as far as possible by sharing file descriptors. They perform better than BA, and PA2 runs faster than PA1. The baseline approach is 3.1-4.0 times slower than PA2, and PA2 is better than PA1 by a factor of up to 20%. For instance, when 10,000 models are accessed, PA2 and PA1 need 230 and 289 seconds respectively, while BA takes 783 seconds.

## 5 Related Work

We have introduced and analyzed Hadoop distributed file system in Section 2. Detailed description and explanation can be found in [2], [3]. Google Bigtable system [8] is a distributed and non-relational database-like storage system for structured data and used in Google’s projects. Facebook [9] describes an object storage system optimized for its photos application. Unlike these work, this paper proposes a solution for 3D models to adapt to Hadoop distributed file system and presents efficient data structures and algorithms to achieve high performance.

## 6 Conclusion

This paper describes our first attempt to solve the big data management problem of 3D models using Hadoop. The key insight is how to arrange 3D model files based on data correlation and access patterns. Two kinds of application-specific storage models are proposed to store and index different types of files. Efficient file accessing and data loading methods are introduced to handle user requests and file storing. The experiments demonstrated the effectiveness and efficiency of our techniques. For future work, we will study parallel strategies on clusters for 3D big data management.

## References

1. ComputerWorld. World’s data will grow by 50X in next decade. IDC study predicts (2011), <http://www.computerworld.com/s/article/9217988>
2. The Hadoop project, <http://hadoop.apache.org/>
3. Ghemawat, S., Gobioff, H., Leung, S.T.: The google file system. In: Proc. ACM Symposium on Operating Systems Principles, pp. 29–43 (2003)
4. Takahiko, F., Ryutarou, O.: Dense sampling and fast encoding for 3D model retrieval using bag-of-visual features. In: Proc. CIVR (2009)
5. Toldo, R., Castellani, U., Fusiello, A.: The bag of words approach for retrieval and categorization of 3D objects. *The Visual Computer* 26, 1257–1268 (2010)
6. Princeton 3D Model Search Engine, <http://shape.cs.princeton.edu/search.html>

7. Dean, J., Ghemawat, S.: Mapreduce: Simplified data processing on large clusters. In: Proc. OSDI, pp. 137–150 (2004)
8. Chang, F., Dean, J., Ghemawat, S., Gruber, et al.: Bigtable: A distributed storage system for structured data. *ACM Trans. Comput. Syst.* 26(2), 1–26 (2008)
9. Beaver, D., Kumar, S., Li, H.C., Sobel, J., Vajgel, P.: Finding a needle in Haystack: Facebook’s photo storage. In: Proc. OSDI, pp. 47–60 (2010)

# Double Hot/Cold Clustering for Solid State Drives

Taedong Jung, Yongmyoung Lee, Junhyun Woo, and Ilhoon Shin\*

NowonGu GongleungDong, Seoul National University of Science and Technology,  
Building 3, 206, Seoul 139-743, South Korea  
ilhoon.shin@snut.ac.kr

**Abstract.** Solid State Drives (SSDs) which connect NAND-flash memory in parallel is going to replace Hard Disk Drives (HDDs). The physical trait of SSD is different from HDD and it uses Flash Translation Layer (FTL) to emulate HDD. Garbage Collection is the process of making an available sphere in SSD which has no function of an over-write. In the case of general policy of FTL, both of valid and invalid pages are randomly mixed, but if it could be separated, an effect of Garbage Collection can be improved. This paper presents a double hot/cold clustering scheme that separates the frequently overwritten region from the opposite. The performance evaluation result shows that the improvement is between 44.3% in maximum and 3.9% in minimum.

**Keywords:** garbage collection, hot/cold clustering, page mapping, flash translation layer, NAND flash memory.

## 1 Introduction

NAND-flash memory is not a new medium of saving but it is a friendly medium around us. NAND-flash memory has a fast speed of I/O and a trait of strong resistance to impact as semiconductor, thus it is used in the various areas. Solid State Drive(SSD) which is connected in the parallel way by NAND-flash memory has a strong point of its memory, besides the problem of high-price is solved recently to low-price, it is used as a saving medium more than before instead of Hard-Disk. However, it cannot be used its way in a general file system, the reason is that NAND-flash memory has some different point about the method of I/O compare to Hard-Disk. NAND-flash memory consists of page-unit for both of reading and writing, but Hard-Disk consists of unit for sector. Also, NAND-flash memory need the process of erasing prior to process of new writing compare to Hard-Disk which has a function of covering. For overcoming this difference, it can be possible to put the software named as Flash Translation Layer (FTL) into class between fine-system and NAND-flash memory and use NAND-flash memory as general saving medium.

NAND-flash memory is impossible to overwrite, therefore it should be via the way of erasing at first for using the data in already-occupied position of data. However, NAND-flash memory is used as a unit for page, but the process of erasing is a unit for

---

\* Corresponding author.

block. It means that Overhead is emerging with the process of erasing in greater scale for the work of writing as small unit in occasion. For overcoming this problem, mapping table as a one of the main function in FTL can be used. Mapping table is the table that saves both of logical address from the order of file-system and physical address of saving the real data. If there is already a written data in the ordering position of file-system, the data should be writing in a different area and saving its pertinent position to the mapping table, not just erasing the pertinent block and rewriting for update. At the moment, the ordering position from file-system is logical address, and the written position for actual data is physical address. Now, the area which had a prior data should be seen as invalid page, if these process of writing goes constantly, there are mixed both of valid and invalid page in the block.

If all pages in NAND-flash memory are used, Garbage Collection will be processing. After Garbage Collection selects the victim block, the valid pages in victim block is transferring to free-block which is administered for Garbage Collection. After that, the free-page is created as much as space that was occupied of invalid page in the victim block, thus new writing process can be possible. If the block which has lots of area of invalid page in the case of selecting the victim block, time for implementing Garbage Collection is decreased because it can be possible to lessen the amount of valid page for free-block, and more writing process can be possible because more free page is created. In conclusion, NAND-flash memory's function is increased as there are more invalid pages in the victim block when Garbage Collection is selected. However, in the occasion for general policy of FTL, both of valid and invalid page is mixed randomly in the block, thus when progress during garbage collection, there is a difficult point for selecting the victim block that has lots of invalid page.

FTL of Double Hot/Cold Clustering scheme is categorizing whether data is updated frequently or not in the process of Garbage Collection, thus its purpose is going to create the block which has lots of invalid page. In the block, the page which has data that is often updated, the probability of being invalid page is more. If there is crated the blocks that have the invalid page overwhelmingly in the process of Garbage Collection via categorizing the data, the efficiency of Garbage Collection is maximized. Garbage Collection process needs a longer time than the process of reading and writing in NAND-flash memory, thus if it can be implemented more effectively, it will help to improve whole function.

Section 2 explains Page Mapping policy. Section 3 explains the structure of Double Hot/Cold Clustering scheme, and section 4 describes the environment and result of experiment. Finally, in section 5, the conclusion is derived.

## 2 Related Work

FTL is classified into Page Mapping scheme, Block Mapping scheme, Hybrid Mapping scheme, etc, depending on the administration by mapping table. Among them, Double Hot/Cold Clustering scheme is based on Page Mapping Scheme. Page Mapping Scheme administers the mapping table by unit of page, thus the mapping

table by unit of page is expected of the fast process because the minimum unit of reading and writing in NAND-flash memory is page. However, it is administering the data by smaller unit compare to both of Block Mapping scheme and Hybrid Mapping scheme, and it has a short point of bigger size of Page Mapping scheme.

In NAND-flash memory, the process of writing is different from the position of saving between via the file-system and in the actual NAND-flash memory as it is described in section 1. The new data will be saved in the block for FTL's writing process of Page Mapping scheme by unit of page, the applicable page is saved on Page Mapping Table. When it is ordered via file-system, FTL will be confirming via Page Mapping Table whether there is already written data or not on the applicable data. If there is already written page on page that is ordered, data is going to be written on another page and Page Mapping Table will be changed into the new written page. The page which has nothing on Page Mapping Table is going to be invalid automatically, because NAND-flash memory tries to find the data while it is looking on Page Mapping Table.

After all the valid pages are used at all in the general Page Mapping Scheme, the new valid page is created via Garbage Collection. After the victim block is selected, and only the valid page in the victim block should be transferring to the extra block for Garbage Collection. The mapping table should be corrected by the changed physical address, because the transferred page's physical address is changed. After erasing the victim block, and then the additional block for Garbage Collection is going to be used. Now, the additional block is used as new writing block, it can be possible to use the sphere for new writing block as much as the amount of occupied space of invalid page in the victim block.

### **3 Double Hot/Cold Clustering**

Garbage Collection has been spending lots of time compare to the process of reading and writing because it operates of erasing the victim block with the process of transferring the multiple valid pages. If the block that has many invalid pages is selected when the time of Garbage Collection choose the victim block, an operation time of transferring the page could be benefited because the amount of valid page for transferring will be lessened, and the new sphere of writing will be created as much as the amount of en-invalid page, then more writing process can be possible until the next Garbage Collection. Therefore, in the general FTL schemes, the victim block which has the largest amount of invalid pages on block. However, the probability of block case that has more invalid page overwhelmingly is very low because both of the valid and invalid page are randomly mixed on block. Double Hot/Cold Clustering scheme is an idea that is creating the block which has more invalid page overwhelmingly then it could be improving the efficiency of Garbage Collection.

Double Hot/Cold Clustering scheme are the scheme that assembles two idea of papers. At first, Woo [1] distributes via process of writing whether it has been updated well or not. Shin [2] separates whether it has been updated well or not its valid page on the victim block. Double Hot/Cold Clustering scheme is distributing



whether it has been updated well or not on the process of Garbage Collection. More updated pages could be saved on certain blocks like categories prior three schemes. The page which updates frequently has lots of probability for the invalidity after the sequential update, certain blocks can be possible to be an invalid page greatly as time passes by.

Double Hot/Cold Clustering scheme administers two blocks simultaneously for current process of writing. The first data of writing is recognized as data of less updated, and it will be used on the cold block. If current data of writing is data of frequently updated, it will be recognized as data of more updated, and it will be used on the hot block. In conclusion, data which has not been updated saves on the cold block, and data which has been updated well saves on hot block. If the pages on cold block or hot block for current process of writing are spent at all, it is going into the waiting line for Garbage Collection, and the new block will be allocated for using the cold or hot block.

The new free pages are created via Garbage Collection like an occasion of using all the free pages of NAND-flash memory. Garbage Collection of Double Hot/Cold Clustering scheme also begins with selecting the victim block. However, it is a different point that the valid page on victim block can be administered separately. On the time of Garbage Collection, the valid page on the victim block is transferring into the additional cold block. The valid page on the victim block is recognized as less updated page, because it was not updated until the appropriate time. After that, the victim block is erased, and the new hot or cold block can be used. If the free page on the additional cold block with the constant Garbage Collection is spent at all, the new type of Garbage Collection will be operated. In this case, the new victim block is selected again, and the valid page on the pertinent block is transferring into the spare block. The existing additional cold block is going into the waiting line for Garbage Collection, the spare block is using as the new additional cold block. The final process erases the new victim process and uses the new spare block.

In conclusion, Double Hot/Cold Clustering scheme is distributing whether it has been updated well or not on writing process and gathering the less updated data independently on the process of Garbage Collection.

## 4 Performance Evaluation

The Open SSD platform [3] is used for evaluating the efficiency of the schemes. Open SSD platform is using the Jasmine Board as an actual SSD Hardware platform. The Jasmine Board is using Barefoot of the corporation of Indilinx, and it has the capacity of 96KB SRAM, SDRAM of 64MB, and NAND-flash memory of 64GB.

The evaluation of efficiency with The Jasmine Board is using two trace files. Trace file is an I/O pattern of users via using the tool of diskmon tool in PC that is installed NTFS file system, windows 7 and XP. It is tested with making two trace files, a trace file is creating the pattern of amount of partition about 45GB, and trace file is extracted, and the another trace file is creating the pattern of amount of partition about

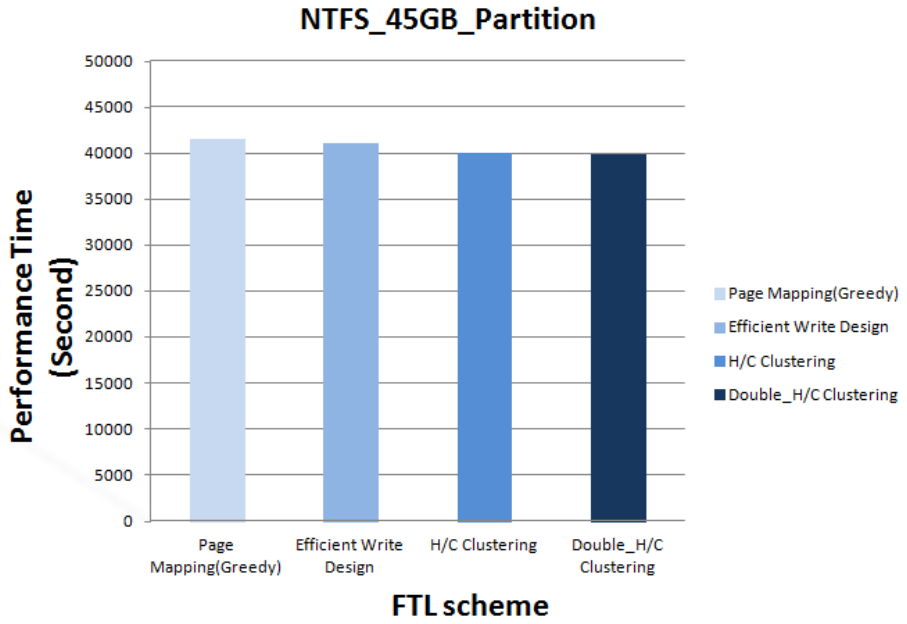


Fig. 1. Elapsed time in NTFS\_45GB partition (window 7)

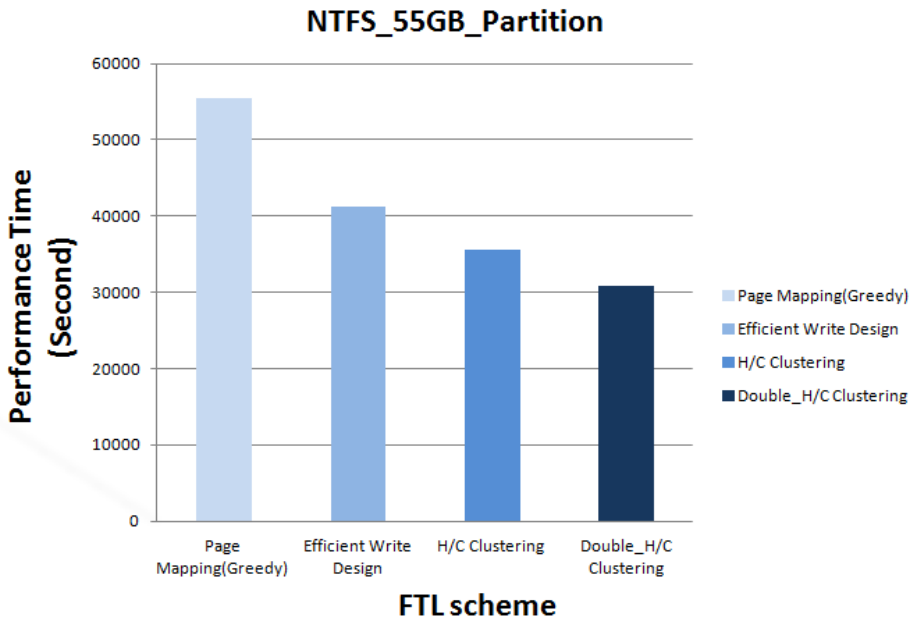


Fig. 2. Elapsed time in NTFS\_55GB partition (window xp)

55GB, and trace file is extracted. The evaluation of efficiency compare and evaluate with FTL schemes of Double Hot/Cold Clustering and two papers('An Efficient FTL design for NAND flash memory', 'Hot/Cold Clustering for Page Mapping in NAND flash memory') which are based on its schemes, and FTL of Paper Mapping scheme(Greedy scheme).

The result of time-measurement with using the trace-file of 45GB partition shows the increasing-ratio of approximately 3.92% compares to Page Mapping Schemes. Relatively, the improvement of function is not good. The result of time-measurement with using the trace-find of 55GB partition shows the great result of the ration of functional improvement about 44.3% compares to Double Hot/Cold Clustering scheme.

## 5 Conclusion

It can be verified that Double Hot/Cold Clustering scheme which this paper proposed is distributing whether is has been updated well or not with the efficiency. It helps to select the victim block for progressing Garbage Collection efficiently while more invalid pages are gathered into some blocks. This result shows the improvement of speed in process about I/O compares to the existent Page Mapping scheme even though it has a difference of function depend on I/O pattern by user. Especially, it is interesting fact that nearly 44.3% of functional improvement was showed depend on I/O pattern by user. The efficiency of Garbage Collection can't be disregarded when the time of using NAND-flash memory.

**Acknowledgments.** This study was financially supported by Seoul National University of Science and Technology.

## References

1. Shin, I.: Hot/Cold Clustering for Page Mapping in NAND flash memory. *IEEE Transactions on Consumer Electronics* 57, 1728–1731 (2011)
2. Woo, J.: An efficient FTL design for the NAND flash memory. Master Thesis. Seoul National University of Science & Technology (2013)
3. The OpenSSD Project,  
[http://www.openssproject.org/wiki/The\\_OpenSSD\\_Project](http://www.openssproject.org/wiki/The_OpenSSD_Project)

# A Case Study for Cyber Physical System with Hybrid Relation Calculus

Shuguang Feng and Lichen Zhang

Shanghai Key Laboratory of Trustworthy Computing  
East China Normal University  
Shanghai 200062, China  
zhanglichen1962@163.com

**Abstract.** Cyber physical systems are a new concept that combines discrete and continuous dynamics to cooperate. The existing modelling language cannot model cyber-physical systems well for that the discrete dynamics is represented by finite state machine and continuous dynamics is represented by differential equations. In this paper, we use a new formal modelling language, Hybrid Relation Calculus, proposed by Jifeng He to model the cyber-physical systems and we apply Hybrid Relation Calculus to specify steam boiler control system.

**Keywords:** Cyber physical systems, Hybrid Relation Calculus, specification, steam boiler control system.

## 1 Introduction

Cyber physical systems combines discrete dynamics and continuous dynamics. Discrete dynamics are represented by the control system and continuous dynamics are represented by the physical world. The control system can be modelled by finite state machine and the physical world can be modelled by differential equations.

Hybrid Relation Calculus, proposed by Jifeng He, can represent cyber-physical systems. Most modelling language can only model the discrete dynamics, for example: UML can model the discrete dynamics with its state machine. The continuous dynamics are generally represented by differential equations. So the new modelling language Hybrid Relation Calculus is proposed. The clock theory also propose by Jifeng He can help the modeling of this system.

In Hybrid Relation Calculus, the program is presented hybrid program. The alphabet of hybrid program has both discrete and continuous variables. The equation uses continuous variables to represent the physical world. The rules in Hybrid Relation Calculus can control the start and stop of a hybrid program.

## 2 An Overview of Hybrid Relation Calculus

In this section, the overview of Hybrid Relation Calculus is showed. The hybrid modelling language proposed is as follows:

$$\begin{aligned}
 A &::= \mathbf{skip} \mid \mathbf{chaos} \mid x := e \mid \mathbf{delay} \mid !s \\
 EQ &::= F(v, \dot{v}) = 0 \mid \mathbf{idle} \mid EQ \parallel EQ \\
 P &::= A \mid P \cap P \mid P ; P \mid P \parallel P \mid P \triangleleft b \mathit{rhd} P \mid EQ \ g \mid \mathbf{whenG} \mid \mu X \bullet P(X) \\
 G &::= g \& P \mid G \parallel G \\
 g &::= s \mid g \cdot g \mid g + g
 \end{aligned}$$

A is the alphabet of the hybrid program. The simple hybrid programs **skip**, **chaos**, **delay** are three basic programs in Hybrid Relation Calculus. EQ is the equation in program. EQ is used to model the physical world. P is the program and can have combination with other program using combinator. G is the guard rule to control the start of a program. While g is the guard value, g can contain signal.

### 3 Case Study

In this section, we describe a Steam Boiler Control System using Hybrid Relation Calculus. As it is in figure 1, the Steam Boiler Control System has a water tank to keep the water from the pump. While the current volume of the water goes up to the maximal water level, the sensor sends a message to the control system which can send commands to the pump. Then the control system sends command to the pump to decrease the velocity of the pump. While the current volume of the water drops to the minimal water level, the sensor sends a message to the control system. Then the control system send a command to the pump to increase the velocity of the pump. There is a steam rate while determined by the environment. Table 1 is the variable used in this case study.

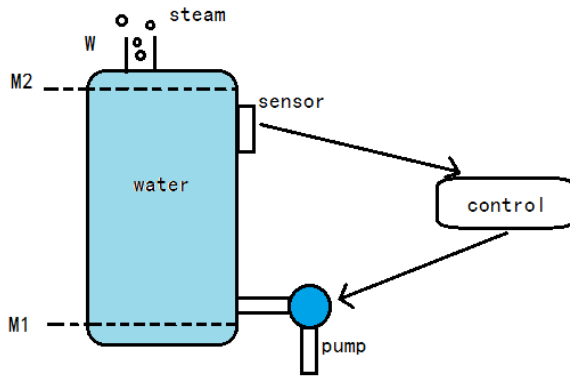


Fig. 1. The Stream Boiler Control System

**Table 1.** Variables in Steam Boiler Control System

Parameter	Value
W	the steam rate
$v_p$	current speed of the pump
e	volume of the current water
$v_s$	current speed of steam
$M_1$	minimal volume of water
$M_2$	maximal volume of water

According to the Hybrid Relation Calculus, the Steam Boiler Control System can be described in this way:

The alphabet of this system is :

$$A ::= \mathbf{skip}|\mathbf{chaos}|\mathbf{delay}|x:=v|!s$$

The system may in a mass while processing, delaying, or right steps. So, **skip**, **chaos** and **delay** are used. While the control system sending a command to the pump, the value of velocity of pump is assigned by  $x:=v$ . The  $!s$  is used by sending a message from sensor to the control system.

The equation of the system is:

$$\begin{aligned} EQ &::= F(e,\dot{e}) = 0|\mathbf{idle} \\ F_{in} &= e + (v \times t + \dot{e}) = 0 \\ F_{out} &= e - (v \times t + \dot{e}) = 0 \end{aligned}$$

The process of the system can be in equation  $F_{in}$  or  $F_{out}$  and the process can also be in **idle** which means doing nothing.

There are mainly two processes in this system. The  $P_{in}$  and  $P_{out}$  which mean increasing pump velocity and decreasing pump velocity.

$$P ::= A | P_{in};P_{out} | P_{in} \triangleleft b \triangleright P_{out} | EQ \mathbf{unit}!g | \mu X \bullet P(X)$$

A process  $p$  can carry out the alphabet action or connect with other processes with the combinator in Hybrid Relation Calculus. The  $P_{in}$  and  $P_{out}$  can be recursive for the requirement.

$$\begin{aligned} G &::= g \ \& \ P \ G || G \\ g &::= s|g \cdot g |g+g \end{aligned}$$

When the guards of  $P_{in}$  and  $P_{out}$  are true, the processes can go on to be executed. The signals are in guard  $g$ . For example, if the guards  $present(s)$  is true, signal  $s$  is sent accurately.

The whole cyber physical system of the case study can be modeled by Hybrid Relation Calculus on the above formulas.

The continuous behaviors are described by the two equations.

$$EQ_{in} \text{ until } present(s_{in})$$

$$EQ_{out} \text{ until } present(s_{out})$$

When the command of increasing the volume of water is sent out, it means the signal  $s_{in}$  is transferred to the pump and the value of  $present(s_{in})$  is true. Then the guard  $g$  of  $EQ_{in}$  is true.

The sensor may send messages to the control system. The behavior is discrete. When the volume of the water runs up to maximal volume of the water  $M_2$ . The sensor is waken up and start to send message to the control system.

$$e + \int_0^t vt - \int_0^t Wt = M_2$$

The value  $t$  can be calculated and sent to the sensor as a trigger which makes the sensor send message to the control system.

$$clock(decrease) \text{ until } clock(t_{in})$$

The next formulas can calculated the time of sending message to the control system. The trigger is described by the clock theory.

$$e + \int_0^t vt - \int_0^t = M_1$$

$$clock(increase) \text{ until } clock(t_{out})$$

## 4 Conclusion

In this paper, we introduce the concept of cyber physical systems and Hybrid Relation Calculus. Cyber physical systems combine discrete and continuous dynamics. The Hybrid Relation Calculus can model the discrete and continuous dynamics of cyber physical systems. We model the Steam Boiler Control System as a case study. The Steam Boiler Control System can be represented by the Hybrid Relation Calculus with hybrid programs and its relations. The hybrid program use discrete variables to model the control system. The continuous variables can be used to model the physical world with the differential equations.

In the future, we focus on developing a tool for the verification of Hybrid Relation Calculus.

**Acknowledgment.** This work is supported by Shanghai Knowledge Service Platform Project (No.ZF1213), national high technology research and development program of China (No.2011AA010101), national basic research program of China (No.2011CB302904), the national science foundation of China under grant (No.61173046, No.61021004, No.61061130541, No.91118008), doctoral program foundation of institutions of higher education of China (No.20120076130003), national science foundation of Guangdong province under grant (No.S2011010004905).

## References

1. Sha, L., Gopalakrishnan, S., Liu, X., et al.: Cyber-physical systems: A new frontier. In: *Machine Learning in Cyber Trust*, pp. 3–13. Springer US (2009)
2. Poroor, J., Jayaraman, B.: Formal analysis of event-driven cyber physical systems. In: *Proceedings of the First International Conference on Security of Internet of Things*, pp. 1–8. ACM (2012)
3. Lee, J., Maeng, J.C., Song, B., et al.: software-Based Fault Detection and Recovery for Cyber-Physical Systems. In (Jong Hyuk) Park, J.J., Ng, J.K.-Y., Jeong, H.Y., Waluyo, B. (eds.) *Multimedia and Ubiquitous Engineering. LNEE*, vol. 240, pp. 1107–1112. Springer, Heidelberg (2013)
4. Jifeng, H.: *Hybrid Relation Calculus*. East China Normal University (2013)
5. Liu, S., et al.: *Formalizing UML State Machine Semantics for Automatic Verification* (2012)
6. Kurpick, T., Pinkernell, C., Look, M., et al.: Modeling cyber-physical systems: model-driven specification of energy efficient buildings. In: *Proceedings of the Modelling of the Physical World Workshop*. ACM (2012)
7. He, J.: *Link Continuous World with Discrete World*. In: *The 10th International Colloquium on Theoretical Aspects of Computing*, Shanghai, China (September 2013) (unpublished) (Keytalk)



# Optimizing Functional Link Neural Network Learning Using Modified Bee Colony on Multi-class Classifications

Yana Mazwin Mohmad Hassim and Rozaida Ghazali

Faculty of Computer Science and Information Technology  
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia  
{yana,rozaida}@uthm.edu.my

**Abstract.** Functional Link Neural Network (FLNN) has emerged as an important tool for solving classification problems and widely applied in many engineering and scientific problems. FLNN is known to be conveniently used as compared to ordinary feed forward network like the Multilayer Perceptron (MLP) due to its flat network architecture which employs less tuneable weights. The standard method for tuning the weight in FLNN is using a Backpropagation (BP) learning algorithm. However, BP-learning algorithm has difficulties such as trapping in local minima and slow convergence especially for solving non-linearly separable classification problems. In this work, a modified Artificial Bee Colony (mABC) is used to recover the BP drawbacks. With modifications on the employed bee's exploitation phase, the implementation of the mABC as a learning scheme for FLNN is expected to give a better accuracy result for the classification tasks.

**Keywords:** Classification; Functional Link Neural Network; Artificial Bee Colony Algorithm.

## 1 Introduction

Classification is one of the most frequent studies in the area of Artificial Neural Networks (ANNs) and mostly involved in decision making task of human activity [1]. One of the best known types of ANNs is the Multilayer Perceptron (MLP). However, difficulties in fixing appropriate number of neurons and determining appropriate number of hidden layers has make the MLP architecture not rather easy to train. To overcome this, a Functional Link Neural Networks (FLNN) [2] which has single layer of trainable connection weights is used. This single layer property also make the learning algorithm used less complicated compared to MLP network [1].

The standard learning scheme of FLNN is Backpropagation (BP) learning algorithm. The BP-learning developed by Rumelhart [3] is the most well-known and widely used for training ANNs where the model would "learn" through a sample set of data. However, one of the crucial problems with the standard BP-learning algorithm is that the gradient search techniques tend to easily get trapped in local minima especially for those non-linearly separable classification problems [4]. To recover the drawback of BP-learning, the Artificial Bee Colony (ABC) optimization [5], is used to optimize the FLNN weights. The ABC algorithm was originally

proposed by Karaboga [6] for solving numerical optimization problem by simulating the intelligent foraging behavior of a honey bee. In this study, a modified ABC is used for training the FLNN. The modification is done on the employed bees' foraging behavior to exploit the entire FLNN dimension vector in order to get better accuracy on classifying the out-of-sample or unseen data for multi-class classification problems.

## 2 Background and Related Works

FLNN is a type of Higher Order Neural Networks (HONNs) created by Pao [7] and has been successfully used in many applications such as classification [8-10], pattern recognition [11, 12] and prediction [13, 14]. FLNN is much more modest than MLP as it has a single-layer of trainable weights compared to the MLP whilst it able to handle a non-linear separable classification tasks.

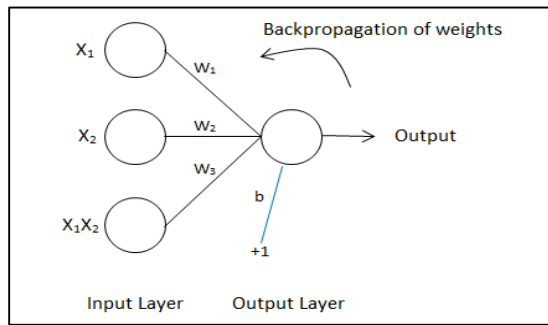


Fig. 1. The 2<sup>nd</sup> order FLNN structure with 2 inputs

Fig.1 presents the FLNN structure up to the second order with 2 inputs. The first order consist of the 2 inputs  $x_1$  and  $x_2$ , while the second order of the network is the extended input based on the product unit  $x_1x_2$ . The learning part of this architecture on the other hand, consists of a standard BP-learning as the training algorithm.

### 2.1 Drawback of FLNN Learning Scheme

Most previous learning scheme used for FLNN training, is the BP-learning algorithm [1, 4, 15]. The idea of the BP-learning algorithm is to reduced error by tuning the connection weights until the networks learned the training data. As shows as in Fig. 1, the weight values between enhanced input nodes and output node  $w_1, w_2, w_3$  and also bias  $b$ , are randomly initialized. The square error  $E$ , between the target output and the actual output will be minimized as:

$$E = \frac{1}{2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

where  $y_i$  is the target output and  $\hat{y}_i$  is the actual output of the  $i$ th input training pattern, while  $n$  is the number of training pattern. During the training phase, the BP-learning algorithm will continue to update  $w$  and  $b$  until the maximum epoch or the convergent condition is reached (minimum error).

Although BP-learning is the mostly used algorithm in the training of FLNN, the algorithm however has limitation; it make FLNN-BP strictly depends on the shape of error surface and prone to stuck in some local minima when moving along the error surface (since a common error surface may have many local minima and multimodal) which affect the performance of FLNN.

## 2.2 Basic of Artificial Bee Colony Optimization

The Artificial Bees Colony (ABC) algorithm is an optimization tool, inspired by foraging behavior of a honey bee swarm introduced by karaboga [6]. This model simulates the interaction between three groups of bees:

- i. **Employed bee:** uses random multidirectional search space in the Food Source area (FS) carry the profitability information (nectar quantity) of the FS based on Eq. (2).

$$v_{i,j} = x_{i,j} + \Phi_{i,j}(x_{i,j} - x_{k,j}) \quad (2)$$

- ii. **Onlooker bee:** evaluate the nectar quantity obtained by the employed and bees and choose FS depending on the probability value base on the fitness base on Eq. (3).

$$p_i = \frac{fit_i}{\sum_{i=1}^{FS} fit_i} \quad \text{where} \quad fit_i = \begin{cases} \frac{1}{1+fit_i}, & \text{if } fit_i \geq 0 \\ 1 + abs(fit_i), & \text{if } fit_i < 0 \end{cases} \quad (3)$$

- ii. **Scout bee:** the employed bee whose food source has been abandoned start to search for finding a new food source randomly using the following:

$$x_i^j = x_{min}^j + rand(0,1)(x_{max}^j - x_{min}^j) \quad (4)$$

## 3 FLNN with Modified ABC Learning Scheme

In this study, a modified ABC (mABC) is used as a learning scheme for training the FLNN. The modification is done on the part of employed bees' foraging phase so that they would exploit all weights and biases in the FLNN weight vector in order to improve the network ability on searching the optimal weights set. In the standard ABC algorithm, the position of a food source (FS) represents a possible solution to the optimization problem, and the nectar amount of a food source corresponds to the profitability (fitness) of the associated solution. In the case of training the FLNN with ABC, the weight,  $w$  and bias,  $b$  of the network are treated as optimization parameters (minimizing error,  $E$ ) presented in Eq. (1). The FLNN optimization parameters is treated as  $D$ -dimensional vector for the solution  $x_{i,j}$  where ( $i = 1, 2, \dots, FS$ ) and ( $j = 1, 2, \dots, D$ ) and each vector  $i$  is exploited by only one employed bee. In order to

produce a candidate food source  $v_{i,j}$  from the old one  $x_{i,j}$  in memory, the ABC uses Eq. (4) where  $k \in \{1, 2, \dots, FS\}$  and both  $k$  and  $j$  are a randomly chosen indexes. The food source of  $x_{i,j}$  can be represented in a form of  $X = FS \times D$  matrix.

$$X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,D} \\ \vdots & \ddots & \vdots \\ x_{FS,1} & \cdots & x_{FS,D} \end{bmatrix} \quad (5)$$

As can be seen from Eq. (4) and matrix representation from Eq. (5), for each row of  $FS$  only one element from  $D$  will be chosen randomly and exploited by the employed bee by using:

$$j = \text{fix}(\text{rand} * D) + 1; \quad (6)$$

However in the case of FLNN mainly for classification tasks which are always deal with large number of optimization parameters (weights + bias), exploiting one element in each solution vector  $x_i$  will cause longer foraging cycle in finding the optimal solution [16]. Random selection of elements in each vector  $x_i$  during employed bee phase also leads to a poor ability for FLNN network in finding the optimal weights set resulted to a poor classification accuracy on unseen data [17]. To overcome this, we eliminate the random employed bee behavior in selecting elements in dimension vector as in Eq. (6). Instead, we direct the employed bee to exploit all elements in  $D$  before evaluating the vector  $x_i$ . The modified ABC (mABC) is performed as shown in pseudo code below, where the box indicates the improvement made to the standard ABC:

- 1) Cycle = 0
- 2) Initialize FLNN optimization parameters,  $D$  and population of scout bee with random solution  $x_i, i=1, 2, \dots, FS$
- 3) Evaluate fitness of the population
- 4) Cycle = 1:MCN
- 5) Form new population ( $v_i$ ) for employed bees
 

<ol style="list-style-type: none"> <li>i. select random neighbor, <math>k</math> in the neighborhood of <math>i</math>,</li> <li>ii. start loop For <math>j = 1:D</math></li> <li>iii. Direct employed bee to exploit nectar values of <math>j</math> in population (<math>v_{i,j}</math>) using Eq. (4) where <math>j = 1, 2, \dots, D</math> is a dimension vector in <math>i</math></li> <li>iv. exit loop when <math>j = D</math>;</li> </ol>
---
- 6) evaluate the new population ( $v_i$ )
- 7) Apply greedy selection between  $v_i$  and  $x_i$
- 8) Calculate the probability values  $p_i$  for the solutions  $x_i$  using Eq. (4)
- 9) Produce the new solutions  $v_i$  for the onlookers from the solutions  $x_i$  selected depending on  $p_i$  and evaluate them
- 10) Apply the greedy selection process for onlookers
- 11) Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution  $x_i$  using Eq. (5)
- 12) Memorize the best solution
- 13) cycle=cycle+1
- 14) Stop when cycle = Maximum cycle number (MCN).

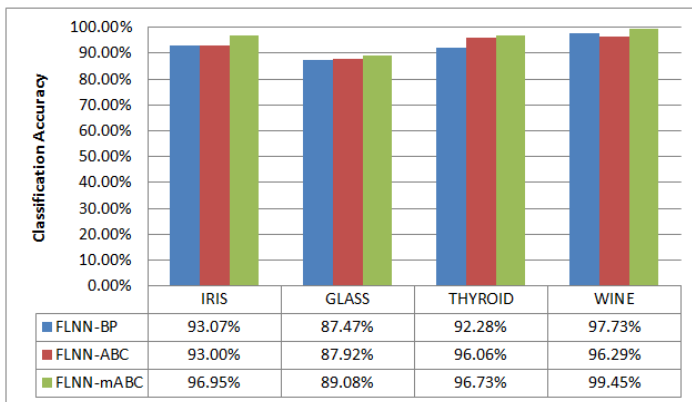
## 4 Experimentation Results

In order to evaluate the performance of the FLNN model trained with mABC (FLNN-mABC), simulation experiments were carried out on a 2.30 GHz Core i5-2410M Intel CPU with 8.0 GB RAM in a 64-bit Operating System. Simulations were performed on FLNN with BP-learning (FLNN-BP), FLNN with ABC (FLNN-ABC) and FLNN-mABC. The parameter setting considered in this experiment is summarized as Table 1 below.

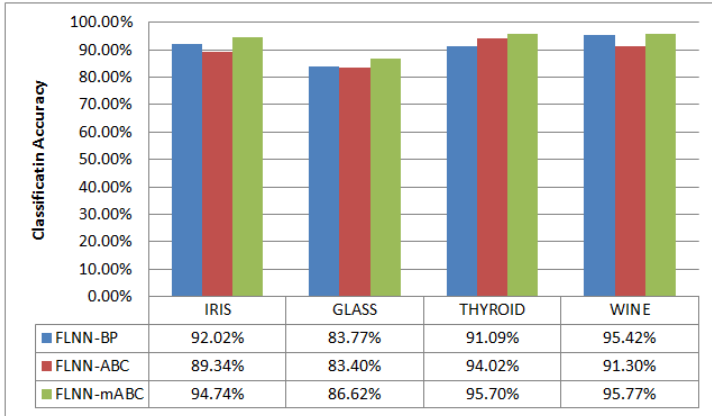
**Table 1.** Parameters considered for FLNN-BP, FLNN-ABC and FLNN-mABC simulation

parameters	FLNN-BP	FLNN-ABC	FLNN-mABC
Learning rate	0.3	-	-
Momentum	0.7	-	-
Maximum epoch/cycle	1000	1000	1000
Minimum error	0.001	0.001	0.001

In this work we considered 4 benchmark of multiclass classification problems obtained from the UCI Machine Learning Repository [18]; IRIS, GLASS, WINE and THYROID datasets. Ten trials were performed on each simulation of the FLNN-BP, FLNN-ABC and FLNN-mABC with the best accuracy result is noted. In order to generate the training and test sets, each datasets were randomly divided into two equal sets (1<sup>st</sup>-Fold and 2<sup>nd</sup>-Fold). Each of these two sets was alternately used either as training set or as a test set. The average accuracy values of each datasets result were then used for comparison. In Fig. 2 and Fig. 3, we will compare the results on multiclass classification problems both on training set and test set respectively. For the sake of convenience we set our FLNN input enhancement up to second order.



**Fig. 2.** Classification accuracy on training set (multiclass problems)



**Fig. 3.** Classification accuracy on test set (multiclass problems)

The plotted graph for training set in Fig. 2 shows that the performance results of FLNN-mABC are better than of FLNN-ABC and FLNN-BP. It gives significant results on IRIS, GLASS and WINE but gives competitive result on THYROID particularly between FLNN-mABC and FLNN-ABC. On the other hand, as depicted in Fig. 3, the FLNN-mABC performance on unseen has shown a clear boundary on IRIS, GLASS, THYROID and WINE test set.

## 5 Conclusion

In this work, we evaluated the FLNN-mABC model for the task of pattern classification for multiclass classification problems. The experiment has demonstrated that FLNN-mABC performs the classification task quite well. For the case of IRIS, GLASS, WINE and THYROID, the simulation result shows that the proposed modified ABC algorithm can successfully train the FLNN for solving classification problems with better accuracy percentage on unseen data. This research work is hopefully would provide a better training scheme for the FLNN in order to give better classification accuracy on unseen data.

## References

- Misra, B.B., Dehuri, S.: Functional Link Artificial Neural Network for Classification Task in Data Mining. *Journal of Computer Science* 3(12), 948–955 (2007)
- Pao, Y.H., Takefuji, Y.: Functional-link net computing: theory, system architecture, and functionalities. *Computer* 25(5), 76–79 (1992)
- Widrow, B., Rumelhart, D.E., Lehr, M.A.: Neural networks: applications in industry, business and science. *Commun. ACM* 37(3), 93–105 (1994)
- Dehuri, S., Cho, S.-B.: A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN. *Neural Computing & Applications* 19(2), 187–205 (2010)

5. Karaboga, D., Akay, B., Ozturk, C.: Artificial Bee Colony (ABC) Optimization Algorithm for Training Feed-Forward Neural Networks. In: Torra, V., Narukawa, Y., Yoshida, Y. (eds.) MDAI 2007. LNCS (LNAI), vol. 4617, pp. 318–329. Springer, Heidelberg (2007)
6. Karaboga, D.: An Idea Based on Honey Bee Swarm for Numerical Optimization. Erciyes University, Engineering Faculty, Computer Science Department, Kayseri/Turkiye (2005)
7. Pao, Y.H.: Adaptive pattern recognition and neural networks (1989)
8. Abu-Mahfouz, I.-A.: A comparative study of three artificial neural networks for the detection and classification of gear faults. *International Journal of General Systems* 34(3), 261–277 (2005)
9. Liu, L.M., Manry, M.T., Amar, F., Dawson, M.S., Fung, A.K.: Image classification in remote sensing using functional link neural networks. In: *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation* (1994)
10. Dehuri, S., Cho, S.-B.: Evolutionarily optimized features in functional link neural network for classification. *Expert Systems with Applications* 37(6), 4379–4391 (2010)
11. Klaseen, M., Pao, Y.H.: The functional link net in structural pattern recognition. In: 1990 IEEE Region 10 Conference on Computer and Communication Systems, TENCON 1990 (1990)
12. Park, G.H., Pao, Y.H.: Unconstrained word-based approach for off-line script recognition using density-based random-vector functional-link net. *Neurocomputing* 31(1-4), 45–65 (2000)
13. Majhi, R., Panda, G., Sahoo, G.: Development and performance evaluation of FLANN based model for forecasting of stock markets. *Expert Systems with Applications* 36(3, Part 2), 6800–6808 (2009)
14. Ghazali, R., Hussain, A.J., Liatsis, P.: Dynamic Ridge Polynomial Neural Network: Forecasting the univariate non-stationary and stationary trading signals. *Expert Systems with Applications* 38(4), 3765–3776 (2011)
15. Dehuri, S., Mishra, B.B., Cho, S.-B.: Genetic Feature Selection for Optimal Functional Link Artificial Neural Network in Classification. In: Fyfe, C., Kim, D., Lee, S.-Y., Yin, H. (eds.) IDEAL 2008. LNCS, vol. 5326, pp. 156–163. Springer, Heidelberg (2008)
16. Mohamad Hassim, Y.M., Ghazali, R.: Using Artificial Bee Colony to Improve Functional Link Neural Network Training. *Applied Mechanics and Materials* 263, 2102–2108 (2013)
17. Mohamad Hassim, Y.M., Ghazali, R.: Training a Functional Link Neural Network Using an Artificial Bee Colony for Solving a Classification Problems. *Journal of Computing Press* 4(9), 110–115 (2012)
18. Frank, A., Asuncion, A.: UCI Machine Learning Repository, School of Information and Computer Science, Irvine, CA. University of California (2010), <http://archive.ics.uci.edu/ml>

# Adult Contents Analysis and Remote Management Framework for Parental Control Based on Android Platform

Jae-Deok Lim, Byeong-Cheol Choi, Seung-Wan Han, and Jeong-Nyeo Kim

Cyber Security Research Department, Electronics and Telecommunications Research Institute,  
218 Gajengno, Yuseong-gu, Daejeon, Korea  
{jdscol92, corea, hansw, jnkim}@etri.re.kr

**Abstract.** This paper proposes a framework that can analyze and manage adult contents remotely on Android-based smart devices. It can be used as a parental control framework that parents can remotely control and manage adult contents of their children's devices with parents' devices without handling the children's device directly. For the purpose, the framework consists of three components: Contents Analysis App. which is running on children's devices, Contents Management App. which is running on parents' devices, and Connection Maintenance Server which plays a role of maintaining the connection between parents' devices and children's devices. The proposed framework can be applied to protect adolescents and children from harmful contents on smart devices.

**Keywords:** Content analysis, Android, Remote management, Parental control.

## 1 Introduction

The number of adolescents and children using mobile internet via smart devices such as smartphone is on the rise. This increase in usage and the unrestricted nature of mobile internet inevitably lead to serious problems that the youth can be exposed to adult contents in easily and the control is difficult. So an efficient control and management framework is desperately demanded to handle the rising problems in an attempt to provide safer and more immaculate environment in mobile internet usage for the adolescents and children.

This paper proposes a framework that can analyze and manage adult contents remotely in Android-based smart devices. It can be used as a parental control framework that parents can remotely control and manage adult contents stored in their children's devices with parents' devices without handling the children's devices directly. For the purpose, we design the framework with three components: Contents Analysis App.(CAA) which is an application running on children's devices, Contents Management App.(CMA) which is an application running on parents' devices, and Connection Maintenance Server(CMS) which plays a role of maintaining the connection between parents' devices and children's devices. Adult contents are analyzed with some MPEG-7 image descriptors as image-based feature [1] and the temporal

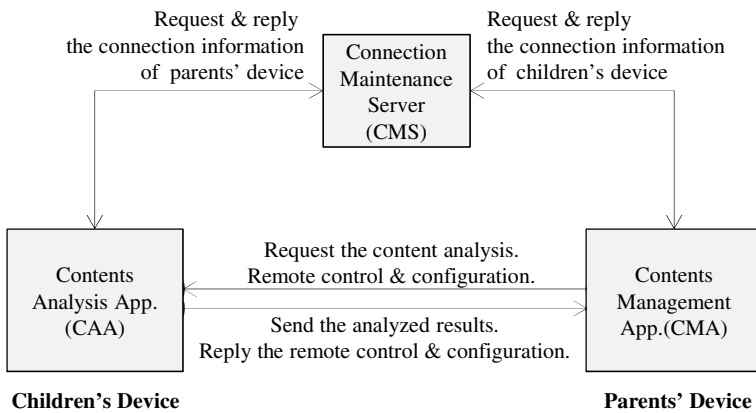


color histogram feature as video-based feature [2,3]. And each feature is classified by the support vector machine(SVM) classifier that is used widely in classification area [4,5,6].

The remainder of this paper is organized as follows. In section 2, we design the proposed framework. And prototype of framework and performance are presented in section 3. Finally, conclusions are summarized in section 4.

## 2 Design of Framework

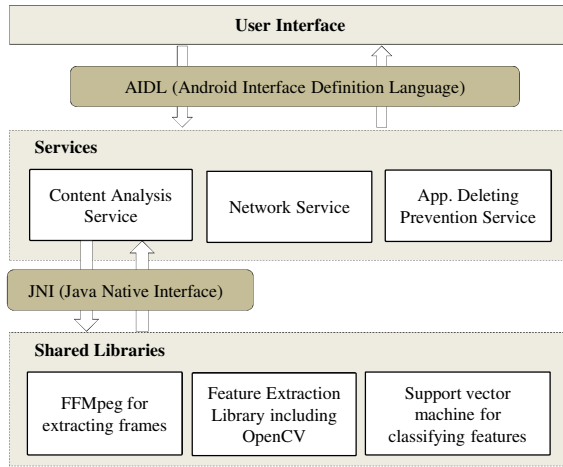
Adults contents analysis and remote management framework consists of three components as shown in Figure 1.



**Fig. 1.** Adult contents analysis and remote management framework

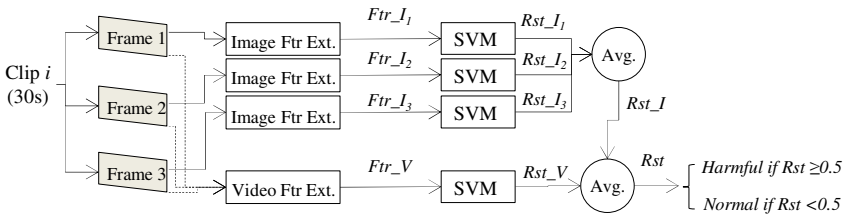
CAA is running on children's device. CAA starts to analyze contents stored in a children's device at the scheduled time or when receiving the analysis request of CMA. After completing to analyze contents, CAA keeps the contents classified as adult contents in isolation and sends the analysis results to CMA that is running on parents' device. CAA runs with three services as shown in Figure 2; Content Analysis Service, Network Service, App. Deleting Prevention Service.

Content Analysis Service that is a core engine of CAA analyzes only new contents that are not analyzed before in order to reduce the processing overhead. The content analysis is done practically with three native libraries via java native interface(JNI) in order to consider the processing performance. FFMpeg library is used for extracting frames that are used in feature extraction library [7]. Feature extraction library computes image-based feature and video-based feature from frames with OpenCV library [8]. And support vector machine(SVM) library is used to classify features. A file to be analyzed is divided into 30 clips with an interval of equal length. Each clip is 30 seconds length and is classified as harmful or not as shown in Figure 3.



**Fig. 2.** Architecture of the Content Analysis App.(CAA)

Three frames are extracted from a clip and image-based features( $Ftr_{I_i}$ ) and video-based feature( $Ftr_V$ ) are computed from these frames. The classification result of image-based feature( $Rst_I$ ) is the average value of each classification result of image-based feature( $Rst_{I_1}, Rst_{I_2}, Rst_{I_3}$ ). A clip is classified as harmful if the average value of the classification result of image-based feature( $Rst_I$ ) and the classification result of video-based feature( $Rst_V$ ) is more than 0.5.



**Fig. 3.** Procedure of feature extraction and classification

Image-based feature is constructed from a single frame by using the color layout descriptors(CLDs) and the edge histogram descriptors(EHDs) for the whole region of the frame and the color structure descriptors(CSDs) and skin information for the inner three fourths region of the same frame [1]. The temporal color histogram feature(TCHF) is used as video-based feature and is constructed from a single video segment of 30s length [2,3]. TCHF uses H-S color histogram in Hue Saturation Value(HSV) of three sampled frames and extract temporal features by averaging of spatial H-S color histograms. The constructed feature vectors are classified by the SVM library [4,6]. In this paper, the ‘*libsvm*’ tools are ported for android platform as an SVM classifier [5], and the radial basis kernel function (RBF) is used as a kernel function. For classifying files, we use a *harmful rate* that is defined as the ratio of

harmful clips among the total analyzed clips(30 clips) within a video file. If a harmful rate is more than a given threshold value, the file is classified as adult content. The threshold value will be determined in section 3.

Network Service manages the connection with CMA by sending the analysis results and receiving the requests. And to avoid the detouring the analysis process by deleting CAA forcibly, App. Deleting Prevention Service is running by monitoring the log messages according to the deleting event.

CMA receives the analysis results and shows them by its user interfaces. User, mainly parents, can check the type of contents that are stored in children’s device and control the adult contents and/or the misclassified contents, for example, deleting adult contents or changing the type of misclassified contents. Whenever CAA and CMA communicate with each other, they send the last IP address of their device to CMS and acquire the other’s last IP address from CMS. From this way, CMS can maintain the last IP addresses of CAA and CMA.

### 3 Prototype of Framework and Performance

We used the Galaxy S2 Smartphone and Galaxy Tap 10.1 Smart Tap with Android 2.3(Gingerbread) as Android platform and prototyped the framework on those devices. Some examples of prototype of CMA for parents are shown in Figure 4. The Top Menu (a) includes ‘Run Analysis’, ‘View Analysis Results’, and ‘Remote Configuration’. When ‘Run Analysis’ is selected, CMA sends the analysis request to CAA and receives the analysis results after completing the contents analysis. When ‘View Analysis’ is selected, Analysis History Lists (b) are shown with the analyzed date, the number of all contents analyzed, the number of contents analyzed as adult and general. Also prototype provides the detailed view of analysis results (c) for each content that user can check if content is analyzed correctly or not. Also user can change the result of misclassified content and delete adult content remotely.

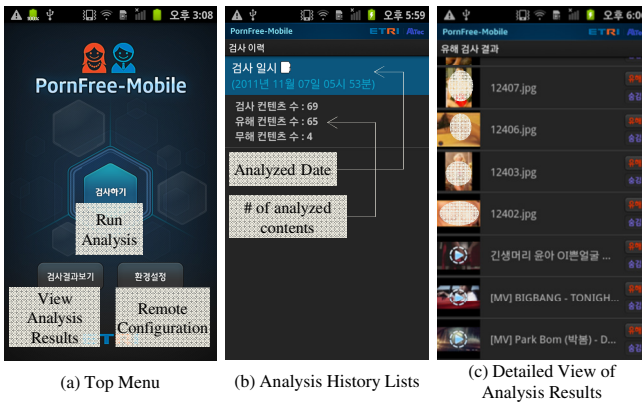


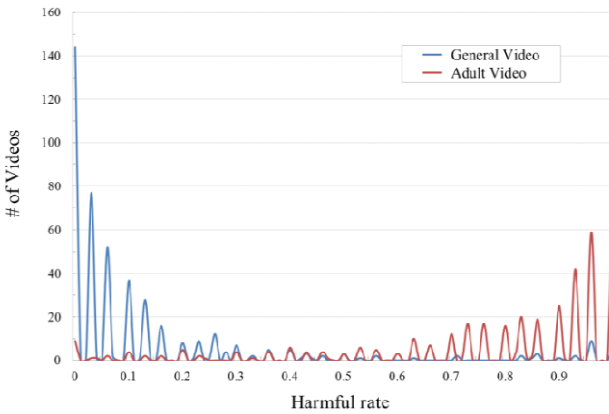
Fig. 4. Some examples of prototypes for Contents Management App

The processing time and classification performance are estimated on Galaxy Tap 10.1 which has 1GHz Dual-core ARM Cortex-A9 CPU and 16GB internal memory. Table 1 shows the processing time of each file that has three types of resolutions. It takes more time to analyze and classify content that has higher resolution because of frame normalization at feature extraction and we can predict that it will take about three hours for analyzing 100 files of 720p resolution. Although the processing time is not enough to apply to real-time service, it can be applied to the general analysis service that performs the analysis periodically at an idle time such as a sleeping time.

**Table 1.** Performance of processing time.

Files	Frame resolution	Playing Time	File Size(MB)	Classification Processing time (s)
File 1	480p(720x480)	28m 10s	465	78.69
File 2	720p(1280x720)	31m 37s	777	104.89
File 3	1080p(1920x1080)	15m 24s	2151	224.93

The classification performances were tested with 452 general video files and 404 adult video files. Figure 5 shows the distribution of harmful rate of general videos and adult videos. From figure 5, we select 30% of harmful rate as the threshold value that determine the adult video.



**Fig. 5.** Distribution of the harmful rate of general videos and adult videos.

The classification performances were estimated in terms of accuracy, precision, and recall, which were computed using Eq. (1), Eq. (2), and Eq. (3), respectively, where  $N_{tp}$  is the number of true positives,  $N_{tn}$  is the number of true negatives,  $N_{fp}$  is the number of false positives, and  $N_{fn}$  is the number of false negatives. Adult video file is used as positive data in this paper. Therefore, a true positive means that an adult video file classified as the adult class, and a false positive means that a video file classified as an adult class is a general video file.

$$Accuracy = (N_{tp} + N_{tn}) / (N_{tp} + N_{fp} + N_{tn} + N_{fn}) \quad (1)$$

$$Precision = N_{tp} / (N_{tp} + N_{fp}) \quad (2)$$

$$Recall = N_{tp} / (N_{tp} + N_{fn}) \quad (3)$$

Classification performances have 89.84% of accuracy, 86.61% of precision, and 92.82% of recall.

## 4 Conclusion

This paper proposes a framework that can analyze and manage adult contents remotely on Android-based smart devices. It can be used as a parental control framework that parents can remotely control and manage adult contents of their children's devices with parents' device without handling the children's device directly. The proposed framework can be applied to protect adolescents and children from harmful contents with smart device. However, the processing time is necessary to be improved in order to reduce the computing overhead.

**Acknowledgments.** This research was funded by the MSIP(Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

## References

1. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7 – multimedia content description interface. John Wiley & Sons Ltd. (2002)
2. Choi, B.C., Han, S.W., Lim, J.D., Chung, B.H., Ryou, J.C.: Temporal visual features and optimal combining rule for video rating system. In: International Conference on Communication and Information Engineering, ICCIE 2011 (2011)
3. Choi, B.C., Han, S.W., Lim, J.D., Ryou, J.C.: Temporal visual features and adaptive decision scheme for large-scale video rating system. Information Journal 15(5), 2321–2332 (2012)
4. Cortes, C., Vapnik, V.: Support Vector Networks. Machine Learning 20, 273–297 (1995)
5. Shigeo, A.: Support Vector Machines for Pattern Classification. Springer, London (2005)
6. Hsu, C.W., Chang, C.C., Lin, C.J.: A Practical Guide to Support Vector Classification, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (accessed 30, October 2011)
7. FFmpeg: a complet, cross-platform solution to record, convert and stream audio and video, <http://ffmpeg.org/> (accessed 30, October 2011)
8. OpenCV: Open Source Computer Vision, <http://opencv.org/> (accessed 30, October 2011)

# A Dual Approach of GeneralMatch in Time-Series Subsequence Matching

Hea-Suk Kim, Minwoo Lee, and Yang-Sae Moon\*

Department of Computer Science, Kangwon National University, Korea  
{hskim, leemw, ysmoon}@kangwon.ac.kr

**Abstract.** In this paper, we propose a dual approach of GeneralMatch, called DualGMatch. GeneralMatch is an efficient time-series subsequence matching method that uses the generalized windows, called *J-sliding* and *J-disjoint windows*. We first investigate the traditional subsequence matching methods. Based on this investigation, we show that DualGMatch can be feasible as a dual approach of GeneralMatch by switching the role of *J-sliding* and *J-disjoint windows*. We analytically and empirically compare the proposed DualGMatch with the previous GeneralMatch.

## 1 Introduction

Time-series data are the sequences of real numbers representing values at specific time points. Typical examples of time-series data include stock prices, biomedical measurements, and multimedia data [5, 6, 9]. The time-series data stored in a database are called *data sequences*. Finding data sequences similar to the given *query sequence* is called *similar sequence matching* [1, 5]. In this paper, we use the Euclidean distance [1, 4, 5, 7, 8] as the similarity model, where two sequences  $X = \{X[1], \dots, X[n]\}$  and  $Y = \{Y[1], \dots, Y[n]\}$  are said to be *similar* if the Euclidean distance  $D(X, Y)$  ( $= \sqrt{\sum_{i=1}^n |X[i] - Y[i]|^2}$ ) is less than or equal to the user-specified *tolerance*  $\epsilon$  [1]. Similar sequence matching can be classified into whole matching and subsequence matching [5]. *Whole matching* [1] finds data sequences similar to a query sequence, where the lengths of data sequences and the query sequence are all identical. *Subsequence matching* [5, 7, 8] finds subsequences, contained in data sequences, similar to a query sequence of arbitrary length. Likewise, subsequence matching is a generalization of whole matching, and thus we focus on subsequence matching in this paper.

Subsequence matching methods [5, 7, 8] consist of index building and subsequence matching algorithms. In the index building algorithm, data sequences are divided into windows of size  $\omega$ , and each window is transformed to a point in an  $f$ -dimensional space and stored in an  $f$ -dimensional index. In the subsequence matching algorithm, a query sequence is divided into windows of size  $\omega$ , and each window is transformed to an  $f$ -dimensional point and constructed to a range query with the tolerance  $\epsilon$ . By evaluating the range queries, the *candidates* that are potentially similar with the query

---

\* Corresponding author.

sequence are identified. The final result is subsequently refined by accessing the database and by selecting only those subsequences that are similar with the query sequence. Faloutsos et al. [5] have proposed a subsequence matching method that divides data sequences into *sliding windows* and the query sequence into *disjoint windows* (FRM in short by taking the authors' initials). FRM, however, causes many *false alarms* by storing MBRs (minimum bounding rectangles) and by not allowing point-to-point comparison, which is called *point-filtering effect* [7]. To solve this problem of FRM, DualMatch [7] has been proposed as a dual approach for FRM in constructing windows: it divides data sequences into *disjoint* windows and a query sequence into *sliding* windows. DualMatch stores individual points directly and improves performance significantly by exploiting the point-filtering effect. Nevertheless, DualMatch has the problem of having a smaller window size. A smaller window increases false alarms, and this effect is called *window size effect* [7].

GeneralMatch [8] has advantages of both FRM and DualMatch: it tries to use large windows like FRM and, at the same time, to exploit point-filtering effect like DualMatch. For this, GeneralMatch generalizes the concept of constructing windows by defining *J-sliding windows* and *J-disjoint windows*. By this generalization, GeneralMatch divides data sequences into *J-sliding* windows and the query sequence into *J-disjoint* windows. In this paper, we investigate a dual approach of GeneralMatch and propose DualGMatch by switching the role of *J-sliding* and *J-disjoint* windows, i.e., by dividing data sequences into *J-disjoint* windows and the query sequence into *J-sliding* windows. However, analytical and experimental evaluations show that DualGMatch is worse than GeneralMatch in matching performance. Through analytical comparison, we first explain why GeneralMatch is superior in performance to DualGMatch. We then empirically compare DualGMatch with GeneralMatch in the number of page accesses.

## 2 GeneralMatch: An Efficient Subsequence Matching Method Using Generalized Windows

Definitions 1 and 2 show how GeneralMatch divides data and query sequences into *J-sliding* and *J-disjoint* windows, respectively.

**Definition 1.** A *J-sliding window*  $s_i^J$  of size  $\omega$  of the sequence  $S$  is defined as the subsequence of length  $\omega$  starting from  $S[(i - 1) \times J + 1]$ .  $\square$

**Definition 2.** A *J-disjoint window*  $q_{(i,j)}^J$  of size  $\omega$  of the sequence  $Q$  is defined as the subsequence of length  $\omega$  starting from  $Q[i + (j - 1) \times \omega]$  in  $Q$ .  $\square$

Fig. 1 shows an example of dividing a sequence  $S$  into 4-sliding windows of length 16. Intuitively speaking, we construct windows by shifting a subsequence of length 16 by 4 entries, and thus, the starting entries of the 4-sliding windows are  $S[1]$ ,  $S[5]$ ,  $S[9]$ , ..., respectively. Similarly, Fig. 2 shows an example of dividing a sequence  $Q$  into 4-disjoint windows of length 16. Intuitively speaking, we construct windows  $Q[i : i + \omega - 1]$ ,  $Q[i + \omega : i + 2\omega - 1]$ , ... by dividing  $Q[i : Len(Q)]$  into disjoint windows for every  $i$  ( $1 \leq i \leq 4$ ).

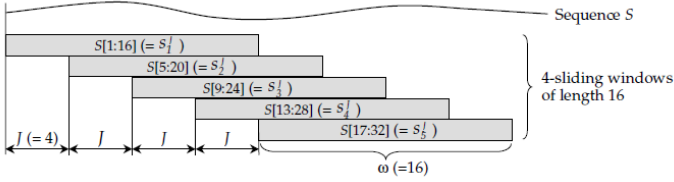


Fig. 1. An example of  $J$ -sliding windows ( $J = 4, \omega = 16$ )

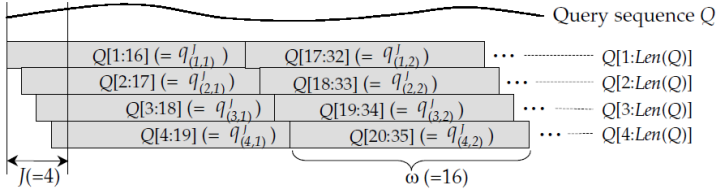


Fig. 2. An example of  $J$ -disjoint windows ( $J = 4, \omega = 16$ )

Theorem 1 states the correctness of GeneralMatch [8].

**Theorem 1.** Suppose the data sequence  $S$  is divided into  $J$ -sliding windows of size  $\omega$ , and the query sequence  $Q$  into  $J$ -disjoint windows of the same size. Then, Eq. (1) holds:

$$D(S[i : j], Q) \leq \epsilon \Rightarrow \bigvee_{n=0}^{\rho-1} D(s_{a+n \times k}^J, q_{(b+1, n+1)}^J) \leq \epsilon / \sqrt{\rho} \quad (1)$$

where  $a$  is  $\lfloor \frac{i-1}{J} \rfloor + 1$ ,  $b$  is  $(a-1) \times J - i + 1$ , and  $\rho$  is  $\lfloor \frac{Len(Q) - b}{\omega} \rfloor$ .

In Eq. (1),  $a$  means that the first  $J$ -sliding window of  $S[i : j]$  is  $s_a^J$ , and  $b$  means that the starting offset of the  $J$ -disjoint window  $q_{(b+1, 1)}^J$  of  $Q$  that will be compared with  $s_a^J$  is  $b + 1$ , which is the difference between the starting offset of  $S[i : j]$  and that of  $s_a^J$ . Theorem 1 guarantees that the candidate set consisting of the subsequences  $S[i : j]$  such that the distance between  $s_{a+n \times k}^J$  and  $q_{(b+1, n+1)}^J$  is in  $\epsilon / \sqrt{\rho}$  (i.e., satisfying the necessary condition of Eq. (1)) contain no false dismissal. Based on Theorem 1, Moon et al.[8] have proposed index building and subsequence matching algorithms of GeneralMatch..

### 3 DualGMatch: A Dual Approach of GeneralMatch

Like that there exists DualMatch as a dual approach of FRM, we can think of DualGMatch as a dual approach of GeneralMatch. For this, DualGMatch switches the role of the generalized windows used in GeneralMatch, i.e., it uses duality in constructing windows by dividing data sequences into  $J$ -disjoint windows and the query sequence into  $J$ -sliding windows. Except the difference in switching the role of  $J$ -sliding and  $J$ -disjoint windows, DualGMatch can use the same index building and subsequence matching algorithms of GeneralMatch [8].

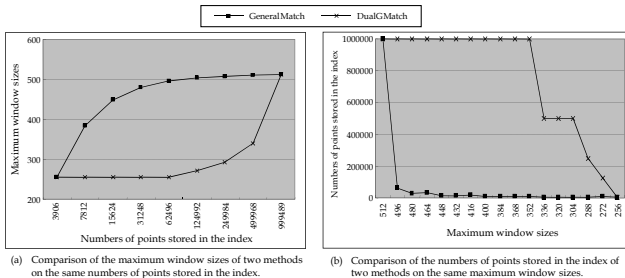


To compare **DualGMatch** with **GeneralMatch**, we analyze the maximum window size and the number of points stored in the index. Table 1 summarizes the maximum window sizes and the numbers of points stored in the index for two methods. Based on [8], we first get the maximum window size and the number of points for **GeneralMatch** in Table 1. We then easily get the size and the number for **DualGMatch** by exploiting the similar processes used in **GeneralMatch** [8].

**Table 1.** The maximum windows sizes and the numbers of points stored in the index

	GeneralMatch	DualGMatch
Maximum window size	$\left\lfloor \frac{Min(Q) - J + 1}{J} \right\rfloor * J$	$\left\lfloor \frac{Min(Q) + J}{2J} \right\rfloor * J$
Number of points stored in the index	$\left\lfloor \frac{Min(S) - \omega}{J} \right\rfloor + 1$	$\sum_{i=1}^J \left\lfloor \frac{Len(S) - i + 1}{\omega} \right\rfloor$

Using the sizes and the numbers in Table 1, we can get Fig. 3 that shows the comparison of the maximum window sizes and the numbers of points stored in the index for **GeneralMatch** and **DualGMatch** when  $Len(S)$  is 1000000. As shown in Fig. 3, in the case where the numbers of points are 3906 (the case of **DualMatch**) and 999489 (the case of **FRM**), both two methods have the same window sizes 256 and 512, respectively. The reason is that **DualGMatch** as well as **GeneralMatch** is the generalized method, which uses the generalized windows, and thus, it would also include **DualMatch** and **FRM**. Except these two special cases, however, the maximum window size of **GeneralMatch** is much larger than that of **DualGMatch** on the same number of points stored in the index. That is, if both two methods have the same number of points stored in the index, the maximum window size of **GeneralMatch** is much larger than that of **DualGMatch**, and thus, **GeneralMatch** would outperform **DualGMatch** due to the window size effect. Moreover, in Fig. 3(b), if both two methods have the same maximum window size, **DualGMatch** would store much more points in the index compared with **GeneralMatch**. That is, to use the same window size, **DualGMatch** would store more points than **GeneralMatch**, and thus, it would incur much overhead in searching the index.



**Fig. 3.** Comparison of the maximum window sizes and the numbers of points stored in the index for **GeneralMatch** and **DualGMatch** when  $Len(Q)$  is 1000000

The reason of incurring differences in Fig. 3 is that, in the case of having the same window size, dividing a sequence into  $J$ -sliding windows generates much smaller points than dividing it into  $J$ -disjoint windows. Since GeneralMatch divides data sequences into  $J$ -sliding windows while DualGMatch divides them into  $J$ -disjoint windows, the number of points for GeneralMatch is much smaller than that for DualGMatch. On the other hand, the number of points contained in a query MBR for GeneralMatch is larger than that for DualGMatch. However, the points in a query MBR can be maintained in main memory, and it would not make a serious effect on the performance. As a result, GeneralMatch would provide better performance than DualGMatch by reducing the number of points stored in the index and by enlarging the maximum window size.

In summary, it is definitely feasible to make DualGMatch as a dual approach of GeneralMatch. However, GeneralMatch is evidently superior in performance to DualGMatch. The reason is that, if two methods store the same number of points in the index, the maximum window size of GeneralMatch is larger than that of DualGMatch, and thus, GeneralMatch better exploits the window size effect. We will confirm this analytical result through experiments in the next section.

### 4 Experimental Evaluation

We have performed experiments using a stock data set used in previous works [5, 7, 8]. The stock data set consists of a long data sequence containing 329,112 entries. We conduct the experiments on a Linux machine with Intel Core2 Duo 2.53GHz CPU, 2GB RAM, and 500GB HDD and use C/C++ language to implement GeneralMatch and DualGMatch. The page size for data and indexes is set to be 4,096 bytes. As the multidimensional index, we use the R-tree [2]. Like [7, 8], we use discrete Fourier transform as the feature extraction function for lower-dimensional transformations [3] and use six features.

Fig. 4 shows the numbers of page accesses of GeneralMatch and DualGMatch. Selectivity [5, 8] of Fig. 4(a) is in  $10^{-6} \sim 10^{-4}$ , called *Low-Range*. We note that GeneralMatch is approximately ten times superior to DualGMatch in the number of page accesses. As we explained in Section 3, this is because GeneralMatch better exploits the window size effect. Selectivity of Fig. 4(b) is in  $10^{-3} \sim 10^{-1}$ , called *High-Range*.

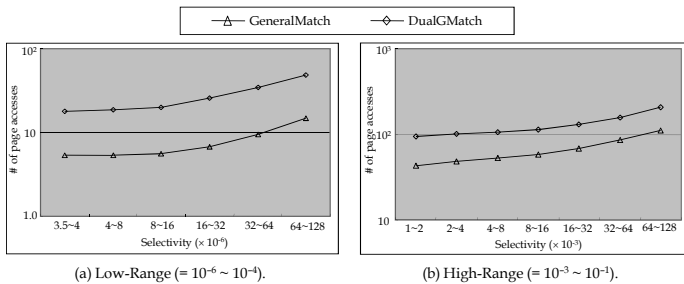


Fig. 4. Comparison of the number of page accesses for GeneralMatch and DualGMatch

We also note that **GeneralMatch** is approximately five times superior to **DualGMatch**. The difference of Low-Range and High-Range also comes from the window size effect. As we can see in [7, 8], the window size effect of Low-Range is much larger than that of High-Range [7, 8].

**Acknowledgement.** This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013.

## References

1. Agrawal, R., Faloutsos, C., Swami, A.: Efficient Similarity Search in Sequence Databases. In: Lomet, D.B. (ed.) FODO 1993. LNCS, vol. 730, Springer, Heidelberg (1993)
2. Beckmann, N., Kriegel, H.-P., Schneider, R., Seeger, B.: The R\*-tree: An Efficient and Robust Access Method for Points and Rectangles. In: Proc. Int'l Conf. on Management of Data, ACM SIGMOD, Atlantic City, New Jersey, pp. 322–331 (May 1990)
3. Berchtold, S., Bohm, C., Kriegel, H.-P.: The Pyramid-Technique: Towards Breaking the Curse of Dimensionality. In: Proc. Int'l Conf. on Management of Data, ACM SIGMOD, Seattle, Washington, pp. 142–153 (June 1998)
4. Chan, K.-P., Fu, A.W.-C., Yu, C.T.: Haar Wavelets for Efficient Similarity Search of Time-Series: With and Without Time Warping. *IEEE Trans. on Knowledge and Data Engineering* 15(3), 686–705 (2003)
5. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast Subsequence Matching in Time-Series Databases. In: Proc. Int'l Conf. on Management of Data, ACM SIGMOD, Minneapolis, Minnesota, pp. 419–429 (May 1994)
6. Han, W.-S., Lee, J., Moon, Y.-S., Hwang, S.-W., Yu, H.: A New Approach for Processing Ranked Subsequence Matching Based on Ranked Union. In: Proc. of Int'l Conf. on Management of Data, ACM SIGMOD, Athens, Greece (June 2011)
7. Moon, Y.-S., Whang, K.-Y., Loh, W.-K.: Duality-Based Subsequence Matching in Time-Series Databases. In: Proc. the 17th Int'l Conf. on Data Engineering (ICDE), pp. 263–272. IEEE, Heidelberg (2001)
8. Moon, Y.-S., Whang, K.-Y., Han, W.-S.: General Match: A Subsequence Matching Method in Time-Series Databases Based on Generalized Windows. In: Proc. Intl Conf. on Management of Data, ACM SIGMOD, Madison, Wisconsin, pp. 382–393 (June 2002)
9. Moon, Y.-S., Kim, B.-S., Kim, M.S., Whang, K.-Y.: Scaling-Invariant Boundary Image Matching Using Time-Series Matching Techniques. *Data & Knowledge Engineering* 69(10), 1022–1042 (2010)

# An Approximate Multi-step $k$ -NN Search in Time-Series Databases

Sanghun Lee<sup>1</sup>, Bum-Soo Kim<sup>2</sup>, Mi-Jung Choi<sup>1</sup>, and Yang-Sae Moon<sup>1,\*</sup>

<sup>1</sup> Department of Computer Science, Kangwon National University, Korea

<sup>2</sup> Advanced Information Technology Research Center(AITrc),

Korea Advanced Institute of Science and Technology (KAIST)

{sanghun,mjchoi,ysmoon}@kangwon.ac.kr, bskim@mozart.kaist.ac.kr

**Abstract.** In this paper, we propose an approximate solution to the multi-step  $k$ -NN search. The traditional multi-step  $k$ -NN search (1) determines a tolerance through a  $k$ -NN query on a multidimensional index and (2) retrieves the final  $k$  results by evaluating the tolerance-based range query on the index and by accessing the actual database. The proposed *tolerance reduction-based (approximate) solution* reduces a large number of candidates by adjusting the tolerance of the range query on the index. To obtain the tight tolerance, the proposed solution forcibly decreases the tolerance by the average ratio of high-dimensional and low-dimensional distances. Experimental results show that the proposed approximate solution significantly reduces the number of candidates and the  $k$ -NN search time over the existing one.

## 1 Introduction

The multi-step  $k$ -NN search [4] retrieves  $k$  nearest time-series (called sequences) using a multidimensional index. In the first step, it obtains  $k$  data sequences by evaluating a  $k$ -NN query on the index and determines a *tolerance* as the  $k$ -th distance from those data sequences to the query sequence. In the second step, it obtains candidate sequences by evaluating the *tolerance*-based range query on the index and returns the final  $k$  sequences by filtering false alarms from those candidate sequences.

In this paper, we address a problem of improving the performance of multi-step  $k$ -NN search in time-series databases. Due to information loss of lower-dimensional transformations, the existing multi-step  $k$ -NN search solution is subject to obtain a large tolerance (i.e., a large search range) as a result of  $k$ -NN search on the multidimensional index. The large tolerance, however, incurs a large number of candidate sequences which are retrieved by a range query. Those many candidates make severe I/O and CPU overheads in the post processing step and eventually degrade the overall search performance.

To alleviate the problem of incurring a large number of candidates, in this paper we propose the *tolerance reduction-based (approximate) solution* that improves

---

\* Corresponding author.

the search performance by intentionally reducing the tolerance. In this approach, how much we reduce the tolerance is important in reducing the number of candidates and, at the same time, avoiding false dismissals. We here use a simple but efficient statistical approach which determines the reduction ratio by the average ratio of high-dimensional and low-dimensional distances.

Experimental results show that the proposed tolerance reduction-based solution significantly reduces the number of candidates and the  $k$ -NN search time over the existing solution. A notable point is that it incurs no or a very few number of false dismissals (less than 3.0% of all experimental cases), which are usually occurred in an approximate solution. This means that, even though the tolerance reduction-based solution is an approximate one, it performs the multi-step  $k$ -NN search very correctly and efficiently.

## 2 Related Work

Time-series data are the sequences of real numbers representing values at specific time points. Typical examples of time-series data include stock prices, exchange rates, biomedical measurements, and multimedia data [2, 3]. The time-series data stored in a database are called *data sequences*. Finding data sequences similar to the given *query sequence* is called *similar sequence matching* [5], which is classified into range search and  $k$ -NN search. In this paper, we focus on the  $k$ -NN search that finds  $k$  nearest data sequences to the query sequence from the time-series database.

For efficient  $k$ -NN search on time-series data, Korn et al. [4] have proposed the multi-step  $k$ -NN search, which consists of two major steps: (1)  $k$ -NN query and (2) range query steps. Fig. 1 shows the working procedure of multi-step  $k$ -NN search consisting of two steps, and each of steps contains two sub-steps. In the  $k$ -NN query of the first step, (1) it retrieves  $k$  candidate sequences by performing a  $k$ -NN query on the multidimensional index; (2) it determines the tolerance as the largest distance  $d_{max}$  among the actual distances from  $k$  candidate sequences to the query sequence ( $= Q$ ). In the range search of the second step, (1) it constructs a range query using the tolerance  $d_{max}$  of the first step and obtains candidate sequences by evaluating the range query on the index; (2) it filters false alarms from those candidates and returns the final  $k$  nearest sequences.

## 3 Tolerance Reduction-Based Approximate Approach

The tolerance reduction-based approach forcibly reduces the tolerance of range queries by the average ratio of  $f$ -dimensional and  $n$ -dimensional distances. (Note that  $n$  is the length of data and query sequences, and  $f$  is the number of features extracted from an  $n$ -dimensional sequence and used for constructing the multidimensional index [4, 5].) Since the smaller tolerance produces the smaller number of candidates in the range query, the reduction approach would improve the overall performance. The reason why we use the average of  $\frac{f\text{-dimensional distance}}{n\text{-dimensional distance}}$ 's as the reduction

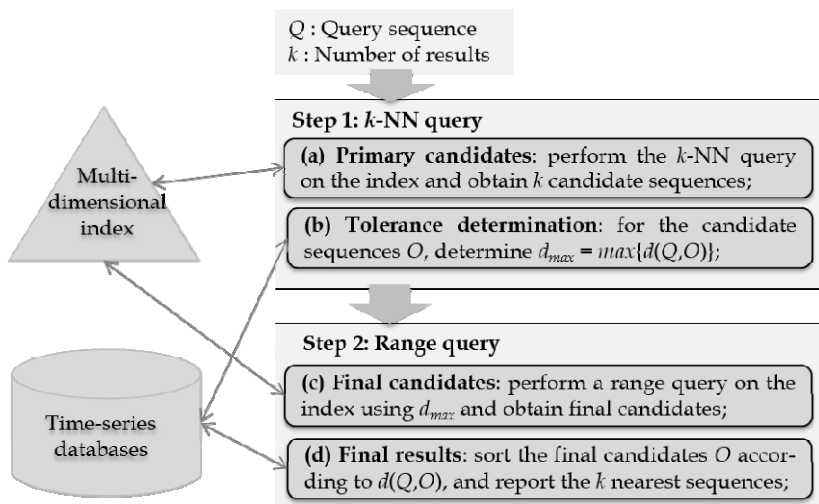


Fig. 1. The working procedure of the multi-step  $k$ -NN search [4]

ratio is that it simply reflects the distance difference before and after lower-dimensional transformations. We call this ratio the *average reduction ratio* and obtain it through the pre-processing step for the actual time-series database. Likewise, reducing the tolerance by the average reduction ratio will evidently improve the overall performance of multi-step  $k$ -NN search, but it has also a risk of causing some false alarms, which will be discussed in Section 4.

More precisely, we empirically obtain *multiple* average reduction ratios and dynamically determine which ratio is the best for the given tolerance. The reason why we use multiple ratios is that, if we apply only one ratio to all possible tolerances, there would be no performance improvement for small tolerances or many false alarms for large tolerances. Thus, through the pre-processing step, we first construct a list  $\mathbf{R}$  of average reduction ratios, and we then choose an appropriate ratio according to the given tolerance. Eq. (1) shows the structure of list  $\mathbf{R}$ .

$$\mathbf{R} = (p_1, m_1, p_2, m_2, p_3, \dots, p_{n-1}, m_{n-1}, p_n) \quad (1)$$

In Eq. (1),  $p_i$  is an average reduction ratio which will be chosen by the tolerance, and  $m_i$  is a delimiter of tolerance sizes. That is, if the tolerance obtained in the first step of Fig. 1 is larger than  $m_1$ , we choose  $p_1$  as the average reduction ratio for that tolerance; if it is in between  $m_{i-1}$  and  $m_i$ , we choose  $p_i$ ; if it is smaller than  $m_{n-1}$ , we choose  $p_n$ . Likewise, after constructing a list  $\mathbf{R}$ , we reduce the tolerance by an appropriate reduction ratio and perform a range query using the reduced tolerance.

Fig. 2 shows an algorithm of computing a list  $\mathbf{R}$  of average reduction ratios. The input to the algorithm is a time-series database  $\mathbf{DB}$  and the number  $T$  of  $n$ -dimensional distances used for measuring average tolerances and average reduction ratios. In Lines (3) to (5), we randomly select two data sequences and compute their real(=  $n$ ) and  $f$ -dimensional distances, i.e.,  $d_n$  and  $d_f$ , respectively. In Line (6), we include

**Procedure** *AvgReductionRate()*

**Input**        **DB**: a database of  $n$ -dimensional data sequences  
                   $T$ : the number of  $n$ -dimensional distances

**Output**      **R**: the result set of average reduction ratios consisting of  $p_i$  and  $m_i$

- (1) **G** =  $\emptyset$ ;
- (2) **for**  $t := 0$  **to**  $T$  **do**
- (3)     Choose two sequences  $A$  and  $B$  randomly from **DB**;
- (4)      $d_n = d_{real}(A, B)$ ;           // actual high-dimensional distance
- (5)      $d_f = d_{feature}(A, B)$ ;       // index-level low-dimensional distance
- (6)     **G** =  $\mathbf{G} \cup (d_n, d_f)$ ;       // construct the initial set of distance pairs
- (7) **end-for**
- (8)  $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_n = \text{GroupSeparation}(\mathbf{G})$ ; //  $\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_n$  are groups separated from **G**
- (9) **for**  $i := 1$  **to**  $n$  **do**
- (10)     $m_i = \text{average of } d_n \text{ in } \mathbf{G}_i$ ;       // average high-dimensional distance of  $\mathbf{G}_i$
- (11)     $p_i = \text{average of } \frac{d_f}{d_n} \text{ in } \mathbf{G}_i$ ;    // average reduction ratio of  $\mathbf{G}_i$
- (12) **end-for**
- (13) **R** =  $(p_1, m_1, p_2, m_2, \dots, p_{n-1}, m_{n-1}, p_n)$ ;   // **R** is a list of average reduction ratios

**Fig. 2.** An algorithm of computing a list **R** of average reduction ratios

( $d_n, d_f$ ) in a set **G** of distance pairs. By repeating Lines (3) to (6)  $T$  times, we obtain an initial set **G** containing  $T$  distance pairs. Function *GroupSeparation()* groups distance pairs of **G** into  $n$   $\mathbf{G}_i$ 's using the binary partition technique. In Lines (10) to (11), for each  $\mathbf{G}_i$ , we obtain an average tolerance  $m_i$  and an average reduction ratio  $p_i$ . In Line (13), we finally construct a list **R** of Eq. (1) by concatenating  $p_i$ 's and  $m_i$ 's. We perform the procedure *AvgReductionRate()* of Fig. 2 in the pre-processing step and use the list **R** in the multi-step  $k$ -NN search of Fig. 1 to reduce the tolerance of a range query.

## 4 Experimental Evaluation

We have conducted the experiment on a Linux machine (CentOS 5.9) with Intel Xeon 3.1GHz CPU, 4GB RAM, and 1TB HDD and used C/C++ language to implement all the algorithms. In the experiment, we have used a real data set consisting 10,000 image time-series [6] of length 360. We extract six features from a sequence of length 360 using PAA (piecewise approximate aggregation) [7], and we use the  $R^*$ -tree [1] as the multidimensional index.

We first construct a list **R** of average reduction ratios by using the algorithm of Fig. 2. Eq. (2) shows a list **R** obtained for six features. We here set  $T$  to 30,000, i.e., we randomly select 30,000 pairs of data sequences and use their distances for constructing **G** of the algorithm.

$$\mathbf{R} = (83\%, 2,893, 75\%, 2,175, 70\%, 1,650, 66\%, 1,230, 61\%, 866, 56\%, 523, 50\%) \quad (2)$$

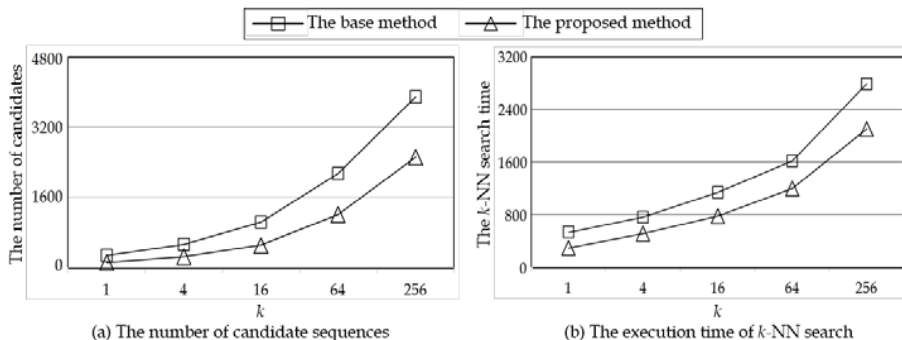


Fig. 3. Performance comparison of the existing and the proposed methods.

In Eq. (2), 83% means that, if the tolerance  $d_{max}$  obtained by a  $k$ -NN query is greater than 2,893, it is reduced to  $d_{max} \times 0.83$  for a range query, and similarly, if  $d_{max}$  is in between 2,893 and 2,175, it is reduced to  $d_{max} \times 0.75$ . We note that the smaller tolerance is, the smaller average reduction is used. This is because the difference between high-dimensional and low-dimensional distances decreases as the tolerance (i.e., the high-dimensional distance) decreases.

Fig. 3 shows the performance comparison of the existing and the proposed multi-step  $k$ -NN search solutions. Fig. 3(a) and (b) show the number of candidate sequences and the execution time, respectively. We use 1, 4, 16, 64, 256 as the coefficient  $k$ . We note that, for all  $k$ 's, the proposed approximate solution significantly reduces the number of candidates in Fig. 3(a), and it thus outperforms the existing solution in the execution time of Fig. 3(b). More precisely, compared with the existing solution, the proposed solution reduces the number of candidates by up to 157.0% and the execution time by up to 81.1%. In Fig. 3, as the coefficient  $k$  increases, the number of candidates and the execution time also increase. This is because the tolerance determined by a  $k$ -NN query increases in proportion to the coefficient  $k$ .

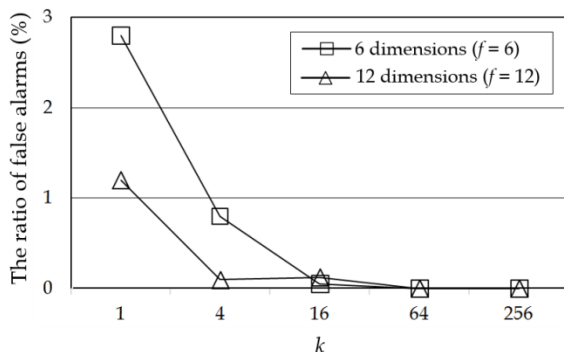


Fig. 4. The ratio of false alarms in the proposed approximate approach



Finally, Fig. 4 shows the ratio of false dismissals, which may occur in the proposed approximate solution. As shown in Fig. 4, the ratio is 1.2% ~3.0% when the coefficient  $k$  is 1, and it is merely less than 1.0% (actually, no false alarm in many cases) when  $k$  is in 4 ~256. Likewise, the ratio of false dismissals can be negligible, and we believe that the proposed tolerance reduction-based approach is an excellent solution that improves the performance of multi-step  $k$ -NN search with slight modification of the existing solution.

## 5 Conclusions

In this paper, we proposed an efficient approach that improved the performance of multi-step  $k$ -NN search. The proposed approach was an approximate one that reduced the tolerance (i.e., the search range) by the average ratio of high-dimensional and low-dimensional distances. For this, we first presented why the reduction method would be efficient and then proposed a formal algorithm of computing the average reduction ratios. Experimental results showed that the proposed approximate solution outperformed the existing one in the execution time as well as the number of candidates. As a future work, we will investigate an exact approach that improves the performance, but incurs no false dismissal.

**Acknowledgement.** This work was partially supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract.

## References

1. Beckmann, N., Kriegel, H.-P., Schneider, R., Seeger, B.: The R\*-tree: An Efficient and Robust Access Method for Points and Rectangles. In: Proc. Int'l Conf. on Management of Data, ACM SIGMOD, Atlantic City, New Jersey, pp. 322–331 (May 1990)
2. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast Subsequence Matching in Time-Series Databases. In: Proc. Int'l Conf. on Management of Data, ACM SIGMOD, Minneapolis, Minnesota, pp. 419–429 (May 1994)
3. Han, W.-S., Lee, J., Moon, Y.-S., Hwang, S.-W., Yu, H.: A New Approach for Processing Ranked Subsequence Matching Based on Ranked Union. In: Proc. of Int'l Conf. on Management of Data, ACM SIGMOD, Athens, Greece (June 2011)
4. Korn, F., Sidiropoulos, N., Faloutsos, C., Siegel, E., Protopapas, Z.: Fast Nearest Neighbor Search in Medical Image Databases. In: Proc. of the 22nd Int'l Conference on Very Large Data Bases, Bombay, India, pp. 215–226 (September 1996)
5. Moon, Y.-S., Whang, K.-Y., Han, W.-S.: General Match: A Subsequence Matching Method in Time-Series Databases Based on Generalized Windows. In: Proc. Intl Conf. on Management of Data, ACM SIGMOD, Madison, Wisconsin, pp. 382–393 (June 2002)
6. Moon, Y.-S., Kim, B.-S., Kim, M.S., Whang, K.-Y.: Scaling-Invariant Boundary Image Matching Using Time-Series Matching Techniques. *Data & Knowledge Engineering* 69(10), 1022–1042 (2010)
7. Roh, G., Roh, J., Hwang, S., Yi, B.: Supporting Pattern Matching Queries over Trajectories on Road Networks. *IEEE Trans. on Knowledge and Data Engineering* 23(11), 1758–1759 (2011)

# The Role of Graph Theory in Solving Euclidean Shortest Path Problems in 2D and 3D\*

Phan Thanh An<sup>1,2</sup>, Nguyen Ngoc Hai<sup>3</sup>, and Tran Van Hoai<sup>4</sup>

<sup>1</sup>Institute of Mathematics, Hanoi, Vietnam

<sup>2</sup>CEMAT, Instituto Superior Tecnico, Portugal

<sup>3</sup>International University, Vietnam National University,  
Ho Chi Minh City, Vietnam

<sup>4</sup>Faculty of Computer Science and Engineering,  
HCMC University of Technology, Ho Chi Minh City, Vietnam  
thanhan@math.ist.utl.pt, nnhai@hcmiu.edu.vn,  
hoai@cse.hcmut.edu.vn

**Abstract.** Determining Euclidean shortest paths between two points in a domain is a fundamental problem in computing geometry and has many applications in GIS, robotics, computer graphics, CAD, etc. To date, solving Euclidean shortest path problems inside simple polygons has usually relied on triangulation of the entire polygons and graph theory. The question: "Can one devise a simple  $O(n)$  time algorithm for computing the shortest path between two points in a simple polygon (with  $n$  vertices), without resorting to a (complicated) linear-time triangulation algorithm?" raised by J. S.B. Mitchell in Handbook of Computational Geometry (J. Sack and J. Urrutia, eds., Elsevier Science B.V., 2000), is still open. The aim of this paper is to show that in 2D, convexity contributes to the design of an efficient algorithm for finding the approximate shortest path between two points inside a simple polygon without triangulation of the entire polygons or graph theory. Conversely, in 3D, we show that graph tools (e.g., Dijkstra's algorithm for solving shortest path problems on graphs) are crucial to find an Euclidean shortest path between two points on the surface of a convex polytope.

**Keywords:** Approximate algorithm, convex hull, discrete geometry, Dijkstra's algorithm, extreme point, Euclidean shortest path, graph theory, shortest path on graph.

## 1 Introduction

Determining Euclidean shortest paths between two points in a domain is a fundamental problem in computing geometry and has many applications in GIS,

---

\* This work received financial support from Portuguese National Funds through FCT (Fundação para a Ciência e a Tecnologia) under the scope of project Pest-OE/MAT/UI0822/2011 and the National Foundation for Science and Technology Development, Vietnam (NAFOSTED) under grant number 101.02-2011.45.

robotics, computer graphics, CAD, etc. To date, all methods for solving this problem in case the domain is a simple polygon, as presented in [9], [11] etc, rely on starting with a rather complicated, but linear-time triangulation of a simple polygon. This leads to the open question below raised by J. S. B. Mitchell [11] "Can one devise a simple  $O(n)$  time algorithm for computing the shortest path between two points in a simple polygon (with  $n$  vertices), without resorting to a (complicated) linear-time triangulation algorithm?" In 1987, the Steiner's problem of finding the in-polygon of a given convex polygon with minimal circumference was solved completely by the method of orienting curves [13]. Efficient algorithms for determining convex ropes in robotics were introduced in [4] and [12]. These problems are variations of the shortest path problem in 2D and thus can be solved without resorting to a linear-time triangulation algorithm or graph theory. In the first part of this paper we recall Lee and Preparata's algorithm that relies on triangulation of the entire polygons and graph theory. Then we present an algorithm for finding the approximate shortest path between two points inside a simple polygon. We show that convexity contributes to the design of this algorithm without triangulation of the entire polygons and graph theory. In the second part of this paper, we investigate Li and Klette's algorithm [10] and show that graph tools (e.g., Dijkstra's algorithm for solving shortest path problems on graphs) are crucial to find the Euclidean approximate shortest path between two points on the surface of a convex polytope. Numerical experiences are presented.

## 2 Lee and Preparata's Algorithm

Lee and Preparata's algorithm [9] for finding the shortest path  $SP(a,b)$  between two points  $a$  and  $b$  in a simple polygon consists of three steps:

Step 1: Triangulate the simple polygon then get a corresponding dual tree.

Step 2: Find shortest path on this tree (by Dijkstra's Algorithm [8]/Thorup's Algorithm [15]) and get corresponding sleeve domain.

Step 3: Using a so-called funnel algorithm to find shortest path on the sleeve.

Thus, graph theory is crucial in Lee and Preparata's algorithm.

## 3 Convexity and Approximate Shortest Paths in a Simple Polygon

We introduce an approximation algorithm for solving the shortest path problem without triangulation and graph theory in [2]. The idea of the direct multiple shooting method [7] is used for the discretization of the shortest path problem based on cut segments parallel to  $y$ -coordinate  $Oy$ . Assume that  $a$  and  $b$  respectively are the first and the final of  $P$  and  $Q$ . The simple polygon  $D=PQ$  is divided by cut segments  $\xi_i=[u_i, v_i]$ ,  $u_i \in P$ ,  $v_i \in Q$  such that

$$\begin{aligned} \xi_i &= [u_i, v_i] \subseteq D \text{ and } ]u_i, v_i[ \subseteq \text{int}D_i \\ [u_i, v_i] & \text{ strictly separates } a \text{ and } b \\ T_i & \text{ is bounded by } P, Q, \text{ cut segments } \xi_i \text{ and } \xi_{i+1} \\ PQ &= \bigcup_{i=0}^k T_i, \text{ int}T_i \cap \text{int}T_j = \emptyset \text{ for } i \neq j \end{aligned}$$

with  $\xi_0 := a, \xi_{k+1} := b$ . Assume that

$$SP(u_i, u_{i+1}) \cap SP(v_i, v_{i+1}) \neq \emptyset \text{ for all } i=1, \dots, k-1.$$

The shortest path  $SP(a, b)$  between two points  $a$  and  $b$  in a simple polygon is due to An and Hoai in [6] without triangulation and theory graph. (To avoid triangulation and theory graph, we do not use Lee and Preparata's algorithm in Sect. 2). Convexity is crucial in An and Hoai's algorithm (see [6]).

*Given:* vertices  $a$  and  $b$  of a simple polygon  $PQ$ , where  $P$  ( $Q$ , respectively) is the polyline formed by vertices of the polygon from  $a$  to  $b$  (from  $b$  to  $a$ , respectively) in counterclockwise order.

*Find:* the shortest path  $SP(a, b)$  between  $a$  and  $b$  inside  $PQ$ .

1. Divide the polygon  $PQ$  into suitable subpolygons  $T_i$  by cut segments  $\xi_1, \dots, \xi_k$  satisfying (1)-(2). Set  $j=0$  and choose initial shooting points  $a^j_i \in \xi_i, i=1, \dots, k$ .
2. Find the shortest path  $Z^j_i := SP(a^j_i, a^j_{i+1})$  in  $T_i$ . Check if all  $Z^j_i (i=1, \dots, k)$  satisfy simultaneously the following
  - (a) If either  $(\pi - \alpha_{-u_i})(\pi - \alpha_{-v_i}) < 0$  and  $\alpha_{-a^j_i} = \pi$ , or  $\alpha_{-u_i} \geq \pi$  (or  $\alpha_{-v_i} \leq \pi$ ) then  $SP(a, b) := \bigcup_{i=1}^k Z^j_i$ . STOP.
  - (b) Else, refine shootings  $a^j_i \in \xi_i$  to ensure that the condition  $\alpha_{-a^j_i} = \pi$  holds true. Set  $j=j+1$ , go to step 2.

Here,  $\alpha_{-u_i}$  ( $\alpha_{-v_i}$ , respectively) is the measure of the angle between two polylines  $SP(u_i, u_{i-1})$  and  $SP(u_i, u_{i+1})$  ( $SP(v_i, v_{i-1})$  and  $SP(v_i, v_{i+1})$ , respectively),  $\alpha_{-a^j_i}$  is the measure of the angle between two polylines  $SP(a^j_i, a^j_{i-1})$  and  $SP(a^j_i, a^j_{i+1})$ , and the conditions (a) and (b) are obtained by convexity of the shortest paths. The number  $k$  of cut segments  $\xi_i, i=1, \dots, k$ , is specified by the user. Thus, we can compute the shortest path between two points inside a simple polygon without triangulation of the entire polygon and graph theory.

#### 4 Shortest Paths on Graphs and Shortest Paths on Polytopes

In this section we describe the use of the ideas of the cut slices parallel to  $xOy$  and the direct multiple shooting method [7] for solving the shortest path problem in 3D. These cut slices are due to Li and Klette's algorithm [10].

We consider the following shortest path problem:

Find the Euclidean shortest path  $Z$  between two vertices  $a$  and  $b$  on the surface of a convex polytope  $D$ .

We split the convex polytope  $D$  into sub convex polytopes  $T_i$  by cut slices  $\xi_1, \dots, \xi_k$  (i.e., a convex polytope shooting grid) as follows:

$\xi_i$  consists of simple polygons  $\xi_i^j = u_i^j u_{i+1}^j \dots u_{n-1}^j \subseteq \text{bd}D \cap \xi_i$

$\xi_i$  is parallel to  $0xy$ ,  $\xi_i$  strictly separates  $a$  and  $b$ ,  $T_i$  is bounded by  $D$ , cut slices  $\xi_i$  and  $\xi_{i+1}$ ,

$\cup_{i=0}^k T_i \subseteq D$ ,  $\text{int}T_i \cap \text{int}T_j = \emptyset$  for  $i \neq j$

with  $\xi_0 := a$ ,  $\xi_{k+1} := b$ .

In the same manner with the method of the algorithm in Sect. 3, instead of cut segments, cut slices parallel to  $0xy$  are used.

These cut slices are constructed via the shortest path on a graph constructed by vertices and edges of the polytope ([10]), where the path is determined by Dijkstra's Algorithm [8]/Thorup's Algorithm [15]. Thus, the number  $k$  of cut slices  $\xi_i$ ,  $i=1, \dots, k$ , is not specified by the user.

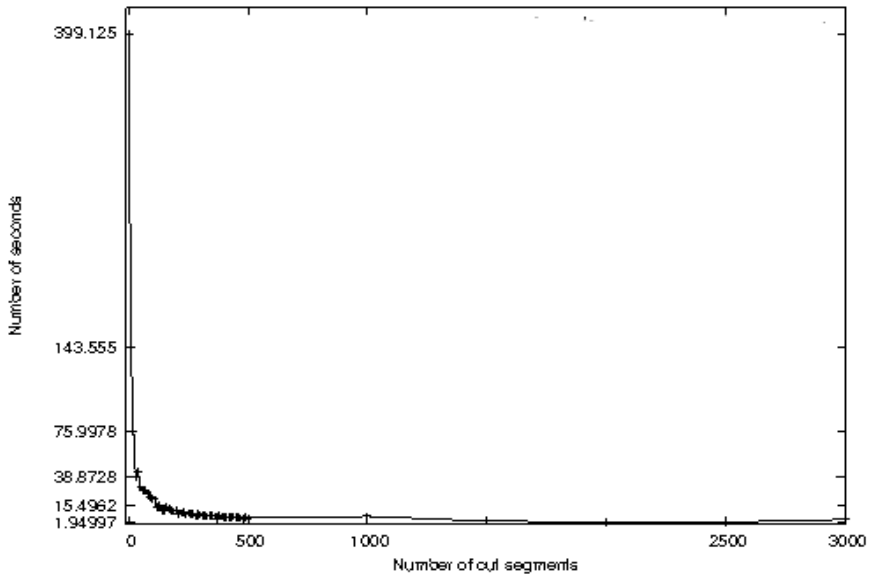
## 5 Numerical Experiences

Both the running time of the algorithm given in Sect. 3 and the peak memory usage of the system significantly reduce as cut segments are used (see Table 1 ([3]) below).

**Table 1.** The peak memory usage of the system significantly reduce as cut segments are used

Number of cut segments	Time (in sec.)	Peak Memory (in MB) (freed on nodes)	Peak Memory (in MB) (freed on nodes)
0	20858.53	0.80	4839.91
5	3641.00	0.13	2040.96
10	1988.84	0.07	610.29
100	298.82	0.00	7.85
1000	283.01	0.00	0.09

The running time of the algorithm given in Sect. 3 decreases significantly from 399.125 seconds to 1.94997 seconds as the number  $\alpha_{a_i} = \pi$  of cut segments increases from 0 to 2500. Here,  $k=0$  means the shortest path between  $a$  and  $b$  inside the polygon  $PQ$  is determined by [5] without cut segments (see Figure 1 [2] below).



**Fig. 1.** The running time of the algorithm given in Sect. 3 decreases significantly as the number  $k$  of cut segments increases

## 6 Conclusion

We have shown that in 2D, convexity contributes to the design of an efficient algorithm for finding the approximate shortest path between two points inside a simple polygon without triangulation of the entire polygons or graph theory. Conversely, in 3D, graph tools are crucial to find an Euclidean shortest path between two points on the surface of a convex polytope.

## References

1. An, P.T., Hai, N.N., Hoai, T.V.: The role of convexity for solving some shortest path problems in plane without triangulation. In: International Conference on Mathematical Sciences and Statistics, Kuala Lumpur, Malaysia, February 5-7. AIP Conference Proceedings, American Institute of Physics, NY (2013)
2. An, P.T., Hai, N.N., Hoai, T.V.: Direct multiple shooting method for solving approximate shortest path problems. *Journal of Computational and Applied Mathematics* 244, 67–76 (2013)
3. An, P.T., Hai, N.N., Hoai, T.V., Trang, L.H.: On the performance of triangulation-based multiple shooting method for 2D shortest path problems. In: International Conference on Advanced Computing and Applications (ACOMP 2013), Ho Chi Minh City, Vietnam, October 23-25 (2013)
4. An, P.T.: Reachable grasps on a polygon of a robot arm: finding convex ropes without triangulation. *Journal of Robotics and Automation* 25(4), 304–310 (2010)

5. An, P.T.: Method of orienting curves for determining the convex hull of a finite set of points in the plane. *Optimization* 59(2), 175–179 (2010)
6. An, P.T., Hoai, T.V.: Incremental convex hull as an orientation to solving the shortest path problem. In: *IEEE Proc. 3rd Int. Conf. Comp. & Auto. Eng.*, Chongqing, China, January 21–23 (2011)
7. Bock, H.G., Plitt, K.J.: A multiple shooting algorithm for direct solution optimal control problems. In: *Proceedings of the 9th IFAC World Congress*, Budapest, pp. 243–247. Pergamon Press (1984)
8. Dijkstra, E.W.: A note on two problems in connexion with graphs. *Numerische Mathematik* 1, 269–271 (1959)
9. Lee, D.T., Preparata, F.P.: Euclidean shortest paths in the presence of rectilinear barriers. *Networks* 14, 393–410 (1984)
10. Li, F., Klette, R.: *Euclidean Shortest Paths: Exact or Approximate Algorithms*. Springer (2011)
11. Mitchell, J.S.B.: Geometric shortest paths and network optimization. In: Sack, J.R., Urrutia, J. (eds.) *Handbook of Computational Geometry*, pp. 633–701. Elsevier Science B. V. (2000)
12. Peshkin, M.A., Sanderson, A.C.: Reachable grasps on a polygon: the convex rope algorithm. *IEEE Journal on Robotics and Automation* 2(1), 53–58 (1986)
13. Phu, H.X.: Ein konstruktives Lösungsverfahren fuer das Problem des Inpolygons kleinsten Umfangs von J. Steiner. *Optimization* 18, 349–359 (1987)
14. O’Rourke, J.: *Computational Geometry in C*, 2nd edn. Cambridge University Press (1998)
15. Thorup, M.: Undirected single-source shortest paths with positive integer weights in linear time. *J. ACM* 46(3), 362–394 (1999)

# Illumination-Robust Local Pattern Descriptor for Face Recognition

Dong-Ju Kim, Sang-Heon Lee, Myoung-Kyu Shon, Hyunduk Kim, and Nuri Ryu

Dept. of Convergence, Daegu Gyeongbuk Institute of Science & Technology (DGIST)  
50-1 Sang-Ri, Hyeonpung-Myeon, Dalseong-Gun, Daegu, 711-873, Korea

**Abstract.** In this paper, we propose a simple descriptor called an extended center-symmetric pattern (ECSP) for illumination-robust face recognition. The ECSP operator encodes the texture information of a local face region by emphasizing diagonal components of a previous center-symmetric local binary pattern (CS-LBP). Here, the diagonal components are emphasized because facial textures along the diagonal direction contain much more information than those of other directions. The facial texture information of the ECSP operator is then used as the input image of an image covariance-based feature extraction algorithm. Performance evaluation of the proposed approach was carried out using various binary pattern operators and recognition algorithms on the extended Yale B database. The experimental results demonstrated that the proposed approach achieved better recognition accuracy than other approaches, and we confirmed that the proposed approach is effective against illumination variation.

**Keywords:** LBP, ECSP, Face Recognition.

## 1 Introduction

Face recognition has become one of the most popular research areas in the fields of image processing, pattern recognition, computer vision, and machine learning, because it spans numerous applications. However, illumination variation that occurs on face images drastically degrades the recognition accuracy [1]. To overcome the problem caused by illumination variation on face images, local texture descriptors such as local binary pattern (LBP) [2], centralized binary pattern (CBP) [3], and center-symmetric local binary pattern (CS-LBP) [4], have recently received increasing interest due to their illumination-robust characteristic. The LBP is a non-parametric kernel which summarizes the local spatial structure of an image. Moreover, it has important properties of robustness against monotonic illumination changes and computational simplicity. Recently, the CBP and CS-LBP were also introduced for face representation. Compared to the LBP, the CBP and CS-LBP produce fewer binary units, so they have the advantage that they can reduce the vector length of corresponding histogram feature. Based on these previous works, this paper introduces a texture descriptor constructed with the proposed extended center-symmetric pattern (ECSP) for illumination-robust face recognition. Unlike previous local texture descriptors that assign a binary code by labelling the pixels toward continuous directions, the proposed ECSP



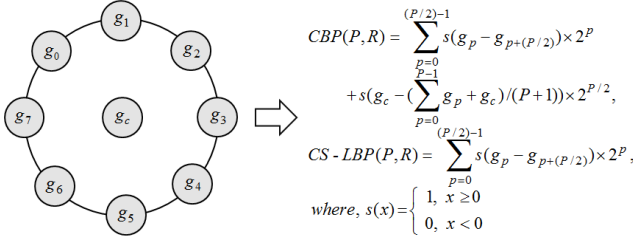
operator encodes the facial textures by reordering the bit priorities as pre-defined directions. Furthermore, this paper proposes a new face recognition approach that directly utilizes a face image transformed by the ECSP descriptor as the input for the image covariance-based feature extraction algorithm, such as two-dimensional principal component analysis (2D-PCA) [5]. This is a new approach compared to most previous research, because previous works utilized the binary pattern descriptors for histogram feature extraction of the face image [2-4]. The proposed approach has the advantages that the illumination effects can be degraded by using the ECSP descriptor and 2D-PCA is more robust against illumination variation than global features such as principal component analysis (PCA) [6], since 2D-PCA is a line-based local feature.

## 2 Methodology

The LBP was originally proposed for texture description, and it has been widely exploited in many applications such as video retrieval, aerial image analysis, and visual inspection. Recently, the LBP has been extensively exploited for facial image analysis, including face detection, face recognition, facial expression analysis, gender/age classification, and so on. The LBP operator labels the pixels of an image by thresholding the 3x3 neighborhood of each pixel with the center value, and considering the results as a binary number, of which the corresponding decimal number is used for labelling [2]. The LBP code is derived by

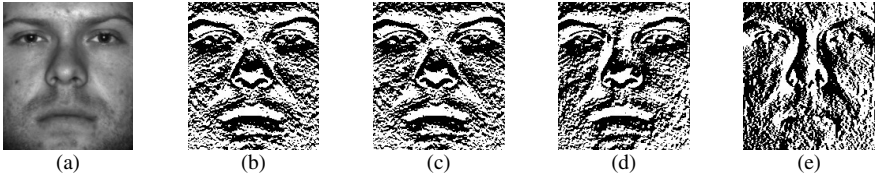
$$LBP(P, R) = \sum_{p=0}^P s(g_p - g_c) \times 2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \quad (1)$$

where  $g_c$  and  $g_p$  denote the center pixel value and neighborhood pixel values, respectively;  $P$  means the number of neighbors; and  $R$  means the radius of the neighborhood. The LBP-based approaches are attracting much attention from researchers due to their advantages of simple computation, robustness to illumination variation, and discriminative ability. However, the LBP operator has several disadvantages: (1) they produce rather long histograms, which affect the recognition speed, especially on large-scale database; (2) under certain circumstance, they miss the local structure as they do not consider the effect of the center pixel; and (3) the binary data produced by them are sensitive to noise [3][4]. Thus, the CBP and CS-LBP have been proposed for face representation more recently. The CBP operator compares pairs of neighbors which are within the same diameter, and compares the central pixel with the mean of all the pixels as shown in Fig. 1. The CS-LBP operator can be also computed by only considering the corresponding patterns of symmetric pixels as shown in Fig. 1. In these methods, instead of comparing the grey level value of each pixel with the center pixel, the center-symmetric pairs of pixels are compared. In addition, they are closely related to the gradient operator. They consider the grey level differences between pairs of opposite pixels in a neighborhood. Thus, they take advantage of both LBP and gradient-based features, and they reduce the computational complexity in comparison with basic LBP.



**Fig. 1.** CBP and CS-LBP operators

To make a more significant pattern against illumination variation, we modified the CS-LBP operator by reordering the bit priorities as pre-defined directions. Generally, the decimal value of most binary pattern operators is created by combining each binary code toward continuous direction. In this view, we can suppose that each binary unit has a different characteristic in terms of facial texture. In other words, previous binary pattern operators have not considered the relation between each bit position and facial texture. Thus, we first investigated their corresponding relation by composing a texture image, in which each texture image is created using only one pair of symmetric pixels. The resultant texture images are shown in Fig. 2, while  $P$  and  $R$  are 8 and 1, respectively. As seen in Fig. 2, the composed texture image (see Fig. 2 (e)) using only horizontal pixels contains more indistinguishable texture than other composed images (see Fig. 2 (b), (c), and (d)). This component has negative effects on the complete binary pattern image of the CS-LBP. Thus, we assign low priority to component of the horizontal directions. In addition, notice that other composed images have similar textures as seen in Fig. 2 (b), (c), and (d).



**Fig. 2.** Relation of bit position and facial texture; (a) Original image, (b) Composed image using only  $g_0$  and  $g_4$  terms, (c) Composed image using only  $g_1$  and  $g_5$  terms, (d) Composed image using only  $g_2$  and  $g_6$  terms, (e) Composed image using only  $g_3$  and  $g_7$  terms

Due to the lack of significant facial textures in the composed image using only horizontal pixels, we rearrange the bit priorities in time of pattern generation as follows:

$$ECSP(P,R) = \sum_{p=0}^{(P/2)-1} s(g_p - g_{p+(P/2)}) \times 2^{w_{p,R}(P)}, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \quad (2)$$

where  $w_{P,R}(p)$  means a weighting function to decide the bit priority. Here, suppose that the  $3 \times 3$  neighborhood pixel positions are set as shown in Fig. 1. When  $P$  and  $R$  are set 8 and 1, respectively,  $w_{8,1}(p)$  for ECSP(8, 1) is defined by

$$w_{8,1}(p) = (3, 1, 2, 0), \quad p = 0, 1, 2, 3. \quad (3)$$

Furthermore, let us suppose that the pixels of an extended  $5 \times 5$  neighborhood are situated like those of the  $3 \times 3$  image patch. Then,  $w_{16,2}(p)$  for ECSP(16, 2) is defined by

$$w_{16,2}(p) = (7, 5, 1, 3, 6, 2, 0, 4), \quad p = 0, 1, 2, 3, 4, 5, 6, 7. \quad (4)$$

In the proposed ECSP operator, we assign the high weight to components of diagonal directions, and we then assign following weight to components of the vertical and horizontal directions as sequential steps. In addition, we expand the image patch to  $5 \times 5$  pixel sizes to obtain a more illumination-robust facial texture, leading to performance improvement as seen in the experimental results. Fig. 3 shows facial texture images transformed by various binary pattern operators, such as LBP, CBP, CS-LBP, ECSP(8, 1), and ECSP(16, 2). As seen Fig. 3, we can confirm that the ECSP operator achieves a more significant facial texture than other operators, since we assign the lowest bit priority to component of the horizontal direction. Moreover, the ECSP(16, 2) image is clearer than the ECSP(8, 1) image as shown in Fig. 3 (e) and (f) in terms of impulse-like noises. Consequently, the proposed ECSP operator seems more stable than other binary pattern images, as shown in Fig. 3, since it has fewer noise components than other images.

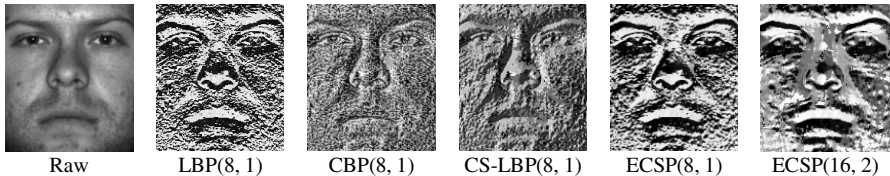


Fig. 3. Example of various binary pattern images

### 3 Experiments

The performance evaluation was carried out using well-known recognition approaches, namely, PCA [6], 2D-PCA [5] and Gabor-wavelets based on LBP [2] on the Yale B database. Note that this work utilized face images transformed by the ECSP descriptor as the input for an image covariance-based feature extraction algorithm, such as 2D-PCA, unlike previous studies. In the Yale B database, we employed 2,414 face images for 38 subjects representing 64 illumination conditions in the frontal pose, in which the subjects comprised 10 individuals in the original Yale B database and 28 individuals in the extended Yale B database. Also, we partitioned the face database

into training and testing sets. Generally, the Yale B database can be subdivided into several subsets depending on the direction of light [7], so we employed several images from various subsets for training, and images from the remaining subsets were used for testing. As a result, the recognition results in relation to the various training sets are shown in Table 1. For the Yale B database, the recognition accuracies of the proposed method using ECSP(16, 2) and 2D-PCA were 95.58%, 96.18%, 97.99%, 98.28% and 98.48%, when the training sets were ‘subset 1’, ‘subset 2’, ‘subset 3’, ‘subset 4’, and ‘subset 5’, respectively. These results demonstrate that the proposed method achieved the best recognition accuracies in most cases. In addition, the proposed method showed performance improvements of 26.08%, 23.07%, 11.20%, 32.44%, and 33.69% over the Gabor-wavelets approach based on LBP when training sets were ‘subset 1’, ‘subset 2’, ‘subset 3’, ‘subset 4’, and ‘subset 5’, respectively. In particular, the recognition results using ECSP images showed significant performance improvements over those of the approaches using CBP and CS-LBP images. Consequently, we experimentally confirmed the robustness and effectiveness of the proposed method under varying lighting conditions.

**Table 1.** Summary of maximum recognition rates

Recognition Approaches		Training Set				
		Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
PCA	LBP	79.89%	68.32%	79.67%	86.14%	85.67%
	CBP	60.57%	26.16%	43.60%	69.12%	47.66%
	CS-LBP	60.86%	37.15%	68.73%	74.97%	55.73%
	ECSP(8,1)	83.74%	57.02%	93.24%	93.98%	91.64%
	ECSP(16,2)	<b>92.86%</b>	<b>92.52%</b>	<b>96.75%</b>	<b>97.53%</b>	<b>95.85%</b>
2D-PCA	LBP	89.24%	83.64%	92.67%	91.25%	96.55%
	CBP	73.68%	63.83%	81.53%	92.16%	77.08%
	CS-LBP	74.30%	70.12%	84.57%	91.46%	80.64%
	ECSP(8,1)	94.97%	92.62%	96.49%	97.31%	97.72%
	ECSP(16,2)	<b>95.58%</b>	<b>96.18%</b>	<b>97.99%</b>	<b>98.28%</b>	<b>98.48%</b>
Gabor-wavelets based on LBP	Raw	57.14%	55.88%	64.75%	48.06%	41.28%
	Histogram	<b>69.50%</b>	<b>73.11%</b>	<b>86.79%</b>	<b>65.84%</b>	<b>64.79%</b>

## 4 Conclusion

This paper proposed a novel face recognition approach that integrated the ECSP image and 2D-PCA under illumination-variant conditions. Through experimental results, the proposed approach showed the best recognition accuracy compared to different approaches; thus, we confirmed the effectiveness of the proposed approach.

**Acknowledgment.** This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (13-IT-03). It was also supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program (Immersive Game Contents CT Co-Research Center).

## References

1. Lian, Z., Er, M.J.: Illumination normalization for face recognition in transformed domain. *Electronics Letters* 46(15), 1060–1061 (2010)
2. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Facial image analysis based on local binary patterns - A survey. *IEEE Trans. Sys.* 41(6), 765–781 (2011)
3. Fu, X., Wei, W.: Centralized binary patterns embedded with image Euclidean distance for facial expression recognition. In: *Int. Conf. Neural Computation* (2008)
4. Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with center-symmetric local binary patterns. In: Kalra, P.K., Peleg, S. (eds.) *ICVGIP 2006*. LNCS, vol. 4338, pp. 58–69. Springer, Heidelberg (2006)
5. Jian, Y., David, Z., Alejandro, F., Yang, J.Y.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(1), 131–137 (2004)
6. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* 3(1), 71–86 (1991)
7. Georghiadis, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* 23(6), 643–660 (2001)

# MAP: An Optimized Energy-Efficient Cluster Header Selection Technique for Wireless Sensor Networks\*

Kanokporn Udompongsuk, Chakchai So-In\*\*, Comdet Phaudphut,  
Kanokmon Rujirakul, Chitsutha Soomlek, and Boonsup Waikham

Department of Computer Science, Faculty of Science, Khon Kaen University  
123 Mitaparb Rd., Naimuang, Muang, Khon Kaen, 40002 Thailand  
{kanokporn.u, chakso, comdet.p, kanokmon.r,  
chitsutha.boonsup}@kku.ac.th

**Abstract.** Recent advances in wireless sensor networks have led to feasibility in implementing a variety of reliable and distributed monitoring and controlling systems used in several areas including environment, healthcare, civil, and military applications. The introduction to novel protocols and their improvements, especially for energy consumption awareness, were due to the major limitations of power-aware tiny sensor nodes. To utilize an overall energy consumption prolonging system lifetime, clustering is one of the promising approaches. By grouping sensors together, the sensor communicates only to its cluster head before gathered, and then forwarded to a base station. In this paper, we evaluate this issue, and then propose an optimization over a well-known hierarchical routing protocol, LEACH, by considering Moving energy window Average and selection Probability (MAP), resulting in an overall energy usage enhancement.

**Keywords:** DCHS, Deterministic Cluster-Head Selection, Hybrid-LEACH, K-LEACH, LEACH, N-LEACH, T-LEACH, W-LEACH, Low Energy Adaptive Clustering Hierarchy, MAP, Wireless Sensor Networks, WSNs.

## 1 Introduction

Advances in micro-electro-mechanical system (MEMS) technology allow the practicality to enable a low power consumption integrated digital electronic device used in multifunctional applications, from military to civil sectors, e.g., security surveillance, disaster management, inventory control, traceability, ambient condition detection, animal tracking, health monitoring, and weather monitoring [1], especially for a tiny sensing, transmitting, and computing logics embedded with dedicated energy power source forming a concept of sensor nodes.

To interact within a sensor farm, in order to scale-up the usability, many wireless communication protocols have been developed/optimized given constraints of

---

\* This work was supported in part by Research Incubation Fund, Khon Kaen University, 2012-2015.

\*\* Corresponding author.

efficiency, reliability, and scalability, forming into Wireless Sensor Networks (WSNs). One of the critical issues in WSNs, network routing, has challenged for years, and to figure out a suitable protocol for communication among a large number of flexible nodes over constraints and limitations, e.g., unknown global addressing, unpredictable and frequent topological change, and most importantly, energy-aware network [2].

Recently, several proposals have been designed to mitigate and overcome some of those issues in various aspects, and generally, there can be classified in three categories [3]: Flat (SPIN, GBR, GOUGAR), Location (GAR, GEAR, SPAN), and Hierarchical (LEACH, TEEN, PEGASIS), and each of which has its own pros/cons.

Consider the first category. All nodes perform equal functions, and so, due to the lack of global identification, network routing follows data-centric routing behavior. In contradictory, the second group considers global addressing information given GPS functions or a signal strength measurement. Finally, a well-known approach based on a cluster routing method providing distinctive advantages on scalability and efficient communication of hierarchical routing for energy-efficient routing in WSNs.

With a hierarchical architecture, proximity sensing is the main task, and then processing and communication tasks are for the high-end cluster head (CH). CH also takes charge of data fusion within each individual cluster range, also exchanges relevant information with other CHs, and finally transmits the data to a base station (BS).

As a result, with a well-organized and efficient cluster construction process contributing to the increase of overall system efficiency; therefore, in this paper, we specifically investigate optimization schemes to enhance the cluster formation based on one of the pioneer hierarchical schemes, LEACH [4], especially the cluster head selection criteria, and then propose a technique for LEACH optimization.

The organization of this paper is as follows: Section 2, we briefly revisit the background of LEACH. Recent proposals over LEACH derivations are discussed, particularly on the cluster head selection criteria in Section 3. Section 4 presents our proposal including Moving window Average and selection Probability (MAP). Later, we comparatively discuss the performance evaluation of LEACH derivations in Section 5. Finally, the conclusions and the future research direction are drawn in Section 6.

## 2 Low Energy Adaptive Clustering Hierarchy (LEACH)

Heinzelman, W. et al. [4] proposed LEACH (Low Energy Adaptive Clustering Hierarchy) as one of the pioneer hierarchical clustering techniques used in WSNs. LEACH focuses on clustering sensor nodes into a group with its CH. To minimize an overall power usage, each sensor within the same group will merely communicate to its CH before gathering, and then forwarding to BS. In addition, the selection of CH could be uniformly rotated in order to evenly distribute the energy load so that an overall network system lifetime could be prolonged.

LEACH performs two main operations. The setup operation involves the clustering formation including the CH selection process; and the steady-state operation includes the actual data transfer from CHs to BS. Consider the setup operation. Each node  $n$

generates a random number in between 0 and 1. Then, node  $n$  performs a threshold computation  $T(n)$ , as stated in equation (1), and makes a decision on CH, i.e., if that number is less than the threshold, then it will be elected CH in the current round  $r$ .

$$T(n) = \begin{cases} \frac{P}{1 - P \times (r \bmod \frac{1}{P})} & , \text{if } n \in G \\ 0 & \end{cases} \quad (1)$$

In this equation,  $P$  denotes as the pre-determined fraction of the elected cluster head node.  $G$  denotes as a set of sensor nodes not being selected to be CHs in the last  $1/P$  rounds. Once a selection process has accomplished, the elected CHs broadcast advertisement messages to other nodes leading to a final decision to join each CH group. Note that the key decision criterion is based on received signal strength from CHs.

After each CH receives the responses back from the sensor nodes, it creates a TDMA schedule and determines time slots for each node for further communication in order to mitigate the collision transmission. CHs notify their member nodes with the schedule. Note that in case of multiple CHs' networks, each cluster applies CDMA codes to reduce the transmission interference from each other.

Consider the steady state operation. Each sensor node can start sensing and transmitting data to its own CH. After that, CHs aggregate all of the data, before finally forward to its corresponding BS. Note that after a given time interval, the network turns back to the setup operation, and then enters another round of selecting new CHs so that uniform energy dissipation in the network can be obtained.

### 3 Related Work in LEACH Cluster Head Selection Techniques

Although LEACH can prolong a network system lifetime by normalizing a node energy uniform distribution during CH selection process, a traditional LEACH leaves rooms for enhancements, e.g., node energy, transmission levels, and protocol modification, which may result in shortening the overall system lifetime. Recently, several proposals have been introduced to overcome these issues, and many of them recommend several optimizations [5]. However, in this paper, we focus on improving the cluster head selection process.

Referring to several surveyed approaches, e.g., A-LEACH, C-LEACH, E-LEACH, K-LEACH, P-LEACH, T-LEACH, and W-LEACH, performing the performance evaluation focusing on the cluster head selection criteria [6]. Here, we will briefly discuss on their outstanding techniques including some additional recent proposals.

For example, Thein, M.C.M. and Thein, T. [7] adopted the optimum number of clusters ( $K_{opt}$ ) to make the product towards the traditional LEACH and the energy factor or current energy ( $E_{residual}$  or  $E_{current}$ ) over initial energy ( $E_{initial}$  or  $E_{max}$ ). Here, the optimum number of clusters is derived which has given the factor of coverage area, the effects of transmission energy consumed in both free-space ( $\epsilon_{fs}$ ) and multi-path ( $\epsilon_{mp}$ ), and distance to BS ( $d_{toBS}$ ); and we call K-LEACH as follows:



$$T(n) = \begin{cases} \frac{P}{1 - P \times \left(r \bmod \frac{1}{P}\right)} \times \frac{E_{residual}}{E_{initial}} \times K_{opt} & , if n \in G \\ 0 & \end{cases} \quad (2)$$

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}} \frac{\sqrt{\epsilon_{fs}}}{\sqrt{\epsilon_{mp}}} \frac{M}{d_{toBS}^2} \quad (3)$$

In addition, Hou, R. et al. [8] proposed T-LEACH which considered the remnant power of sensor nodes in order to balance network loads and changes the round time by introducing a new probability ( $P_{head}$ ), the optimum number of clusters over all nodes instead of that used in the traditional LEACH. This probability took the distance between CHs and BS into account including the area coverage ( $M$  is the length of node distributing field) and the number of existing nodes as follows:

$$T(n) = \begin{cases} \frac{P_{head}}{1 - P_{head} \times \left(r \bmod \frac{1}{P_{head}}\right)} \times \frac{E(t)}{E_{total}(t)} & , n \in G \\ 0 & \end{cases} \quad (4)$$

$$P_{head} = \frac{\sqrt{N}}{\sqrt{2\pi}} \frac{\sqrt{\epsilon_{fs}}}{\sqrt{\epsilon_{mp}}} \frac{M}{d_{toBS}^2 \times N} \quad (5)$$

Recently, in early 2013, So-In, C. et al. [6] proposed the improvement over K-LEACH by introducing the use of moving average window ( $w$ ) to smooth the fluctuation of current energy in each round, the so-called W-LEACH as follows:

$$T(n) = \begin{cases} \frac{P}{1 - P \times \left(r \bmod \frac{1}{P}\right)} \times \frac{E_{mov}}{E_{initial}} \times K_{opt} & , if n \in G \\ 0 & \end{cases} \quad (6)$$

$$E_{mov} = \frac{E_{r-w} + \dots + E_{r-2} + E_{r-1} + E_r}{w} \quad (7)$$

As stated in [6], W-LEACH outperforms other LEACH derivations when considering only the energy factor enhancement; however, there are also some proposals focusing on the improvement given the selection probability. For example, Azim, A. et al. [9] proposed the modification over energy selection probability of initial and current energy given number of nodes ( $N$ ) and number of cluster ( $k$ ) modified in the traditional LEACH, the so-called Hybrid-LEACH by as follows:

$$T(n) = \begin{cases} \frac{k}{N - k \times \left(r \bmod \frac{N}{k}\right)} \times \frac{E_{current} - E_{initial}}{E_{current}} & , if n \in G \\ 0 & \end{cases} \quad (8)$$

Similarly, Li, Y. et al. [10] proposed the modification over Azim, A. et al [9] and the traditional LEACH (using probability  $P$  instead of factors  $k$  and  $N$ ), and we call N-LEACH as follows:

$$T(n) = \begin{cases} \frac{P}{1 - P \times \left(r \bmod \frac{1}{P}\right)} \times \frac{E_{initial} - E_{current}}{E_{initial}} & , if n \in G \\ 0 & \end{cases} \quad (9)$$

Furthermore, Handy, M.J. et al. [11] proposed an enhancement scheme over LEACH, the so-called Deterministic Cluster-Head Selection (DCHS) by considering the selection probability based on the increase probability to become cluster head. Here,  $r_s$  is used as the number of consecutive rounds in which a node has not been cluster-head, and so when  $r_s$  reaches the value  $1/P$ , the threshold is reset to the value it had before the inclusion of the remaining energy as follows:

$$T(n) = \begin{cases} \frac{P}{1 - P \times (r \bmod \frac{1}{P})} \times \left[ \frac{E_{current}}{E_{max}} + \left( r_s \operatorname{div} \frac{1}{P} \right) \times \left( 1 - \frac{E_{current}}{E_{max}} \right) \right], & \text{if } n \in G \\ 0 & \end{cases} \quad (10)$$

## 4 Moving Energy Window Average Probability (MAP)

As we briefly discussed on the various threshold criteria, several modifications and optimizations have been proposed, and primarily based on three main factors: energy, perimeter, and number of clusters.

Principally, the energy factor applying the current node energy over either total energy for all nodes or initial/maximum node energy was considered, i.e., K-LEACH and T-LEACH. In order to smoothen out and absorb the probable fluctuation of energy consumption in each round, our previous work focused on the reduction of energy fluctuation by introducing the (weighed) moving average window or W-LEACH.

However, W-LEACH lacks of considering the CH selection probability, similarly stated in N-LEACH, Hybrid-LEACH, and DCHS, and that may make W-LEACH energy distribution un-uniformly resulting into shortening the network system lifetime. Note that as shown in the intensive simulation later, only modifying the selection probability over either N-LEACH and Hybrid-LEACH or DCHS may not yield a significant improvement. Thus, with the optimization over N-LEACH and DCHS, we propose the improvement over W-LEACH by introducing the Moving energy window Average and selection Probability (MAP) factor ( $E_{MAP}$ ) as follows:

Note that for other smoothing derivatives are also applicable, e.g., exponential, modified, and adaptive filtering moving averages [12-13]; however, with only simple moving average results in high performance with low computational complexity.

$$T(n) = \begin{cases} \frac{P}{1 - P \times (r \bmod \frac{1}{P})} \times K_{opt} \times E_{MAP}, & \text{if } n \in G \\ 0 & \end{cases} \quad (11)$$

$$E_{MAP} = 1 - \left( \frac{E_{mov} - E_{current}}{E_{mov}} \right) \quad (12)$$

$$E_{mov} = \frac{E_{r-w} + \dots + E_{r-2} + E_{r-1} + E_r}{w} \quad (13)$$

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}} \frac{\sqrt{\varepsilon_{fs}}}{\sqrt{\varepsilon_{mp}}} \frac{M}{d_{tobs}^2} \quad (14)$$

## 5 Performance Evaluation

In this section, we discuss the evaluation process to illustrate the performance of MAP over LEACH derivations focusing on the cluster head selection criteria.

**Table 1.** Configuration Parameters

Parameters	Symbol	Values
Number of Nodes	$N$	100
Initial Node Energy	$E_{init}$	0.5 J
Percentage of Cluster Head Selection	$P$	0.05
Maximum Number of Rounds	$R_{max}$	5000
Energy Required in Sending/Receiving	$E_{TS}$	50 nJ/bit
Sensing Area	$M \times M$	100m $\times$ 100m
Data + Control	$k + L_{ctrl}$	4000 bits + 100 bits

### 5.1 Simulation Setup

To comparatively show the performance of MAP, parameters and configurations are similar to what described in a traditional LEACH [4] including some recommendations from the surveyed proposals. Here, due to the space limitation, there are two main evaluation metrics: left-over energy (Joules - J) over times (number of rounds) and number of dead nodes (zero in energy or not enough energy to transmit).

Consider the network architecture. We limited a hierarchical architecture to two levels: one from member nodes to its own CH, and the other one is from CH to BS. Note that when all member nodes are dead, it denotes as network failure. We conducted the simulation over three trials resulting in mean and standard deviation.

Table 1 shows the baseline simulation parameters [4]. In general, the number of nodes ( $n$ ) was initialized to 100 with random placement over 100m $\times$ 100m area. BS is located at the center. All nodes are no longer mobile once they are randomly deployed. The baseline simulation tool is LEACH module-based on MATLAB [4].

Note that each node initially has the same value of energy ( $E_{init}$ ) which will be draining over times depending upon the probability to be elected as CHs which then functions as the aggregator/forwarder from its members to BS. In addition, every node transmits a  $k$  bits data packet per round ( $r$ ) including  $L_{ctrl}$  bits for control.

We performed the evaluation of MAP over other LEACH derivations including a traditional LEACH: 1) T-LEACH, 2) K-LEACH, 3) W-LEACH, 4) N-LEACH, 5) Hybrid-LEACH and 6) DCHS according to two main metrics. Note that we also did the intensive simulation to figure out the suitable window size, and we selected 10 for our proposal (outstanding performance); and 500 for W-LEACH [6]; however, with window size in range over 10 and 1000, the results remain superior over the others.

### 5.2 Simulation Results

Figures 1 and 2 show the evaluation results. Consider the energy load dissipation over times (number of rounds) shown in Fig. 1. Generally, all LEACH derivations can

dissipate the energy consumption over all sensor nodes over times. Similarly to what discussed in [6], W-LEACH outperforms T-LEACH and K-LEACH, accordingly, and especially, all of those techniques are superior to a traditional LEACH. In addition, when the selection probability is included, although Hybrid-LEACH and DCHS performances are similar to that of LEACH, N-LEACH has higher energy efficiency gain than the others. In addition, finally, the performance of MAP is outstanding which can efficiently balance the overall energy dissipation resulting in the increasing of system lifetime.

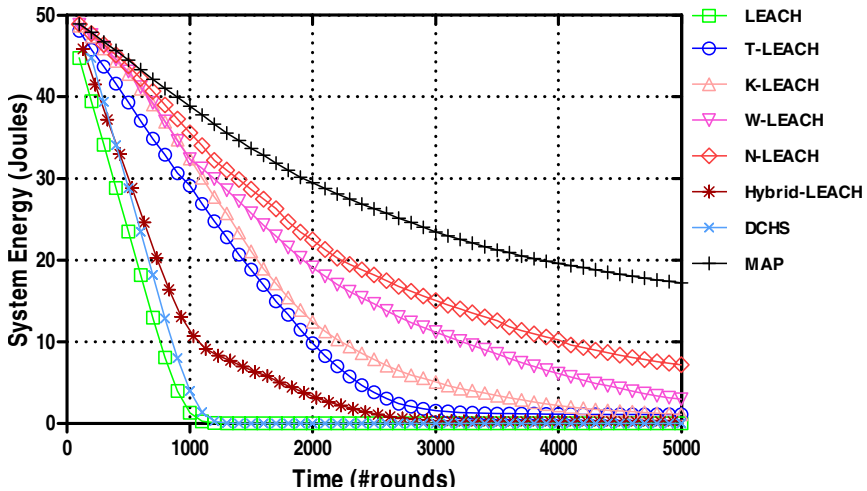


Fig. 1. LEACH derivations: energy consumption over times: y axis (J) and x axis (#rounds)

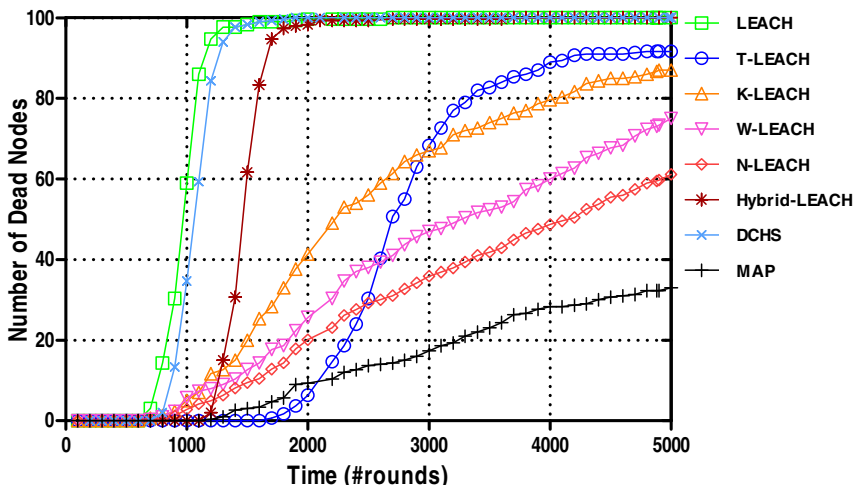


Fig. 2. LEACH derivations: number of dead nodes over times: y axis (n) and x axis (#rounds)

Consider the number of dead nodes over times shown in Fig. 2. Likewise, the increasing trend of dead nodes are prolonging for all LEACH derivations vs. a traditional one. Moreover, although, in the long run, W-LEACH (vs. T-LEACH) are outstanding when the only energy factor is considered, MAP can efficiently maintain the number of dead-nodes in each round uniformly resulting in the superior clustering technique to the others, i.e., N-LEACH, W-LEACH, K-LEACH, T-LEACH, Hybrid-LEACH, DCHS, LEACH, respectively, when the selection probability is involved. Note that the averages of standard deviation for all simulations are less than 1.5 for both two scenarios applying for all LEACH derivations.

## 6 Conclusions

LEACH is one of the pioneer hierarchical clustering techniques developing to prolong network system lifetime by balancing the probability being a cluster head among sensor nodes so that the energy distribution tends to be equally consumed; however, several aspects could be improved, especially, in terms of energy factor; therefore, in this paper, we investigated and performed the evaluation, especially on the threshold criteria to select the cluster head over recent LEACH optimizations, i.e., K-LEACH, T-LEACH, W-LEACH, N-LEACH, Hybrid-LEACH, and DCHS including a traditional LEACH.

To enhance LEACH derivation further, we proposed a modification over W-LEACH by including a selection probability over moving energy window average (MAP) resulting in a superior energy load distribution performance leading to the increase of network system lifetime. Note that although MAP can achieve the performance improvement, more investigation could be performed in other scenarios and constraints including QoS aware mechanism, data aggregation technique, and multi-level transmission. Additionally, as for other factors, which may affect the energy consumption, comprehensive simulation and analysis could be performed including network density and diversity, network dimension, and heterogeneous data traffic.

## References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* 40(8), 102–114 (2002)
2. Pantazis, N.A., Nikolidakis, S.A., Vergados, D.D.: Energy-Efficient Routing Protocols in Wireless Sensor Networks: A Survey. *IEEE Commun. Survey & Tutorials* 15(2), 551–591 (2012)
3. Al-Karaki, J.N., Kamal, A.E.: Routing Techniques in wireless sensor network: a survey. *IEEE Wireless Commun.* 11(6), 6–28 (2004)
4. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Trans. on Wireless Commun.* 1(4), 660–670 (2002)
5. Tyagi, S., Kumar, N.: A systematic review on clustering and routing techniques based upon LEACH protocol for wireless sensor networks. *J. of Network and Computer Applications* 36(2), 623–645 (2013)

6. So-In, C., Udompongsuk, K., Phudphut, C., Rujirakul, K., Khunboa, C.: Performance Evaluation of LEACH on Cluster Head Selection Techniques in Wireless Sensor Networks. In: Meesad, P., Unger, H., Boonkrong, S. (eds.) IC2IT2013. AISC, vol. 209, pp. 51–61. Springer, Heidelberg (2013)
7. Thein, M.C.M., Thein, T.: An Energy Efficient Cluster-Head Selection for Wireless Sensor Networks. In: Inter. Conf. on Intelligent Systems, Modeling and Simulation, pp. 287–291. IEEE Press, UK (2010)
8. Hou, R., Ren, W., Zhang, Y.: A wireless sensor network clustering algorithm based on energy and distance. In: Inter. Workshop on Computer Science and Engineering, pp. 439–442. IEEE Press, USA (2009)
9. Azim, A., Mohammad, M.I.: Hybrid LEACH: A Relay Node Based Low Energy Adaptive Clustering Hierarchy for Wireless Sensor Networks. In: Inter. Conf. on Commun., pp. 911–916. IEEE Press, Kuala Lumpur (2009)
10. Li, Y., Ding, L., Liu, F.: The Improvement of LEACH Protocol in WSN. In: Inter. Conf. on Computer Science and Network Technology, pp. 1345–1348. IEEE Press, Harbin (2011)
11. Handy, M.J., Haase, M., Timmermann, D.: Low Energy Adaptive Clustering Hierarchy with Deterministic Cluster-Head Selection. In: IEEE Conf. on Mobile and Wireless Commun. Networks, pp. 368–372. IEEE Press, USA (2002)
12. Makridakis, S., Wheelwright, S.C.: Adaptive Filtering: An Integrated Autoregressive/Moving Average Filter for Time Series Forecasting. *Opl Res. Q.*, Pergamon Press 28(2), 337–425 (1977)
13. Hyndman, R.J., Athanasopoulos, G.: Forecasting Methods and Applications (2012), <http://www.otexts.com/fpp/>

# MRP-NEP: A Non-Equal-Probability Multicast Routing Protocol for Target Tracking in Wireless Sensor Networks

Boyu Diao<sup>1,2</sup>, Peipei Li<sup>1</sup>, Zhulin An<sup>1</sup>, Fei Wang<sup>1,2</sup>, and Yongjun Xu<sup>1</sup>

<sup>1</sup> Institute of Computing Technology, Chinese Academy of Sciences,  
Beijing 100190, China

{diaoboyu2012, lipeipei, anzhulin, wangfei, xyj}@ict.ac.cn

<sup>2</sup> University of Chinese Academy of Sciences,  
Beijing 100049, China

diaoboyu12@mails.ucas.ac.cn

**Abstract.** Target tracking is regarded as one of the key applications of wireless sensor networks (WSNs). Due to the limited energy in sensor nodes, reducing energy consumption while maintaining acceptable accuracy is an important issue in target tracking. Previous work usually awakens only those nodes adjacent to the target to save power. However, simple full-flood routing protocol used in those work for awakening brings significant communication overhead, which consumes energy two more orders of magnitude compared with computing. In this paper, we proposed a novel routing protocol with non-equal-probabilistic forwarding to reduce communication overhead. In our protocol, sensor nodes forward packets with a probability which is mainly determined by how much the node's location deviates from the direction of target motion. The simulation results show that our protocol can save energy by 30% while tracking a moving target.

**Keywords:** mobicast routing protocol, non-equal-probability, target tracking, wireless sensor networks.

## 1 Introduction

Wireless Sensor Networks (WSNs) is becoming increasingly important in both civilian and military applications. With various sensors equipped, WSNs can provide comprehensive sensing information of the environment. Target tracking is one of the key applications in WSNs. A large number of nodes deployed in a particular zone, once a target steps into the zone, sensor nodes will work cooperatively to achieve the data of interest, such as velocity, direction, number of targets, and etc. With the advantages of flexible deployment and self-organizing, target tracking in WSNs has been widely used in fields like habitat monitoring, battlefield monitoring, vehicle tracking and so on. As the tiny sensor nodes cannot provide sufficient energy, low-power consumption is a serious challenge.

Target tracking protocols are widely researched in last decade. If energy consumption is not considered, all nodes can be set awake to monitor the environment [1-2]. Of course this approach may waste large amount of energy and is not practical. There is a variety of energy management approaches for target tracking in WSNs, for example, we can only set cluster heads in awake state and others asleep. When targets approaching, cluster heads will wake up their members as soon as possible [3-4]. This is a feasible method but cluster heads suffer a high risk of running out of energy much earlier. Another approach is dynamically setting nodes around targets in wake state to track targets and other nodes asleep [5-6]. This approach is obviously effective to leverage between energy consumption and tracking accuracy. In order to wake up nodes in exact positions when targets approaching, sensor nodes must work in a cooperative way by communicating with each other, thus an energy efficiency routing protocol is necessary as support. Multicast routing protocol can provide a good guarantee for both time and space, that is, the protocol can guarantee packets forwarded only in a specific area at a particular moment. In view of the advantages, various multicast routing protocols are applied for target tracking in WSNs.

Qingfeng Huang et al. proposed a just-in-time multicast protocol in constraints of spatiotemporal for target tracking in wireless sensor networks [7]. They introduce the conception of k-coverage zone. But the k-coverage zone which is based on  $\Delta$ -compactness of the whole networks is difficult to calculate. This protocol has low scalability and high computational cost. To improve scalability and reduce computational cost, Qingfeng Huang et al proposed another multicast protocol for wireless sensor networks [8] which optimized the algorithm to calculate local compression value, but this protocol cannot provide spatiotemporal guarantee. To overcome the disadvantages in [7-8], Qingfeng Huang et al. proposed a planar-based reliable multicast protocol [9] which introduces the concept of a node's planar spatial neighborhood, the mechanism of *Greedy Forwarding* and *Timed Forwarding*. It provides a high scalability, but each node must update its planar spatial neighborhood list periodically, and this will bring a large amount of extra energy consumption.

All aforementioned protocols [7-9] have their own advantages and disadvantages, but they all share a common disadvantage, that is, nodes in transfer zone employ a full flooding routing protocol to wake up nodes in forwarding zone. In scenarios of low speed (below 10 m/s) target tracking, energy wasted by full flooding is not obvious. But in scenarios of high speed (above 50 m/s) target tracking, the situation is different. For UAVs tracking as an example, UAVs have a general speed between 50 m/s and 100 m/s. If the sampling frequency is 1 Hz, network must wake up nodes in right direction in advance. As the target speed is as high as 100 m/s, sensor network has to forward the wake-up packets in range of 100 meters. If full flood protocol is employed, it will waste huge energy which is limited.

And this is precisely the problem addressed by this paper.

In scenarios of high speed targets tracking in WSNs, we bring geographical-position-based forwarding probability to decide whether a node forwards packages or not, and use this idea for delivering information of moving targets from nodes to nodes in certain directions. With this idea we present a novel target tracking protocol for target tracking in WSNs.



Rest of the paper is organized as follows: In section 2, we illustrate the basic idea of proposed protocol and clarify some main conceptions. In Section 3, the protocol is formulated. In Section 4, we will present how the protocol works in detail. In Section 5, simulation and results are given. And in Section 6, we conclude the paper with summary.

## 2 Basic Idea

Multicast routing protocols in [7-9] are only suitable for a low speed targets. Under the condition of network nodes location self-aware, aiming at tracking high speed targets like UAVs, we will propose a novel low power consumption protocol: non-equal-probability multicast routing protocol (MRP-NEP) in this chapter

Some concepts must be clarified here. Target sensing area is a circular area which takes the target as the center of the circle and a certain length  $D$  as radius (Area  $A$  in Fig. 1). Target related area is the area between the target sensing areas of two adjacent moments, likes Area  $B$  in Fig. 1 which is made up of two symmetry right triangles.

In order to ensure the continuity of tracking, nodes in area of the following moment should be waken up earlier. When wake-up packages transferring in the target related area. To reduce the unnecessary energy consumption caused by full-flooding, we introduce geographical-position-based forwarding probability to decide whether a node in target related area forwards the wake-up package or not. The forwarding probability is set as follows: nodes close to the direction of targets' motion have a higher forwarding probability, because packages forwarded by nodes close to the direction of targets can be received more easily by nodes in predicted target sensing area.

## 3 Protocol Formulation

In this paper, the application scenario is based on the following assumptions:

1. Nodes are homogeneous, and are randomly deployed in a rectangle area.
2. Every node is aware of its own position. (We can use trilateration with some anchor nodes during network initialization phase)
3. Every target maintains uniform liner motion from time  $t$  to time  $t+I$ .

As shown in Fig.1, two solid black spots represent target positions at two adjacent moments. As mentioned above, Area  $A$  in Fig.1 is target sensing area in time  $t$ . Area  $B$  in Fig.1 is target related area in time  $t+I$ . Area  $C$  in Fig.1 is target sensing area in time  $t+I$  which is predicted by nodes in Area  $A$ . Nodes in Area  $A$  will work cooperatively both tracking the target and make prediction of the following trajectory.

Then with the help of nodes in Area  $B$ , wake-up packages will be delivered to nodes in Area  $C$ . When a target steps into Area  $C$ , the nodes will repeat the steps mentioned above, and nodes in Area  $A$  will go back to sleep to save energy.

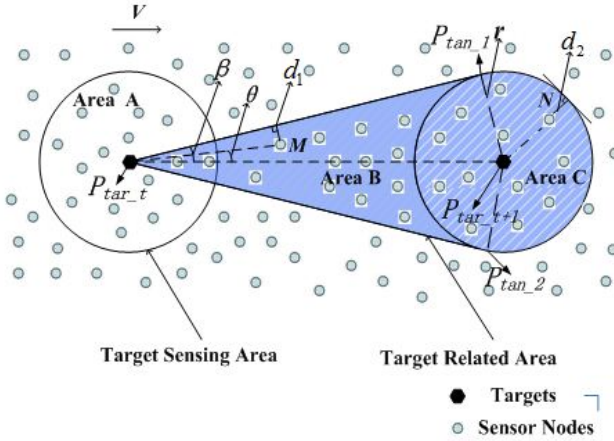


Fig. 1. Model of MRP-NEP

A wake-up package is formed in this manner:

$$\langle M_t, M_{t+1}, P_s, T \rangle$$

Where  $M_t$  is the target position in time  $t$ ,  $M_{t+1}$  is the target position in time  $t+1$ ,  $P_s$  is position of the sender node,  $T$  is the time when package generated. During the package delivery, the contents of the package will change correspondingly.

## 4 Protocol Description

MRP-NEP is a gradually screening process to decide which node to forward the wake-up package and has following main steps. We will present what is forwarding probability made up of and how to calculate it in detail. Notations used in MRP-NEP are listed in Table 1.

- Step 1. Predict the target position of next moment and target sensing area of next moment. Prediction algorithm has a variety of choices, like unscented Kalman filter [10], etc.
- Step 2. Determine the target related area and target sensing area of next moment. Calculate the coordinates of tangency points. As shown in Fig. 1, target related area is determined by target positions  $P_t, P_{t+1}$  and two tangency points  $P_{tan1}, P_{tan2}$ .  $P_{tan1}, P_{tan2}$  can be Calculated by the following formulas. In Formula (1) and (2),  $X_{tan1}$  and  $Y_{tan1}$ , represent abscissa and ordinate of point  $P_{tan1}$ . After two tangency coordinates are determined, the target related area is formed.

$$(X_{tan1} - X_{t+1})^2 + (Y_{tan1} - Y_{t+1})^2 = r^2 \tag{1}$$

$$\frac{Y_{tan1} - Y_t}{X_{tan1} - X_t} \times \frac{Y_{tan1} - Y_{t+1}}{X_{tan1} - X_{t+1}} = -1 \tag{2}$$

**Table 1.** Parameters Instruction

Parameters	Instruction
$P_t(X_t, Y_t)$	Target position at time
$P_{t+1}(X_{t+1}, Y_{t+1})$	Target position at time $t+1$
$P_{tan1}(X_{tan1}, Y_{tan1})$	Coordinates of two tangency points
$P_{tan2}(X_{tan2}, Y_{tan2})$	
$R$	Radius of target sensing area
$M$	A certain sensor node in target related area
$\theta$	Angle between Line $P_m P_t$ and Line $P_{t+1} P_{tan1}$
$\beta$	Angle between Line $P_t P_{tan1}$ and Line $P_t P_{t+1}$
$d_l$	Distance from $M$ to Line $P_t P_{tan1}$

- Step 3. Determine whether a node in the target related area. If in, turn to Step 4; if not, drop the package and keep asleep.
- Step 4. Determine whether nodes located in target relates area to send packages or not. This is the key process of protocol, we will describe it in more detail.
- Step 4-1. Determine the *forwarding probability* of the nodes in target related area. Forwarding probability is consisted by two parts: *angle probability* which depicting how well the node fits the direction of the target, and *distance probability* which depicts how near the node is to the edge of target related area. Taking Fig. 1 as an example, we can get values of angle probability and distance probability from Formula (3) and (4).
- Step 4-2. Get the *forwarding probability* from *angle probability* and *distance probability*. Forwarding probability is a linear combination of the angle probability and distance probability, according to Formula (5)
- Step 4-3. Determine whether to forward the package or not. Generate a random float number  $P$  between 0 and 1. If  $P_{send} > P$ , forward the package, otherwise drop the package.

$$P_{angle} = \frac{\theta}{\beta} \quad (3)$$

$$P_{edge} = \frac{d_l}{r} \quad (4)$$

$$P_{send} = ratio \times P_{edge} + 1 - ratio \times (1 - P_{angle}) \quad (5)$$

## 5 Simulation and Results

To verify the validity of our protocol, we simulated it on OPNET 14.5, the MAC protocol is based on IEEE 802.15.4. 1000 sensor nodes are uniformly randomly

distributed over a  $600 \times 250$  m<sup>2</sup> fit field. Communication and sensing radius is randomly set as 35 m. We ran the simulator 1000 times, and take the average as our final results to reduce error.

In the model of our protocol, packet forwarding probability is affected by probability combination coefficient ratio, we firstly determine the optimal value of ratio and we can get it when the average forwarding angle is the smallest.

We ran our simulator in the following two scenarios: first is velocity changing over time, from 50 m/s to 80 m/s, increasing by 5 m per second. Second is direction changing over time, angle between target direction and horizontal line changing from 5° to 35°, increasing by 5° per second. In both two scenarios, we mainly care about following three system parameters:

1. Energy consumption: total energy consumption of target related area.
2. Package cost in target related area: Total packages forwarded in target related area.
3. Wake-up rate: The ratio of nodes wakened up and total number of nodes in target sensing area of the following moment. The more nodes be wakened, the more nodes participating to track targets, and the higher accuracy we get [5]. That is to say, wake-up rate is significantly positively related to tracking accuracy.

We will compare our protocol, a non-equal-probability multicast protocol for target tracking, which is recorded as MRP-NEP for short, with the full flood protocol in target related protocol, which is recorded as MRP-FF. Results and analysis are shown below.

### 5.1 Get Optimal Value of Ratio

First of all, we examined how probability combination coefficient will affect average angle of nodes delivering packages. As shown in Fig. 2(a), it could be seen the node average angle is minimum when  $ratio = 0.6$ , average angle is 0.16283. So, all the results below are in the condition of  $ratio = 0.6$ .

### 5.2 Total Energy Consumption

From Fig. 2(b) and Fig. 2(c), we can see MRP-NEP reduce package cost by 30% compared to MRP-FF and also saves about 30% energy compared to MRP-FF in both scenarios, because energy consumption is mainly made up of package forwarding, we reduce the package cost and energy is saved as well.

### 5.3 Wake-Up Rate

Finally, we will analysis wake-up rate. Fig. 2(d) shows how wake-up rate changes over time in two scenarios. We can see that wake-up rate is more than 50% when target velocity changes over time which can provide wonderful support for the collaborative tracking. In scenarios of direction changing over time, 40% awake nodes can also ensure the accuracy of tracking.

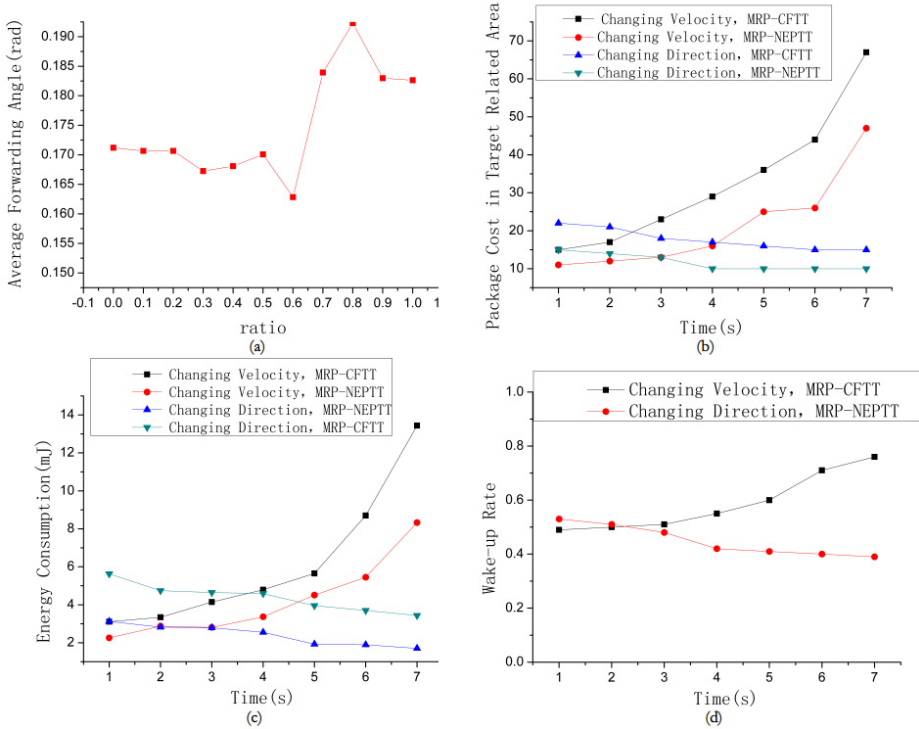


Fig. 2. (a) Comparing ratio due to average forwarding angle. Comparing Package cost (b) and energy consumption (c) due to two protocols in two different scenarios, Wake-up rate in two different scenarios (d).

## 6 Conclusion

Wireless sensor networks are generally deployed in critical environments, energy consumption is a severe problem, especially in target tracking. Predicting and just waking up nodes close to the direction of target motion is an effective way to save network energy resource. A non-equal-probability multicast protocol, proposed in this paper, is an effective routing protocol supporting the solution above. The simulation results show the proposed protocol reduces energy consumption by 30% compared with traditional flooding routing protocol [7-9], to save network energy even further and prolong the network lifetime. At the same time, the idea of spatial-probability-based multicast routing protocol will be a useful reference for design of wireless sensor network routing protocol.

**Acknowledgments.** This paper was supported in part by Important National Science & Technology Specific Projects under grant No.(2010ZX03006-002), the National Basic Research Program of China (973 Program) (No. 2011CB302803), National

Natural Science Foundation of China (NSFC) under grant No.(61173132, 61003307), the National Science Foundation for Young Scientists of China under grant No. (61202430, 61303245), and Internet of Things Technology R&D and Industrialization Specific Projects entitled “R&D and Industrialization of Conformance Testing Smart Instruments for IPv6 Network Protocol in Wireless Sensor Networks”.

## References

1. Kung, H.T., Vlah, D.: Efficient location tracking using sensor networks. In: 2003 IEEE Wireless Communications and Networking, WCNC 2003, vol. 3, pp. 1954–1961 (2003)
2. Lin, C.Y., Peng, W.C., Tseng, Y.C.: Efficient in-network moving object tracking in wireless sensor networks. *IEEE Transactions on Mobile Computing* 5(8), 1044–1056 (2006)
3. Guo, M., Olule, E., Wang, G., Guo, S.: Designing energy efficient target tracking protocol with quality monitoring in wireless sensor networks. *The Journal of Supercomputing* 51(2), 131–148 (2010)
4. Chen, W.P., Hou, J., Sha, L.: Dynamic clustering for acoustic target tracking in wireless sensor networks. In: Proceedings of the 11th IEEE International Conference on Network Protocols, pp. 284–294 (2003)
5. Zhang, W., Cao, G.: Dctc: dynamic convoy tree-based collaboration for target tracking in sensor networks. *IEEE Transactions on Wireless Communications* 3(5), 1689–1701 (2004)
6. Zhao, F., Shin, J., Reich, J.: Information-driven dynamic sensor collaboration. *IEEE Signal Processing Magazine* 19(2), 61–72 (2002)
7. Huang, Q., Lu, C., Roman, G.-C.: Mobicast: Just-in-time multicast for sensor networks under spatiotemporal constraints. In: Zhao, F., Guibas, L.J. (eds.) *IPSN 2003*. LNCS, vol. 2634, pp. 442–457. Springer, Heidelberg (2003)
8. Huang, Q., Lu, C., Roman, G.C.: Spatiotemporal multicast in sensor networks. In: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, pp. 205–217. ACM (2003)
9. Huang, Q., Lu, C., Roman, G.C.: Reliable mobicast via face-aware routing. In: Twenty-Third Annual Joint Conference of the IEEE Computer and Communications Societies, *INFOCOM 2004*, vol. 3, pp. 2108–2118. IEEE (2004)
10. Wan, E.A., Van Der Merwe, R.: The unscented kalman filter for nonlinear estimation. In: The IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, *AS-SPCC*, pp. 153–158. IEEE (2000)

# Predicting Osteoporosis by Analyzing Fracture Risk Factors and Trabecular Microarchitectures of the Proximal Femur from DXA Images

Eun Byeol Jo<sup>1</sup>, Ju Hwan Lee<sup>1</sup>, Sung Yun Park<sup>2</sup>, and Sung Min Kim<sup>1,\*</sup>

<sup>1</sup>Department of Medical Bio Technology, Dongguk University, Seoul, Korea  
{eunbyeol127, ykjhlee}@gmail.com, smkim@dongguk.edu

<sup>2</sup>Dongguk Medical device Innovation Center, Ilsan, Korea  
bmepark@gamil.com

**Abstract.** This study aimed to identify the optimal threshold ranges for predicting osteoporosis and osteoporotic fractures by analyzing the correlations between trabecular patterns and fracture risk factors. We selected 85 postmenopausal women as experimental subjects and classified them into 29 normal and 56 osteoporotic groups. We proposed a novel thresholding algorithm that divides the threshold ranges from 0 to 95% based on trabecular bone area and assessed osteoporosis predictability for each range. Evaluation parameters were categorized into morphological parameters (Tb.Area, Sk.Length and fractal dimension) and fracture risk factors (MSCT, LSCT, FNW, TW, FNAL and HAL). Consequently, we found the clinical usefulness of our algorithm for discriminating patients with osteoporosis from those with normal bone. The significances between the morphological parameters and the fracture risk factors improved as bone mineral density decreased. Based on these results, we selected the optimal threshold conditions for predicting osteoporosis and osteoporotic fractures at thresholds of 40-80%.

**Keywords:** Osteoporosis, Osteoporotic Fracture, BMD (Bone Mineral Density), Trabecular Bone, DXA (Dual-energy X-ray Absorptometry), Threshold.

## 1 Introduction

Osteoporosis, which is characterized by loss of bone mass and deterioration of bone quality, commonly, increases the risk of fracture as bone strength is weakened [1]. Gender difference is one of the most common primary factors for determining osteoporosis. In general, osteoporosis occurs more frequently in elderly women compared with men, since bone loss rapidly increases in elderly women 5-10 years after menopause. In addition, osteoporotic fracture risk arises in elderly females due to the decrease in bone mass. It is therefore known that osteoporosis and fractures are the most common diseases in menopausal women [2].

---

\* Corresponding author.

The current definition of osteoporosis is based on a BMD (Bone Mineral Density). Osteoporosis is determined, when standard deviations fall below the peak bone mass less than 2.5. BMD measurement is used in clinical practice to determine if osteoporosis treatment is appropriate [3]. BMD based diagnosis, however, reveals the low accuracy regarding different races, especially in Asian women including Korean [4].

To cope with this limitation, a number of studies have been performed to analyze the bone architecture of the trabecular bone by using various segmentation approaches. Liu et al. [5] suggested the optimum threshold value as a fixed value of 40% of the maximal grey scale by using a global segmentation approach. Kang et al. [6] improved the predictability of osteoporosis by employing local adaptive threshold to the original image. On the other hand, Ito et al. [7] analyzed risk factors from the femoral CT image, and observed that the variance of the bone mass in the trabecular bone can be differentiated from those of the cortical bone as BMD decreases. Gregory et al. [8] analyzed both BMD and femur structure to find a correlation between osteoporosis and fractures. In addition, Kanis et al. [9] analyzed relationships between BMI (Body Mass Index), gender and diseases, and suggested a novel diagnosis tool based on FRAX (Fracture Risk Assessment Tool). Despite these efforts, the existing approaches can be influenced by grey scale of the original image due to a fixed threshold value. Additionally, most of the existing methods are performed by 3D analysis, whereas there are no sufficient approaches which associated with 2D analysis.

This study aims to find optimal threshold ranges for predicting osteoporosis and osteoporotic fractures by analyzing the relationships between trabecular patterns and fracture risk factors of the proximal femur. This paper is organized as follows: Section 2 suggests the extraction procedure of the trabecular patterns and risk factors, novel thresholding technique and statistical analysis. Section 3 gives the experimental results, and we analyzed our experimental results in Section 4. Lastly, we conclude in Section 5.

## **2 Material and Methods**

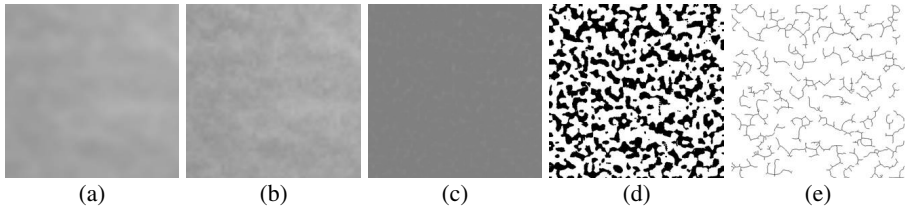
### **2.1 Experimental Subjects**

We selected a total of 85 menopausal women as experimental subjects who approved this study with informed consent. According to the standard of the WHO (World Health Organization), experimental subjects were classified into 29 normal and 56 osteoporotic groups. We measured BMD on the femoral neck, trochanter and ward's triangle using DXA (Dual-energy X-ray Absorptiometry) (Hologic, Inc., Waltham, MA, USA). Among these measured areas, we only analyzed the femoral neck of total subjects to perform more direct comparison.

### **2.2 Extraction of the Trabecular Patterns**

We utilized a modified version of the White and Rudolph's method [10] to extract the trabecular patterns from the original DXA image. We first obtained the blurred image





**Fig. 1.** Image processing procedure for obtaining the trabecular patterns from 2D DXA scans

(Fig. 1(b)) after applying Gaussian filter (sigma=10 pixel) to the ROI (Region of Interest) image (Fig. 1(a)). We then subtracted the obtained image from the original image, and added 128 grey levels to normalize the image signals (Fig. 1(c)). After normalization, the resulting image was binarized; thereby differentiating the trabecular bone from the background (Fig. 1(d)). In addition, we employed erosion and dilation methods to remove shot noises. Lastly, the skeletonized image was acquired by eroding the image until only the central line remained in the image (Fig.1(e)). Fig. 1 shows the entire procedure for extracting the trabecular patterns from DXA scans. All image processing was performed by Matlab software (R2011b, Mathworks Inc., Natick, MA, USA).

We selected Tb.Area, Sk.Length and FD (Fractal Dimension) as the morphological parameters for assessing osteoporosis predictability [11]. Tb.Area indicates the mean area of trabecular bone (Fig. 1(d)). Sk.Length is the mean length of the skeletonized elements (Fig. 1(e)). FD is known as a useful way for quantifying fractal pattern complexity. It is possible to utilize the FD as an evaluation index for predicting osteoporosis, since the trabecular bone presents typical features of the fractal pattern. We computed the FD of the trabecular bone using the box counting method.

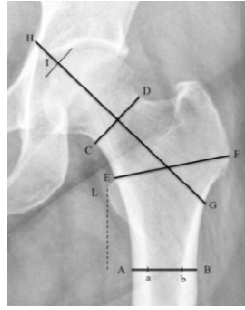
### 2.3 Proposed Thresholding Technique

We proposed a novel thresholding technique to find the optimum range that minimized the influence of the original grey level. The proposed method divided the thresholding range from 0 to 95% corresponding to the trabecular area. The thresholding ranges applied to each image can be derived by equation (1):

$$P(T) = \frac{1}{A} \int_0^T a(x) dx \tag{1}$$

where  $a(x)$  indicates the area function of the trabecular bone with gray value  $x$ .  $T_p$  is the threshold value,  $A$  is total area (number of pixels) of the trabecular bone, and  $P$  is the percentage function.

After computing each threshold value, all values above  $T$  were considered trabecular bone, and those below were considered non-bone. In other words, the proposed method excluded a certain percentage of the trabecular area ( $P$ ) from the morphological analysis. According to these characteristics, the white areas increased, whereas black areas decreased, as the threshold increased from 0 to 95%. We selected the threshold ranges as 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 95% based on trabecular bone area and assessed osteoporosis predictability for each range.



**Fig. 2.** The measuring area of the fracture risk factors (A-a: MSCT, b-B: LSCT, C-D: FNW, E-F: TW, H-G: HAL, I-G: FNAL)

## 2.4 Extraction of the Fracture Risk Factors

We extracted six fracture risk factors composed of MSCT (Medial femoral Shaft Cortical bone Thickness), LSCT (Lateral femoral Shaft Cortical bone Thickness), FNW (Femoral Neck Width), TW (Trochanteric width), FNAL (Femoral Neck Axis Length) and HAL (Hip Axis Length) from the original femur image (Fig. 2). MSCT and LSCT are the cortical bone thicknesses of the medial femoral shaft and lateral femoral shaft, 3cm below the center of the lesser trochanter, respectively (Fig. 2(A-a), (B-b)) [12]. FNW represents the width of femoral neck and its most narrow distance perpendicular to the neck axis (Fig. 2(C-D)) [13]. TW is the width of the intertrochanteric region across the greater trochanter from the lesser trochanter (Fig. 2(E-F)) [14]. FNAL is the distance from the base of the greater trochanter to the top of the femoral head (Fig. 2(I-G)) [15]. HAL indicates the distance from the greater trochanter to the inner pelvic brim, called the hip axis length (Fig. 2(H-G)) [16].

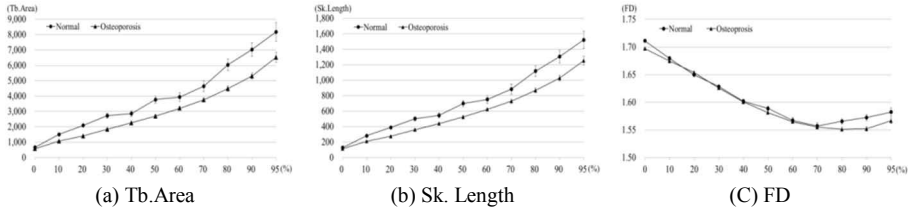
## 2.5 Statistical Analysis

All dataset were analyzed by one-way ANOVA and paired t-tests using SPSS ver. 12.0 for Windows (Chicago, IL, USA). A p-value of less than 0.05 was considered significant.

# 3 Experimental Results

## 3.1 Comparison of the Morphological Parameters for the Different Thresholding Conditions

Tb.Area and Sk.Length increased significantly as the threshold value increased ( $p < 0.05$ , Fig. 3). All parameters, except FD, showed relatively small differences between the normal and osteoporosis groups under the no-threshold condition. However, the differences increased drastically after applying the thresholds, thereby improving the significance between the experimental groups. On the other hand, FD



**Fig. 3.** Comparison of the morphological parameters for the different experimental groups

**Table 1.** Correlation of the morphological parameters between the normal and the osteoporosis groups

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
Tb.Area	0.000	0.000	0.000	0.000	0.004	0.000	0.019	0.025	0.001	0.001	0.009
Sk.Length	0.001	0.000	0.000	0.000	0.000	0.000	0.022	0.032	0.002	0.004	0.019
FD	0.001	0.230	0.349	0.559	0.800	0.094	0.657	0.728	0.019	0.004	0.081

showed slightly different results. FD declined continuously as threshold increased, whereas it increased suddenly at thresholds of 70% and 80% for the normal and osteoporosis groups, respectively ( $p < 0.05$ ). Consequently, Tb.Area and Sk.Length significantly separated the patients with osteoporosis from the normal group for all threshold conditions ( $p < 0.05$ ). All morphological parameters, except FD, showed significance for threshold ranges of 10-90%. Table 1 presents the statistical results for each threshold.

### 3.2 Correlations of Fracture Risk Factors between the Normal and Osteoporosis Groups

Among all fracture risk factors, only MSCT and the LSCT were larger in the normal group than those in the osteoporosis group ( $p < 0.05$ ). On the other hand, the remaining fracture risk factors produced opposite results ( $p < 0.05$ ). Moreover, MSCT and LSCT increased continuously as BMD increased ( $p < 0.05$ ), whereas the remaining fracture risk factors decreased significantly as BMD increased ( $p < 0.05$ ). Consequently, we found reliable differences in fracture risk factors between the normal and the osteoporosis groups. Table 2 compares the differences in the fracture risk factors between the normal and osteoporosis groups.

### 3.3 Correlations between Morphological Parameters and Fracture Risk Factors

The correlations between the morphological parameters and the fracture risk factors were not significant in the normal group for all threshold conditions ( $p > 0.05$ ). In contrast, MSCT in the osteoporosis group showed relatively high significance levels for certain thresholds. MSCT showed significant correlations with Tb.Area for the threshold condition of 30-90%. Relatively high correlations were also revealed with Sk.Length for the threshold of 20-90%. However, FD did not show any significant

**Table 2.** Comparison of the fracture risk factors between the normal and the osteoporosis group

	Normal group	Osteoporosis group	P-value
MSCT	53.448±6.384	41.214±8.383	0.000
LSCT	55.379±8.377	43.232±8.809	0.000
FNW	207.319±12.268	217.051±16.518	0.006
TW	358.970±21.865	372.353±27.536	0.017
FNAL	639.760±40.217	659.029±38.304	0.034
HAL	751.726±41.026	772.631±46.653	0.045

association with any of the fracture risk factors irrespective of threshold conditions. Therefore, we found that all morphological parameters, except FD, presented significant differences only for the MSCT within the thresholds of 40-80%.

## 4 Discussion

All morphological parameters for the different threshold conditions, except FD, increased continuously as the threshold increased. In addition, differences between the groups increased drastically when thresholds were used. Particularly, the threshold condition of 10-90% yielded significant differences for all parameters, except FD, between the experimental groups. FD revealed low significance levels for all threshold ranges. Particularly, FD was larger in the normal group with no threshold, whereas the threshold condition of 20% produced completely opposite results. On the other hand, FD decreased continuously as threshold increased, and then FD increased suddenly at a certain threshold percentage. In their previous study, Corroller et al. reported that FD significantly discriminates patients with osteoporosis from normal subjects [17]. However, there are still conflicting results whether FD decreases [18] or increases [19] as BMD decreases. In addition, FD tends to be influenced by various internal-external factors, since the differences are relatively small. Therefore, FD yielded insignificant differences for all threshold conditions.

MSCT and LSCT were larger in the normal group than those in the osteoporosis group, whereas the remaining risk factors were larger in the osteoporosis group than those in normal subjects. These results may be because MSCT and LSCT represent cortical bone characteristics, whereas the others represent trabecular bone [20]. The thickness of the cortical bone typically tends to decrease as BMD decreases [21]. On the other hand, thickness of the trabecular bone increases under the same conditions, since the distance between trabecular bones is lengthened [22].

No significant correlations were found between the morphological parameters and the fracture risk factors for any threshold conditions in the normal group. However, MSCT in the osteoporosis group revealed relatively high significance levels for certain threshold ranges. MSCT was significantly correlated with all morphological parameters, except FD, within the thresholds of 40-80%. Although both MSCT and LSCT represent characteristics of cortical bone, the reason that only MSCT was

significantly correlated was mainly due to the different decreasing rates in bone thickness. Previous studies have reported that the MSCT and the LSCT show different rates of decrease due to different physical loading [23]. In addition, decreasing rates were shown as 3% and 5.1% for MSCT and LSCT, respectively, as BMD decreases [24]. Therefore, only the MSCT yielded a significant correlation with the morphological parameters.

## 5 Conclusions

The results of the present study demonstrated the clinical usefulness of our algorithm for discriminating patients with osteoporosis from normal subjects. Moreover, the correlations between the morphological parameters and the fracture risk factors improved as BMD decreased. Based on these experimental results, we selected the optimal threshold conditions for predicting osteoporosis and osteoporotic fractures at thresholds of 40-80%. In future study, we will focus on additional evaluations for the greater trochanter and Ward's triangle to improve the reliability of our findings. Furthermore, we will analyze the correlations between the trabecular patterns and the fracture risk factors with a wide range of ages and with different genders.

## References

1. Lin, J.T., Lane, J.M.: Osteoporosis: a Review. *Clin. Orthop. Rel. Res.* 425, 126–134 (2004)
2. Riggs, B.L., Melton, L.J.: Evidence for Two Distinct Syndromes of Involutional Osteoporosis. *Am. J. Med.* 75, 899–901 (1987)
3. Kanis, J.A.: Assessment of Fracture Risk and its Application to Screening for Postmenopausal Osteoporosis: Synopsis of a WHO report, WHO study group. *Osteoporosis Int.* 4, 368–381 (1994)
4. Dalstra, M., Huiskes, R., Odgaard, A., Erning, L.V.: Mechanical and Textural Properties of Pelvic Trabecular Bone. *J. Biomech.* 26, 523–535 (1993)
5. Liu, X.S., Cohen, A., Shane, E.: Individual Trabeculae Segmentation(ITS)-ased Morphological analysis of high-resolution Peripheral Quantitative Computed Tomography Images detects abnormal Trabecular plate and Rod Microarchitecture in Premenopausal Women with Idiopathic Osteoporosis. *J. Bone Miner. Res.* 25, 1496–1505 (2010)
6. Kang, Y., Engelke, K., Kalender, W.A.: A new Accurate and Precise 3-D Segmentation Method for Skeletal Structures in Volumetric CT data. *IEEE Trans. Med. Imaging* 22, 586–598 (2003)
7. Ito, M., Wakao, N., Hida, T., Matsui, Y., Abe, Y., Aoyagi, K., Uetani, M., Harada, A.: Analysis of Hip Geometry by Clinical CT for the Assessment of Hip Fracture Risk in elderly Japanese Women. *Bone* 46, 453–457 (2010)
8. Gnudi, S., Malavolta, N., Testi, D., Viceconti, M.: Differences in Proximal Femur Geometry Distinguish Vertebral from Femoral Neck Fractures in Osteoporotic Women. *Br. J. Radiol.* 77, 219–223 (2004)
9. Kanis, J.A., McCloskey, E.V., Johansson, H., Strom, O., Borgstrom, F., Oden, A.: Case Finding for the Management of Osteoporosis with FRAX-Assessment and Intervention Thresholds for the UK. *Osteoporosis Int.* 19, 1395–1408 (2008)

10. White, S.C., Rudolph, D.J.: Alterations of the Trabecular Pattern of the Jaws in Patients with Osteoporosis. *J. Bone Miner. Res.* 19, 1640–1650 (1999)
11. Croucher, P.I., Garrahan, N.J., Compston, J.E.: Assessment of Cancellous Bone Structure: Comparison of Strut Analysis, Trabecular Bone Pattern Factor, and Marrow Space Star Volume. *J. Bone Miner. Res.* 11, 955–961 (1996)
12. Chappard, C., Bousson, V., Bergot, C., Mitton, D., Marchadier, A., Moser, T., Benhamou, C.L., Laredo, J.D.: Prediction of Femoral Fracture Load: Cross-Sectional Study of Texture Analysis and Geometric Measurements on Plain Radiographs versus Bone Mineral Density. *Radiology* 255, 536–543 (2010)
13. Faulkner, K.G., Wacker, W.K., Barden, H.S., Simonelli, C., Burke, P.K., Ragi, S., Rio, L.D.: Femur Strength Index Predicts Hip Fracture Independent of Bone Density and Hip Axis Length. *Osteoporosis Int.* 17, 593–599 (2006)
14. Pulkkinen, P., Partanen, J., Jalovaara, P., Jämsä, T.: Combination of Bone Mineral Density and Upper Femur Geometry Improves the Prediction of Hip Fracture. *Osteoporosis Int.* 15, 274–280 (2004)
15. Theobald, T.M., Cauley, J.A., Gluer, C.C., Bunker, C.H., Ukoli, F.A., Genant, H.K.: Black-White Differences in Hip Geometry. *Osteoporosis Int.* 8, 61–67 (1998)
16. Dincel, V.E., Sengelen, M., Sepici, V., Cavusoglu, T., Sepici, B.: The Association of Proximal Femur Geometry with Hip Fracture risk. *Clin. Anat.* 21, 575–580 (2008)
17. Corroller, T.L., Halgrin, J., Pithioux, M., Guenoun, D., Chabrand, P., Champsaur, P.: Combination of Texture Analysis and Bone Mineral Density Improves the Prediction of Fracture load in Human Femurs. *Osteoporosis Int.* 23, 163–169 (2012)
18. Chen, S.K., Chen, C.M.: The Effects of Projection Geometry and Trabecular Texture on Estimated Fractal Dimensions in Two Alveolar Bone Models. *Dentomaxillofac. Radiol.* 27, 270–274 (1998)
19. Bollen, A.M., Taguchi, A., Hujoel, P.P., Hollender, L.G.: Fractal Dimension on Dental Radiographs. *Dentomaxillofac. Radiol.* 30, 270–275 (2001)
20. Napoel, N., Jin, J., Peters, K., Wustrack, R., Burch, S., Chau, A., Cauley, J., Ensrud, K., Kelly, M., Black, D.M.: Are Women with Thicker Cortices in the Femoral Shaft at Higher Risk of Subtrochanteric/Diaphyseal Fractures? The Study of Osteoporotic Fractures. *J. Clin. Endocrinol. Metab.* 97, 2414–2422 (2012)
21. Holzer, G., Skrbensky, G.V., Holzer, L.A., Pichl, W.: Hip Fractures and the Contribution of Cortical versus Trabecular Bone to Femoral Neck Strength. *J. Bone Miner. Res.* 24, 468–474 (2009)
22. Rivadeneira, F., Zillikens, M.C., De Laet, C.E., Hofman, A., Uitterlinden, A.G., Beck, T.J., Pols, H.A.: Femoral Neck BMD is a Strong Predictor of Hip Fracture Susceptibility in Elderly Men and Women because It Detects Cortical Bone Instability: The Rotterdam Study. *J. Bone Miner. Res.* 22, 1781–1790 (2007)
23. Peacock, M., Turner, C.H., Liu, G., Manatunga, A.K., Timmerman, L., Johnston, C.C.: Better Discrimination of Hip Fracture using Bone Density, Geometry and Architecture. *Osteoporosis Int.* 5, 167–173 (1995)
24. El-Kaissi, S., Pasco, J.A., Henry, M.J., Panahi, S., Nicholson, J.G., Nicholson, G.C., Kotowicz, M.A.: Femoral Neck Geometry and Hip Fracture Risk: The Geelong Osteoporosis Study. *Osteoporosis Int.* 16, 1299–1303 (2005)

# Malicious Web Page Detection: A Machine Learning Approach

Abubakr Sirageldin, Baharum B. Baharudin, and Low Tang Jung

Computer & Information Science Department, University Technology Petronas  
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia.  
abubakrsirag@gmail.com,  
{baharbh, lowtanjung}@petronas.com.my

**Abstract.** Due to the rapid growth of the internet, websites have become the intruder's main target. An intruder embeds malicious contents in a web page for the purpose of doing some bad and unwanted-activities such as: credential information and resource theft, luring a user to visit a dangerous website, downloading and installing software to join a botnet or to participate in distributed denial of service, and even damage the visitor system. As the number of web pages increases, the malicious web pages are also increasing and the attack is increasingly become sophisticated. In this paper, we provide a framework for detecting a malicious web page using artificial neural network learning techniques. In addition to the significant detection rate, our objective is to find also which discriminative features characterize the attack and reduce the false positive rate. The algorithm is based on two features group, the URL lexical and the page content features. The experiments has shown the expected results and the high false positive rate which produced by machine learning approaches is reduced.

**Keywords:** Malicious, Benign, Feature, Detector, URL, Content, Script, Learning, Lexical.

## 1 Introduction

The internet has become an essential in our daily life; it's the base of banking transactions, shopping, entertainment, resource sharing, news, and social networking. The growth of the web rewarded the cyber criminals towards it ,with this growth, also the design and the use of the malware scenario has changed, its more stealthier and polymorphic than damaging the machines. The majority of malware is intended to either steal the user's private data, or force the victim system to join a malware distribution network. Web is a common method for spreading malware; the attackers exploit the vulnerabilities of web browsers, web application, and operating system to gain control of a victim's machine, which is used to host various malicious activities, such as heap spray, botnet, key loggers, sending spam emails, and so on.

The attack occurs when a user visits a suspected website, therefore attacker focuses on a web site that has become the centre of attention, and then exploits the

vulnerabilities in both client and server to launch the attack. Web attack is a challenge; it necessitates a careful understanding of the details and the behavior. Internet browsers supported active client-side content with many techniques, among them JavaScript, which is the most widely used as a primary component of active and dynamic web content, while it provides a great chance for exploits. Further techniques such as PHP language, adobe flash, and visual basic script are in common have capability of download and execute code from the Internet [8]. In addition most browsers have a feature of plug-ins, which allow third parties to extend the functionality of the browser. Although several solutions have been proposed to fight malicious software, but web site exploits has not received much attention so far. In this paper, we provide a framework for detecting a malicious web page based on two groups of features using artificial neural network (ANN). This work is continuing to, add some values to the field malware combat, mitigate some threats, and improve performance by enhancing the detection rate.

## 2 Related Work

Existing detection approaches can be classified into three categories: signature-based, behavior monitoring and machine learning, and here we grouped them in three categories.

**URL and Domain Analysis:** Ma et.al [1] introduced malicious web site detection model based on URL lexical and host-based features using online classifiers algorithms, the model was able to achieve 99% accuracy. Yoshiro Fukushima et.al [2] analyzed characteristics of malicious web sites by their domain information. They studied the IP address, IP address, domain, and registrar with reputations, and then they proposed a blacklisting scheme derived from the combination of IP address and registrars. A hierarchical structure graph is constructed using the extracted information. Zhang et.al [3] provided a model based on secondary URLs, and redirect chains recorded by a high-interaction client honeypot to determine a malware distribution networks. The method also encompasses a drive-by download detection mechanism based on a set of regular expression-based signatures.

**Page Content Analysis:** Cove et.al [4] introduced an approach for detecting malicious JavaScript code based on assigning probability to a feature that reflects the likelihood that a given feature value occurs using anomaly detection with emulation. Yuan-Tsung Hou et.al [5] proposed a malicious web page detection model based on dynamic HTML and some Java script native functions using the boosted decision tree algorithm. Van Lam Le et.al [6] [ have introduced a two-stage classification model based on 52 features to detect malicious web page using information gain values. In the first stage, they examined only static feature, and only potential malicious page is scanned to extract run-time features in the second stage. Mario Heiderich et.al [7] introduced an approach based on 8 classes attack features to detect attacks against the HTML document object model (DOM) tree using support vector machine. The



approach is able to mitigate attacks against web browser and protect the DOM integrity also. Chen et.al [8] designed a web-based malware detection model named as WepPatrol. The model used the improved PHoneyC and Libemu ShellCode detection to collect malware scenario, and the result outperformed Google's safe browsing. S. Chitra et.al [9] have provided a dynamic approach based on 21 features for detecting a malicious web page using genetically evolved fuzzy rules, and the approach has shown good result.

**Behavior Analysis:** Xu et.al [10] proposed a model for detecting the infection delivered through vulnerable applications and web browser based on the dependency between a user's action and system's event. Two components were used to record user behavior; one at kernel level, and another as a Firefox extension. Hsu, F.H., et al [11] have proposed a runtime, behavior-based solution to protect the system against drive-by-download attacks. The approach utilized the browser helper object (BHO) mechanism of Windows operating system to implement the framework on internet explorer 7.0. The experimental results have shown low performance overhead and good performance. As a conclusion the URL and domain analysis method is a very useful idea, but the limitation of this method is that the malicious URL does not mean that the corresponding page holds malicious contents and vice versa. Page content-based faces a lot of challenges, such as continuous change of the attack features, Obfuscation and so on.

### 3 The Approach

#### 3.1 Model Structure

The model mainly composed of three components, Feature Extractor, Learning and Model Selector, and the Detector as shown in fig. 1. It starts with the feature extraction process via feature extractor component, and then the extracted features are sent to the detector.

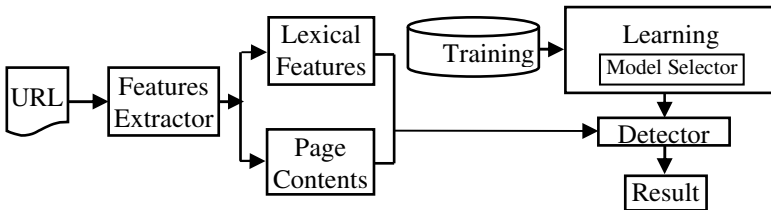


Fig. 1. The general structure of the approach

**Feature Extractor:** This is the first component, which is responsible for capturing the web page features. It consists of: Jericho3.3 HTML parser, JavaScript engine and URL lexical class using Java. This part is important mainly because, it's essential to collect data set.

**Learning and Model Selector:** It necessitates the data set for learning to build the model, and according to the performance, the learning algorithm with the best result is selected.

**Detector:** The detector is using the final classification model generated from the Learning and Model Selector component.

### 3.2 Feature Selection

There are a lot of candidate features that can support the degree of maliciousness of a web page and considered as the target of intruders. In our study we have selected 39 features, 21 are new or modified and the rest is reused from the previous work.

**URL Lexical Features:** The uniform resource locator (URL) textual form was used to get some signs of the potential maliciousness in malicious web page detection approaches [1]. Our framework is based on 10 features 6 of them are reused from previous work and we came up with 4 new features. The reused features are: number of words, number of directories, length of host, length of URL, number of digits in host, and number of parameters.

**Page-Contents Features:** HTML tags in general have been known as a common method used to load the malware from the outside, such as body, iframe, img, meta, applet, frameset, style, layer, ilayer, embed, script, object, link, and normal hyperlink. The HTML reused features are: number of local links, number of external links, number of redirects, number of tags, number of applets, number of objects, number of frames, number of forms, and abnormal visibility. On the other hand we countered some dangerous JavaScript keywords such as: escape, unescape, eval, exec, and unbound. In addition, the obfuscation and encoding such as: obfuscation and encoding attempt, length and number of scripts, harmful keyword, internal and external scripts, and number of methods. The total number of page content features that we have selected is 29.

### 3.3 The Classification Algorithm

**Support Vector Machine (SVM):** is a classification method introduced in 1992 by Boser, Guyon, and Vapnik [12]. The original optimal hyperplane algorithm proposed by Vladimir Vapnik in 1963 was a linear classifier. However, in 1992, Bernhard Boser, Isabelle Guyon and Vapnik suggested a way to create non-linear classifiers by applying the kernel trick to maximum-margin hyperplanes, by replacing dot product by a non-linear kernel function. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space.

**Decision Tree (DT):** is a machine learning classifier based on the tree structure. Each node in the tree is associated with a particular feature, and the edges from the node separate the data based on the value of the feature. Each leaf node binds to a class in

the classifier model. The training data is key point for the information gain (IG) of the feature selection policy [13].

**Naive Bayes (NB):** (Maron & Kuhns, 1960) is a probability based classifier, which simplify the learning process by assuming all features are independent. Let  $x=(x_1, x_2, \dots, x_n)$  be the features, and  $y$  be the target or the class of the vector  $X$ . The posterior probability of class  $y$  given  $x$  is calculated as  $P(X|Y) = \pi \prod P(x_i|y)$  where  $i=1 \dots n$ . The probability of the class  $y$  and the probability of the feature given a class  $P(x_i|y)$  is calculated from the training frequencies.

**K Nearest-Neighbor (KNN):** is a simple algorithm for predicting a class of an example. This classifier is supervised learning based on the distance of the example. The training stage simply stores all training examples with their labels. To predict the class for a new test example, first it computes its distance to every training example and then, keeps the  $k$  closest training examples, where  $k \geq 1$ . Finally, it looks for the label that is most common among these examples. This label is assigned to this test example as the predicted result [14].

**Artificial Neural Network (ANN):** is a solution inspired by biological neural networks. An artificial neuron is a computational model receives signals through synapses located on the dendrites. When the signals received are exceeded a certain threshold value, the neuron is activated and emits a signal through the axon. The emitted signal might be sent to another synapse, or activate other neurons [15]. ANN can be viewed as directed graphs with weights, which artificial neurons are nodes and directed edges represent the connections between neuron outputs and neuron inputs. According to the connection pattern, ANNs can be put into two groups: feed-forward networks, and feedback networks [16].

## 4 Experiment and Results

In this section, we detail data source, datasets, and experiments scenario. And then we evaluate our approach targeting the accuracy of detection and feature significance.

### 4.1 Data Set and Data Source

The dataset is composed of benign and malicious instances; the benign set is collected via web search and Alexa website ranking verified by Google safe browsing [6], and the malicious set is collected from some of the common public announced malware and exploited websites [17],[18] such as malwaredomainlist.com, StopBadWare, mwsl.org.cn, PhishTank, and malwareurl.com. The collected features are divided into two groups training and test. The training set consists of 29500 instances (21500 benign, 8000 malicious), the test set consists of 451 instances (126 benign, 325 malicious).

**Experimental Procedure:** The experiment is carried out in four steps. Firstly, malicious and benign URL lists are prepared. Secondly, features are extracted and grouped into training and test. Thirdly, generate the model using training set. And finally, the generated model is tested using the test data set.

## 4.2 The Evaluation

**The Accuracy:** The accuracy is defined as the ratio of correctly identified examples over all examples in the test set. The test set is applied to five classifiers SVM(kernel=rbf,sigma=4), KNN(distance=Jaccard), NB(distribution=multivariate multinomial), DT(type=bagger), and ANN(type=perceptron, layers=2, neurons=5) and the performance is shown in Table1. The false positive rate is also in the desired range above 0.01 and below 0.1.

**Table 1.** The Accuracy With The New Features

Algorithm	Recall (%)	Specificity (%)	Accuracy (%)
SVM	93.85	92.86	93.57
KNN	90.77	98.41	92.90
NB	88.00	89.68	88.47
DT	93.23	100.00	95.12
ANN	95.08	98.41	96.01

**The Features Significance:** In this study we presented some relevant attack features that remain resilient against possible anticipated future attack. The URL lexical features can give distinguishing values and rise up the true positive rate, where JavaScript features have a significant effect on the true positive rate. HTML features are easier to be used by the attacker and easier to be captured by detection engines; therefore the attackers usually encode the target code to circumvent the signature-based detection engine. The experiment shows that the combination of the feature groups has shown the higher true positive rate and lower false positive. Table2 shows how the accuracy went down when using only the reused features.

**Table 2.** The Accuracy With Existing Features

Algorithm	Recall (%)	Specificity (%)	Accuracy (%)
SVM	94.15	89.68	92.90
KNN	83.08	96.83	86.92
NB	91.08	92.06	91.35
DT	93.23	98.41	94.68
ANN	90.77	95.24	92.02

## 5 Conclusion and Future Work

The idea behind doing this work is to provide lightweight early assessment model in order to reduce the threats of the web-based attacks. In this paper we provide a mostly static identification mechanism based on two feature groups, the URL lexical and the page content to detect malicious web page using machine learning algorithms. The experiment has shown the expected result, and the model is able to reach 97% accuracy and the false positive rate can be reduced down to zero. The drawback of this work is the partial rendering method used for feature collection. One of the solutions is to integrate a dynamic behavior monitoring unit to trace the client-server interactions. Moreover the updated model could be integrated with the internet browser. There are some challenges on the way such as: the change of the attack feature, which might required a combination of different levels, the efficient algorithms, and effective comparative measures, therefore the future research may focus on tracing the attack vector features and pay a lot of attention on feature selecting methods, which might incorporate several disciplines to achieve the best performance.

## References

1. Ma, J., Saul, L.K., Savage, S., Voelker, G.M.: Learning to detect malicious URLs. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2(3), 30 (2011)
2. Fukushima, Y., Hori, Y., Sakurai, K.: Proactive Blacklisting for Malicious Web Sites by Reputation Evaluation Based on Domain and IP Address Registration. *IEEE* (2011)
3. Zhang, J., Seifert, C., Stokes, J.W., Lee, W.: ARROW: GenerAting SignatuRes to Detect DRive-By DOWNloads. *ACM* (2011)
4. Cova, M., Kruegel, C., Vigna, G.: Detection and analysis of drive-by-download attacks and malicious JavaScript code. *ACM* (2010)
5. Hou, Y.T., Chang, Y., Chen, T., Laih, C.S., Chen, C.M.: Malicious web content detection by machine learning. *Expert Systems with Applications* 37(1), 55–60 (2010)
6. Van Lam Le, I.W., Gao, X., Komisarczuk, P.: Two-Stage Classification Model to Detect Malicious Web Pages. *IEEE* (2011)
7. Heiderich, M., Frosch, T., Holz, T.: ICESHIELD: Detection and mitigation of malicious websites with a frozen DOM. In: Sommer, R., Balzarotti, D., Maier, G. (eds.) RAID 2011. LNCS, vol. 6961, pp. 281–300. Springer, Heidelberg (2011)
8. Chen, K.Z., Gu, G., Zhuge, J., Nazario, J., Han, X.: WebPatrol: Automated collection and replay of web-based malware scenarios. *ACM* (2011)
9. Chitra, S., Jayanthan, K., Preetha, S., Shankar, R.N.U.: Predicate based Algorithm for Malicious Web Page Detection using Genetic Fuzzy Systems and Support Vector Machine. *International Journal of Computer Applications* 40(10) (2012)
10. Xu, K., Yao, D., Ma, Q., Crowell, A.: Detecting infection onset with behavior-based policies. *IEEE* (2011)
11. Hsu, F.H., Tso, C.K., Yeh, Y.C., Wang, W.J., Chen, L.H.: BrowserGuard: A Behavior-Based Solution to Drive-by-Download Attacks. *IEEE Journal on Selected Areas in Communications* 29(7), 1461–1468 (2011)

12. Lee, Y.J., Huang, S.Y.: Reduced support vector machines: A statistical theory. *IEEE Transactions on Neural Networks* 18(1), 1–13 (2007)
13. Kohavi, R., Quinlan, R.: C5. 1.3 Decision Tree Discovery
14. Elkan, C.: Nearest neighbor classification. University of California, San Diego (2007)
15. Gershenson, C.: Artificial neural networks for beginners. arXiv preprint cs/0308031 (2003)
16. Jain, A.K., Mao, J., Mohiuddin, K.M.: Artificial neural networks: A tutorial. *Computer* 29(3), 31–44 (1996)
17. Tao, W., Shunzheng, Y., Bailin, X.: A Novel Framework for Learning to Detect Malicious Web Pages. *IEEE* (2010)
18. Zhang, W., Ding, Y.X., Tang, Y., Zhao, B.: Malicious web page detection based on on-line learning algorithm. *IEEE* (2011)

# Proposal for Virtual Web Browser by Using HTML5

Tomokazu Hayakawa and Teruo Hikita

School of Science and Technology, Meiji University  
Kasawaki, 214-8571, Japan  
{t\_haya,hikita}@cs.meiji.ac.jp

**Abstract.** We propose a method of a virtual Web browser that enables safer Web-browsing environment. The method uses an HTML5 compliant Web browser as client environment and JavaScript-related technologies in server environment. The idea of the method is that (1) the server works as the HTTP/WebSocket proxy, (2) it transforms requested Web pages from clients into equivalent images, and (3) it returns the images to the clients, thereby making Web pages that contain malicious software (e.g., viruses, worms, and so on) harmless and protecting the clients against being infected with such malware. The virtual Web browser also supports other features such as keyboard events, mouse events, Cookies, and so on. The evaluation shows that the method provides a safer way of browsing Web pages without increasing network traffic. Finally, we conclude that the method is useful to realize safer Web-browsing environment.

**Keywords:** security, proxy, Web browser, HTML5, JavaScript.

## 1 Introduction

As the World Wide Web becomes an essential part in our business, threats of malicious Web pages that contain viruses, worms, or others are increasing gradually. For example, there is an attack called “Drive-by download” that causes unintended download of computer software from the Internet.

To protect our computers from malicious Web pages, it is required to use the latest Web browsers, the latest browser plugins (e.g., Java Applet, Flash Player, Adobe Acrobat Reader, and others), the latest anti-viruses, and the latest Operating Systems.

In spite of these efforts, there still remains another kind of threats called “Zero-day attack”, because they use vulnerability in computer software before we aware and fix the vulnerability. Unfortunately, as far as we know, there seem few effective solutions for this kind of threats.

To solve the problem, we propose a method, named “virtual Web browser”, that enables safer Web-browsing environment. The virtual Web browser consists of two parts: an HTML5 compliant Web browser as client environment and JavaScript-related technologies in server environment. The client side of the virtual Web browser is written in HTML5, and it transparently runs on any Web browser that is compatible with HTML5, as if the browser were not virtualized. The server side of the virtual

Web browser consists of two parts: the HTTP/WebSocket proxy server and the rasterization server. The idea of the method is that (1) the server works as the HTTP/WebSocket proxy, (2) it transforms requested Web pages from clients into equivalent images, and (3) it returns the images to the clients. This makes Web pages that contain viruses, worms, or others, harmless and protects the clients against being infected with them. We have evaluated the method from the point of view of network traffic, and the result shows that the method rarely increases network traffic.

The rest of this paper is organized as follows: Section 2 and 3 describe the design and the implementation of the method, respectively. Section 4 reports the result of the evaluation. Section 5 introduces related work. Section 6 gives the conclusion.

## 2 Design of Virtual Web Browser

### 2.1 Objectives

The virtual Web browser aims to be a light-weight secure browser that strengthens user security. Its main objective is to protect computers that run Web browsers against malicious Web pages with as good user experience as modern Web browsers without any additional cost. We define the user experience as follows: let the user of the virtual Web browser be able to (1) keep using his/her current Web browser as runtime environment without installing any additional software, (2) bookmark Web pages into his/her Web browser, (3) operate the browser with the keyboard and/or the mouse, and (4) use Web applications that use persistent features such as Cookies, Local Storage, and WebSQL. Moreover, there is also another objective: to solve a compatibility problem of RIAs (Rich Internet Applications). The problem is caused by the fact that there is no compatibility among RIA technologies. Hence, already existing RIAs cannot be ported to other environment even if the Web browsers or the plugins on which the RIAs run become obsolete.

### 2.2 How to Make Malicious Web Pages Harmless

The virtual Web browser makes malicious Web pages harmless by transforming them into equivalent images, i.e., “rasterization”, in controlled and isolated environment. Although it is not impossible to embed malicious software in images in some situations, we believe that this method undoubtedly strengthens user security.

Although there are a large number of methods that realize virtualization such as application virtualization, desktop virtualization, OS virtualization, and so on, we have decided to use HTML5 as client environment and several Unix’s APIs, namely `setuid(2)`, `setgid(2)`, and `chroot(2)`, in server environment to realize a safer browser without any cost nor installing any software<sup>1</sup>.

---

<sup>1</sup> The rasterization process, described in Section 2.5, is launched under a jailed directory with a limited privilege in the server, and the contents of Cookie, Local Storage, and WebSQL are stored under the directory, which minimizes and contains the influence of malware.



By this decision, there exist advantages and disadvantages. Some of the advantages are: users can use the virtual Web browser transparently through their browsers, and they can bookmark any Web page as their browsers' bookmark, not the virtual Web browser's bookmark. Some of the disadvantages are: users cannot download/upload any file because the virtual Web browser does not allow any file system access, and users cannot view Web pages that require any plugin. In spite of these disadvantages, we consider them acceptable to increase security in exchange for usability.

### 2.3 Screenshots of Virtual Web Browser

Fig. 1 shows a screenshot of the virtual Web browser that shows the CSA 2013 Web site, indicating that there is no "View Page Source" in the context menu. This is because the entire Web page is rasterized as a single image. Since the virtual Web browser handles user events, any link shown in the browser is clickable; if a link of an already-loaded page is clicked, the virtual Web browser moves to the new URL, and then it shows a new image of the new page indicated by the URL.

### 2.4 Network Model and Assumptions

We assume that the network structure in which our proposed method is applied as shown in Fig. 2. The client computers are to be in the internal network and the network to which they belong is required to contain at least one DMZ (Demilitarized Zone) that is located between the two firewalls. The HTTP/WebSocket proxy server and the rasterization server are required to be placed in the DMZ. This network structure reduces the risk of the client computers being cracked even if one or both of the servers are cracked, because the firewall between the internal network and the DMZ does not allow access from the DMZ to the internal network.

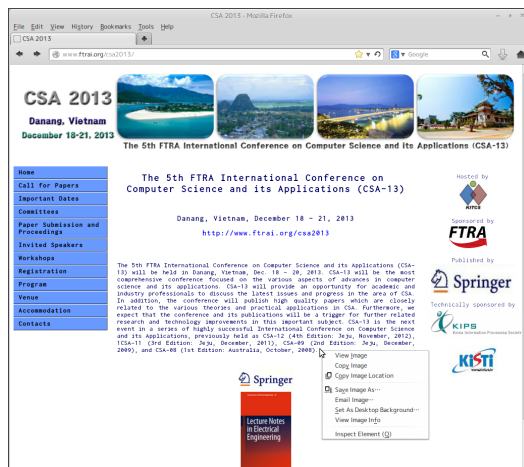
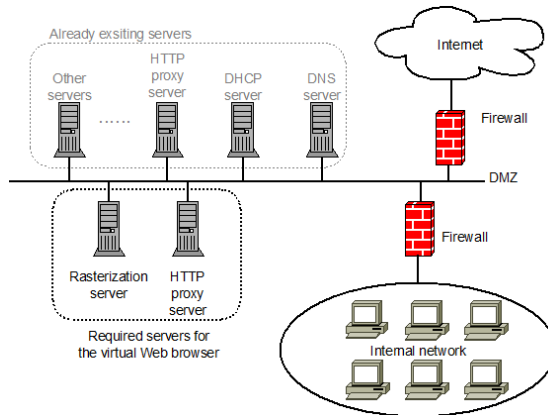


Fig. 1. Screenshot of Virtual Web Browser that Shows CSA 2013 Web Site



**Fig. 2.** Network Structure of Virtual Web Browser

## 2.5 Behavior of Virtual Web Browser

Fig. 3 shows the behavior of the virtual Web browser, which indicates as follows<sup>2</sup>:

- (1) Each Web browser of the clients requests a Web page by entering a URL, by selecting a bookmark, or by clicking a link contained in an already-loaded Web page.
- (2) In response to the first request<sup>3</sup>, the proxy server requests the digest access authentication that queries a username and a password; both are used to launch the rasterization process in the rasterization server under the isolated environment.
- (3) The browser re-requests the Web page with the proxy authentication information.
- (4) The proxy server launches the rasterization process on the rasterization server according to the information of proxy authentication.
- (5) The proxy server returns the virtual Web browser written in HTML5.
- (6) The virtual Web browser sends a request to the Web page by using a WebSocket.
- (7) The proxy server transfers the request to the rasterization server.
- (8) The rasterization server transfers the request to the original destination server or an upstream HTTP proxy server.
- (9) The rasterization server receives the response from the server.
- (10) The rasterization server transforms the response into an equivalent image.
- (11) The rasterization server returns the image to the proxy server as a binary image.
- (12) The proxy server returns the binary image to the browser as a base64 image.
- (13) Finally, the browser shows the base64 image onto its `<img>`.

<sup>2</sup> Whenever the size of the user's Web browser changes, the virtual Web browser sends the new size to the server, so that the rasterization process rasterizes requested Web pages into equivalent images with the correct resolution.

<sup>3</sup> Once the proxy authentication succeeds, the Web browser caches the authentication information, so that the user will never be asked to input his/her username/password anymore.

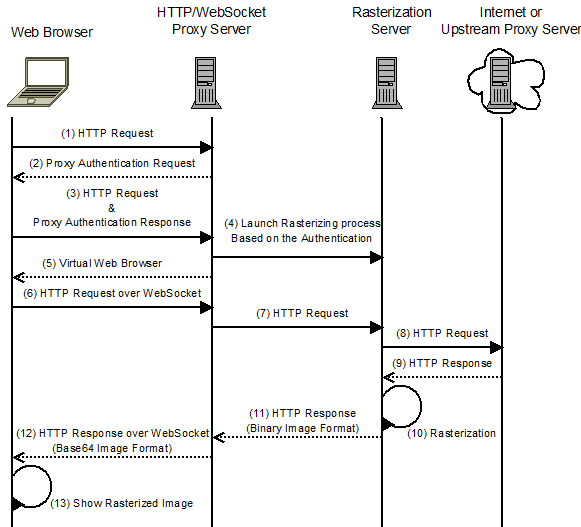


Fig. 3. Behavioral Overview of Virtual Web Browser

It is important to use WebSocket instead of HTTP to communicate with the proxy server, because it can communicate bidirectionally and asynchronously. Hence, the virtual Web browser can almost completely be synchronized with the launched rasterization process. For example, if a loaded Web page in the rasterization process moves to other URL, then the process sends a URL-changed notification to the virtual Web browser, which causes the browser to move to the new URL. For another example, if any of supported events such as `onresize`, `onscroll`, `onmouseup`, `onmousedown`, `onmousemove`, `onclick`, `ondblclick`, `onkeyup`, `onkeydown`, and `onkeypress` is occurred, the information of the event is sent to the process through the WebSocket, which means that the virtual Web browser supports user interaction events such as keyboard, mouse, scroll, and resize events.

### 3 Implementation of Virtual Web Browser

#### 3.1 Used Software and Implementation

We have used the software products shown in Table 1 to implement the virtual Web browser. JavaScript is used as the single programming language of the system to reduce development costs. jQuery makes our virtual Web browser portable among modern Web browsers. PhantomJS [6] is a CUI-based Web browser that is used as the rasterization engine. Apache HTTP server is used to act as the HTTP/WebSocket proxy. Node.js [5] is used to implement the HTTP/WebSocket proxy software that runs behind the Apache and communicates with PhantomJS.

To reduce network traffic, we have decided to compress the contents between the virtual Web browser and the proxy server. Table 2 shows the compression methods

that current Web browsers support. We choose *deflate* over *gzip*, because it is supported by the Apache module named `mod_deflate`.

Fig. 4 shows the skeleton of the virtual Web browser. As figure shows, it has only one `<img>` element that shows a rasterized image. All events related to user interaction such as keyboard and mouse events are handled by the embedded JavaScript and are notified to the server through the WebSocket. In addition, all events fired in the rasterization process are also notified to the virtual Web browser through the WebSocket. This event-handling lets the virtual Web browser act as though it were the browser itself on which the virtual Web browser runs.

**Table 1.** Used Software for Virtual Web Browser Implementation

Software	Version	Description
jQuery	2.0.3	JavaScript library.
PhantomJS	1.9.1	Used as rasterization engine.
Node.js	0.10.15	Used for HTTP/WebSocket proxy implementation.
Apache HTTP Server	2.4.6	Used as HTTP/WebSocket proxy server.
CentOS	6.4	Operating System.

```

<!DOCTYPE html>
<html>
  <head>
    <meta charset="UTF-8">
    <title>Virtual Web Browser</title>
    <style type="text/css">
      * {
        margin: 0px;
        padding: 0px;
      } /* Remove any space. */
    </style>
    <script type="text/javascript">
      /* (1) Load jQuery.
       * (2) Add event handlers for user interaction events.
       * (3) Open a WebSocket to communicate with the proxy server.
       * (4) Send/Recv event notifications through the WebSocket.
       */
    </script>
  </head>
  <body>
    <!-- This is the only element that shows a rasterized image. -->
    <img id="mybrowser">
  </body>
</html>

```

**Fig. 4.** Skeleton of Virtual Web Browser

**Table 2.** HTTP 1.1 Compression Methods Supported by Web Browsers

Web Browser	Version	gzip	compress	deflate
Internet Explorer	10.0	Yes	No	Yes
Firefox	22.0	Yes	No	Yes
Chrome	28.0	Yes	No	Yes
Opera	12.15	Yes	No	Yes

### 3.2 Limitations

The virtual Web browser does not support any Web browser plugin such as Java Applet, Flash Player, Adobe Acrobat Reader, and others. This fact is generally considered to be a disadvantage that decreases user experience, but we consider this acceptable as a trade-off between usability and security. We intend to use the virtual Web browser with a PAC (Proxy Auto Configuration) file [4]. Since the PAC file controls what URLs should be browsed directly or through a proxy, the virtual Web browser will be used only to browse URLs that are not listed in the file. This design minimizes deterioration of user experience and enables users to use the virtual Web browser with other proxy and/or other solutions for the Zero-day attack. In addition, the virtual Web browser does not support HTTPS, because the behavior of the virtual Web browser is recognized as the MITM (Man-In-The-Middle) attack by Web browsers, and modern browsers are designed to prevent the attack from occurring.

## 4 Evaluation

To evaluate our proposed method, we measured the amount of transferred bytes between the virtual Web browser and the proxy server. We use the top 5 Web sites ordered by traffic volume (© Alexa Internet, Inc.). Table 3 shows the size of the images of the top pages of the Web sites. The result shows that (1) transforming Web pages into equivalent images rarely increases network traffic<sup>4</sup>, (2) encoding the images with base64 increases their size, and (3) by using the deflate, the size of the base64-encoded images becomes almost the same as the original image size.

## 5 Related Work

There are several virtualization technologies such as desktop/application/OS virtualization. One of the differences between them and ours is that such virtualizations often require dedicated software and/or OS, but our proposed method requires only an HTML5 compliant Web browser as client environment, which means that our method can be widely used without any additional cost.

**Table 3.** Image File Sizes of Top Pages of Web Sites.

Web Site	Raw Size	PNG	PNG (base64)	PNG (base64, deflate)
Facebook	520,832	88,554	119,626	86,476
Google	184,125	45,392	61,321	42,490
YouTube	1,279,892	814,113	1,099,767	829,920
Yahoo!	978,704	812,161	1,097,133	821,183
Amazon.com	1,316,049	1,114,448	1,505,484	1,132,271

<sup>4</sup> Strictly, it strongly depends on the contents of the Web page.

Palanques et al. [2] has proposed the model and architecture called “Secure Cloud Browser” that supports secure Web navigation. Their work and ours are similar in that both methods rasterize the contents of requested Web pages in controlled environment to protect client computers against malicious software. On the other hand, both methods are different in that their method is based on the assumption that an attacker has administrative privileges on a victim’s computer and their method requires a Web browser and JRE (Java Runtime Environment) to run; our method aims to protect client computers against being infected with malware, and our method requires only an HTML5 compliant Web browser.

Grier et al. [1] has proposed the “OP web browser” to enable more secure Web browsing. Their and our methods are similar in that both rasterize requested Web pages in isolated processes and send them back to the clients. However, both methods are different in that their method requires JRE and the dedicated Web browser; our method requires an HTML5 compliant Web browser only.

Wang et al. [3] has proposed “SafeFox”, to create a safe browsing environment. They need light-weight virtualization to protect each Web browser process. On the other hand, we do not need virtual environment; instead, we use Unix’s APIs to isolate the behavior of the rasterization process.

## 6 Conclusion

In this paper, we have proposed the virtual Web browser that enables safer Web-browsing environment. As a result, we conclude that the method strengthens user security of client computers to some extent, and it can be one of the solutions for the malware threats. We plan to enhance our browser to increase usability. For example, partial rasterization of requested Web pages is one idea, since rasterizing entire pages loses some information including links, texts, animations, and others. Or enhancing the browser by using the tag `<canvas>` is another idea, which enables links to be noticeable, texts to be copied, animations to be runnable, and so on.

## References

1. Grier, C., Tang, S., King, S.T.: Secure Web Browsing with the OP Web Browser. In: IEEE Symposium on Security and Privacy, Oakland, pp. 402–416 (2008)
2. Palanques, M., Dipietro, R., del Ojo, C., Malet, M., Marino, M., Felguera, T.: Secure Cloud Browser: Model and Architecture to Support Secure WEB Navigation. In: 31st IEEE Symposium on Reliable Distributed Systems (SRDS), Irvine, pp. 402–403 (2012)
3. Wang, J., Huang, Y., Ghosh, A.: SafeFox: A Safe Lightweight Virtual Browsing Environment. In: 43rd Hawaii International Conference on System Sciences (HICSS), Honolulu, pp. 1–10 (2010)
4. Microsoft TechNet, <http://technet.microsoft.com/library/Dd361918>
5. Node.js, <http://nodejs.org/>
6. PhantomJS: Headless WebKit with JavaScript API, <http://phantomjs.org/>

# Design and Implementation of Task Context-Aware E-mail Platform for Collaborative Tasks

Masashi Katsumata

Department of Computer and Information Engineering, Nippon Institute of Technology  
4-1 Gakuendai, Miyashiro-machi, Minamisaitama-gun, Saitama, Japan  
katumata@nit.ac.jp

**Abstract.** E-mail communication is important for collaborative tasks in an enterprise. However, the management of collaborative tasks using existing e-mail management applications has certain problems. In this paper, we present a task context-aware e-mail platform that helps task users to send e-mails quickly and efficiently. This platform also automatically extracts data from reply e-mail messages. It enables task users to automatically classify task-related information and user support services by using a task context model. We have built the task context model as a semantic representation model of the associations between the task and the task-related e-mail process. This paper describes the design and implementation of this platform system on the basis of the task context model.

**Keywords:** Collaborative Task, Task Context, Ontology, E-Mail Platform.

## 1 Introduction

E-mail-based communication is important for collaborative tasks in an enterprise. In other words, e-mail communication among group members for a given task is required to achieve effective task management for reusing an e-mail message and its related resources (schedule, attached file, contact list, etc.). Knowledge workers can efficiently search and use e-mail messages and the corresponding resources by organizing these messages according to individual tasks [1]. Hence, multi-tasking knowledge workers often set up automatic filtering into folders or manually move e-mail messages into folders. Recently, enhanced existing task management systems and task-centric mail clients have been used for this purpose. Further, several research studies that support the discovery of e-mail messages and the related resources by adding meta-data to e-mails and the related resources have been conducted [1-4].

In this paper, we present a task context-aware e-mail platform that helps task users to send e-mails quickly and efficiently. This platform also automatically extracts data from the reply e-mail messages. In order to realize the concept of this platform, we have built a prototype system. This platform enables task users to automate task-related information classification and user support services by using a task context model. We have built the task context model as a semantic representation model of

the conceptual associations between a task and the task-related e-mail process. By using the task context model, this platform system can provide a context-aware service for mail form composition and automatic mail data extraction.

The rest of this paper is organized as follows: In Section 2, we review the related work. Section 3 discusses the design of the task context-aware e-mail platform. In Section 4, we address the system implementation. In Section 5, we present conclusions and future research directions.

## 2 Related Work

TaskMaster [2] enhances an e-mail client to function as a task management system and manages the resources as e-mail messages and file attachments for each task. Further, a useful user interface with both browsing and operating resources is provided. This system makes it easy to search through the resources of a task. KASIMIR [3] and OntoPIM [4] are ontology-based personal task-management systems. These systems provide semi-automated functions for retrieving and registering the task-related information within e-mail messages according to an ontology-based model. Topika [5] enhances the existing e-mail client to provide suggestions about the relevant shared space such as Wiki. This system facilitates the transition management of a user's collaborative activities to appropriate collaboration tools. All these works were primarily concerned with improving the management of and the search for task related information. Our primary goal is to provide the support function of reusing the managed data for accomplishing a collaborative task.

## 3 Design of Task Context-Aware E-mail Platform

### 3.1 Task Context Model

We have created an ontology-based semantic representation model that represents the conceptual associations between a task and the e-mail processes. We call this the task context model. The task context-aware e-mail platform performs the services required by a task user on the basis of a task context model. This model relates the conceptual associations between a task and an e-mail process to physical context entities (e-mail messages, attached files, group members, mail form items, etc.).

The semantic representation of the task context model is based on the Resource Description Framework (RDF)<sup>1</sup>. RDF is a collection of triples, each of which consists of a resource, a property, and a literal. A set of such triples is called an RDF graph. Figure 1 shows a task context model represented by an RDF graph.

The task context is represented as a property of a "Task" resource. In the task context model, task contexts are classified as files, schedules, participants, memos, mail form items, etc. Further, we define the concept of Action. The concept of Action refers to the type of collaborative task. These collaborative tasks involve the

---

<sup>1</sup> RDF, <http://www.w3.org/RDF>



characteristic used for retrieving the contents from a reply mail for many users in an organization. We have considered the following three Actions for a collaboration task:

1. Event Notification. This Action represents the concept of a mail process for the target of attendance confirmation toward an event being held in an organization.
2. Questionnaire Request. This Action represents the concept of a mail process for the target of questionnaire requests and collection.
3. File Collection. This Action represents the mail process for attached file collection.

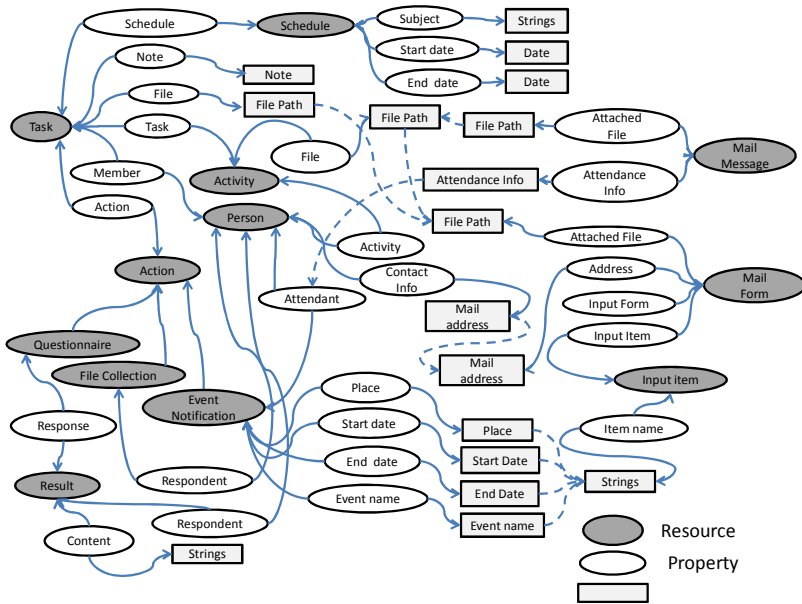


Fig. 1. Example of task context model

### 3.2 Task Context Management

The task context model manages the task context’s property and its value for each task. This property represents a task-related file, member, contact information, and schedule data, as shown in Fig. 1. The value automatically retrieved from an e-mail message is stored as the task context data in an XML data format.

### 3.3 Service for E-Mail Process

The aims of the task context-aware e-mail platform are to support 1) the composition of e-mail forms and 2) the extraction of the data contained in reply e-mails. The support for creating e-mail forms is provided when a task owner creates an e-mail form

for a task request and a task member creates an e-mail form for replying to a task request. On the other hand, the support for the extraction of data contained in a reply e-mail is provided when a task owner receives a reply mail from a task member.

## 4 Implementation

### 4.1 System Overview

We have implemented a prototype system that executes a service for task users on the task context-aware e-mail platform. This prototype system is composed of the following three systems: task context server, mail server, and mail client (see Fig. 2). The task context server manages the task context (file path, schedule, contact information, etc.) and its value. The task context server can accept a request command (create, refer, update, and delete) from the mail server and client via TCP/IP. By accepting a request command, the task context server can update the task context model by using the Jena API<sup>2</sup>. The mail server is built on Apache James<sup>3</sup>. Apache James is a mail application platform that enables users to program the code of custom applications for e-mail processing. Apache James provides e-mail filtering through a function called Matcher and provides e-mail processing through a function called Mailet. We have introduced extended e-mail headers for realizing the service (see Table 1). After Matcher refers to the extended e-mail headers, Mailet can be executed according to the purpose of these extended e-mail headers. The client can connect with both the

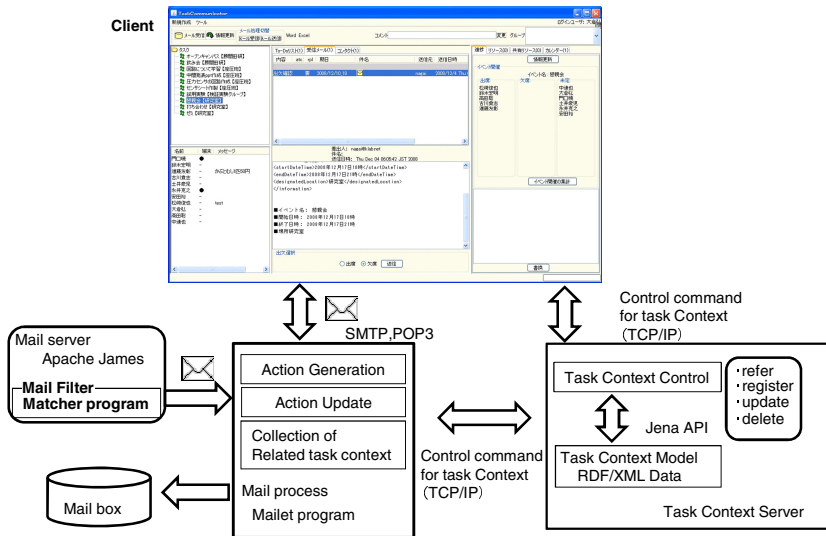


Fig. 2. Task context-aware e-mail platform

<sup>2</sup> Jena API, <http://jena.sourceforge.net>

<sup>3</sup> Apache James, <http://james.apache.org>

mail server and the task context server. In addition to general e-mail operations, the client provides the user interface that manages the task context data. The e-mail message submitted by the client is automatically added to the extended e-mail header. In our prototype system, the client displays the structured mail form by referring to the extended e-mail header.

**Table 1.** Types of extended mail headers

Header name	Role
X-Task-Name	Indicates the name of task
X-Task-Owner	Indicates the task owner
X-Action-Model-Type	Indicates the request of generating Action and the type of Action
X-Action-Update-Type	Indicates the request of updating the status of Action and the type of Action
X-Action-Retrieve-Type	Indicates the request of retrieving task context data and the type of Action

## 4.2 E-mail Form Composition Service

When a task owner selects the type of Action on the client, an e-mail form for the selected Action is displayed. In the e-mail form for the Action, the task-related data are provided as a list of suggestions of possible input. As a result, the task owner spends less time typing and querying for information related to the task. When the task client receives an e-mail from the task owner, the replying form is displayed by referring to the extended e-mail header X-Action-Model-Type. The displayed input fields on the reply form are the elements of the Action type corresponding to “reply mail form composition,” as shown in Table 2. A task member can type the value according to the displayed input field on the reply form.

**Table 2.** Services for Action type

Function	Event Notification	Questionnaire Request	File Collection
Request mail form composition	Schedule form (Start date, End date, Place, Event name)	Questionnaire form (Support of form composition)	File name and stored folder name
Reply mail form composition	Attendace form(Yes, No)	Questionnaire form (reply form)	Selection of attached file
Automated extraction	Attendance data	Questionnaire response data	Attached file

## 4.3 Data Extraction Service

When the mail server receives an e-mail according to the type of Action, the contents of the reply e-mail are automatically retrieved as the task context. In the task context server, the retrieved task context data are managed as RDF/XML format data that are based on the conceptual model for Action. When the task context server receives a

reply e-mail, the value of the task context in Action is updated, and the state of Action is displayed on the client's State panel. Thus, the task owner can confirm the state of the task intuitively without checking each reply e-mail. Automated processes such as the generation or update of Action are performed via Maillet according to the value of the extended mail header. The attendance data and the questionnaire data can be written to a Comma Separated Value (CSV) file on the assumption that task context data might be used by a spreadsheet application (e.g., Microsoft Excel). In the case of File Collection, files are automatically renamed according to a predefined file name from the attached file in the task member's reply e-mail.

#### 4.4 Evaluation

We conducted an experiment to verify the realization of our platform. We obtained qualitative data compiled from 13 university students (male, 21–22 years old) who used the prototype system for a month. The prototype system was set up in our laboratory. The results confirmed a high valuation for service functionality for a given task. Further, we obtained some beneficial comments for improving the platform and service functionality. More extensive evaluations will be performed in the near future.

## 5 Conclusions

In this paper, we described the design and implementation of a task context-aware e-mail platform for collaborative tasks. In order to provide a task context-aware service for a task user, we introduced the task context model that represents conceptual associations between a task and the related mail process. Using the prototype system, we confirmed that the task context-aware platform executed the required service on the basis of the task context model. From an operational experiment, we obtained beneficial comments for improving the prototype system for practical use from the point of view of an actual user environment. In our future work, we will further develop the prototype to support more tasks.

## References

1. Krämer, J.-P.: PIM-Mail: Consolidating Task and Email Management. In: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 4411–4416 (2010)
2. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I.: Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 345–352. ACM, New York (2003)
3. Grebner, O., Ong, E., Riss, U.V.: KASIMIR—Work Process Embedded Task Management Leveraging the Semantic Desktop. In: Proceedings of Multikonferenz Wirtschaftsinformatik, Workshop Semantic Web Technology in Business Information Systems, pp. 715–726 (2008)

4. Lepouras, G., Dix, A., Katifori, T., Catarci, T., Habegger, B., Poggi, A., Ioannidis, Y.: On-toPIM: From Personal Information Management to Task Information Management. In: Proceedings of SIGIR Workshop on Personal Information Management, pp. 78–81 (2006)
5. Mahmud, J., Matthews, T., Whittaker, S., Moran, T.P., Lau, T.: Topika: Integrating Collaborative Sharing with Email. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 3161–3164. ACM, New York (2011)

# Empirical Analysis on the Users' Reply Behaviors of Online Forums

Guirong Chen<sup>1</sup>, Wandong Cai<sup>1</sup>, Aiwang Chen<sup>2</sup>, Huijie Xu<sup>1</sup>, and Rong Wang<sup>2</sup>

<sup>1</sup> School of Computer Science, Northwestern Polytechnical University,  
710068 Xi'an, People's Republic of China

<sup>2</sup> School of Information and Navigation, Air Force Engineering University of PL,  
710077 Xi'an, People's Republic of China  
{guirongchen315,xachenaiwang}@163.com, caiwd@nwpu.edu.cn  
xhj004@gmail.com, xxmnl1hl@sina.com

**Abstract.** Quantitative understanding of human behaviors is of huge significance to uncover the origins of many social and economic phenomena. This paper focuses on users' reply behaviors of online forums. The statistical results show that users' one-day activities and one-post interests both follow heavy-tailed distribution not only on population level but also on individual level. Specifically, we found that users' one-day activities obey a power law distribution with exponential cutoff and users' one-post interests obey a power law distribution. Further, we observe a positive correlation between users' total activities and the power law exponents of one-day activities. In addition, we study the content scatter degree of users' one-day replies and the time scatter degree of users' one-post replies, and find some suspicious user behaviors, which provide new clues to the detection of cyber-space hype and network water army.

**Keywords:** Empirical Analysis; Reply Behavior; Heavy-tailed Distribution; BBS.

## 1 Introduction

Human activities are the driving force of the development of the social. Understanding human behaviors is of huge significance which has attracted more and more attentions. However, due to the complexity of human behaviors and the lack of real data, traditional studies usually assume that the human behaviors are random in time and the temporal statistics of human activities can be described by a Poisson process. With the popularity of the Internet and the rapid development of database technology, more and more human behaviors are recorded, quantitative study of human behaviors become possible. Barabási extracted the statistical laws of human behaviors from the historical records of human actions [1] and found that both surface mail communication and email communication follow non-Poisson statistics, characterized by bursts of rapidly occurring events separated by long periods of inactivity. In order to verify whether the non-Poisson statistics is prevalent in human

behaviors, more and more researchers conducted empirical analysis on other human activities, including online movie watching [2], web browsing [3], blog commenting [4], short message communications [5-6], library-loaning [7], mobile communications [8] and so on, and found that human behavior patterns are all different from Poisson statistics. Those empirical statistical analysis results indicate the invalidity of Poisson process in mimicking the human behaviors in many real life systems.

Online forum (also called the Bulletin Board System, BBS for short) is an excellent platform for people to communicate and share information. Compared with other social network platforms such as blogs on which people use real names, BBS has greater openness and concealment. People can register new identifications freely and arbitrarily without revealing their real names, and when someone has logged on BBS, he or she can browse the content submitted by all the others but not be limited to his or her friends. Therefore, BBS has become one of the most important platforms for ordinary people to share knowledge, discuss problems and express their views. On the other hand, it also facilitates criminals to swindle and scam. One typical example is cyber-space hype. In order to hype some character, topic or product in the Internet, for instance, a common trick is to hire a large number of network water army [9] to submit a huge number of posts and replies on various major online forums in a very short period of time, concocting network hot topics to praise or create the illusion of support for it, or on the contrary, slandering it. Research on user behaviors of BBS is not only helpful to the understanding of online human dynamics, but meaningful to the detection of cyber-space hype and network water army. Recently, some researchers conducted empirical statistical analysis on BBS and found that the number of views and replies of users and the number of replies in posts all follow power law distributions [10-13].

In this paper, we investigate the statistics of users' reply behaviors based on the real data of one famous online forum, including users' one-day activity distribution, one-post interest distribution. The empirical analysis results prove that users' one-day activities and one-post interests both follow heavy-tailed distributions not only on population but also on individual level. Further, we study the content scatter degree of users' one-day replies and the time scatter degree of users' one-post replies, and find some interesting results. The findings in the paper provide some new clues to the detection of cyber-space hype and network water army.

## 2 Data Set

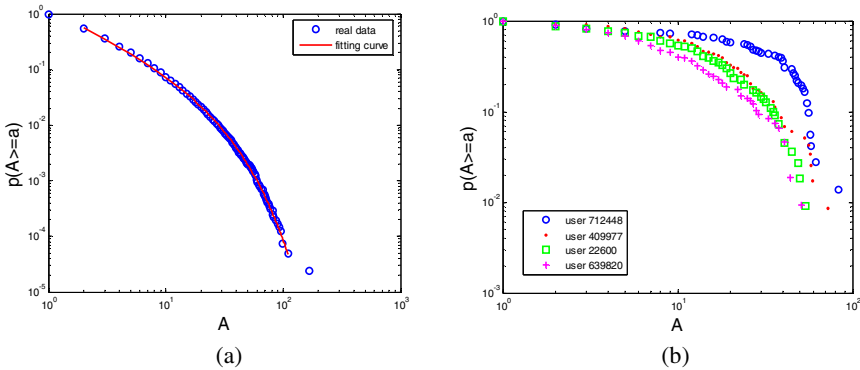
Our data set, obtained from <http://bbs.cnhubei.com/portal.php>, is collected using a web forum crawling system [14]. Donghu Community is the largest and most influential online forum in Hubei province of China, which has a large number of users. Its themes cover culture, life, society, current affairs, entertainments, sports, etc. We collected all the posts and replies from ZhongBuJueQi Forum, a sub-forum of Donghu Community, ranging from January 1, 2011 to May 17, 2011. During this period, there are totally 8155 posts, 145284 replies and 5112 users. Users who didn't release any posts or replies during that period are not considered.

### 3 Distribution of One-Day Activity

Recently, Yu etc. conducted empirical analysis based on two online forums and showed that the total number of views and replies of users obey power law distribution with different power exponents [10]. The result indicates that most people are inactive and they view and reply rarely, while some people are very active and they view and reply frequently. This is meaningful to the understanding of human dynamics on BBS.

Here, we focus on the activities of users on each day. The *one-day activity* of user  $i$  on date  $d$ , denoted by  $a_{i,d}$ , is defined as the number of replies submitted by user  $i$  on date  $d$ . Tracking the records of user  $i$ , we can calculate all the one-day activities of user  $i$  on all the dates during the period of time. One-day activities of all the users on the period of time are calculated in the same way, and we get a large series containing 55531 records.

Fig.1 (a) is the cumulative distribution of one-day activities of all the users on population level. As shown in Fig. 1(a), the distribution shows heavy-tailed characteristics and can be well fitted by a power-law distribution with exponential cutoff  $p(a) \propto a^{-\lambda} e^{-\delta a}$ , where  $\lambda = 1.05$  is the power-law exponent and  $\delta = 0.048$  is the exponential exponent.



**Fig. 1.** The cumulative distribution of one-day activities. (a) On population level. (b) On individual level. The red curve in panel (a) is the fitting curve. The blue circles, red dots, green squares and pink crosses in panel (b) correspond to user 712448, 409977, 22600 and 639820.

The result proves that most users on most dates are inactive, releasing few of replies, while some users are very active on some date, releasing a lot of replies, even larger than 150. Further statistics show that more than 95% of the one-day activities are smaller than 10. The users one of whose one-day activities is larger 10 add up to a total of 555, accounting to 7.2% of the total, which means that 92.8% users were inactive on all the days of the period and they should not be online water army.

In order to further understand the users' one-day reply behaviors, we choose fours active users from our data set, who released a large number of replies during the



period of time. Some basic properties of the four users are listed in Table 1. Each line in the table shows the minimum, maximum, average one-day activity and the total number of replies. Here, we define *overall activity* of user  $i$ , denoted by  $A_i$ , as the total number of replies submitted by user  $i$  during the period. As Table 1 shows, the overall activities of the users increase gradually from user 712448 to user 639820.

Fig.1 (b) is the one-day activity distributions of the four users in Table 1. As what we can see from Fig.1 (b), the one-day activity distributions all follow a power law distribution with exponential cutoff. The power law exponents of the four users increase gradually from user 712448 to 639820. That is to say, there is a positive correlation between users' overall activities and the power law exponents of one-day activities. Furthermore, we find other users display very similar statistics, thus we believe the findings shown here are common for the most of BBS users.

**Table 1.** Properties of the four users we chose.

User ID	Min One-day Activity	Max One-day Activity	Aver One-day Activity	Overall Activity
712448	1	83	27	1993
409977	1	72	16	1977
22600	1	54	14	1570
639820	1	51	11	1224

## 4 Distribution of One-Post Interest

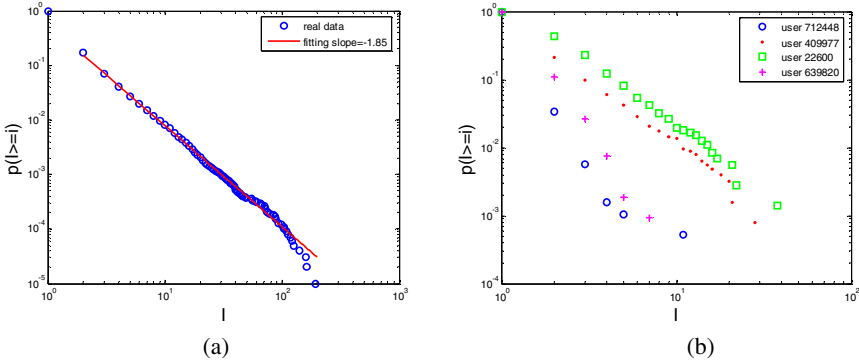
Similarly, the *one-post interest* of user  $j$  on post  $p$ , denoted by  $i_{j,p}$ , is defined as the number of replies submitted by user  $j$  to post  $p$ . Tracking the records of user  $j$ , we can calculate all the one-post interests of user  $j$  on all the posts. We compute all the one-post interests of all the users in a similar way and get a large series.

Fig.2 shows the one-post interest distributions. As Fig.2 (a) shows, the cumulative distribution of one-post interests of all the users also follows heavy-tailed distribution on population level and can be well fitted by a power law distribution  $p(i) \propto i^{-\gamma}$ , where  $\gamma = 1.85$  is the power law exponent.

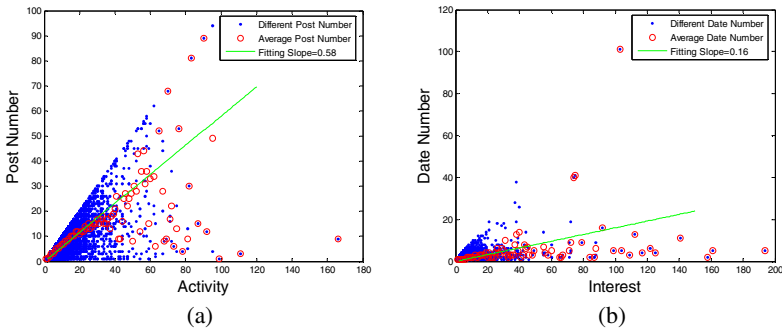
Fig.2 (b) reports the one-post interest distributions of the four users in Table 1 on individual level. As we can see, the one-post interest distributions of the four users all follow a power law distribution with different power law exponents. That is to say, the one-post interest distributions show some common characteristics on population level and individual level. The results indicate that most users have small interest in most posts, while some users are very interested in some posts and release a large number of replies, even larger than 200. We infer these behaviors are abnormal and we should pay close attention to them. Further statistics show that more than 99.3% of the one-post interests are smaller than 10.

To understand users' reply behaviors more deeply, we study the content and time concentration of the one-day and one-post reply behaviors respectively, as Fig.3

shows. The *content scatter degree* of user  $i$  on day  $d$ , denoted by  $csd_{i,d}$ , is defined as the number of different posts user  $i$  replied to on that day. The *time scatter degree* of user  $i$  on post  $p$ , denoted by  $tsd_{i,p}$ , is defined as the number of different dates user  $i$  released replies to post  $p$ .



**Fig. 2.** The cumulative distribution of one-post interests. (a) On population level. (b) On individual level. The red line in panel (a) is the fitting line. The blue circles, red dots, green squares and pink crosses in panel (b) correspond to user 712448, 409977, 22600 and 639820.



**Fig. 3.** (a) The relationship between users' one-day activity and the content scatter degree. (b) The relationship between users' one-post interest and the time scatter degree. The blue dots and red circles in Fig.3 are both empirical data. One blue dot in panel (a) represents one user's reply behavior on one day, horizontal ordinate represents the one-day activity; vertical coordinate represents the content scatter degree. One red circle in panel (a) stands for a distinct one-day activity and the average content scatter degree. One blue dot in panel (b) represents one user's reply behaviors about one post, horizontal ordinate represents the one-post interest, vertical coordinate represents time scatter degree. One red circle in panel (b) stands for a distinct one-post interest and the average time scatter degree. The green lines in panel (a) and (b) are both fitting lines.

In Fig.3, the blue dots on the lower right part of panel (a) (under the green line) mean that some one are active on some day, but his or her replies are concentrated on few posts, while the blue dots on the same area of panel (b) mean that some one has great interest on some post, and his or her replies are concentrate on few days. These two types of behaviors are doubtful, and should be further studied in the detection of cyber-space hype and network water army.

## 5 Conclusion

In this paper, we empirically studied the statistical regularities of one famous online forum in China: Donghu Community. We found that the one-day activity follow a power law distribution with exponential cutoff both on population level and individual level, and the one-post interest obey a power law distribution. These results prove that users' reply behaviors have apparent burst characteristics, and some users are active on some days while most users are inactive on most days; some user show great interest on some post while most users just have little interest on most posts. We also studied the content scatter degree of users' one-day replies and the time scatter degree of users' one-post replies and found a few of reply behaviors show some characteristics such as "short period of time", "a large number of" and so on, which are common for cyber-space hypes. In the next step, we will conduct deeper analysis to the suspect behaviors.

## References

1. Barabási, A.L.: The Origin of Bursts and Heavy Tails in Human Dynamics. *J. Nature* 435, 207–211 (2005)
2. Zhou, T., Kiet, H.A.T., Kim, B.J., et al.: Role of Activity in Human Dynamics. *J. EPL (Europhysics Letters)* 82, 28002-p1–28002-p5 (2008)
3. Radicchi, F.: Human Activity in the Web. *J. Physical Review E* 82, 026118-1–026118-8 (2009)
4. Guo, J.L., Fan, C., Guo, Z.H.: Weblog Patterns and Human Dynamics with Decreasing Interest. *J. The European Physical Journal B* 81, 341–344 (2011)
5. Wei, H., Xiao-Pu, H., Tao, Z., et al.: Heavy-tailed Statistics in Short-message Communication. *J. Chinese Physics Letters* 26, 028902-1–028902-3 (2009)
6. Zhi-Dan, Z., Hu, X.I.A., Ming-Sheng, S., et al.: Empirical Analysis on the Human Dynamics of a Large-scale Short Message Communication System. *J. Chinese Physics Letters* 28, 068901-1–068901-4 (2011)
7. Fan, C., Guo, J.L., Zha, Y.L.: Fractal Analysis on Human Dynamics of Library Loans. *J. Physica A: Statistical Mechanics and its Applications* 391, 6617–6625 (2012)
8. Jiang, Z.Q., Xie, W.J., Li, M.X., et al.: Calling Patterns in Human Communication Dynamics. *J. Proceedings of the National Academy of Sciences* 110, 1600–1605 (2013)
9. Information on, <http://baike.baidu.cn/view/3098178.htm>
10. Yu, J., Hu, Y., Yu, M., et al.: Analyzing Netizens' View and Reply Behaviors on the Forum. *J. Physica A: Statistical Mechanics and its Applications* 389, 3267–3273 (2010)

11. Ding, F., Liu, Y., Shen, B., et al.: Modeling Reading and Replying Activities in a BBS Social Network. *J. Journal of Electronics (China)* 8, 300–306 (2010)
12. Si, X.M., Liu, Y.: Empirical Analysis of the Interpersonal Interacting Behavior in Virtual Community. *J. Acta. Phys. Sin.* 60, 073908-1–073908-8 (2011)
13. Xiong, F., Liu, Y.: Empirical Analysis and Modeling of Users' Topic Interests in Online Forums. *J. PLoS ONE* 7, e50912-1– e50912-7 (2012)
14. Peng, D., Cai, W.D.: The Web Forum Crawling Technology and System Implementation. *J. Computer Engineering & Science* 33, 157–160 (2011) (in Chinese)

# On Companding Transform Techniques for OFDM Visible Light Communication over Indoor Dispersive Channels

Kasun Bandara, Pararajasingam Niroopan, and Y.H. Chung\*

Department of Information and Communications Engineering,  
Pukyong National University,  
Busan, Republic of Korea  
yhchung@pknu.ac.kr

**Abstract.** The  $\mu$ -law companding transform (CT) is proposed for peak-to-average power ratio (PAPR) reduction in orthogonal frequency division multiplexing visible light communication (VLC-OFDM) systems. The proposed PAPR reduction technique based on the  $\mu$ -law CT can minimize nonlinear signal distortions and LED chip overheating problems that occur due to nonlinear characteristics of the transmitter LEDs. The performance of the proposed PAPR reduction is studied via simulations. It is verified that the proposed  $\mu$ -law CT outperforms the previously proposed DFT spread (DFTS) method in VLC-OFDM systems in terms of PAPR reduction capability. The bit error rate (BER) performance of the VLC-OFDM system with the  $\mu$ -law CT is also analyzed over an indoor dispersive VLC channel model.

**Keywords:** Companding transform, PAPR reduction, dispersive visible light communication channel, OFDM.

## 1 Introduction

Visible light communication (VLC) is an emerging communication technology that has been being developed since the past decade [1], [2]. A number of VLC applications have been presented for both indoor and outdoor VLCs [2-5].

In order to have consistent and error-free indoor communication between the transmitter and the receiver, channel characteristics of the indoor environment should be properly studied as well as the properties of the system components. In an indoor environment, the main impairment that increases the bit error rate (BER) is the inter-symbol interference (ISI). ISI is caused by the multipath propagation in the indoor environment due to multiple transmit light emitting diodes (LEDs) and reflection from walls and other surfaces in the indoor environment. Therefore, it is crucial to mitigate the ISI problem in VLC, in order to have a reliable VLC transmission. Orthogonal frequency division multiplexing (OFDM) is a favorable technique to reduce multipath propagation induced effects in the VLC systems.

---

\* Corresponding author.

When OFDM is employed, the signal undergoes a high peak-to-average power ratio (PAPR) that subsequently causes nonlinear signal distortions and chip overheating in light emitting diodes (LEDs) at the transmission, due to the nonlinear voltage-current (VI) characteristics of the LEDs. Recently, a companding transform technique for fiber optic based OFDM systems has been proposed [6]. However, this PAPR reduction method derived for fiber optical communications is not directly applicable to VLC-OFDM. In VLC-OFDM, a DFT spread (DFTS) method has been proposed [7]. This DFTS method creates higher complexity in the system and also offers a limited PAPR reduction. Therefore, simpler and better PAPR reduction methods still need to be documented for VLC-OFDM systems.

In this paper, we investigate the use of  $\mu$ -law CT in indoor VLC-OFDM. With the CT, the OFDM signal is compressed and expanded, while reducing the PAPR. We analyze the PAPR reduction performance of the  $\mu$ -law CT and the BER performance of the VLC-OFDM system with an indoor dispersive channel model.

Most of the works in the past for VLC-OFDM have been done in the additive white Gaussian noise (AWGN) channel [8]. Although the performance analysis with AWGN channel gives an insight about the effect of noise channel on the VLC system, it does not provide the effect of channel dispersion on the performance. In the literature, the performance of indoor VLC-OFDM systems with CT over dispersive channels has not been reported yet. Therefore, the present study of the CT technique for the PAPR reduction as well as the performance analysis over dispersive VLC-OFDM channels is justified.

From simulation results and comparative studies, we verify that the  $\mu$ -law CT is an attractive technique to reduce PAPR in VLC-OFDM systems.

## 2 The Indoor Channel Model and Channel Characteristics

The proposed channel model consists with an empty room of the dimensions 5 m  $\times$  5 m  $\times$  3 m (length  $\times$  width  $\times$  height). For the simulation, the  $x$ - $y$  plane (the working plane) is defined to be on the floor of the room and  $z$  axis is defined in the height direction. Each of the four transmitters contains 100 LEDs (10  $\times$  10), with 4 cm gap between each two. The centers of each transmitter are positioned on the ceiling at the coordinates (1.25, 1.25, 3), (1.25, 3.75, 3), (3.75, 1.25, 3), (3.75, 3.75, 3). The receiver is placed at the point (0.1, 2.1, 0). The receiver position is chosen considering the root mean square (rms) delay spread at the position, as described in Section IV. The illumination characteristics of similar indoor VLC systems are well studied in the literature [1], [2].

An important measure of a dispersive channel is the rms delay spread. When  $h(t)$  represents the channel impulse response and  $\tau_0$  represents the average delay, the rms delay spread of the channel is obtained by [1]

$$\tau_{RMS} = \sqrt{\frac{\int_{-\infty}^{\infty} (t - \tau_0)^2 h^2(t) dt}{\int_{-\infty}^{\infty} h^2(t) dt}} \quad (1)$$

The rms delay spread on the working plane of the proposed indoor dispersive channel is shown in Fig. 1.

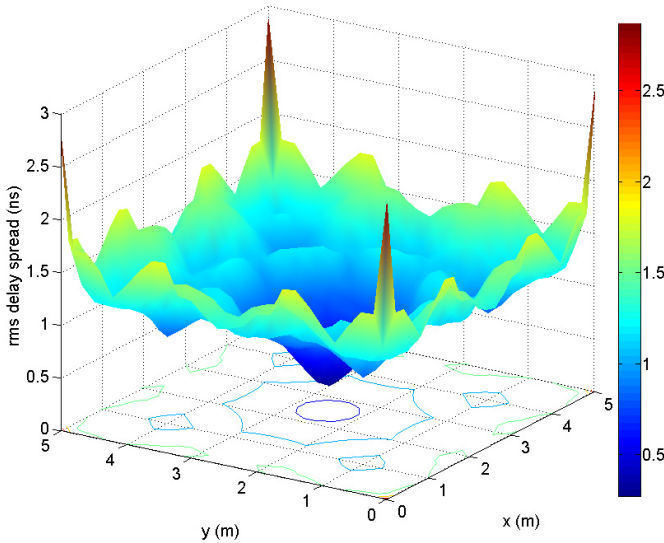


Fig. 1. Distribution of rms delay spread on the working plane

### 3 VLC-OFDM and the $\mu$ -law Companding Transform

The PAPR of the OFDM signal is defined as in (7), where  $s_n$  is the samples of OFDM symbols output from the IFFT, and  $E\{x\}$  is the expected value of  $x$ .

$$PAPR = \frac{\max \{|s_n|^2\}}{E\{|s_n|^2\}} \tag{2}$$

We apply the  $\mu$ -law CT,  $f$ , to the magnitude of the real-valued OFDM signal. The companded signal,  $y_n$ , is then

$$y_n = \text{sgn}(s_n)f(|s_n|) \tag{3}$$

where  $\text{sgn}(\cdot)$  is the Signum function. The distortions made by the CT transform can be reduced by applying the inverse CT to the received signal the receiver

The  $\mu$ -law CT is expressed as [6]

$$f(t) = k \text{sgn}(t) \left[ \frac{\ln(1 + \mu|t/k|)}{\ln(1 + \mu)} \right] \tag{4}$$

In (4),  $k$  is the mean amplitude of the input, whereas  $\mu$  determines the shape of the transform function as in Fig. 2. At the receiver, the inverse of the CT function is applied to remove any distortions made by the CT.

In the VLC-OFDM, the OFDM signal should be a real valued signal, since the signal is transmitted via LEDs using intensity modulation. To create the real VLC-OFDM signal, we make the input to the IFFT to have Hermitian symmetry ( $X_n = X_{N-n}^*$ ). This would waste a half of the bandwidth for the complex conjugate subcarriers. In VLC, however, since there is theoretically an unlimited bandwidth, this inefficient bandwidth usage is not a considerable issue. After that, cyclic prefix (CP) is added to the signal. At this point, a high PAPR value is observed in the OFDM signal. Hence, we perform the  $\mu$ -law CT at this stage. Now, the OFDM signal is suitable to transmit vial nonlinear LED characteristics. We design the nonlinear VI characteristics of the LEDs using Rapp model [9].

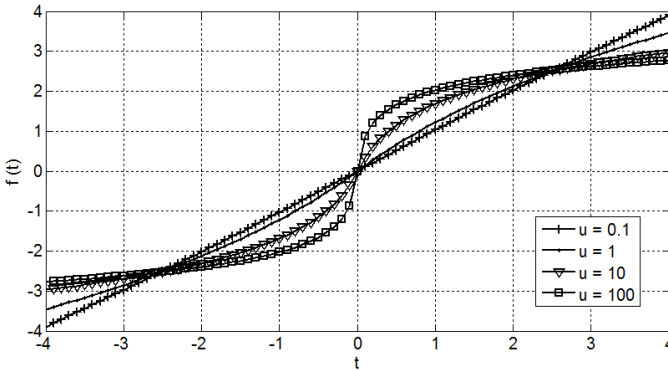


Fig. 2. The curves of transform function with different  $\mu$  values

## 4 Simulation Results and Discussion

In order to evaluate the PAPR reduction capability of the proposed companding function, we use the complementary cumulative distribution function (CCDF) of the probability of PAPR. The CCDFs of both uncompanded and the  $\mu$ -law CT OFDM signals is depicted in Fig. 3. It illustrates how the PAPR is affected by the value of  $\mu$ . As seen in Fig. 2, the shape of the CT function changes with  $\mu$ . Therefore, the signal expansion and compression ratios change with  $\mu$ , giving different PAPR reductions. It can be observed in Fig. 3 that the PAPR reduction increases as the value of  $\mu$  increases.

In comparison with the DFTS method [7], we achieve nearly 3 dB and 5 dB PAPR at the CCDF of  $10^{-3}$ , when  $\mu$  is 100 and 10, respectively. However, the PAPR achieved by the DFTS is only about 7 dB at the same CCDF. Hence, it is apparent that the proposed  $\mu$ -law companding outperforms the DFTS method in terms of the PAPR reduction capability.

To evaluate the BER performance of the indoor dispersive VLC-OFDM system, we choose a receiver position where there exists comparatively high delay spread. From the rms delay spread in Fig. 1, it is found that the maximum rms delay spread occurs at the corners of the room and the value is 2.7 ns. However, in the practical



situations, there is low probability for the receiver to exist at the corners of the room. Hence, we choose the point (0.1, 2.1, 0), which is a practical position for the receiver. The rms delay spread at the chosen receiver position is found to be 1.48 ns.

The BER performance of the indoor VLC-OFDM system with the  $\mu$ -law CT is evaluated and the results are shown in Fig. 4. The system employs both BPSK and QPSK modulations over the proposed indoor dispersive channel. The BER performances with  $\mu = 1$  and  $\mu = 10$  are better than the ones with  $\mu = 0.1$  and  $\mu = 100$ . However, as shown in Fig. 4, it is important to note that a higher value of  $\mu$  yields a higher PAPR reduction. Therefore, it can be concluded that there is a tradeoff between PAPR reduction and BER performance in the VLC-OFDM system with the  $\mu$ -law CT.

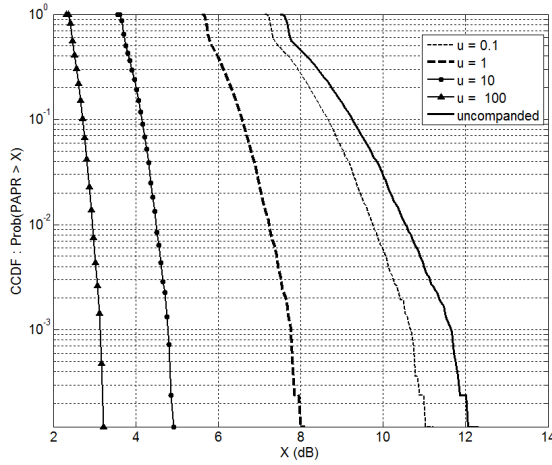


Fig. 3. CCDFs of uncompanded signal and  $\mu$ -law CT signal with different  $\mu$  values

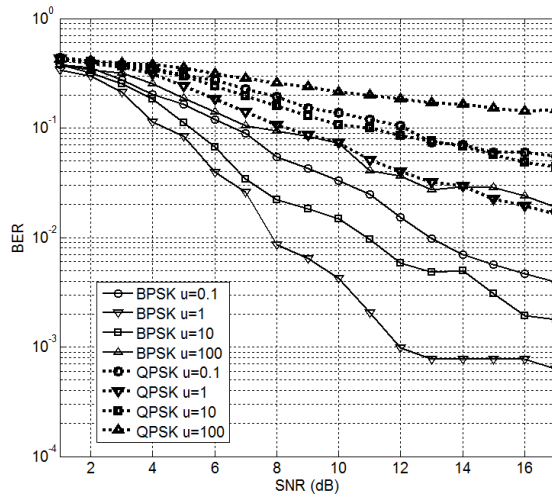


Fig. 4. BER performance with  $\mu$ -law CT for different values of  $\mu$

## 5 Conclusion

The high PAPR in the VLC-OFDM signals distorts the transmitted signals, due to the nonlinear VI characteristics of the transmitter LEDs. To reduce these adverse effects in the indoor VLC-OFDM systems, the  $\mu$ -law CT is proposed to reduce the high PAPR of the real-valued VLC-OFDM signals. It is demonstrated that the proposed strategy outperforms the conventional method in terms of PAPR reduction capability in the VLC-OFDM systems. Moreover, the PAPR reduction capability increases with the value of  $\mu$ . The BER performance of the system with the proposed  $\mu$ -law CT is analyzed over an indoor dispersive channel model. The results of the BER performance exhibit that the higher the value of  $\mu$  is, the higher the BER will be. Therefore, the tradeoff between the PAPR reduction strength and the BER performance should be taken into account for the  $\mu$ -law CT based VLC-OFDM systems.

## References

1. Tanaka, Y., Komine, T., Haruyama, S., Nakagawa, M.: Indoor visible light data transmission system utilizing white LED lights. *IEICE Trans. Commun.* E86-B(8), 2440–2454 (2003)
2. Bandara, K., Chung, Y.H.: Reduced training sequence using RLS adaptive algorithm with decision feedback equalizer in indoor visible light wireless communication channel. In: *International Conference on ICT Convergence (ICTC)*, Jeju, Korea, pp. 149–154 (2012)
3. Wada, M., Yendo, T., Fujii, T., Tanimoto, M.: Road-to-vehicle communication using LED traffic light. In: *IEEE Intelligent Vehicles Symposium*, pp. 601–606 (2005)
4. Park, S.B., Jung, D.K., Shin, H.S., Shin, D.J., Hyun, L.K., Oh, Y.J.: Information Broadcasting System based on Visible Light Signboard. In: *Wireless and Optical Communications* (2007)
5. Grubor, J., Gaete, J.O., Waleski, J., Randel, S., Langer, K.: High-speed wireless indoor communication via visible light. In: *ITG Fachbericht*, pp. 203–208 (2007)
6. Li, X., Tao, L., Zhang, J., Wang, Y., Fang, Y., Zhu, J., Wang, Y., Shao, Y., Chi, N.: Companding Transform for PAPR Reduction in Coherent Optical OFDM System. In: *Wireless and Optical Communications Conference*, pp. 46–47 (2012)
7. Ryu, S.B., Choi, J.H., Bok, J., Lee, H.K., Ryu, H.G.: High Power Efficiency and Low Nonlinear Distortion for Wireless Visible Light Communication. In: *IFIP Int. Conf. on New Technologies, Mobility and Security*, pp. 1–5 (2011)
8. Elgala, H., Mesleh, R., Haas, H.: A study of LED nonlinearity effects on optical wireless transmission using OFDM. In: *IFIP International Conference on Wireless and Optical Communications Networks*, pp. 1–5 (2009)
9. Elgala, H., Mesleh, R., Haas, H.: An LED model for intensity-modulated optical communication systems. *IEEE Photonics Technology Letters* 22(11), 835–837 (2011)

# Head Pose Estimation Based on Random Forests with Binary Pattern Run Length Matrix

Hyunduk Kim, Sang-Heon Lee, Myoung-Kyu Sohn, Dong-Ju Kim, and Nuri Ryu

Dept. of Convergence, Daegu Gyeongbuk Institute of Science & Technology (DGIST)  
50-1 Sang-Ri, Hyeonpung-Myeon, Dalseong-Gun, Daegu, 711-873, Korea

**Abstract.** In this paper, a novel approach for head pose estimation in gray-level images is presented. In the proposed algorithm, there were two techniques employed. In order to deal with the large set of training data, the method of Random Forests was employed; this is a state-of-the-art classification algorithm in the field of computer vision. In order to make this system robust in terms of illumination, a Binary Pattern Run Length matrix was employed; this matrix combined a Local Binary Pattern and a Run Length matrix. Experimental results show that our algorithm is robust against illumination change.

**Keywords:** Head pose estimation, Random Forests, Binary Pattern, Run Length matrix.

## 1 Introduction

Determining head pose is one of the most important topics in the field of computer vision. There are many applications with accurate and robust head pose estimation algorithms, such as human-computing interfaces (HCI), driver surveillance systems, entertainment systems, and so on. For this reason, many applications would benefit from automatic and robust head pose estimation systems. Accurately localizing the head and its orientation is either the explicit goal of systems like human computer interfaces or a necessary preprocessing step for further analysis, such as identification or facial expression recognition. Due to its relevance and to the challenges posed by the problem, there has been considerable effort in the computer vision community to develop fast and reliable algorithms for head pose estimation [1]. The several approaches to head pose estimation can be briefly divided into two categories: model-based approaches and appearance-based approaches.

The model-based approaches combine the location of facial features (e.g. eyes, mouth, and nose tip) and a geometrical face model to calculate precise angles of head orientation [2]. In general, these approaches can provide accurate estimation results for a limited range of poses. However, these approaches have difficulty dealing with low-resolution images due to invisible or undetectable facial points. Moreover, these approaches depend on the accurate detection of facial points. Hence, these approaches are typically more sensitive to occlusion than appearance-based methods, which use information from the entire facial region [3].

The appearance-based approaches discretize the head poses and learn a separate detector for each pose using machine learning techniques that determine the head poses from entire face images [3]. These approaches include multi-detector methods, manifold embedding methods, and non-linear regression methods. Generally, multi-detector methods train a series of head detectors each attuned to a specific pose and assign a discrete pose to the detector with the greatest support [1, 4]. Manifold embedding based methods seek low-dimensional manifolds that model the continuous variation in head pose. These methods are either linear or nonlinear approaches. The linear techniques have an advantage in that embedding can be performed by matrix multiplication; however, these techniques lack the representational ability of the nonlinear techniques [1, 5]. Non-linear regression methods use nonlinear regression tools (e.g. Support Vector Regression, neural networks) to develop a functional mapping from the image or feature data to a head pose measurement. These approaches are very fast, work well in the near-field, and give some of the most accurate head pose estimates in practice. However, they are prone to error from poor head localization [1, 6].

In this paper, we propose a new head pose estimation method for gray-level images; this method consists of two techniques. First, random forests were employed; these are powerful tools capable of mapping complex input spaces into discrete or respectively continuous output space. A tree achieves highly non-linear mappings by splitting the original problem into smaller ones, solvable with simple predictors [7]. Second, binary pattern run length matrix was employed for binary test function in random forests; this method is a combination of a local binary pattern and a run length matrix. The experimental results show the efficiency of the proposed algorithm.

## 2 Related Work

### 2.1 Local Binary Pattern

Recently, the Local Binary Pattern has been extensively exploited for facial image analysis, including face detection, face recognition, facial expression analysis, gender/age classification, and so on. The Original LBP operator labels the pixels of an image by thresholding a 3x3 neighborhood of each pixel with the center value and considering the results as a binary number, of which the corresponding decimal number is used for labeling. Formally, given a pixel at  $(x_c, y_c)$ , the resulting LBP can be derived by:

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n \quad , \quad (1)$$

where  $n$  runs over the 8 neighbors of the central pixel,  $i_c$  and  $i_n$  are gray-level values of the central pixel and the surrounding pixels, respectively, and the function  $s(x)$  is defined as:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

According to the definition above, the LBP operator is invariant to the monotonic gray-scale transformations that preserve the pixel intensity order in local neighborhoods. The histogram of LBP labels calculated over a region can be exploited as a texture descriptor [8].

## 2.2 Gray Level Run Length Matrices

The Gray Level Run Length (GLRL) method is a way of extracting higher order statistical texture features. This technique has been described and applied by Galloway and by Chu et al. A set of consecutive pixels with the same gray level, collinear in a given direction, constitutes a gray level run. The run length is the number of pixels in the run, and the run length value is the number of times such a run occurs in an image.

A Gray Level Run Length Matrix (GLRLM) is a two-dimensional matrix in which each element  $p(i, j|\theta)$  gives the total number of occurrences of runs of length  $j$  at gray level  $i$ , in a given direction  $\theta$ . Let  $G$  be the number of gray levels,  $R$  be the longest run and  $n$  be the number of pixels in the image. Galloway introduced statistical texture features (Run Percentage) to be extracted from the GLRLM as follows:

$$RP = \frac{1}{n} \sum_{i=1}^G \sum_{j=1}^R p(i, j|\theta) \quad (2)$$

This feature is a ratio of the total number of runs to the total number of possible runs if all runs had a length of one [9].

## 3 Proposed Head Pose Estimation Algorithm

### 3.1 Training

A Tree  $T$  in a forest  $F = \{T_i\}$  is built from the set of annotated patches  $P = \{I, c\}$  randomly extracted from the training images.  $I$  and  $c$  are the intensity of patches and the annotated head pose class labels, respectively. Starting from the root, each tree is built recursively by assigning a binary test  $\phi_{p, q, \tau}(I) \rightarrow \{0, 1\}$  to each non-leaf node. When a patch satisfies the test it is passed to the right child; otherwise, the patch is sent to the left child. Such a test is derived using the Binary Pattern Run Length matrix, which can be calculated by the following steps. First, calculate the binary patterns at  $I(p)$  and  $I(q)$  similar to LBP. Second, construct the Run Length matrices from the binary patterns. Third, calculate the Run Percent from the Run Length matrices. A binary test  $\phi_{p, q, \tau}(I)$  is derived by:

$$\frac{1}{n} \sum_{i=1}^G \sum_{j=1}^R (p_p(i, j | 0) - p_q(i, j | 0)) > \tau, \tag{3}$$

where  $G$  is the number of the binary pattern level,  $R$  is the longest run, and  $n$  is the number of pixels in the image, i.e. in this case  $G = 2$ ,  $R = 8$ ,  $n = 1$ . Fig. 1 shows an example of a Binary Pattern Run Length matrix.

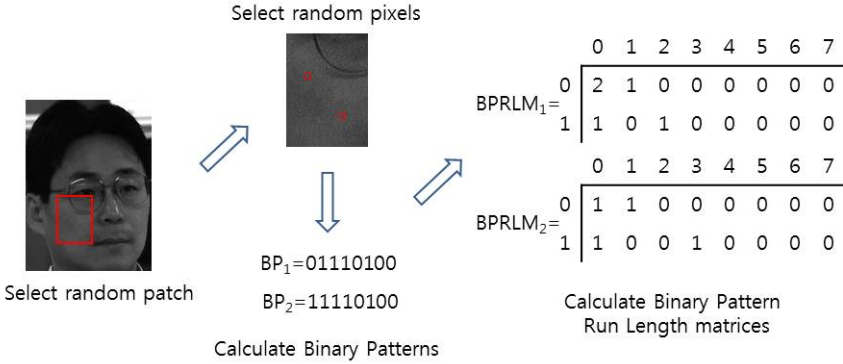


Fig. 1. Example of Binary Pattern Run Length matrices

The best test  $\phi^*$  is chosen from a pool of randomly generated ones. All patches arriving at the node are evaluated by all tests in the pool and compute the information gain which is predefined as follows:

$$\phi^* = Arg \max_{\phi} IG(\phi) \tag{4}$$

$$IG(\phi) = \frac{\sum_{i \in \{L,R\}} (\mu_i - \mu)^2}{\sum_{i \in \{L,R\}} \sum_{j=1}^{n_i} (c_{ij} - \mu_i)^2}, \tag{5}$$

where  $n_i$  and  $\mu_i$  are the number of samples and the mean of class at the child node  $i$ , respectively,  $c_{ij}$  is the head pose class label of the  $j$ -th patch contained in child node  $i$ , and  $\mu$  is the mean of class at the parent node. The information gain  $IG(\phi)$  indicates the ratio of within variance to between variance. The Process continues with the left and the right child using the corresponding training sets  $P_L(\phi^*)$  and  $P_R(\phi^*)$  until a leaf  $l$  is created when either the maximum tree depth is reached or less than a minimum number of training samples are left. Each leaf stores the mode of all class labels together with their covariance.

### 3.2 Testing

Given a new gray image of a head, patches that have the same size as the ones used for training are densely sampled and passed through all trees in the forest. Each patch is guided by the binary tests stored at the nodes. At each node of a tree, the stored

binary test evaluates a patch, sending it either to the right or left child, all the way down until a leaf. Arriving at a leaf, a patch gives an estimate for the pose class and its variance.

Because leaves with a high variance are not very informative and mainly add noise to the estimate, we discard all votes with a variance greater than an empiric threshold  $max_v$ . Finally, we estimate the head pose from the mode of the remaining class labels.

## 4 Experiments

We evaluate the performance of our algorithm based on the CMU Multi-PIE database, which contains more than 750,000 images of 337 people recorded in up to four sessions over the span of five months. Subject were imaged under 15 view points and 19 illumination conditions while displaying a range of facial expressions. In our paper, neutral expression, 19 illumination and 7 view points, which consist of  $0^\circ$ ,  $\pm 15^\circ$ ,  $\pm 30^\circ$ , and  $\pm 45^\circ$ , were employed. All of these face images were cropped to  $48 \times 64$ . Among these images, 50% were used for training and the rest for testing. Fig. 2 shows an example of the CMU multi-PIE data bases.



**Fig. 2.** Example of CMU Multi-PIE databases

In order to compare the performance of the proposed head pose estimation, we employed a combination of several methods. First, the Local Binary Pattern was employed for preprocessing. Second, Principal Component Analysis and Linear Discriminant Analysis were employed for feature extraction. Finally, a Support Vector Machine was employed for the classifiers. Table 1 shows the comparison results of the classification accuracies (CA) of the different algorithms. Because of the illumination change, the results of the LBP image were better than those of the raw image. Furthermore, the proposed method has performance better than that of other methods, about 12% higher than that of LBP+PCA+SVM, and 7% higher than that of LBP+LDA+SVM.

**Table 1.** Comparison of classification accuracies (CA) of different algorithms

Algorithm	Raw image	LBP image
PCA + SVM	64.4%	78.1%
LDA + SVM	70.5%	83.7%
Proposed	90.4%	-

## 5 Conclusion

In this paper we proposed to use a Binary Pattern Run Length matrix based on the random forests method for head pose estimation. In order to make this method robust in terms of illumination, the Binary Pattern Run Length matrix was employed; this matrix combined a Local Binary Pattern and a Run Length matrix. In order to evaluate the discriminative power of the random tree method, a novel information gain was employed. Experiments on public databases show the advantages of this method over other algorithm in terms of accuracy and illumination invariance.

**Acknowledgment.** This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (13-IT-03). It was also supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program (Immersive Game Contents CT Co-Research Center).

## References

1. Murphy-Chutorian, E., Trivedi, M.M.: Head pose estimation in computer vision: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 31(4), 607–626 (2009)
2. Gee, E., Cipolla, R.: Determining the gaze of faces in images. *Image and Vision Computing* 12(10), 639–647 (1994)
3. Li, Y., Wang, S., Ding, X.: Person-independent head pose estimation based on random forest regression. In: *IEEE International Conference on Image Processing*, pp. 1521–1524 (2010)
4. Huang, C., Ai, H., Li, T., Lao, S.: High-performance rotation invariant multiview face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 29(4), 671–686 (2007)
5. Raytchev, B., Toda, I., Sakaue, K.: Head pose estimation by nonlinear manifold learning. In: *IEEE International Conference on Pattern Recognition*, pp. 462–466 (2004)
6. Li, Y., Gong, S., Liddell, H.: Support vector regression and classification based mult-view face detection and recognition. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 300–305 (2000)
7. Fanelli, G., Gall, J., Ban Gool, L.: Real time head pose estimation with random regression forests. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 617–624 (2011)
8. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Facial image analysis based on local binary patterns – A survey. *IEEE Trans. Sys., Man, and Cyber., - Part C* 41(6), 765–781 (2011)
9. Galloway, M.M.: Texture analysis using gray level run lengths. *Computer Graphics and Image Processing* 4(2), 172–179 (1975)
10. Gross, R., Matthews, I., Cohn, J.F., Kanade, T., Baker, S.: Multi-PIE. *Image and Vision Computing* 28(5), 807–813 (2010)



# Triggers That Increase Co-Creation Risks: A Consumer Perspective

Freida Palma and Soon Goo Hong\*

Department of Management Information Systems, College of Business Administration,  
Dong-A University, Bumin Campus, Busan, 602-816, Korea  
palmita@freidapalma.com, shong@dau.ac.kr

**Abstract.** The purpose of this paper is to identify the triggers that increase the likelihood of consumer participation and collaboration in the co-creation of value with producers by focusing on the effects these triggers have on consumers' participation and on a company's reputation, products, and services within a community. The paper utilizes previous literature review to identify the triggers that increase the risks of consumers who co-create value through collaboration with stakeholders and companies. Research to date has illustrated the positive impact of co-creation, but with limited information on the risk factors that cause negative effects on co-creation. This paper discovered six triggers (transparency, dialogue, time and effort, financial compensation, social benefits, and intellectual stimulation) associate with the risks consumers face during co-creation that increases the probability of impeding consumers from co-creating value with producers and other stakeholders. It also illustrates a number of real-world examples to assist future co-creators to reduce the triggers that deter consumers from co-creating with other stakeholders. Foreseeing potential triggers that can hinder such progress which contributes to successfully co-creating value.

**Keywords:** Co-Creation, Triggers, Consumers, Collaborators, Stakeholders, Value co-creation.

## 1 Introduction

Globalization has affected the ways in which businesses function on a daily basis. Resulting in the decline of businesses that are not able to transform and compete in this new global competitive business era. New businesses and companies surface on a daily basis with innovative products and services, while old ones decline because of their inability to be innovative quickly. Successful companies maintain their successes by creating innovations consistently. However, innovation now is not a task that can be solely accomplished by a single company. It requires collaboration and cooperation from both the company and its customers through co-creation. Stern (2011) explained *co-creation* as working on new product and service ideas together

---

\* Corresponding author.

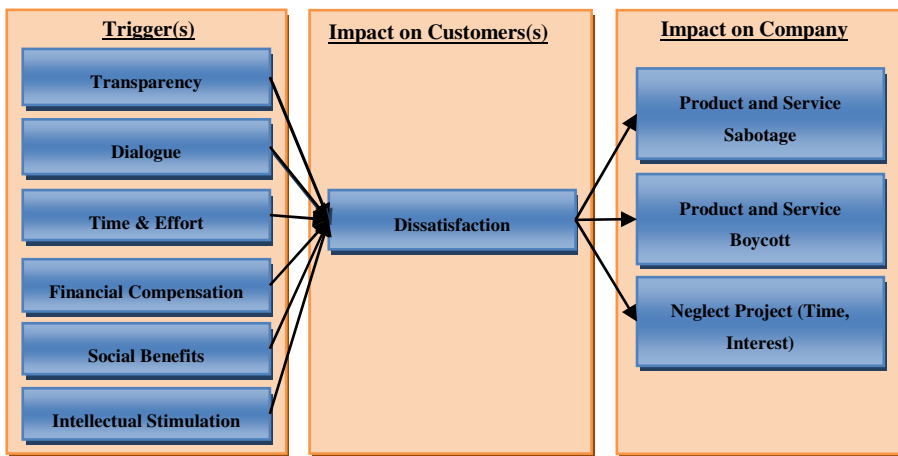
with the customers who will be purchasing them. In addition, Brown and Miller (2011) emphasize, "When you need to transform a brand or product, you can't just do the same things better. You need to do something new. You tap into the creativity of your consumers." Frigo and Ramaswamy (2009) add that companies no longer create value and wealth by themselves alone, rather, customers, suppliers, investors, and others play active roles in the process. Comparatively, the pioneers of co-creation, Prahalad and Ramaswamy (2000), describe co-creation as an emerging relationship between customers and companies. They explain that customers are fundamentally changing the dynamics of the market place, which has become a forum in which consumers play an active role in creating and competing for a company's value. Furthermore, they stated that the change involves the co-creation of value through personalized interactions that are meaningful and sensitive to a specific consumer, and that the co-creation experience is the basis of unique value for each individual. Therefore, co-creation is a vital element used to harness the power of the consumers to collaborate with companies to create value through the innovation of new products and services.

However, with great power comes a tremendous number of risks, from both the consumers' and the producers' perspectives. Thus, far very little is known about what makes a consumer "high-risk," and how to better co-create value with consumers to mitigate those risks. As with every successful negotiation, both parties need to leave the bargaining table feeling a level of satisfaction from the transaction that occurred. Buell (2007) emphasizes that successful bargaining means looking for positives in every possible circumstance. Moreover, they explain that the process of trading-off of issues that results in a positive outcome for both parties is creating value in a transaction. A negotiation should be a win-win situation for both parties involved or a negative outcome will present itself. Hence, this paper aims to provide "triggers," or attributes, that increase the likelihood of consumer co-creation whilst collaborating to create value, and it identifies those triggers related to the risks consumers face in the process of co-creation. Equally, it focuses on reducing consumers risks to maximize consumer participation in value creation by identifying the risks and plausible solutions to minimize them. It also focuses on the impact, or results, the risk factor(s) have on the company's reputation, products, and services within the community after co-creating with the consumer.

The approach to be taken will provide the reader with a conceptual overview of the triggers associated with risk factor(s) that consumers face when they participate with stakeholders and companies to co-create value through collaboration. This paper will focus solely on the consumer perspectives on the risks involved in co-creation and on the ways in which the risks can be mitigated to improve the likelihood of a positive outcome for the co-creation process. This paper is organized as follows. In section 2, we focus on triggers affecting co-creation with consumers, to recognize the reasons why a consumer would face a risk in participating in co-creation activities. In section 3, there is a summary of the paper.

## 2 Triggers Impacting Co-Creation with Consumers

In the co-creation process risks are involved for both consumers and producers. However, this paper focuses solely on the consumer perspectives of the risks involved and the ways in which the risks can be mitigated to have a positive outcome for the co-creation process. Hence, sustainable wealth creation requires balanced risk taking by focusing on co-creation opportunities that can generate superior returns while simultaneously reducing risks for companies and their stakeholders [10]. For this reason, in Figure 1, six triggers are presented along with their impact on consumers in co-creation activities and the after effects on the companies. The six triggers were derived based on previous literature review.



**Fig. 1.** Triggers affecting co-creation when creating value with consumers

The co-creation process can become more fruitful from the consumers perspective through mitigating triggers such as: transparency, dialogue, time and effort, financial compensation, social benefits, and intellectual stimulation. In order to trigger change by innovation, co-creation needs to be implemented as disruptively as necessary and as nondisruptively as possible [7]. Hoyer et al. (2010) state that through risk reduction, consumers will be more motivated to participated in co-creation processes, and that moderation is a vital concern when it comes to risk. Risk, if not properly balanced, can diminish control over strategic planning and increase the complexity of managing a firm's objectives, its misperformance, and its selection of consumers' ideas. Hence, it is vital for producers to manage risk return in co-creation, particularly for consumers, to create a win-win environment.

Figure 1 illustrates that if any one of the six triggers occurs at a given time or at the same time, it could possibly lead to a negative outcome for the consumer, which, in this case, is total or partial dissatisfaction with the co-creation partnership. This can result in either (1) product and service sabotage or boycott or (2) project neglect. Gebauer et al. (2012) add that there is a dark and a bright side of co-creation; they

further conducted an empirical study to prove their findings. They found that perceived unfairness and dissatisfaction with the outcome could cause negative reactions of participants such as negative word of mouth. In addition, other studies on user misbehavior in offline settings showed that both dissatisfaction and perceived unfairness are among the most relevant triggers for dysfunctional user behaviors [2]. In addition, Fisk et al. (2010) discussed the various economic, situational, or personal as well as cognitive and emotional reasons that may trigger such behaviors. Perceived injustice as well as dissatisfaction with a company's action and offerings may unleash customer misbehavior [16]. Once the consumer feels dissatisfied from a transaction in the activity it can result in a win-lose situation, which could lead to the failure of a project. In this case, it is vital for both producers and consumers to leave the table feeling satisfied from a negotiation to improve the likelihood of success in a project: particularly, because the idea behind co-creation is to create wealth through active collaboration with people that share the same vision. In order to collaborate successfully with a consumer in a co-creation activity, all six triggers, which will be discussed in the subsections, are required as they play a vital part in the success of any future projects.

## 2.1 Transparency

In any given transaction, transparency is vital when establishing and maintaining trust. This trigger is an important one for the success of any co-creation project. In the ideal world, having access to the necessary data to complete a project would make everyone's life easier. Prahalad and Ramaswamy (2004) define transparency, in the four building blocks of co-creation, as information that is necessary to create trust between institutions and individuals.

One of the main problems that arise from transparency is that producers are not forthcoming and clear with information. This can result in a loss of trust in the producers, from the consumers' perspectives, which can lead to customer dissatisfaction. Brown (2009) stated that "while experts are very good at fueling thinking, consumers bring value by grounding that thinking." For this reason, it is essential to establish trust within the producers-to-consumers relationships to successfully co-create value. To accomplish this task, the flow of information is necessary for consumers to make decisions, particularly, because their access to unprecedented amounts of information, knowledgeable consumers can make more informed decisions [20].

In addition, to the availability of information that is necessary for consumers to make appropriate decisions, information should also be clear and understood by both the producers and consumers for transparency. Unclear information can lead to confusion and can result in consumers networking to influence a company's product or service in a negative aspect. Gebauer et al. (2012) provided an example of this through research done on the company, SPAR Austria, which carried out a SPAR bag design contest. This was an international online design contest for shopping bags with the participation of nearly 2500 community members. The result of it was not a positive one. The company received negative reactions from the members after the winner was selected. This occurred because the members did not agree and were not satisfied with the outcome of the contest. Even though, the jury selected the winner

based on predefined criteria's. The process used to select the winner was not clear or transparent to the consumers, and, more importantly, the consumers were not involved in this process. Another example to illustrate the importance of transparency within a company occurred when Kraft conducted an idea contest to choose a new name for its new cheese-vegemite spread. They received over 40,000 entries but were forced to abandon the original name because of the community dissatisfaction and the negative reaction that was received by *isnack 2.0* [28]. As a result, consumer influence can do more harm than good when a consumer is dissatisfied with a decision, especially because consumers exercise their influence through social medias such as Facebook, Twitter, and many others. Consequently, it is crucial that when co-creating with consumers, producers exercise transparency.

## 2.2 Dialogue

Communication is imperative when conducting business, as Prahalad and Ramaswamy (2004) state, "dialogue encourages not just knowledge sharing but even more important, qualitatively new levels of understanding between companies and consumers. It also allows consumers to interject their views of value into the value-creation process." Dialogue is an important element in the co-creation view. Markets can be viewed as a set of conversations between the customer and the firm [15]. In addition, dialogue involves two parties who are willing to listen, understand, compromise, and communicate effectively with each other despite the circumstances.

Effective dialogue will result in the success of any endeavor. Nevertheless, it is a challenging that both producers and consumers must overcome to attain success in co-creating value. Dialogue can be a difficult to accomplish, particularly, if the consumers do not have access to necessary data or information required to make suitable decisions. In order for consumers and producers to co-create value effectively, there must be a clear line of communication along with the transparency of information. In addition, consumers should be treated as equal partners in such endeavors, to foster understanding between the customers and the companies. This provides customers with the opportunity to provide their feedback in the creative process.

Many of the problems associated with dialogue occur because the consumers involved are not treated as equal partners, and companies are reluctant to release control, which is a common human tendency. In most cases, they perceive giving up control as weakness. Since it is deemed a privilege to have information, information is being held back from consumers, thus reducing the probability that consumers would be able to make informed decisions that could benefit the company. In addition, ineffective dialogue leads to more negative consequences than positive, because it reduces the likelihood of the company being able to understand the consumers' work style, and emotional, social, and cultural values, which form who the consumer is and the ways in which he/she/they can contribute to the company's value. Communication today is like globalization: without communication, innovation, which is a necessary tool for the survival of any given business, is impossible. An example of communication failure can be found in the music industry. Now, companies are fighting against "illegal downloading," while avoiding the inevitable, which is

modifying their business models. This has occurred because those in the music industry are not listening to the customers who are now interested in creating their own personal musical experience. Another example is with Macintosh and Apple. Prahalad and Ramaswamy (2002) state, "Dialogue was what kept a loyal community of Macintosh users together when Apple Computer Inc.'s product development began to wane. And it is dialogue that is helping the personal-computer manufacture to recover with the introduction of the new iMac." Therefore, dialogue is an important element for successful co-creation of activities within companies.

### 2.3 Time and Effort

Another major trigger that increases dissatisfaction among consumers is time and effort. The time that consumers spend is expected to produce fruitful results. Co-creation products are often shown to possess high expected benefits and novelty, which ultimately increases commercial attractiveness [9]. On the contrary, when fruitful results are not obtained, the customers are left feeling dissatisfied, as their hopes and dreams have been diminished; particularly, because of the investment of time, which is a huge risk when there is complete failure of a co-creation project. Most co-creation projects, bringing new innovations to life, fail because of the inability to adequately assess and fulfill consumer needs [18].

An example of this is the case where Coca Cola crowd sourced blind taste tests that led them to bring in the "New Coke" while trying to dispose of their main brand, which is a national icon. The idea was developed through the public's input; however, Coke took it too far. Erasing tradition was something that could not have been done, leading Coke to eventually bring back their "Classic Coke" and sell both products [6]. Another example would be with Skittles, who drastically changed their website from a Twitter to a Facebook fan page because of the negative feedback that was posted by critics. For the true fans who had contributed positively, their original posts were ignored because they were overshadowed by the negative comments that were detrimental to Skittles' reputation [23]. Another case would be Doritos and Pepsi MAX, who received over 5600 submissions, where the winner would receive a one-million-dollar top prize and a total payout of five million dollars, an advertising contract, and airtime during Super Bowl XLV. The winning ad, "Feed the Flock," had over 100,000 views on YouTube. However, the winning ad, because it showed a parish where attendance was dropping and in response the priest decides to distribute Doritos and Pepsi as Holy Communion to bring in the people, the ad created a lot of controversy among Catholic activists who considered the commercial "terribly blasphemous" [6]. Because of the complaints received, the ad still would not be used after all the effort and time put in by the consumer.

As a result, reducing the risk of consumers wasting their time and effort on a co-creation project will greatly motivate consumers to participate in co-creation activities. Hoyer et al. (2010) state that firms can also stimulate co-creation by reducing the costs, time, effort, and foregone opportunities for consumers participating in co-creation. Von Hippel and Katz (2002) looked at one approach to reducing consumer cost, providing user kits, which help in easing the process of creating new ideas, products, and marketing materials for potential participants.

## 2.4 Financial Compensation

Financial compensation is another major trigger that leads to consumers experiencing dissatisfaction. Many individuals experience financial dissatisfaction when they believe they are not properly compensated for the work, time, and expertise that they provide to a task. Therefore, it is vital before embarking on any co-creation activity to establish if there will be any monetary involvement, and, if there is, to establish clearly the compensation amount. Brown (2009) explains that setting the right tone for co-creation means being completely clear with all parties with respect to the business aims of the project. He further states, “Collaborators should understand that challenge and be given a clear explanation of your reasons for involving them, the part they will play, and what they are ultimately hoping to achieve.”

Hoyer et al. (2010) described co-creation from a consumer perspective as monetary and nonmonetary costs of the time, resources, and physical and psychological effort to learn and participate in the co-creation process. For those consumers who are motivated by financial compensation, not receiving what falls within their expectation range will yield dissatisfaction on the consumers’ part. Some co-creating consumers are motivated by financial rewards, either directly in the form of monetary prizes or profit sharing from the firm that engages in co-creation with them, or indirectly, through the intellectual property that they might receive or through the visibility that they might receive from engaging in co-creation competitions [14]. Consequently, reducing this risk of financial dissatisfaction through being transparent from the initial start of the project can assist in increasing the success of co-creation.

## 2.5 Social Benefits

Similar to those who participate in co-creation activities because of the financial rewards are those who participate in co-creation to uplift their social status within their communities to gain recognition and increase their contact list. Many others are not simply motivated by money; they choose to “free reveal” ideas and freely share their efforts in the postideation stages of co-creation [14]. Many people prefer to be recognized for the work they have done through awards, titles, promotions, and many others. Nambisan and Baron (2009) explain that some may receive social benefits from titles or other forms of recognition that a firm might bestow on particularly valuable contributors. Social benefits of co-creation comprise increased status, social esteem, “good citizenship,” and strengthening of ties with relevant others. If customers feel that they have not been appropriately recognized for the work they have done, whether internally or externally, they are left feeling emotionally cheated, which results in dissatisfaction.

An example of this would be Netflix, a famous way to watch movies and TV shows online. It crowd sources its research and development (R&D) by providing one million dollars to anyone who could improve their recommendation engine. As a result, they ended up with a winning team of researchers in 2009, who were awarded one million dollars. However, the solution presented was never used by Netflix because, according to Netflix, the “additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production

environment” [13]. The researchers who completed this project and became the winners, because their gratification was linked to Netflix actually using their work, left feeling a level of dissatisfaction. Other examples would be Starbucks and Dell, which allowed their users to provide ideas but not fully interact, much like the case of Adobe’s [ideas.adobe.com](http://ideas.adobe.com). This platform provides users with the ability to submit ideas and receive feedback for their ideas. Further, the engineering staff made sure that their feature lists were aligned with the users’ requests. The author states that dozens of ideas have been implemented and more are in the development pipeline [12]. This is vital for consumers, particularly, because they feel a level of satisfaction in seeing their contribution acknowledged.

Consumers want to and will continue to participate in co-creation if they feel that they make a difference, and, most importantly, if they are recognized for the difference they have made, whether big or small. Brown (2009) states that most people are likely to feel that they have made a positive difference every now and then, no matter how small. In co-creation, the people who are truly driven by this desire on a day-to-day basis are invaluable. This is because in the co-creative approach, both parties benefit, one from the desire and the other by feeding that desire, ensuring a level of engagement from collaborators that no financial incentive could ever achieve.

## **2.6 Intellectual Stimulation**

Another major factor influencing the success of co-creation projects is intellectual stimulation. People want to participate in a co-creation project to learn and grow as an individual by gaining knowledge and experience. Nambisan and Baron (2009) claim that “others might be motivated by a desire to gain technology knowledge by participating in forums and development groups run by the manufacturer. Co-creators might reap important cognitive benefits of information acquisition and learning.” If consumers leave the activity feeling as if it there was not any “give and take” in the project, they will depart the project feeling a level of dissatisfaction. Some people participate in an activity so that they can interact and learn more to improve their personal growth.

An example is with Michelle Obama’s Twitter Q&A for “Let’s Move!” the main purpose of which was to address obesity in America. However, it drew unrelated negative feedback from President Obama’s critics, and failed to fulfill the true purpose for which it was intended. The experience may have left the contributors of suggestions and ideas to address childhood obesity in America feeling a level of dissatisfaction from a knowledge perspective [30]. The same situation occurred with Barack Obama when he tried to collect questions through crowd sourcing for a press conference on jobs, energy reform, and health care. His website did not serve its intended purpose for those who were serious about the press conference; instead, enthusiast took over the page to discuss legalizing marijuana.

Therefore, it is vital for producers to understand that consumer-to-consumer communication provides consumers with an alternative source of information and perspective [21]. In addition, it is important to be transparent from the initial stages of the project to allow innovators the freedom of creativity and expression through their inventions with a team working toward the same goal. Communicating a clear



ambition encourages collaborators to trust you and gives them a sense of purpose in their work [3].

### 3 Conclusion

In conclusion, through the utilization of the six triggers (transparency, dialogue, time and effort, financial compensation, social benefits, and intellectual stimulation) co-creation risks can be minimized, yielding a more positive reaction to the co-creation process and thus encouraging more consumers to participate actively. To avoid failure, it is important to understand the risks consumers are taking and ways to manage them more effectively. With consumers being an integral part of the system for value creation, they can influence, where, when, and how value is generated through networking with other consumers like themselves whom have similar interest. Mitigating these triggers can produce more fruitful results from co-creation activities. Further, they can reduce consumers' attempts to damage an organizations reputation, products, or service within a community, particularly, through networking with other consumers and negatively influencing them.

This paper provided a consumers' perspective on the triggers associated with risk factor(s) faced by consumers who co-create value with producers. It provided an insight on the risks involved and the ways in which the risks can be mitigated to produce more positive outcomes for the co-creation process to decrease the probability of impeding consumers in co-creating value with producers and other stakeholders. By identifying those risks that exist, producers will be able to form the proper course of action to have consumers more willingly participate in co-creation activities. In addition, a number of real-world examples were used to assist future co-creators in reducing the triggers associated with the risks factors that deter consumers from co-creating with other stakeholders. Another key point was the social aspect of co-creation, which requires a network of skilled individuals to collaborate on a project from start to completion. This allows for the open innovation of new products and services that would best serve a specific target audience in the community, and improves the overall value for all collaborators involved. Working with others and keeping good communication from the conception phase is vital for transparency and decision making in large groups. Avoiding these potential triggers that can hinder progress helps in co-creating value.

This paper has summarized the six triggers that lead to decreasing consumer participation in co-creation, from the perspective of consumers. Producers being able to fully recognize and identify these triggers will be able to reduce or prevent their negative effect on consumers and thus achieve success within their co-creation endeavors. Further, this paper serves as additional research material on the triggers that lead to consumer risk, particularly consumer dissatisfaction, which is a common denominator used to determine whether a consumer will or will not participate in co-creation. The limitation of this paper is the lack of data to support effectively the triggers presented. Consequently, future research should focus on an empirical study to support our findings on the triggers associated with co-creation risk. Proving further information about these triggers will contribute to the academic research currently available on co-creation risk.

## References

1. Producers. Wikipedia (2012)
2. Blodgett, J.G., Granbois, D.H.: Toward an Integrated Conceptual Model of Consumer Complaining Behaviour. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour* (5), 93–103 (1992)
3. Brown, J.: The Spirit of Co-Creation Risk-Managed Creativity for Business. *Sense Worldwide* (2009)
4. Buell, B.: Negotiation Strategy: Six Common Pitfalls to Avoid. *Stanford Business* (2007)
5. Chafkin, M.: The Customer is the Company Inc. *Magazine* (2008)
6. Cooper, V.: Crowdsourcing's eighth deadly sin! Don't mess with Pepsi's Super Bowl ad campaign! (2011)
7. Cruz-Valdivieso, E., Humphreys, P., Roser, T., Samson, A.: Co-Creation New Pathways to value an overview. *Promise Corporation* (2009)
8. Fisk, R.S., Grove, S., Harris, L.C., Keefe, D., Reynolds, K.L., Russell-Bennett, R., Witz, J.: Customers behaving badly: A state of art review, research agenda and implications for practitioners. *Journal of Services Marketing* 4(6), 417–729 (2010)
9. Franke, N., Von Hippel, E., Schreier, M.: Finding Commercially Attractive User Innovations: A test of Lead-User Theory. *Journal of Product Innovation Management* 23(4), 301–315 (2006)
10. Frigo, M.L., Ramaswamy, V.: Co-Creating Strategic Risk-Return Management. *Strategic Finance* (2009)
11. Gebauer, J., Fuller, J., Pessei, R.: The Dark and the Bright Side of Co-Creation: Triggers of member behaviour in online innovation communities. *Journal of Business Research* (2012)
12. Gouillart, F.: Why Starbucks and Dell get the wrong ideas (2009), <http://francisgouillart.com/wordpress/?p=587>
13. Holiday, R.: What the Failed \$1M Netflix Prize Says About Business Advice (2012)
14. Hoyer, W.D., Chandy, R., Dorotic, M., Krafft, M., Singh, S.S.: Consumer Cocreation in New Product Development. *Journal of Service Research* 13(3), 283–296 (2010)
15. Levine, R., Locke, C., Searls, D., Weinberger, D.: *The Cluetrain Manifesto: The End of Business as Usual*. Perseus Publishing, Cambridge (2001)
16. Lovelock, C.: Complaint Handling and Service Recovery. In: Lovelock, C., Witz, J. (eds.) *Services Marketing: People Technology, Strategy*. Pearson Education, Upper Saddle River (2010)
17. Nambisan, S., Baron, R.A.: Virtual Customer Environments: Testing a Model of Voluntary Participation in Value Co-Creation Activities. *The Journal of Product Innovation Management* 26(4), 388–406 (2009)
18. Ogawa, S., Piller, F.: Reducing the Risks of New Product Development. *Sloan Management Review* 47, 65–72 (2006)
19. Prahalad, C.K., Ramaswamy, V.: The Co-Creation Connection. *Strategy + Business* (27) (2002)
20. Prahalad, C.K., Ramaswamy, V.: Co-Creation Experiences: The Next Practice in Value Creation. *Journal of Interactive Marketing* 1 (2004)
21. Prahalad, C.K., Ramaswamy, V.: *The Future of Competition*. Harvard Business School Publishing (2004)
22. Prahalad, C.K., Ramaswamy, V.: Co-opting Customer Competence 78, 79 (2000)
23. Schroeder, S.: Skittles Swaps Homepage from Twitter Search to Facebook Page (2009)

24. Sheth, J., Sisodia, R.S., Sharma, A.: The Antecedents and Consequences of Customer-Centric Marketing. *Journal of the Academy of Marketing Science* 28, 55–66 (2000)
25. Stern, S.: *A Co-creation Primer* (2011)
26. Tapscott, D., Williams, A.D.: *Macro Wikinomics: Rebooting Business and the World*. Penguin Group (2010)
27. Vargo, S.L., Lusch, R.F.: Evolving to a New Dominant Logic for Marketing. *Journal of Marketing* 68 (2004)
28. Vasek, L.: Kraft renames iSnack 2.0 - Cheesybite (2009),  
<http://www.news.com.au/business/kraft-renames-isnack-20-cheesybite/story-e6frfm1i-1225783684238#ixzz2M3gAnM1U>
29. VonHippel, E., Katz, R.: Shifting innovation to users via toolkits. *Management Science* 48(7), 821–833 (2002)
30. Young, D.: Michelle Obama's Twitter Q&A For 'Let's Move!' To Miss The Mark (2013)

# A New Lightweight Protection Method against Impersonation Attack on SIP

Yoon Mi Koh and Kyung Hee Kwon

Department of Computer Science, Dankook University, Korea  
{alice8105, khkwon}@dankook.ac.kr

**Abstract.** Impersonation attack, based on vulnerable authentication of SIP, facilitates an attacker to take malicious actions such as toll fraud and session hijacking. In this paper, we propose a new lightweight protection mechanism against above actions. As soon as an attacker sends a REGISTER message to SIP registrar, it creates a record called *binding* in a location server. At this point, our proposing mechanism can detect the attacker by just comparing *binding* with the REGISTER message received from the attacker. Our protection method can easily enhance user authentication of SIP with little overhead.

**Keywords:** SIP, Impersonation Attacks, Registrar, Location Server, Binding.

## 1 Introduction

The security of VoIP communications became gradually important as the most of private phone conversations are carried out on insecure IP networks [1][2][3]. The most severe threat in VoIP environment is probably the easy access to communication channel due to the weak authentication scheme of SIP.

While there are many approaches in SIP authentication, they can be roughly categorized into two methods. One is HTTP digest which SIP just relies on default. The other method is run by powerful encryption techniques which is commonly used in IPsec, TLS and S/MIME[4]. Even though HTTP digest can easily suffer from password hacking [5] due to the lack of securing all headers and parameters, it is most favoured and used method as SIP authentication because others can suffer from serious service delay due to big overhead[6].

For this reason, an attacker can easily intercept and forge the SIP packets to disguise as a legitimate user [7]. Attack scenarios may be as follows. Initially, an attacker may sniff the REGISTER message of a legitimate user and send a forged REGISTER message with the same header fields of To, From, and password as the legitimate user to impersonate. The registrar, receiving the forged message, then may recognize the attacker as a legitimate user and create a *binding* that associates SIP URI with contact address [8] in a location server. Then, the attacker will be able to receive normal service, hijack a session between legitimate users and even avoid billing.

In this paper, we propose a strong and lightweight mechanism to detect and block impersonation attack with little overhead.

## 2 Proposed Mechanism

In order for all of SIP UA to join SIP communication system, they must be registered at registrar. After the procedure of registration shown in Fig.1, a *binding* is supposed to be created in a location server. After initial registration, SIP UA periodically keeps registering to maintain the *binding* in a location server. At this point, an attacker may send a forged REGISTER message at the registrar and impersonate as a legitimate user.

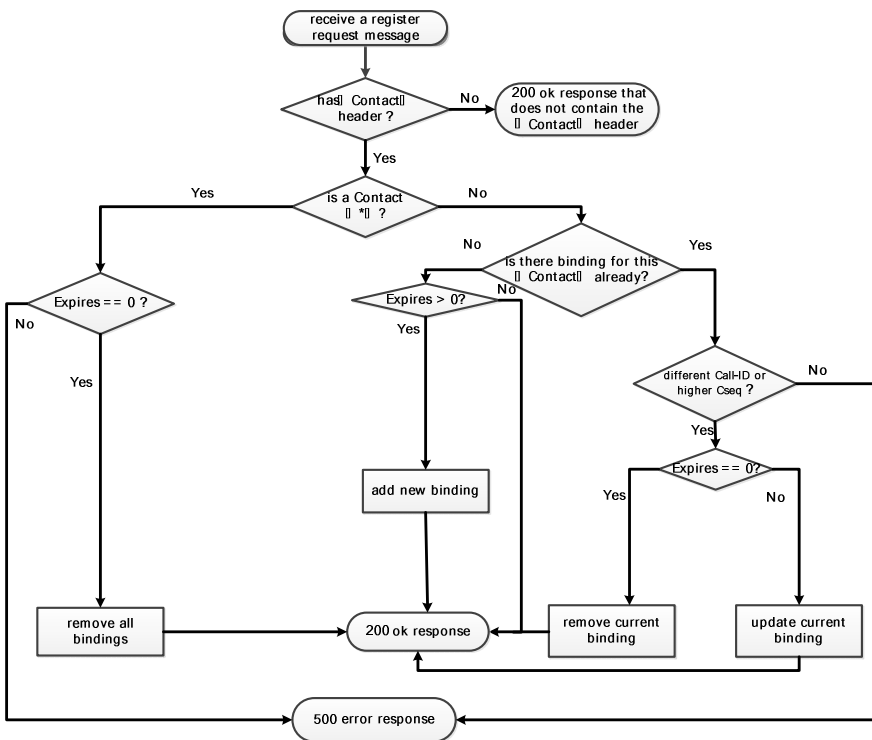


Fig. 1. Procedure of registration at registrar

While this impersonated attempt also creates a *binding* in a location server, we can detect the forged message by comparing the *binding* with contact address that is the value of *contact* field in a REGISTER message received from the attacker as shown in the dotted circle in Fig. 2. Here, we need to notice that IP address of callee is included in the contact address, and it can be altered even in case the legitimate user moves.

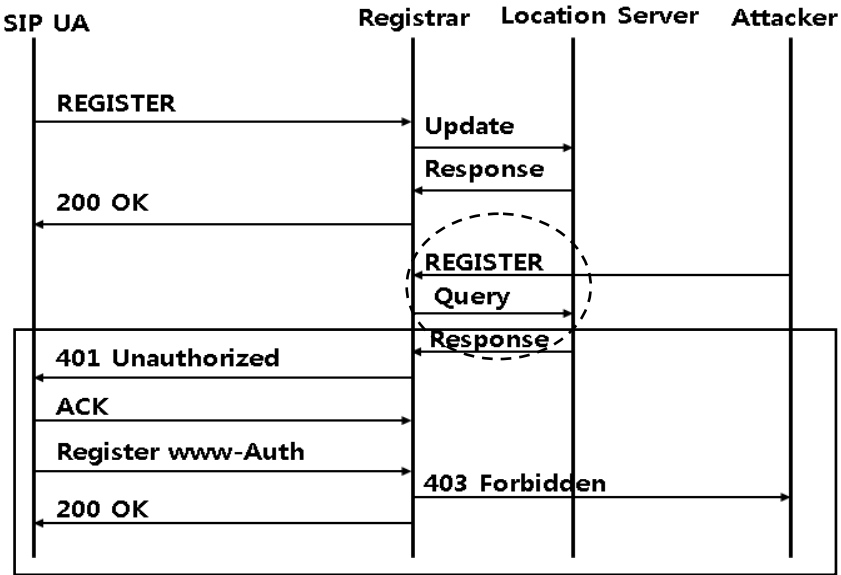


Fig. 2. Process to detect and block impersonation attack

So to speak, the contact address is specified in both the *binding* in location server and the *contact* field in a REGISTER message. Our proposed mechanism detects the impersonation attack by checking if these two values of the contact address are same or different. If two values are same, it shall mean no attack is detected.

Even though two values are different, there are three cases of legitimate registration as follows:

- Case1: To restart SIP UA by legitimate user.
- Case2: To start another SIP UA by legitimate user.
- Case3: To change contact address by legitimate user.

Any attempt of registration to have two different values of the contact address can be recognized as an impersonation attack except above three cases.

When the registrar detects a suspected impersonation attacks, it transmits '401 unauthorized message' to legitimate UA of which the contact address is stored in a location server. If the *binding* of legitimate UA is kept in a location server, the legitimate UA sends a REGISTER message including Auth headers as shown in the rectangle in Fig.2. After then, it sends '403 forbidden message' to the attacker. Therefore, we can detect and block an impersonation attack with little overhead. If the binding of legitimate UA is not kept in a location server, the legitimate UA can not send a response message because it is turned off. Therefore, the registrar should wait for the acknowledgement message of '401 unauthorized message'. If the waiting time exceeds four seconds, that REGISTER message can be recognized to be sent by legitimate UA as shown in the rectangle of Fig. 3.

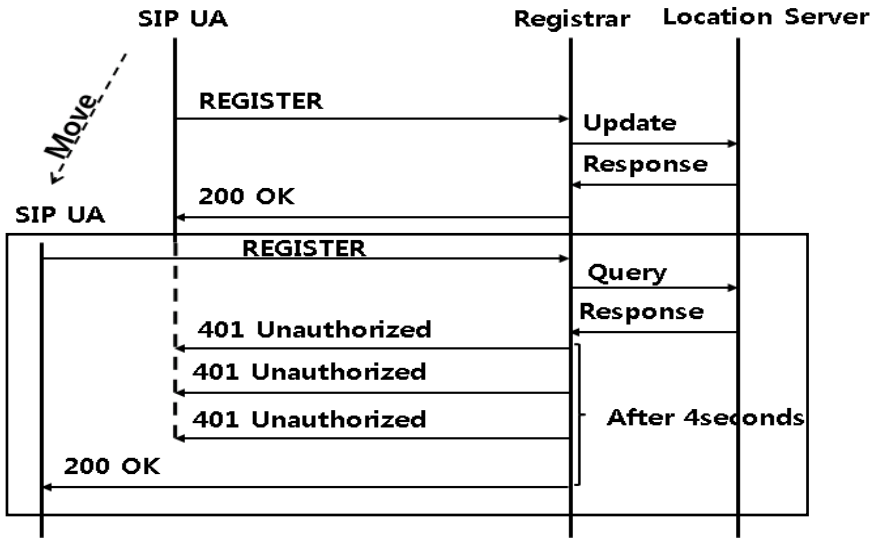


Fig. 3. Case that response messages are not transmitted

### 3 Conclusion

VoIP has rapidly come into wide use because of the convenience and low communication costs. However, impersonation attacks may have fatal effect on the VoIP industry due to the vulnerability of SIP authentication. In this paper, we proposed a new lightweight mechanism to detect and block an impersonation attack very easily without any additional user authentication or encryption technique.

### References

1. Keromytis, A.D.: A Comprehensive Survey of Voice over IP Security Research. *IEEE Communications Surveys & Tutorials* 2, 514–537 (2012)
2. Keromytis, A.D.: A Look at VoIP vulnerabilities. *Login* 1, 41–50 (2010)
3. Angelos, D.: Keromytis: Voice over IP Security: Research and Practice. *IEEE Security Privacy* 2, 76–78 (2010)
4. Go, Y.-M., Kwon, K.-H.: Countermeasure of SIP Impersonation Attack Using A Location Server. *Journal of the Korea Contents Association* 4, 17–22 (2013)
5. Yoon, E.-J., Yoo, K.-Y., Kim, C., Hong, Y.-S., Jo, M., Chen, H.-H.: A Secure and Efficient SIP Authentication Scheme for Converged VoIP Networks. *Computer Communications* 14, 1674–1681 (2010)

6. Go, Y.-M., Kwon, K.-H.: Expanding the User Authentication Scheme in SIP. *Journal of the Korea Contents Association* 12, 88–93 (2013)
7. Geneiatakis, D., Kambourakis, G., Dagiuklas, T.: SIP Security Mechanisms: A state-of-the-art review. In: *Proc. 5th Int. Network Conference*, pp. 147–155 (2005)
8. Rosenberg, Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.: SIP: Session Initiation Protocol. RFC 3261, Internet Engineering Task Force (2002)



# Measuring Syntactic Sugar Usage in Programming Languages: An Empirical Study of C# and Java Projects

Donghoon Kim<sup>1</sup> and Gangman Yi<sup>2,\*</sup>

<sup>1</sup> Department of Computer Science, North Carolina State University, Raleigh, NC, USA  
dkim2@ncsu.edu

<sup>2</sup> Department of Computer Science & Engineering, Gangneung-Wonju National University,  
Wonju, South Korea  
gangman@cs.gwnu.ac.kr

**Abstract.** Syntactic sugar is introduced to existing programming languages to improve their readability and brevity. There have been many debates about pros and cons of using syntactic sugar. However, it is now an integral part of programming languages. No existing work studies the usage of syntactic sugar in real-world software development. In order to fill this gap, we conducted the first empirical study to examine the usage of syntactic sugar in 20 open source software projects written in either Java or C#. Our study results show that syntactic sugar is generally used more than an corresponding feature in the wild, but there may be a limit to what syntactic sugar cannot replace an corresponding feature completely. In this paper, we make several suggestions as to why this occurs.

**Keywords:** empirical study, syntactic sugar, C#, Java, static analysis.

## 1 Introduction

Syntactic sugar extends existing programming languages by allowing concise syntax of frequently occurring language usage patterns. An example of syntactic sugar is the `foreach` feature. The `foreach` allows a developer to iterate every element in a collection, which is a common usage pattern for `for`-loop, in a more concise manner than using `for` statement. Moreover, syntactic sugar improves the readability of source codes. Also taking `foreach` as an example, understanding a `foreach` loop is much easier than the `for`-loop since one has no need to interpret the explicit loop conditions.

However, everyone does not have the same opinion of using syntactic sugar. There are criticisms in both academia and industry. One famous saying goes "Syntactic sugar causes cancer of the semicolon" in Epigrams on Programming by Allen Perlis [6]. The great computer scientist alludes that the excessive pursuing of convenience blurs some usage of programming languages. We also found criticisms raised by industrial developers on popular online forums<sup>1</sup>. The developers used expressions

---

\* Corresponding author.

<sup>1</sup> For example, see <http://stackoverflow.com/>

such as "affect the formal structure of a language" and "interact in an unpleasant way" to describe the drawbacks of using syntactic sugar. Although we do not have a formalized survey about developers' attitude toward syntactic sugar, we do believe there is a lot of debating concerning whether or not using syntactic sugar is beneficial.

However, to the best of our knowledge, there is no existing study about how syntactic sugar is used in real software projects. To fill this gap, we conducted an exploratory study on 20 open source projects written in either Java or C#. We found that syntactic sugar is generally accepted, but there may be a limit depending on languages. We found that introducing new syntactic sugars to existing projects does not negatively affect software quality. Taking `foreach` (enhanced `for`) as an example, developers use it even more frequently than ordinary `for` statement.

This paper made the following contributions:

- An empirical study of syntactic sugars in real world open source projects
- Evidence showing that syntactic sugars are generally used, but there may be a limit depending on programming languages
- Enumeration of several causes that not all developers welcome syntactic sugar.

## 2 Background

In this study, we selected several syntactic sugars which are used in C# and Java programming language because the two programming languages have several similarities, but also important differences [7].

C# and Java are similar from a developer's point of view since both languages have very similar syntax; the syntax at the statement and expression level is almost identical, but there are some minor differences in how syntactic sugars are implemented. For example, the `foreach` feature is for traversing items in a collection or an array, which avoids potential off-by-one-errors and makes code simpler to read compared a standard `for` statement [2]. However, C# and Java have relatively different characteristics in terms of the syntax of `foreach`; Both languages introduced `foreach` feature at different times: C# in 2003 as part of .NET 1.1 and Java in 2004 as part of J2SE 5.0. C# uses `foreach` keyword while Java uses `for` keyword which is different syntax from `for` statement [1], [3]. Another example is `using` feature in C#, which makes it possible to automatically release the memory (resources) used to store objects that are no longer required. This feature can be done more concise and clean instead of using `try-finally` with `dispose()` statement [8]. Java also introduced the same feature, `try-with-resources` in Java 1.7 (Java SE 7) in July 2011. Instead of introducing a new keyword as C# `using`, Java allows developers to declare resources that are part of the `try` block so Java developers do not need to add `close()` in `finally` block since the resources developers defined in `try` block close automatically after the execution of the `try` block. Thus, these syntactic sugar features provide not only concise syntax but also improved programming quality. Along with above syntactic sugar, we analyze one more `lock` feature in C# for this study.

## 3 The Study

### 3.1 Research Questions

In order to understand how syntactic sugar is used in the wild, we formulated two fundamental research questions:

- **RQ1:** Will syntactic sugar be used more than its corresponding features?
- **RQ2:** Will project members broadly use syntactic sugar after introduction into the project?

### 3.2 Methodology

We analyzed 20 open source projects to answer the research questions in previous section. We selected C# and Java projects that were the "most used", according to Ohloh.net. Table 1 displays the name of each project with short name inside brackets and lines of code (LOC) in the project measured by ohloh.net on April 2012.

We used our existing programming language analysis framework, written in C#, Java and Python [5]. Our framework analyzes the code using the following procedure: download the full history of each project from a remote development repository using Git or Subversion to a local machine; check out every version of every file from a project's repository, store the different file revisions in an intermediate format and transfer this information to a database; extract the information of language features from each file revision and populate the database server with this information; and analyze the data in the database about each research question.

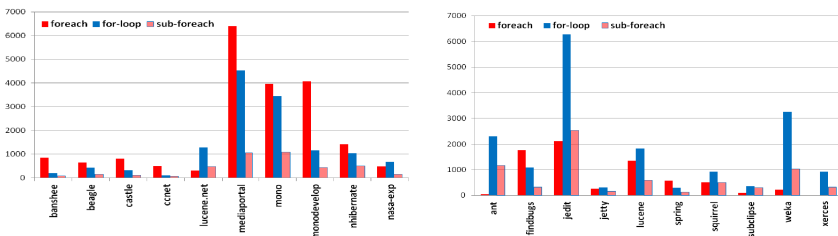
**Table 1.** The 20 projects under investigation

Language	Project Name	LOC
C#	Banshee (banshee)	130,440
	Beagle (beagle)	174,611
	Castle (Castle)	235,487
	CruiseControl.NET (ccnet)	175,042
	Lucene.Net (lucene.net)	154,984
	MediaPortal (mediaportal3)	592,214
	Mono (mono)	3,125,097
	MonoDevelop (monodevelop)	164,710
	NHibernate (nhibernate)	292,379
	Worldwind (nasa-exp)	417,803
Java	Apache Ant (ant)	128,878
	FindBugs (findbugs)	205,571
	jEdit (jedit)	211,019
	Jetty - Java HTTP Servlet Server (jetty)	541,816
	Apache Lucene (lucene)	370,584
	Spring Framework (spring)	741,441
	Squirrel SQL Client (squirrel)	619,249
	Subclipse (subclipse)	94,717
	WEKA (weka)	375,618
Apache Xerces2 J (xerces)	139,051	

## 4 Results

### 4.1 RQ1: Will Syntactic Sugar Be Used More?

Our first research question is whether or not syntactic sugar is used more than a corresponding language feature. To test RQ1, we measured the number of both syntactic sugar and a corresponding language feature to observe how syntactic sugar is used. Figure 1 shows the usage of `foreach`, `for-loop` and `sub-foreach`. The `sub-foreach` denotes a `for-loop` which can be converted to `foreach` such as a substitute. In C# projects (on the left of Figure 1), eight out of 10 C# projects used `foreach` rather than `for-loop`. In contrast, two out of 10 Java projects have more `foreach` than `for-loop` (on the right of Figure 1). One Java project (*xerces*) never used `foreach` at all. The average percentage of `sub-foreach` out of `for-loop` is about 35.5% in C# projects, while 45.2% is the average in Java projects. Figure 2 shows the number of `foreach` and `for-loop` over time for C# projects and Java projects. C# projects show a steady increase in the number of both `foreach` and `for-loop` such as *mediaportal* and *mono* projects on the left in Figure 2. We have not found any conversion from `for-loop` to `foreach` in C# project since `foreach` was added in C# 1.2 in December 2002 right after C# was launched in December 2001. However, we have found that three Java projects show a conversion effort such as *findbugs* and *subclipse* in August 2005 and August 2011, respectively on the right in Figure 2. We note that Java projects have not embraced `foreach` widely although almost half of Java `for-loop` could be converted to a `foreach` statement.



**Fig. 1.** The usage of `foreach` vs. `for-loop` in C# projects (left) and Java projects (right)

We analyzed two more languages such as `lock` and `using` in C#. We found that C# projects have more `lock` than `try-finally` with `Monitor.Exit()`. Also, we found that all C# projects have more `using` than `try-finally` with `dispose()`. Java introduced the `try-with-resources` feature which is the same as `using` in C# was added in Java 1.7, July 2011. We found that none of Java projects have not used `try-with-resources` feature yet. Throughout our previous empirical study of language features, we found that it takes more than one year for a new feature to be adopted in open source projects. Instead, we investigated a potential conversion code from `try-finally` to `try-with-resources` feature. We found that about 10% `try-finally` code would be able to be converted `try-with-resources` feature.

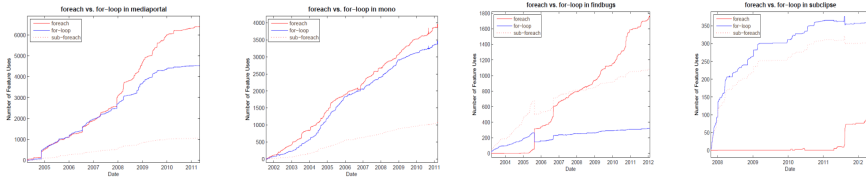


Fig. 2. The number of foreach and for-loop over time

Overall, our analysis indicate that **syntactic sugar is used more than its corresponding feature in C# projects, not in Java projects**; the usage of syntactic sugar may be different in different programming languages. We will investigate why this occurs in the future work. Instead, we enumerate several reasons in Section 5.

### 4.2 RQ2: Will Project Members Broadly Use Syntactic Sugar?

Recall that C# projects embrace more the usage of foreach than Java projects. Figure 3 shows the percentage of developers who used foreach, for-loop and sub-foreach in projects. More C# developers used foreach than for-loop, while four out of 10 Java projects show that more Java developers used foreach than for-loop. It seems that the usage of foreach in Java projects is relatively associated with the number of developers who use foreach based on the visual inspection of Figure 1 and 5. We assessed the strength of the relationship between the number of usages and developers using Spearman's rank correlation coefficient [4]. Spearman's coefficient is 0.56 which indicates a direct mild relationship. In other words, Java developers may not embrace foreach widely as we have found the same fact in Java projects in Section 4.1.

As we have analyzed individual developers' introduction and removal of foreach and sub-foreach from five of the most developers who used foreach per project, most C# developers who used foreach prefer foreach than for-loop (sub-foreach). However, the more than half of Java developers who used foreach do not seem to be willing to use foreach, while some Java developers prefer foreach.

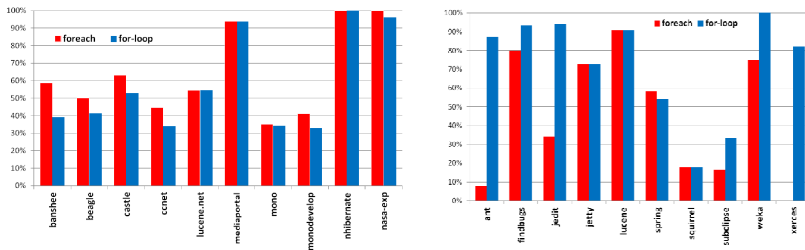


Fig. 3. The percentage of developers who used foreach and for-loop in C# projects (left) and Java projects (right)

Overall, the data and our analysis indicate that **C# developers use syntactic sugar more while Java developers relatively use less syntactic sugar**. As we mentioned in Section 4.1, we have not analyzed why this occurs in this paper; our research is in progress. Instead, we will discuss several reasons in the following section.

## 5 Conclusion

We have analyzed the usage of syntactic sugar in 20 open source projects. We found that syntactic sugar is used more widely in C#, not in Java and that C# developers prefer syntactic sugar while Java developers do not seem to be willing to use `foreach` in projects. However, we are unsure why not all developers embrace syntactic sugar. Instead, we enumerate several reasons as follows: (1) **Different adoption time**: The `foreach` was adopted at different times: C# in 2003 as part of .NET 1.1 and Java in 2004 as part of J2SE 5.0. (2) **Different keyword**: `foreach` in C# and `for` in Java. Java developers may not recognize `foreach` feature because of using the same keyword such as `for` statement. The `for` keyword is simple but may not be not recognizable than `foreach`. We also found one web-page which proposes new syntax for enhanced `for`-loop to be more readable and easy to understand by using `eachof` keyword<sup>2</sup>. (3) **Developer's unaltered tendency**: Developers may dislike to learn a new feature since an existing one works.

The future work is needed to find out the reasons we enumerated in this paper. We hope this work potentially contributes to a design of programming language features.

## References

1. `foreach`, in Visual Studio .NET (2003), <http://msdn.microsoft.com/en-us/library/ttw7t8t6v=vs.71.aspx>
2. `foreach` loop, [http://en.wikipedia.org/wiki/Foreach\\_loop](http://en.wikipedia.org/wiki/Foreach_loop)
3. The `For-Each` Loop, <http://docs.oracle.com/javase/1.5.0/docs/guide/language/foreach.html>
4. Myers, J., Well, A., Lorch, R.F.: *Research Design and Statistical Analysis*, 3rd edn. Routledge Academic (2010)
5. Pamin, C., Bird, C., Murphy-Hill, E.R.: Java generics adoption: how new features are introduced, championed, or ignored. In: MSR 2011: Proceedings of the 8th International Working Conference on Mining Software Repositories, pp. 3–12 (2011)
6. Perlis, A.J.: *Epigrams in programming*. ACM's SIGPLAN Publication (September 1982)
7. Skeet, J.: *C# in Depth*, 2nd edition. Manning Publications, Stamford, CT 06901 (2010)
8. `using` Statement, <http://msdn.microsoft.com/en-us/library/yh598w02v=vs.80.aspx>

---

<sup>2</sup> [http://onjava.com/onjava/2003/09/24/readable\\_java.html](http://onjava.com/onjava/2003/09/24/readable_java.html)

# Log Pre-processor for Security Visualization

Jinwon Seo<sup>1</sup> and Jin Kwak<sup>2</sup>

<sup>1</sup> ISAA Lab., Department of Information Security Engineering,  
SoonChunHyang University, Korea  
jinseo@ebay.com

<sup>2</sup> Department of Information Security Engineering,  
SoonChunHuyang University, Korea  
jkwak@sch.ac.kr

**Abstract.** Security Visualization which is largely used in dashboards is to formulate logs saved in text. However, with Internet volumes' rapid growth, analyzing a vast amount of logs has become quite a challenging job.

For that purpose, this paper proposes a method that prior to the visualization of a whole log, implements a pre-processing log, based on a security matrix and accomplishes an inference of refined data for visualization. Through this process, the time required to analyze a normal log can be curtailed, and with its intuitive visualization function, this process will be useful to detect any Advanced Persistence Threats.

**Keywords:** Log pre-processing, Security Visualization, Application Protocol, Security Metric, Advanced Persistence Threat.

## 1 Introduction

The development of network infrastructure and the increased volumes of Internet based services give rise to a sharp increase in the network communications on the Internet. At the same time, however, their adverse effects including the threats and frequency of hacking, malicious codes, and DDoS have also risen, creating many difficulties in dealing with incident responses.

Most security control systems carry out a threat detect based on an attack signature. However, the capability of the attack signature is limited to a known attack and there is significant problem to detect an evolving threat like APT (Advanced Persistence Threat), which is a shortcoming.

As a solution to that, this paper proposes a Log pre-processor for Security Visualization that extracts a specified value for security, targeting a collected log. Carrying out a Log pre-processor prior to data visualization will contribute to obtaining refined data specified for visualization and this will enable us to implement Security Visualization and to detect anomaly symptoms and to respond to security threats.

This paper is composed as follows; analyze on existing studies on a variety of event visualization technology and their implementations in Chapter 2, followed by Visualization based system's structures and Pre-processing methods based on

Security Matrices in Chapter 3, the results of experiments of the proposed methods in Chapter 4 and the conclusion in Chapter 5.

## 2 Related Studies

There are many precedent studies on the attempts and technology of network protocol visualization in place. However, they are mostly from mid-year 2000 and not many active studies are found afterwards.

The studies of Kasemsri[1] and Denise Ferebee[2] are considered to be an exemplary survey materials in the Security Visualization area. These studies suggest the study trends and various techniques of security visualization. Sven Krasser[3]'s studies propose the implementation of network traffic in 2D and 3D and how to perceive a malicious network activity. However, that method which is related to mapping collected network traffic, using Source IP/Port or Destination IP/Port, has a shortcoming which is only effective to detect the threats of DDoS or Netorok Scans. In the studies of Anita Komlodi[4], an Intrusion Detection toolkit including IDtk visualization function is proposed and yet it is a method which simply visualizes the IDS result could bring about massive confusion, given the fact that IDS has a high False Positive Rate. In the studies of Glenn A. Fink[5], he attempts the visualization of data up to host level. He collects data while classifying the scope of visibility by Network Visibility, Host Visibility, Application Visibility, and Overall Visibility. However, too much data could be a hindrance to achieving visualization. Andrew Stewart[6] proposes a visualization method comparing a difference between the past and the present. In that study, an IP is identified as '.' and according to the difference of the past and the presence of a specific IP, use characters such as '+', '-', 'X' and implemented a visualization. This method has strength in terms of an easy identification where a company managing a large volume of network hosts is able to achieve visualization by adopting the dot system (.). However, this has a shortcoming that detecting a particular host in the event of security incidence where host is indicated as .(dot) is a difficult task.

Barry V.W.Irwin[7] proposes a Visualization tool that helps detecting Network Scanning. This method, however, has a shortcoming that provides only a function detecting Network Scans and is restricted to analyze universal anomaly symptoms.

In Beom-Hwan Chang[8] studies, he attempts to achieve Security Situational Awareness, using a Traffic Pattern-Map. However, this method has a shortcoming to perceive a threat producing mass traffic. Chi Yoon Jeong[9]'s studies suggest a method to attain visualization by collecting flow information from a Network switch. However, this also has a limitation that can barely detect network anomaly symptoms like a traffic spike, network scan because the input data for visualization are not based on flow information.

Despite the variety of precedent studies on security visualization, most studies focus on Network Activities rather than security aspects. The reality is that there is lack of security visualization in security aspects. I'd like to summarize these studies following table.



**Table 1.** Summary of related Studies

Ref.	A Strong point	A Fault point
3	Log Visualization with 2D, 3D	Only describe massive connection
4	Visualization using IDS logs	Problem a False Positive of IDS
5	Visualization via Hosts log	Too much node, decrease the intuition of visualization
6	Easy to understand on status change	it's hard to figure out a real host
7	Good to detect a Network Scan	Only can detect Network Scan
8	Propose a new type such as Traffic Pattern map	Only describe massive connection like DDoS, Network Scan
9	Visualization via Network flow data from network devices	Optimization for Network Activity not Security

To resolve various problems, we need follow capabilities.

1. Visualize security anomaly context instead of Network activities.
2. No dependency various logs from any security devices
3. Not require an additional environment and no dependency of host OS
4. Can acquire a traceability and visibility from visualization results

To effect, this paper suggests a Security Visualization System using application protocol data. The log pre-processor method proposed the system which carries out an interaction between application protocol log generation and a visualization system, while acting as a security index group identifying security anomaly symptoms in vast logs.

### 3 Designs of Security Visualization System

#### 3.1 High-Leveled Design of the Security Visualization System

The Log pre-processor that has been proposed in this paper carries out an interaction between an Application protocol analyzer module and Security visualization module. The application protocol analyzer module captures packets in the network and generates a log file in application protocol unit. The Log pre-processor module brings the generated log and carries out Log Pre-processing that generates a specialized visualization data, based on security metric. The security visualization module then eventually implements visualization on the basis of the result of Log pre-processor. The regular implementation of this serial course will contribute to the visualization of result analysis on network anomaly symptoms.

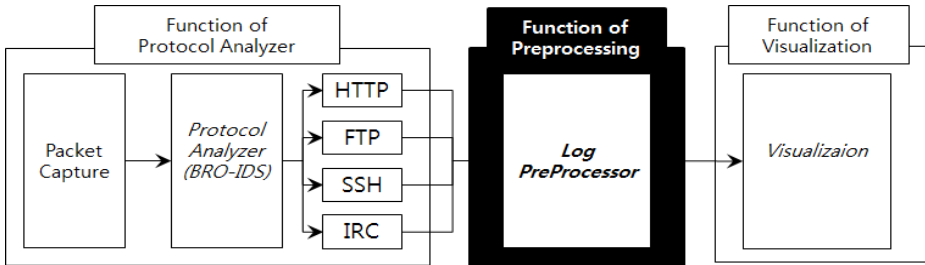


Fig. 1. High Level Design of Security Visualization System

### 3.2 Design of the Application Protocol Analyzer Module

Application protocol analyzer directly collects packets in a network and classifies them by application protocol and carries out the generating of each log file. Its core function is to classify by protocols in a network. For this process, the following functions are required.

- To be able to directly collect packets in a network
- To be able to classify collected packets by protocols
- To be able to save the data in session based form, not packet based
- To be able to save protocols that can be analyzed with security aspect.

### 3.3 Design of the Log Pre-processor Module

A Log Preprocessor carries out a function to extract specialized data for visualization in a generated log. For this process, following functions are required.

- To be able to pass a log per protocol
- To be able to easily manage security matrices
- To be able to generate outputs in a format recognizable in a visualization system

### 3.4 Design of the Security Visualization Module

Visualization module carries out a function to visualize an extracted data by Pre-processor. Visualization module carries out while considering directions that requires following functions.

- To be able to filtering data excluded by users
- To be able to extract data targeting visualization, according to security regulations identified by users

The ultimate objective of visualization is to deliver intuitive data. Therefore, data visualization should not include too many results.

The major strength of visualizing a Text log is a linking activity between source and destination. In the example of a server Scan, a compromised host tries to scan other hosts within the same network range. For this reason, the source host is one and the destination hosts are many. In other words, it is a good example showing a link that source to destination is equal to 1 to N (Source: Destination = 1: N).

## 4 Implementation of Log Pre-processor for Security Visualization

### 4.1 Whole System Implementation

An Application Protocol Analyzer needs to be able to generate logs per protocol in network packets. The Protocol Analyzer which has been proposed in this paper uses Bro-IDS[10] which is well known to be open source. Bro-IDS alongside with Snort is one of the two master open source based Intrusion Detection System(IDS) but is capable of extracting session based data that is composed after communication, unlike Snort.

Log Preprocessor extracts a specified value per security metric, based on a generated application protocol. As it is in shell script, it is easy to manage and alter and has the strength that enables prompt new security matrices in the event of a new security threat. The characteristics of each protocol for pre-processing are extracted as shown in table 1. These are for identifying anomaly symptoms found in a companies' internal network and extracted from known threat incidents.

The extracted values from the Pre-processor are taken to the Visualization Module which generates a link based graph. This paper adopts a linked list visualization method. It also uses a kind of open source, afterglow[11] program as a visualization engine. Afterglow can assign a user preferred format through a configuration file and has the strength to utilize ASCII format CSV file as input data.

**Table 2.** Security Metric for Anomaly Detection

No.	Anomaly Activities	Required Protocol Value
R1	{HTTP FTP SSH IRC SMTP} over non-standard port	sIP, dIP, dPort, Protocol ID
R2	{HTTP FTP SSH IRC SMTP} port over non-standard protocol	sIP, dIP, dPort, Protocol ID
R3	{HTTP FTP SSH} Service Scan	sIP, dIP, dPort, Protocol ID
R4	Massive outer connection via {HTTP FTP SSH SMTP IRC}	sIP, dIP, dPort, Protocol ID
R5	Login failure via {HTTP SSH FTP}	sIP, dIP, dPort, Login Result
R6	Suspicious File Up/Down via {HTTP FTP IRC}	sIP, dIP, mime_type, Filename
R7	Access to C2 server via {HTTP DNS IRC SMTP}	sIP, dIP, dDomain, Protocol ID

### 4.2 Security Visualization Experiment

The log files used in experiments are the logs occurred during the time of 12 to 3 pm in week days that generated application protocol logs like HTTP, FTP, SSH, and IRC.

Fig. 2 shows using FTP service, not using FTP standard port, 21/TCP by applying R1. Fig. 3 , applying R3 shows the detection of hosts that use IRC service.

In SSH protocol, the application of R5 infers Figure 4 and Figure 5. In Figure 5, a prompt check of login failure in SSH server from a specific host is identified. As demonstrated, refined data extracted from a Log pre-processor can be utilized to achieve visualization of contents which cannot be perceived in text log.

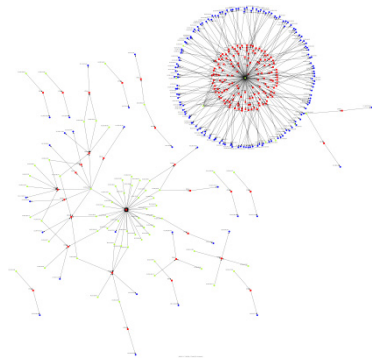


Fig. 2. FTP over non-standard port

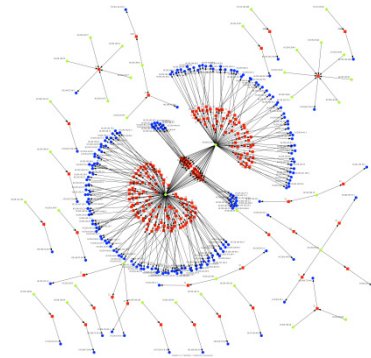


Fig. 3. Login Failure of SSH

By pre-processor, we can resolve the problem for security visualization like following these.

Table 3. Improvement of Fault point by Pre-processor

No.	A Fault point	Resolving
1	Visualize security anomaly context instead of Network activities.	Pre-processor extracts a security anomaly logs based on security aspect
2	Not require an additional environment and no dependency of host OS	Pre-processor handles data which get from not host level but network level
3	No dependency various logs from any security devices	Pre-processor uses a data from protocol analyzer instead of security logs
4	Can acquire a traceability from visualization results	Pre-processor generates a paired data between source and destination

## 5 Conclusion

This paper illustrates Log Pre-processor method applying Security Matrices to achieve visualization of an application protocol log extracted through a Protocol Analyzer. To carry out security visualization, generating a log per application protocol in a network is made firstly. However, while visualizing a vast volume of logs, detecting anomaly symptoms in a network becomes very tricky. A Visualization Pre-processor which applies anomaly symptoms emerged in analyzing the threat activities in real incidents rather than an IDS signature shows a good result to identify each threat's anomaly symptom.

In addition, this enables a security manager to perceive a threat through a specifically generated log for visualization while a massive log such as the Web goes through a Pre-processor. This method can be made good use of by perceiving security situations as the inference of visualization data per threat and can be obtained by simply adding an index to the Log Pre-processor in the event of a new threat.

**Acknowledgment.** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No. 2012-010886).

This work was supported by the Soonchunhyang University Research Fund.

## References

1. Kasemsri, R.R.: A survey, taxonomy, and analysis of network security visualization techniques. Computer Science Theses 17 (2006)
2. Ferebee, D., Dasgupta, D.: Security visualization survey. In: Proc. of the 12th Colloquium for Information Systems Security Education (2008)
3. Krasser, S., et al.: Real-time and forensic network data analysis using animated and coordinated visualization. In: Proceedings from the Sixth Annual IEEE SMC Information Assurance Workshop, IAW 2005. IEEE (2005)
4. Komlodi, A., et al.: A user-centered look at glyph-based security visualization. In: IEEE Workshop on Visualization for Computer Security (VizSEC 2005). IEEE (2005)
5. Fink, G., et al.: Network Eye: End-to-End Computer Security Visualization. In: Submitted for consideration at ACM CCS Workshop on Visualization and Data Mining for Computer Security, VizSec/DMSec (2004)
6. Stewart, A.: Efficient visualization of change events in enterprise networks. In: Securecomm and Workshops. IEEE (2006)
7. Irwin, B.V.W., van Riel, J.-P.: Inetvis: a graphical aid for the detection and visualisation of network scans. In: Conference on Visualization Security, VizSec2007 (2007)
8. Chang, B.-H., Na, J.-C., Jang, J.-S.: Network Security Situational Awareness using Traffic Pattern-Map. Journal of the Korea Industrial Information System Society 11(3), 34–39 (2006)
9. Jeong, C.Y., Sohn, S.-G., Chang, B.-H., Na, J.-C.: An Efficient Method for Analyzing Network Security Situation Using Visualization. Journal of the Korea Institute of Information Security and Cryptology 19(3), 107–117 (2009)
10. Bro Network Security Monitor, <http://www.bro-ids.org>
11. Afterglow, <http://afterglow.sourceforge.net>

# Fast Coding Unit (CU) Depth Decision Algorithm for High Efficiency Video Coding (HEVC)

Chan-seob Park<sup>1</sup>, Byung-Gyu Kim<sup>1</sup>, Gwang-Soo Hong<sup>1</sup>, and Sung-Ki Kim<sup>2</sup>

<sup>1</sup> Dept. of Computer Engineering, SunMoon University,

<sup>2</sup> Dept. of IT Education, SunMoon University

A-san city, Rep. of Korea

{wing0819, bg.kim, hongzolv}@mpcl.sunmoon.ac.kr,

skkim@sunmoon.ac.kr

**Abstract.** In this paper, we propose a fast CU depth decision algorithm for high efficiency video coding (HEVC) technology to reduce its computational complexity. In  $2N \times 2N$  prediction unit (PU), the proposed method compares to rate-distortion (RD) cost and determines the depth using the compared information. Moreover, in order to speed-up the encoding time, the efficient merge SKIP detection method is developed additionally based on the contextual mode information of neighboring CUs. Experimental result shows that the proposed algorithm achieves the average time-saving factor of 43.67% in the random access (RA) at Main profile configuration with the HEVC test model (HM) 10.0 reference software. Compared to HM 10.0 encoder, a small BD-bitrate loss of 0.55% is also observed without significant loss of image quality.

**Keywords:** HEVC, correlation, Merge SKIP Mode, CU depth.

## 1 Introduction

Block based predication and compression has been utilized in most video coding standards such as MPEG-2 and H.264/AVC [1]. Especially, the variable block size prediction and compensation in H.264/AVC is a key factor which contributes to significant bit reduction with the same image quality. To improve performance of video coding efficiency, the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) recently formed the Joint Collaborative Team on Video Coding (JCT-VC) [2]. The JCT-VC is finalizing the next-generation video coding standard, called high efficiency video coding (HEVC). JCT-VC has taken the 14th Meeting up to this day.

The video encoding and decoding processes in HEVC are composed of three units: a) a coding unit (CU) for the root of the transform quadtree, as well as a prediction mode for the INTER/SKIP/INTRA prediction, b) a prediction unit (PU) for coding the mode decision, including motion estimation (ME) and rate-distortion optimization, and c) a transform unit (TU) for transform coding and entropy coding. Initially, a frame is divided into a sequence of non-overlapped largest coding units, called a Coding Tree Unit (CTU). A CTU can be recursively divided into smaller coding units

(CU) and made flexible using quadtree partitioning, which is called a Coding Tree Block (CTB). It is clear that the new CTB structure with larger coding block size in HEVC greatly increases the computational complexity to achieve the high efficiency of coding gain in HEVC standard, comparing to the H.264/AVC video standard.

To reduce the computational complexity of the HEVC, there are several algorithms with high speed-up factor keeping negligible losses of BD-bitrate. In [3], an early termination scheme called coded block flags (CBF) fast method (CFM) was used the CBF of luminance and chrominance in order to reduce the complexity of the inter mode decision. If the CBFs of luminance and chrominance are both zeros, the search process for the next PU modes in the current depth level is not performed.

Choi et al. [4] proposed a tree-pruning algorithm that makes an early determine CU. To reduce computational complexity, it uses mode information of the current CU. When the best PU mode of the current CU selects the SKIP mode, the current CU is not divided into sub-CUs in the sub-depth level of the current CU. This process was adopted in HEVC test model 4.0 reference software [5]. Zhang et al. [6] proposed an algorithm by reducing the depth search range. Their method was based on the depth information correlation between spatio-temporal adjacent CUTs and the current CTU.

Similar to [3] and [4], Yang *et al.* [7] proposed an early detection algorithm for the SKIP mode. Their motivation was early determination of skip conditions from fast method decision schemes in the H.264/AVC [8-10]. They utilized differential motion vectors (DMV) and coded block flags (CBF) of the inter  $2N \times 2N$  mode as skip conditions.

To reduce the computational complexity of HEVC encoding system, we propose an effective CU selection algorithm for HEVC based on RD cost of  $2N \times 2N$ . In the proposed algorithm, we use also early merge SKIP mode detection technique based on correlation of neighboring CUs including depth level.

This paper is organized as follows: In Section 2, an overview of the HEVC is described, simply. Section 3 presents the suggested algorithm including a merge SKIP detection. Simulation results and some discussion will be given in Section 4. Concluding comments are given in Section 5.

## 2 Proposed Fast CU Decision Algorithm

### 2.1 Adaptive CU Depth Decision Method

In order to enhance the encoding speed, the depth information of CTU is found by using  $2N \times 2N$  PU information. Table 1 indicates the probability that  $2N \times 2N$  PU is determined as the best mode. In this table, you can see the probability of  $2N \times 2N$  PU determined is very high. It shows almost 90% in Class A, and 79% in Class B case. This means that  $2N \times 2N$  PU decision has large portion. From this result, if we find a fast scheme to decide it early, then the overall consumed time for encoding may be decrease effectively.

In this study, the experimental environment to calculate  $2N \times 2N$  PU selection probability was set as: the sequences from the Class A for the Class D and 50 frames

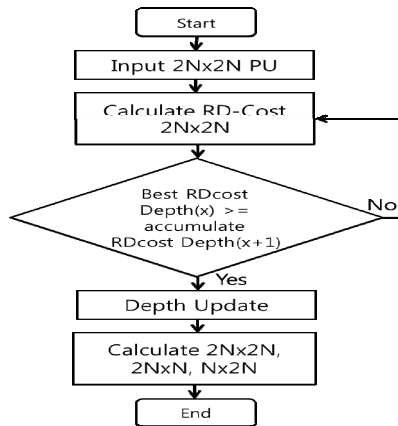
for each sequence. Also, the selection probability has been computed by averaging of various QP values (22, 27, 32, 37).

As shown in Table 1, the maximum 91.2% and minimum 73.7% of  $2N \times 2N$  PU portion are shown. By using the depth information selected from  $2N \times 2N$  PU, we are able to predict the proper CU in the relevant depth. When the  $2N \times 2N$  is given as input in the current frame, depth information of each CU is calculated in CTU [2]. Depth is defined CTU in the quad-tree structure that represents the split of the CU.

**Table 1.** The probability of  $2N \times 2N$  PU as the best mode

Class A		Class B		Class C		Class D	
Traffic	85.7%	Kimono	82.1%	Basketball Drill	86.3%	Basketball Pass	86.5%
PeopleOn Street	87.3%	ParkScene	76.4%	BQMall	89.6%	BQSquare	78.7%
Nebuta	92.1%	Cactus	82.3%	PartyScene	77.8%	Blowing Bubbles	81.2%
SteamLocomotive	91.2%	Basketball Drive	73.7%	RaceHorses	79.1%	RaceHorses	74.8%
-	-	BQTerrace	80.9%	-	-	-	-
Average	<b>89%</b>	Average	<b>79%</b>	Average	<b>83.2%</b>	Average	<b>80.3%</b>

After finding the residual with the best motion vector obtained through  $2N \times 2N$  PU, finally, the RD is calculated through the Full RQT. By using this depth information, the depth of CTU is determined through the RD cost competition. The current depth( $x$ ) with the best RD cost will be determined as the current depth level. If the RD cost of depth( $x+1$ ) is smaller than that of the current depth( $x$ ), the calculation to find the minimum RD cost is repeated recursively for upper depth. After that, the best sub-partitioned mode is determined by calculating and comparing the RD cost from the depth  $2N \times 2N$ ,  $2N \times N$ ,  $N \times 2N$  to go to the detail search.



**Fig. 1.** Flowchart of the proposed CU depth decision algorithm



Figure 1 illustrates the overall procedure of the proposed CU depth decision. The proposed algorithm is performed as follows: Firstly, as input, if the  $2N \times 2N$  PU size is given at the current depth( $x$ ), our algorithm calculates the RD cost from the given  $2N \times 2N$  PU size. After that, the best RD cost of the current depth( $x$ ) is compared with the accumulated RD cost of next depth( $x+1$ ). The accumulated RD cost can be calculated from the previous encoded modes (PUs). If the given condition is satisfied, the current CU depth is selected as the best depth level. Otherwise, go to next CU depth ( $2N \times 2N$ ) and perform in the same manner.

In the flowchart, after selecting the best CU depth, the detailed partition mode is determined by calculating and comparing the RD cost from the depth  $2N \times 2N$ ,  $2N \times N$ ,  $N \times 2N$  as the detailed search.

### 2.2 Early Merge SKIP Decision Method

We also develop an early merge SKIP decision to increase the encoding. According to the SKIP of HEVC standard, the merge SKIP has been adopted [2] for providing more coding efficiency. The proposed merge SKIP detection method utilizes the information of neighboring blocks.

Figure 2 shows the position of adjacent CUs relative to the current CU. Neighboring CUs such as CU1, CU2 and CU3 (above-left, above, and left CU from the current CU) have a high degree of spatial correlation. CU4 and CU5 are used as CUs for temporal correlations, and CU6 and CU7 are used for depth correlations. Identify the each CUs Merge information. Each encoding units have the information of the flag generated in  $2N \times 2N$ , and after checking the flags merge and merge SKIP mode is decided.

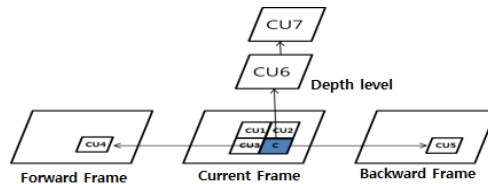


Fig. 2. A relationship between the current and adjacent blocks

When encoding the current block, the spatial and temporal neighboring blocks can provide useful information, because they have much similarity in terms of texture and motion. In HEVC, depth concept has been introduced as described in Section 2. So, the optimal block mode of the current block can also be deduced from the neighboring blocks which is composed of the spatial, temporal, and depth relationship.

In Fig. 3, the proposed early merge SKIP decision method is displayed. When  $2N \times 2N$  is the current block in usual HEVC coding, the merge process of motion information is performed to achieve more coding gain. In the merge process, the proposed method checks on the mode types of spatial, temporal, and depth neighboring blocks (as shown in Fig. 2) in the first. If all modes of neighboring blocks

are SKIPs, then SKIP is selected as the best mode for merge process. The remained mode search is omitted directly. Otherwise, usual search for merge process is employed. With the merge SKIP detection technique, the complexity of the motion estimation can be more reduced while keeping the image quality.

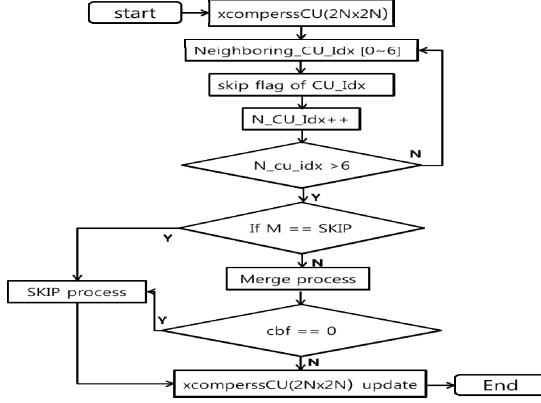


Fig. 3. Flowchart for the early merge SKIP detection method

### 3 Experimental Results and Discussion

The proposed algorithm was implemented on HM 10.0 (HEVC reference software). Test conditions were random access using RA-Main. Standard sequences with 100 frames were used from Classes A to D with various QP values (22, 27, 32, 37). To evaluate performance, we used a total of 100 frames for test sequences in Class A to D and defined the measurement of  $\Delta Bit$ ,  $\Delta PSNR_Y$ , and  $\Delta Time$  as:

$$\Delta Bit = Bit_{proposed} - Bit_{original}, \quad (1)$$

$$\Delta PSNR_Y = PSNR_{Y_{proposed}} - PSNR_{Y_{original}}. \quad (2)$$

$\Delta Time$  is a complexity comparison factor used to indicate the amount of total encoding time saving (Eq. (3)).

From Eq. (3),  $Time_{(x)}$  means the total consumed time of the method  $x$  for encoding.

$$\Delta Time = \frac{Time_{Proposed} - Time_{Anchor}}{Time_{Anchor}} \times 100. \quad (3)$$

The results in Table 2 show the performance of our algorithm when comparing to the original HM 10.0 encoder. The proposed algorithm achieves 43.67% of time – saving factor on average with only a 0.05 (dB) loss in PSNR and a 0.55% increment in total bits. For Class A (Full HD image), about 48% of time-saving factor was observed with very small loss of quality. For smaller size of image (Class D), the

speed-up gain is slightly decreased, but the quality loss is still negligible. From this result, we can deduce that the proposed algorithm is able to obtain more speed-up gain for very high quality video (Full HD video).

**Table 2.** The performance of the proposed scheme

Test Seqs.	$\Delta Bit(\%)$	$\Delta PSNR_Y$	$\Delta Time(\%)$
Class A	0.36%	-0.06	-47.74%
Class B	0.31%	-0.06	-45.60%
Class C	0.61%	-0.06	-41.74%
Class D	0.95%	-0.05	-39.61%
<b>AVERAGE</b>	<b>0.55%</b>	<b>-0.05</b>	<b>-43.67%</b>

## 4 Conclusions

In this paper, we have proposed a fast CU depth decision algorithm based on the RD cost comparison for high efficiency video coding (HEVC) technology to reduce its computational complexity. In addition, merge SKIP extraction method was developed and integrated with CU depth decision algorithm. Experimental result shows that the proposed algorithm achieves the average time-saving factor of 43.67% in the random access (RA) at Main profile configuration with HM 10.0 reference software while keeping small loss of quality.

## References

- [1] Wiegand, T., Sullivan, G.J.: The H.264/AVC Video Coding Standard. IEEE Signal Processing Magazine II, 148–153 (2007)
- [2] Sullivan, G.J., Ohm, J.R., Han, W.-J., Wiegand, T.: Overview of the High Efficiency Video Coding (HEVC) Standard. IEEE Transaction on Circuits and Systems for Video Technology (December 2012)
- [3] Lee, Y.-L., Lim, J.: Early Termination of CU Encoding to Reduce HEVC Complexity. presented at JCTVC-F045, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 6th Meeting, JCTVC, Torino, Italy, July 14-22
- [4] Choi, K., Jang, E.S.: Fast coding unit decision method based on coding tree pruning for high efficiency video coding. Optical Engineering Letters (March 2012)
- [5] Choi, K., Park, S.-H., Jang, E.S.: Coding tree pruning based CU early termination. presented at JCTVC-F092, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 6th Meeting, JCTVC, Torino, Italy, July 14-22 (2011)
- [6] Zhang, Y., Wang, H., Li, Z.: Fast Coding Unit Depth Decision Algorithm for Interframe Coding in HEVC. In: Data Compression Conference, DCC (2013)

- [7] Yang, J., Kim, J., Won, K., Lee, H., Jeon, B.: Early SKIP Detection for HEVC, JCTVC-G543, JCT-VC (2011)
- [8] Jeon, B., Lee, J.: Fast method decision for H.264. JVT-J033, Hawaii, USA (December 2003)
- [9] Lee, J., Jeon, B.: Fast method decision for H.264. In: IEEE International Conference Multimedia and Expo (ICME) (June 2004)
- [10] Choi, I., Lee, J., Jeon, B.: Fast coding mode selection with rate-distortion optimization for MPEG-4 Part-10 AVC/H.264. IEEE Trans. Circuits Syst. Video Technol. 16(12), 1557–1561 (2006)

# A Study on Using OWC with Battery Storage for Providing Power to Sensor Nodes in a Fish Farm

Olly Roy Chowdhury<sup>1</sup>, Hong-Geun Kim<sup>1</sup>, Dae-Heon Park<sup>2</sup>, Chang-Sun Shin<sup>1</sup>,  
Yong-Yun Cho<sup>1</sup>, and Jang-Woo Park<sup>1,\*</sup>

<sup>1</sup> Dept. of Information and Communication Engineering, Suncheon National University,  
Suncheon, Republic of Korea

{ollyroy, khg\_david, csshin, yycho, jwpark}@suncheon.ac.kr

<sup>2</sup>Department of IoT Convergence, Electronics and Telecommunications Research Institute,  
Daejeon, Republic of Korea  
dhpark82@etri.re.kr

**Abstract.** This paper describes the concept and idea of a project which addresses the use of an oscillating water column (OWC) to produce electricity from wave energy. This device is a partially submerged chamber with air trapped above the water surface. The aim of our project is to provide power to sensor nodes in a floating fish farm. The system uses Well's Turbine as power take off (PTO) system, which is a bi-directional air turbine. Here we briefly discussed the working principle of OWC, Well's turbine, power storage system, sensor nodes and given general idea of the system designed. As the system will be used in the sea area that is why it will be developed in a well-structured and more strengthened way and of course with less complexity that it can tolerate the harsh weather condition of the sea.

**Keywords:** Wave energy harvester, OWC, sensor nodes, storage system.

## 1 Introduction

The main challenge for offshore floating energy devices is to build a structure capable of withstanding the ruthless ocean environment in such a way that costs are competitive in global energy markets. Principle Power Inc. and National Renewable Energy Lab (NREL) has done a survey by integrating wave energy converters into the WindFloat, resulting in a new concept called the WindWaveFloat (WWF). This leads to some benefits:

1. Wind and wave energy converters can share the electrical cable and power transfer equipment to transport the electricity.
2. Reduction in cost due to shared construction.
3. Wave energy conversion continues even in storm conditions, when wind turbines may be shut down.

---

\* Corresponding author.

NREL have completed a contract to assess the technical and economic feasibility of WWF. Three types of devices were studied within the scope of that work, oscillating water column, single or multiple point absorbers, oscillating wave surge converter [1]. We have a view to integrate Wind energy converter with wave but as wind energy conversion system is quite mature one now a days, this paper will focus about wave energy converter using OWC. In this study, the power take off system will be developed by an air turbine which is a bidirectional Wells turbine that is most common in this type of application.

The aim of our project is to develop the energy harvesting structure around a fish farm to power the fish farm sensor nodes as fish farms floating in the sea needs small supply of power and connection of this to the national grid system would be impractical and also economically unrealistic also. As the structure will be constructed with a fish farm floating in a sea, it will be designed in such a way that would not disturb the fishes and can withstand the sea weather as well. Due to the variation of power from the generator of the wave energy converter (WEC) with weather conditions, WEC with battery storage system is proposed in this paper to feed the output power constantly to sensor nodes.

## 2 Theoretical Background

Ocean waves encompass two forms of energy: the kinetic energy and the potential energy of elevated water particles. On the average, the kinetic energy in a linear wave equals its potential energy. The energy flux in a wave is proportional to the square of the amplitude and to the period of the motion. The total potential and kinetic energy of an ocean wave can be expressed as

$$E = \frac{1}{2} \rho g A^2 \quad (1)$$

To obtain the average energy flux or power of a wave period, energy  $E$  is multiplied by the speed of wave propagation,  $v_g$ , where,  $v_g = L/2T$

$$P_w = \frac{1}{2} \rho g A^2 L/2T \quad (2)$$

The dispersion relationship describes the connection between the wave period  $T$  and the wave length  $L$  as

$$L = \frac{gT^2}{2\pi} \quad (3)$$

The power or energy flux of an ocean wave can be calculated as

$$P_w = \frac{\rho g^2 T A^2}{8\pi} \quad (4)$$

Can be rewritten as a function of wave height,  $H$  (Considering that the wave amplitude is half of the wave height) [2].

$$P_w = \frac{\rho g^2 T H^2}{32\pi} \tag{5}$$

### 3 System Model

#### 3.1 Oscillating Water Column Technology

Among all the types of wave energy extractors, Oscillating Water Column (OWC) is the most successful and extensively studied technology for extracting energy from ocean waves that can be located on the shoreline, nearshore or offshore [3]. OWC can actually be quite efficient and present point absorbing characteristics. A particular case of this category of which is a floating OWC [4]. Both OWC and Floating OWC works on the same working principles.

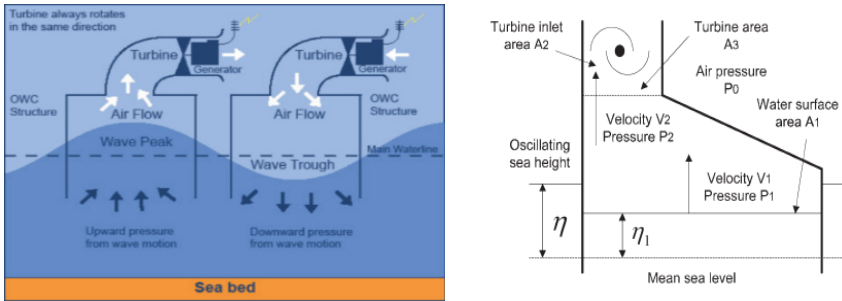


Fig. 1. (a) OWC working principle

(b) OWC chamber parameter

Figure 1(a) shows working principle of OWC. Water enters through a subsurface opening into a chamber with air trapped above it. The wave action causes the captured water column to move up and down like a piston and the air trapped above the water level is compressed and decompressed by this movement to generate an alternating stream of high-velocity air through an exit blow hole. This air is channeled through a turbine-generator to produce electricity [5].

Figure 1(b) describes the OWC chamber and chamber parameters for deriving the equations of power available at the turbine [6]. According to the theory described in [6] the total chamber power  $P_{total}$  available at the turbine is

$$P_{total} = C_p(P_t + P_a) \tag{6}$$

Where  $C_p$  is conversion rate.  $P_a$ , the power acting on the turbine due to the airflow is

$$P_a = \rho A_2 V_2^3 / 2 \tag{7}$$

And the power  $P_t$  due to the pressure difference across the turbine becomes

$$P_t = \left[ \rho \frac{A_1}{A_2} \frac{\partial \varphi_1}{\partial t} - \rho \frac{Q}{A_2} (V_2 - V_1) \right] Q \quad (8)$$

Here  $Q$  is the volume rate of airflow across the turbine.  $V_1$  and  $V_2$  are related to water height in the chamber as

$$V_1 = \frac{d\eta_1}{dt} \quad (9)$$

And 
$$V_2 = \frac{A_1}{A_2} V_1 \quad (10)$$

$\phi_1$  and  $\phi_2$  are velocity potential. 
$$\varphi_1 \approx V_1 \eta_1 \quad (11)$$

and 
$$\varphi_2 \approx \frac{A_1}{A_2} \varphi_1 \quad (12)$$

Thus from (6), (7) and (8) the total power available at the turbine is,

$$P_{total} = C_p \left\{ \left[ \rho \frac{A_1}{A_2} \frac{\partial \varphi_1}{\partial t} - \rho \frac{Q}{A_2} (V_2 - V_1) \right] Q + \rho A_2 V_2^3 / 2 \right\} \quad (13)$$

From equation (5) and (13) the ratio of the power input from the waves and the output power available can also be obtained.

### 3.2 Wells Turbine

OWCs mostly use self-rectifying Wells turbines which eliminates the need for expensive and delicate valve systems to rectify the direction of the airflow. The Wells turbine has a low drag and can be driven at high rotational speeds of several hundred rpm without the need for a gearbox. The efficiency of Wells rotor is the highest when the air pressure is corresponding to 2–3 m of water rise in the column, which is the typical ocean wave height [2]. Prof. A.A. Wells developed this turbine specifically for direction-changing airflows such as the wave motion induced airflow in an OWC [7].

### 3.3 Battery Storage

In renewable energy systems: lead acid, lithium, and nickel batteries are used. . They act as a constant voltage source in the power systems [8]. Our system will generate electricity for the sensor nodes of a fish farm which is approximately of 100 W. About 50 W of power at 12 V, will be provided by the wave energy converter. So for the storage system of the above power, for example the lithium ion battery of a laptop charger with 5.2 Ah and 10.8 V can be a good solution.



### 3.4 Sensor Nodes

Sensor nodes have different structure, connectivity with different power consumption. Among these the suitable ones that can be served with our harvester system are mentioned here. Fig.2 (a) shows an example of sensor nodes in which various communication techniques such as Zigbee, Dash 7, or Wifi etc can be used according to architecture of a fish farm and distance from land. Also, sensor nodes have to be equipped with different sensors like Electric Conductivity (EC), Oxidation/Reduction Potential (ORP), Biochemical Oxygen Demand (BOD) and Dissolved Oxygen (DO) to monitor the environment conditions. These sensors have different operating voltage and consume different power. We have estimated power consumption per day by the sensor nodes including transmitting, receiving and all sensing will be about 20W, which can be supplied by our harvesting system.

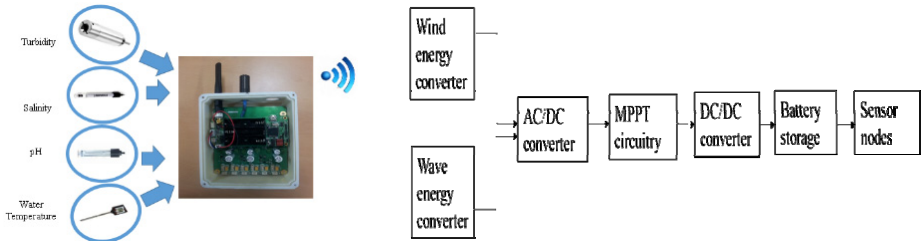
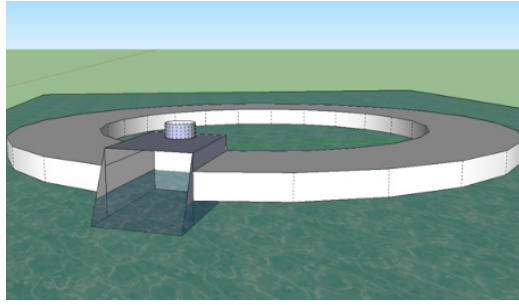


Fig. 2. (a) Sensor node (b) Block diagram of energy conversion system

## 4 Experimental Setup

Wave energy is a developing concept that needs more analysis, modeling and experimental observation comparing to wind energy converter. Papers related to modeling and numerical analysis of WEC, mostly used Linear radiation-diffraction software WAMIT by computing Response Amplitude Operators (RAOs)[1], [7]. Several numerical methods like Runge-Kutta-Nystrom method and Computational Fluid Dynamics with a software tool CFX, has also been used in other papers [9-10] for deriving the air pressure and kinetic energy terms at turbine end to calculate the power output and to analyze the effect of wave height, wave period on output power. The study showed effective results [6]. For our project, any one of the above mentioned methods may be suitable but with more accuracy. Fig 2(b) shows simple block diagram of our system.

Our WEC will be attached to the side of floating fish farm platform. The chamber will be built externally in the outer space of the periphery of the fish farm and will be submerged in water. Fig 3 showing a 3D image of the OWC with fish farm platform.



**Fig. 3.** 3D image of the floating fish farm OWC

## 5 Conclusion

Providing grid connection for electricity to offshore structures like floating fish farm, is impractical. Our project's target is to serve the sensor nodes of an on sea floating fish farm with electricity produced from wave dynamics. This paper describes OWC as a simple and efficient way to feed power to the fish farm sensor nodes. Since the power requirement of the sensor node is very small about 100w, the size of our structure will be small in agreement with the power. Finally OWC can be the best solution to match the situation. We believe perfect analysis, precise modeling and careful test model can show effective results.

**Acknowledgements.** This work was supported by the Industrial Strategic technology development program, 10041766, Development of energy management technologies with small capacity based on marine resources funded by the Ministry of Knowledge Economy (MKE, Korea).

## References

1. Principle Power Inc., WindWaveFloat (WWF): Final Scientific Report. DOE Grant DE-EE0002651
2. Khaligh, A., Onar, O.C.: Energy Harvesting: Solar, Wind, and Ocean Energy Conversion Systems. Taylor&Francis Group, Boca Raton (2010)
3. Dresser-Rand HydroAir Variable Radius Turbine,  
<http://www.dresser-rand.com/products/hydroair/>
4. Lorc Knowledge, <http://www.lorc.dk/wave-energy/types-of-devices>
5. Lorc knowledge, Oscillating Water Column,  
<http://www.lorc.dk/wave-energy/oscillating-water-column>
6. Dorrell, D.G., Hsieh, M.F., Lin, C.C.: A Multichamber Oscillating Water Column Using Cascaded Savonius Turbines. IEEE Transactions on Industry Applications 46(6), 2372–2380 (2010)
7. Aubault, A., Alves, M., Sarmiento, A., Roddier, D., Peiffer, A.: Modeling of an Oscillating Water Column on the Floating Foundation Windfloat. In: Proceedings of the ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering (2011)

8. Carrasco, J.M., Franquelo, L.G., Bialasiewicz, J.T., Galván, E., Guisado, R.C.P., Prats, M.A.M., León, J.I., Moreno-Alfonso, N.: Power-Electronic Systems for the Grid Integration of Renewable Energy Sources: A Survey. *IEEE Transactions on Industrial Electronics* 53(4) (2006)
9. Dorrell, D.G., Hsieh, M.F., Lin, C.C.: A Small Segmented Oscillating Water Column Using a Savonius Rotor Turbine. *IEEE Transactions on Industry Applications* 46(5), 2080–2088 (2010)
10. Hsieh, M.F., Lin, I.H., Dorrell, D.G., Hsieh, M.J., Lin, C.C.: Development of a Wave Energy Converter Using a Two Chamber Oscillating Water Column. *IEEE Transactions on Sustainable Energy* 3(3), 482–497 (2012)

# Location-Based Intelligent Robot Management Service Model Using RGPSi with AoA for Vertical Farm

Hong-Geun Kim<sup>1</sup>, Dae-Heon Park<sup>2</sup>, Olly Roy Chowdhury<sup>1</sup>, Chang-Sun Shin<sup>1</sup>,  
Yong-Yun Cho<sup>1</sup>, and Jang-Woo Park<sup>1,\*</sup>

<sup>1</sup> Department of Information and Communication Engineering, Sunchon National University,  
Suncheon, Republic of Korea

{khg\_david, ollyroy, csshin, yycho, jwpark}@sunchon.ac.kr

<sup>2</sup> Department of IoT Convergence, Electronics and Telecommunications Research Institute,  
Daejeon, Republic of Korea  
dhpark82@etri.re.kr

**Abstract.** The purpose of this paper is that intelligent robot control service model is one of the service model for a complete auto control system in vertical farm. This model provides environment monitoring service based on context-aware without human intervention in vertical farm. These service need to provide necessary accurate locations indoor environment based on Ubiquitous Sensor Network (USN). It will precisely monitor crop growth environment by intelligent robot using the indoor localization algorithm based on Ratiometric Global Positioning System iteration (RGPSi) with Angle of Arrival (AoA). Therefore, this model makes it possible to shorten operate time and to decrease workforce in vertical farm. It also will contribute to the complete auto control system of vertical farm on a large scale.

**Keywords:** intelligent robot, vertical farm, control, localization, USN.

## 1 Introduction

Lately occurrence of unpredictable abnormal climate and environmental pollution due to nuclear facility, influence crop production in outdoor culture a lot. Also, solving a food shortage caused from lack of crop yield due to increase of future population is overriding issue. The agriculture system called vertical farm that enables production of clean and fresh crop is proposed for solution of these problems. Vertical farm is the agriculture system that enables fast growth and planned production by controlling nutrient solution of crop and growth environment based on hydroponics mode. This vertical farm could be utilized as precision agriculture system that enables planned acquisition of yield and safe crop cultivation by monitoring and controlling growth environment, elements needed in growth of crops using various sensors [1].

Lately, vertical farm attempts fusion with various advanced technologies for reduction of labor, improvement of productivity and high quality. Since advanced

---

\* Corresponding author.

vertical farm with integrated radio wire communication technology and sensor technology is gradually enlarging currently, automatized and intellectualized growing condition management and control service provision technology is required [1, 2].

This paper propose intelligent service model that could provide monitoring service of vertical farm environment and crop condition by robot in the condition where human intervention is minimized by using various state information occurred in real time from Ubiquitous Sensor Network (USN) based sensors. This model control vertical farm that get all context information by Service Provider. This context information consist all environment information, status of actuators, growth condition of crops and low-data of Intelligent Robot based on localization. Key point of this model is Intelligent Robot that will obtain more precise monitoring and controlling by self-moving based on localization [3].

Indoor localization method for self-moving use RGPSi with AoA to provide high accuracy based on USN. This method will be applied in various communication environments with low cost and high accuracy. It provide Path Planning that analyze context information about Indoor Mapping of Intelligent Robot based on localization by Service Provider. We discuss some of the main issues in section 2 with vertical farm and intelligent robot based on localization. In Section 3 we present main model with concepts of intelligent robot control services. In Section 4 we present service scenario. Conclusion and future work are presented about service model in last section.

## 2 Related Works

### 2.1 Intelligent Robot Based on Localization Algorithm by Using RGPSi with AoA

Localization based services are now introduced for outdoor scenarios, but such services also have a large potential for certain indoor scenarios, such as shopping malls, convention centers, medical centers, large museums and university complexes[4 - 6].

There are several issues making the localization challenging. A localization technique by which the sensor nodes determine their position with respect to a set of fixed beacon nodes that are capable of covering the entire network area by wireless transmissions using powerful directional antenna. It use sensor node for collecting all environment information in vertical farm. Therefore, localization of Intelligent Robot use Ubiquitous Sensor Network (USN) techniques. This localization method utilizes the Ratiometric Global Positioning System iteration (RGPSi) with Angle of Arrival (AoA) algorithm. This method deployed three sensors which forms a triangular to contribute to find the target's position. And, this method will be implemented in various communication environments with low cost and high accuracy. And due to low computation complex, low performance processors can adopt this algorithm with no big burden. Transmission devices for localization use Zigbee or Wi-Fi or Ultrasonic Wave [7].

Intelligent Robot's navigation perform path planning that analyze context information through Indoor Mapping to assist with utilizing Acceleration, Gyroscope, Proximity and Terrestrial Magnetism sensors [8 – 10].

## 2.2 Vertical Farm Monitoring System

Monitoring factor of vertical farm is growth environment factor of crop. Therefore, it should confirm current growth condition of crop by using sensors related to growth of crop, and artificially control growth environment using various sensors to match the environmental condition needed for growth of crop.

Table 1 show elements for growth environment monitoring in vertical farm. Such monitoring must involve vertical farm's growth data, environment data, real sense data and status of actuator [1 - 3].

**Table 1.** Elements of vertical farm monitoring

Category	Elements
Growth	Leaf Area, Leaf Number, Plant Height, Fresh Weight, Fruit number, Fruit Color, Fruits Size, etc.
Environment	Air Temperature, Air Humidity, Air CO <sub>2</sub> , Leaf Wetness, Illumination, PPFD(Photosynthetic Photon Flux Density), Soil Water, Soil Temperature, Crops Image, etc.
Sensor	Air/Soil Temperature, Air Humidity, Illumination, Air/Soil CO <sub>2</sub> , Leaf Wetness, etc.
Actuator(Device)	Air Conditioning, CO <sub>2</sub> Generator, Irrigation, Heater, Cooler, Ventilation Fan, Humidifier, Dehumidifier, Low Pressure Fogging System, Artificial Light, Camera, etc.

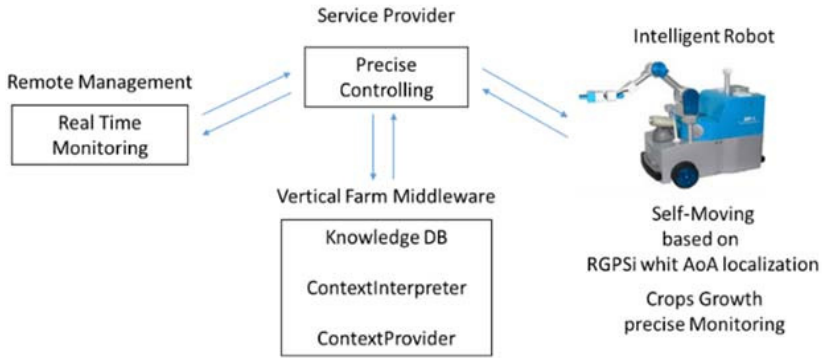
## 3 Location-Based Intelligent Robot Management Service Model

Vertical farm is managed by maintaining fixed growth environment by crop cultivation plan. Vertical farm intelligent robot control service is as one of service model required for realization of complete auto control system of vertical farm, sensors and actuators are installed in each area, and being monitored and controlled. Since inner environment of vertical farm is managed by auto control system for acceleration of crop production, operation of sensor and actuator which are components needed for growth of crop is very important.

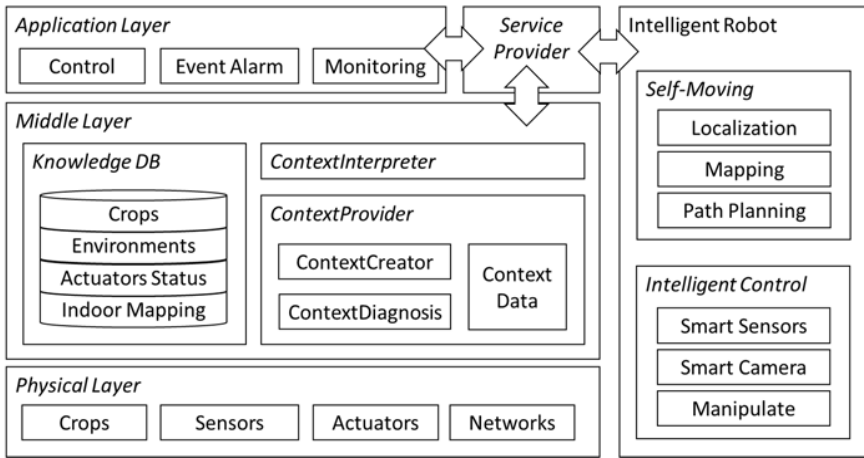
Figure 1 shows our service model concept. The concept is intelligent robot by request of service provider performed itself precise monitoring of crop's growth environment in vertical farm.

Vertical farm is managed by maintaining fixed growth environment by crop cultivation plan. Figure 2 is the architecture of this model comprise Application Layer, Middle Layer and Physical Layer.

The Physical Layer generates low-level data, which are real sensed data, crops condition, status of actuators and devices in vertical farm. The Middle Layer performed management the growth data of crops, real-time environment information, status of actuators, indoor mapping data for path planning of intelligent robot, etc. by Knowledge DB, and this layer analyzed context information by ContextInterpreter and ContextProvider. The Application Layer performed Remote Management with all monitoring of vertical farm. Intelligent Robot generates low-level data, which real



**Fig. 1.** The concept of the intelligent robot monitoring service



**Fig. 2.** Architecture for the intelligent robot monitoring service model

smart sensed data and smart image data Manipulated. It also performs localization of itself based on RGPSi with AoA algorithm for Self-Moving.

All process of this service model processed by Service Provider. This Process, ContextProvider analyze context information based on obtained low-data by Knowledge DB, Physical Layer and Intelligent Robot. Intelligent Robot performs Self-Moving and Intelligent Controlling using knowledge DB and ContextProvider by Service Provider. Therefore, it uses Intelligent Robot that will give more precise monitoring and Controlling.

## 4 Service Scenario

In this paper intellectualized environment monitoring service model using intelligent robot is proposed for auto control of vertical farm. When the event about crop growth environment occurs during the operation of vertical farm, start precise monitoring

using robot, analyze the state information using this information, then control actuator again.

This section propose the scenario for intelligent robot`s intellectualized environment monitoring service. This scenario is about sensing abnormal situation in one area where crop is growing during remote monitoring and controlling the vertical farm, intelligent robot moves to concerned area itself for precise monitoring, send the sensing and detail video, analyze according to state information of concerned area, then perform the control of actuator.

Vertical farm manager James harvested crops in one area of vertical farm. In this time, James will grow the lettuce in the area. In addition to James`s vertical farm operated with the help of Intelligent Robot even more precise control of the crop growth environment is performing. Operating Remote Management of vertical farm, the incident occurred that lettuce leaf changed the color to yellow. The following example service scenario.

### **Example Service Scenario**

1. Remote Management requests analysis of context about situation that the leaf color of lettuce change to yellow via Service Provider. (Assumption: It is possible that the situation analyzed crops nutritional disorder or crops disease and insect)
2. Intelligent Robot turn on. First, Intelligent Robot transmit itself location information to Service Provider after performing localization. Second, Service Provider perform to analysis path planning of Intelligent Robot. Finally Service Provider transmit to Intelligent Robot, and it move in the area.
3. After the Intelligent Robot arrived in the area, it collect environment information of lettuce via Smart Sensor, Smart Camera and Manipulate. And it analyzed this context information by Service Provider. (This information analyzed crops nutritional disorder.)
4. Service Provider request to control actuators of the area in order to replenish the nutrient via Remote Management.
5. This process completes after returning to original position of the Intelligent Robot and turn off.

## **5 Conclusion**

This paper proposed intelligent robot control service model for complete auto control system of vertical farm. It provides environment monitoring service by intelligent robot based on localization approach to the area where event happened and perform the precise monitoring based on state information produced from various sensors installed in vertical farm. It is expected to be able to provide highly reliable product to consumer by robot accurately performing growth monitoring of cultivating crop on its own in minimized human intervention, and by accurately controlling actuator which controls environmental factor needed for growth of crop. It is also expected to reduce the expense for first establishment of vertical farm by minimizing the installment of



detail sensor and reduce the provision of sensors installed in areas separated by crops. This service model could minimize the labor time needed for management of vertical farm. It will be also expected to greatly contribute on auto control service of vertical farm which will enlarge.

It is planned to continue the research about the system that could control growth environment by ontology based state information offering service and ontology based intelligent robot reason the situation on its own based on this service model which is one of the element of complete auto control system of vertical farm that could auto control growth environment suitable for growth of crop within vertical farm.

**Acknowledgements.** This work was supported by the Industrial Strategic technology development program, 10040125, Development of the Integrated Environment Control S/W Platform for Constructing an Urbanized Vertical Farm Funded by the Ministry of Knowledge Economy (MKE, Korea). And, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MEST) (No. 2012-0003026).

## References

1. Kim, Y.: International Technology Trends and Automation of Plant Factory. BION Special ZINE 18 (2010)
2. Lee, M.B., Kim, T., Kim, H.G., Bae, N.J., Baek, M., Park, J.W., Cho, Y.Y., Shin, C.S.: Implementation of the Closed Plant Factory System Based on Crop Growth Model. In: Park, J.J., Ng, J.K.-Y., Jeong, H.Y., Waluyo, B. (eds.) *Multimedia and Ubiquitous Engineering*. LNEE, vol. 240, pp. 83–89. Springer, Heidelberg (2013)
3. Kim, J., Im, J.: A Design of Intelligent Plant Factory Control Structure based on Ontology for Growth Environment. *Korean Society for Internet Information* 11(2), 107–108 (2010)
4. Pack, R.T.: IMA: The Intelligent Machine Architecture. PhD thesis, Vanderbilt University, Nashville, Tennessee (2003)
5. Bouwer, A., Visser, A., Nack, F., Terwijn, B.: Location Awareness, Orientation and Navigation: Lessons Learned from the SmartInside Project. In: *IUI Workshop on Location Awareness for Mixed and Dual Reality* (2013)
6. Rodney, A.B.: A Robust Layered Control System for a Mobile Robot. *IEEE J. of Robotics and Automation* RA-2(1), 14–23 (1985)
7. Park, J.W., Kim, H.G.: Ratiometric GPS Iteration Localization Method Combined with the Angle of Arrival Measurement. *International Journal of Smart Home* 7(3), 197–205 (2013)
8. Kim, H., Lee, M., Kim, T., Bae, N., Beak, M., Shin, C., Cho, Y., Park, J.: The Efficient Indoor Localization Algorithm Through the Communication between IoT Devices. In: *Conference on Korean Institute of Information & Telecommunication Facilities Engineering*, pp. 27–32 (2012)
9. Imad, A., Cyril, R., Christophe, C.: Spatial models for context-aware indoor navigation system. *Journal of Spatial Information Science* 4, 85–123 (2012)
10. Lee, I.K., Kwon, S.H.: Ontology-based Control of Autonomous Robots. In: *Korean Institute of Intelligent System Conference*, vol. 19(1), pp. 69–74 (2009)

# Design and Implementation of Intelligent Video Surveillance System Using Dual Cameras

Seung-Hyeok Yoo<sup>1</sup>, Mi-Jeong Park<sup>1</sup>, Chang-Gyun Lim<sup>2</sup>,  
Heon Jeong<sup>3</sup>, and Eung-Kon Kim<sup>4,\*</sup>

<sup>1</sup> Dept. of Computer Science, Sunchon National University, Korea  
{sh-yu,mj21}@sunchon.ac.kr

<sup>2</sup> Gwang-Ju Technopark Home Robot Support Center, Korea  
cglim@jnu.ac.kr

<sup>3</sup> Dept. of Fire Administration, Chodang College, Korea  
hjeong@cdu.ac.kr

<sup>4</sup> Dept. of Computer Engineering, Sunchon National University, Korea  
kek@sunchon.ac.kr

**Abstract.** Existing CCTV video surveillance systems have difficulty in accurately identifying tracked objects. In a pan-tilt-zoom (PTZ) video surveillance system for accurate object identification, the monitoring region has a blind spot because it only tracks a specific object. Multiple PTZ cameras can be used for removing a blind spot, but this increases the delay of camera operation due to increased computational complexity. In this paper, we propose a system to remove a blind spot occurring in an existing surveillance system, using wide-area surveillance cameras and PTZ cameras. In wide-area cameras, a particle filter is used for identifying an object. Further, the moving direction of an object is predicted using motion templates, and the delay time of camera operation is reduced by controlling the PTZ cameras efficiently.

**Keywords:** Video surveillance system, Particle filter, motion Templates.

## 1 Introduction

Recently, considerable progress has been made to improve picture quality, video transmission, and storage technology in the field of video surveillance systems. Further, network cameras directly connected to the Internet have been developed to monitor objects remotely. The application areas of these systems have been widened by integrating them with other technologies, and the system cost has been decreased, making these systems relatively affordable and increasing the use of efficient video surveillance systems [1].

Most of the current video surveillance systems use a single wide-area camera or pan-tilt-zoom (PTZ) camera. In the case of a video surveillance system using a

---

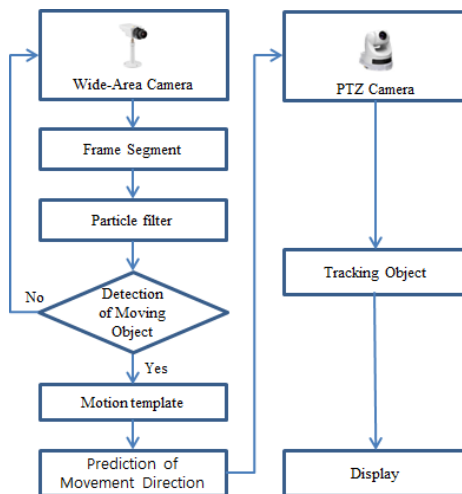
\* Corresponding author.

wide-area camera, multiple objects can be monitored through wide-area video monitoring; however, the system cannot accurately identify a single tracking object.

In the case of a video surveillance system using a PTZ camera, continuous video monitoring on a single object and accurate identification can be achieved using the PTZ movements. However, when multiple PTZ cameras are used for eliminating a blind spot, the computational complexity to calculate the camera movement schedule and interoperation between cameras is increased, thereby delaying the object tracking operation and leading to a difficulty in tracking the object accurately. To compensate for these problems, in this paper, we propose an intelligent video surveillance system using wide-area cameras and PTZ cameras. Continuous video surveillance can be carried out on a set region by using a wide-area camera, and objects can be tracked using a particle filter [2][3]. Upon the occurrence of an event, zooming in on and fixed monitoring of the object are carried out using the interoperated PTZ cameras. In the case of these cameras, motion templates [4] are used for predicting the moving direction of an object, thereby reducing the delay time of the camera operation.

## 2 Intelligent Video Surveillance System using Dual Cameras

A wide-area camera and a PTZ camera are interoperated to monitor the set region. During monitoring, once an event takes place, the wide-area camera detects the object's movement and identifies moving objects using the particle filter. The moving direction of an identified object is predicted using the motion templates. According to the predicted result value, the delay time of operation of the PTZ camera is set, and the object of interest is monitored using PTZ operations. The system allows a user to monitor the video of the wide-area camera as well as that of the PTZ camera in a specific region continuously. Figure 1 is a block diagram of the proposed system.



**Fig. 1.** Block diagram of the proposed system

### 2.1 Prediction of Moving Direction of an Object Using Motion Templates

Figure 2 is a display for the prediction of the moving direction of the object using the motion templates. While the display in (a) shows that the object is moving from left to right, the display in (b) shows that the object is moving from right to left.

Using the diagram in the display, we can predict an object’s moving direction.

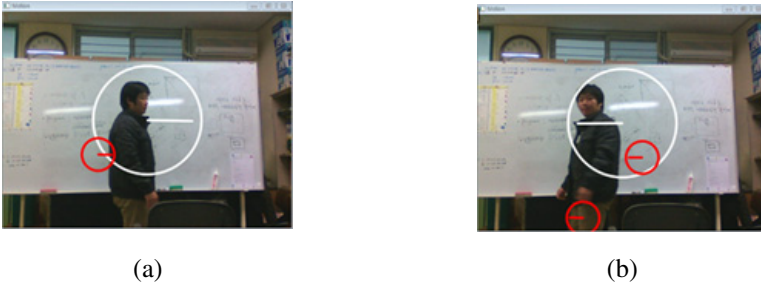


Fig. 2. Prediction of moving direction using the motion templates

### 2.2 Technique of Synchronization between Dual Cameras and Control System

In a system using dual cameras, the wide-area camera monitors the specified region fixedly, and the PTZ camera monitors a moving object using the PTZ movements.

As shown in Figure 3, the width and the height of the wide-area camera are defined. Here, the field-of-view (FOV) of the wide-area camera is 66°. The object tracking region is defined so that the coordinates of the left upper end are (x, y), and its width and height are ObjectWidth and ObjectHeight, respectively. For the PTZ camera, the value of PAN is set as 0~360°, that of TILT is 0~90°, and the ZOOMlevel is 0~9999.

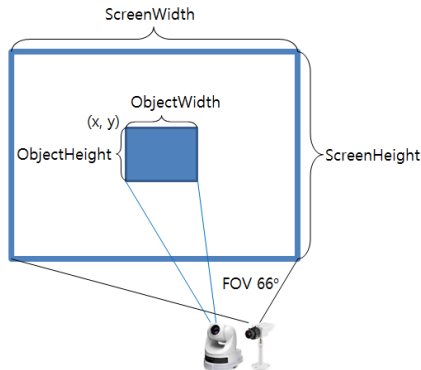


Fig. 3. Synchronization and PTZ control system

The following Formulas show the parameter values for the PTZ operations of the PTZ camera.

$$PAN = \left( x + \frac{ObjectWidth - ScreenWidth}{2} \times \frac{FOV}{ScreenWidth} \right) \tag{1}$$

$$Tilt = - \left( x + \frac{ObjectHeight - ScreenHeight}{2} \times \frac{FOV \times (3/4)}{ScreenHeight} \right) \tag{2}$$

$$Zoom = \frac{MaxZoomLevel - MinZoomLevel}{MinObjectWidth - MaxObjectWidth} \times (ObjectWidth - MinObjectWidth + MaxZoomLevel) \tag{3}$$

The implementation of the proposed video surveillance system is shown in Figure 4. It shows an example display where the specified region was monitored continuously by interoperating between the wide-area camera and the PTZ camera. The display on the left was obtained from the wide-area camera, and the one on the right was obtained from the PTZ camera.

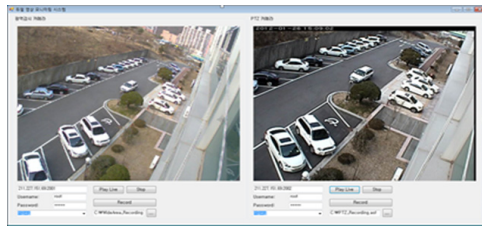


Fig. 4. Implementation of the proposed system

The operation control signal of the PTZ camera was reduced by predicting the object’s moving direction using the motion templates. The control signal frequency and the tracking rate with respect to persons and vehicles were tested as shown in Figure 5, and the results are presented in Table 1. The screen resolution was set as  $640 \times 480$  pixels for the PTZ camera. Considering that objects move because of natural phenomena, when a pixel change in an object was more than 10, it was considered an object movement.

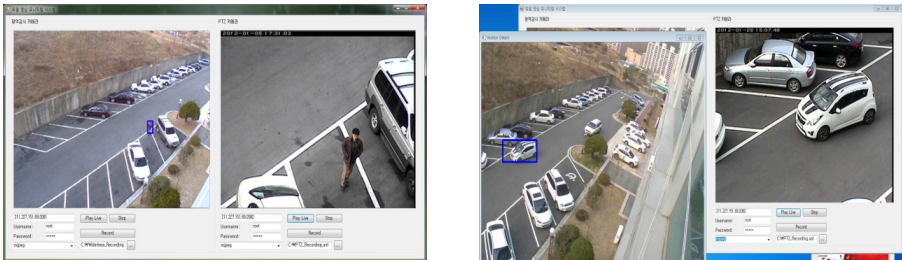


Fig. 5. Moving object tracking of the proposed system

As for the experiment method, moving persons and vehicles were observed and the frequency of signal transmission was considered the PTZ camera's movement. With respect to the criteria of the failure of tracking, if the object image could not be found in the PTZ camera display over 40 frames, it was considered a tracking failure.

**Table 1.** Control signal frequency and object tracking rate

	Tracking object			
	Person		Vehicle	
	Transmission frequency (times)	Tracking rate (%)	Transmission frequency (times)	Tracking rate (%)
Before motion templates applied	78	89	62	84
After motion templates applied	31	92	22	90

The experimental results with respect to the moving persons showed that the video surveillance system without using the motion templates transmitted a total of 78 control signals; one signal was transmitted per 10 pixels on average.

The system that applied the motion templates transmitted a total of 22 control signals by predicting the object's moving direction; that is, one signal per 46 pixels, approximately, was transmitted. Similarly, the experimental results with respect to vehicles showed that the system that applied the motion templates had better performance than the system not using the motion templates.

Further, when the motion templates were applied to the system, a high tracking rate was found because of the reduction of the operation delay of the PTZ camera. As shown in the experiment and implementation display, the proposed system provides videos of both the wide-area camera and the PTZ camera. The proposed video surveillance system solves the problem of inaccurate object identification found in the existing closed-circuit television (CCTV) video surveillance systems.

At the same time, it resolved the problem of blind spot occurrence, which was found in the existing video surveillance systems using the PTZ camera, by performing the wide-area video monitoring.

### 3 Conclusion

A CCTV video surveillance system using a single camera can have a monitoring blind spot due to an environmental factor of the camera installation location, thereby generating inaccurate identification of a monitoring object. Further, a video surveillance system using the PTZ camera can zoom in/out, track, or monitor an object by using the PTZ camera's movability, but it cannot monitor wide areas other than an object and generates operation delay due to the computational complexity of the control signal. In this paper, a novel intelligent video surveillance system was proposed to solve these problems. To implement the proposed system, object detection techniques

for video analysis and the theories on the prediction of the movement direction of an object were reviewed in order to implement the novel intelligent video surveillance system using dual cameras. Further, an experiment on the operation control of the PTZ camera was performed using object detection and an algorithm for predicting the moving direction of an object.

It was verified by the experiment and the implementation that the problem of inaccurate identification of an object, which is considered a problem in the existing CCTV video surveillance systems, was solved by the proposed intelligent video surveillance system. At the same time, the problem of the monitoring a blind spot, which is found in existing PTZ cameras during the operations of horizontal rotation, vertical rotation, and zoom in/out, was solved by continuously monitoring the video of the wide-area camera.

Future research will focus on a system that notifies the user by an alarm sound when an event occurs, as well as a mobile implementation of this system through which a user can check the system display using his/her mobile handset.

**Acknowledgements.** This research was financially supported by the Ministry of Education (MOE) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation (No. 2013H1B8A2032217).

## References

1. Chang, F.: PTZ Camera Target Tracking in Large Complex Scenes. In: Intelligent Control and Automation(WCICA), pp. 2914–2918 (2010)
2. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing* 50(2), 174–188 (2002)
3. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. *European Conference on Computer Vision* 1, 661–675 (2002)
4. Davis, J., Bradski, G.: Real-time Motion template gradients using Intel CVLib. In: *ICCCV Workshop on Frame-rate Vision*, pp. 1–20 (1999)

# Context-Aware Control Service Model Based on Ontology for Greenhouse Environment

Nam-Jin Bae, Kyung-Hun Kwak, Saraswathi Sivamani, Chang-Sun Shin,  
Jang-Woo Park, Kyungryong Cho, and Yong-Yun Cho\*

Information and Communication Engineering, Sunchon National University,  
413 Jungangno, Suncheon, Jeonnam 540-742, Korea  
{bakkepo, catania-k, saraswathi, csshin, jwpark, jkl,  
ycho}@sunchon.ac.kr

**Abstract.** IT fusion service of every kind is having the limelight in various fields lately. Agriculture area which inevitable in development of human civilization is also trying development of various IT fusion service, and the method that efficiently cultivates various crops with automatic control as an example of these service. However, most of these services have weakness that intervention of human is inevitable for exception with various variables. Therefore, this thesis proposes ontology based context aware environment control service model that could actively deal with exceptions that could happen in actual greenhouse cultivation environment without human intervention by defining relationship between each environmental factors and control factors by utilizing ontology.

**Keywords:** Ubiquitous Computing, context-aware, control service, greenhouse, ontology, u-Agriculture.

## 1 Introduction

The whole world is urbanizing due to fast development of science, environmental pollution is in serious condition. As a result, food shortage became an international issue due to reduction of site for food production due to urbanization and environmental pollution, an increase of demand for food due to increase of population, and Korea is also in serious situation that has only 30% of self-sufficiency rate of grain.

In order to solve these issues, various demonstration projects and studies are being attempted to enhance added value and productivity of agriculture by applying IT technology such as various sensors and radio communication technology to agriculture lately. Among them, Service provided by using IT fusion technology could be roughly divided by monitoring service and control service. These services control monitoring and environment of crop growth environment in real time which collects and connects environmental factors that influence the growth and development of crop through

---

\* Corresponding author.



temperature sensor, soil sensor and moisture sensor. Also it increase the production efficiency of farm by enabling analysis and provision of optimized growth environment of crop, and forecast growth based on accumulated data [1].

However, most of existing systems are subordinated to particular system and user control and monitor directly or remotely, or set the growth environment value. This has disadvantage that is hard to deal actively with exception of system and environment with various variables. In order to solve these disadvantages, this thesis proposes ontology based context aware environment control service model that could actively deal with exceptions that could happen in actual greenhouse cultivation environment without human intervention by defining relationship between each environmental factors and control factors by utilizing ontology.

Also, in order to confirm if proposed context aware control service model could actively deal with exceptional situation that happens in efficient environment control service and greenhouse cultivation environment, the experiment was preceded with virtual scenario and delivered the result.

## **2 Related Work**

### **2.1 Context-Aware Service**

Context aware means recognizing the change of system by realizing and understanding user and surrounding backgrounds in computing environment, and controlling environment according to individual preference and intention [2]. In other words, it changes according to environment on its own, and the same service provides and adapts the optimized service in different form according to its role to the same user who provides user-oriented service.

Human's daily life for example, it performs various roles according to the person who meet and circumstance facing with, and different service is needed according to different role. At this point, although the service performs the same role, the form of service changes by various surrounding environment.

The main research task to realize context-aware service could be defined; first, modeling of context about how to express and store situation information for machine to understand. Second, deduction of context about how to understand context based on information obtained from physical environment. Third, interoperability for sharing of context knowledge among independent and dispersed information equipment.[3] In the typical context modeling methods are Key-Value Models, Markup Scheme Models, Grapical Models, Object Oriented Models and Logic Based Models.

### **2.2 Ontology-Based Services**

In order to specify the concept and knowledge of certain domain, ontology defines the standard term that explains that knowledge, defines the relations between terms, and additionally includes the deduction rule that could expand this. One of the role of ontology is, solve when separate databases use the different term or identifier for the

same concept, and it is very important element that enables web based knowledge processing, knowledge sharing between application programs, and reuse [4].

The research about ontology based service is in progress in various fields such as best path search, medical service and knowledge map service by conceptualizing context data for particular situation, express and deduct that relation [5-7].

### 3 Context-Aware Control Service Model for Greenhouse Environment

#### 3.1 Control Service Architecture Based on Context-Aware

In this paper, the proposed control service architecture based on context-aware for greenhouse environment is shown in Figure 2.

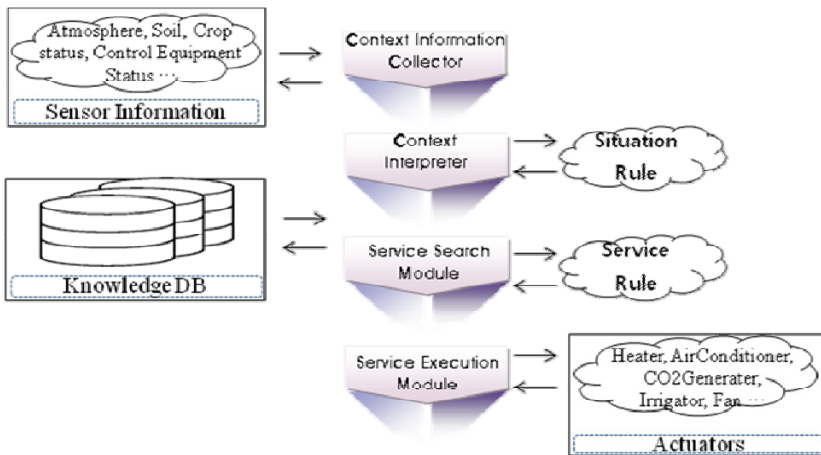


Fig. 1. Control Service Architecture based on Context-Aware

In general, the Crop required according to each growth stage have different optimal growth environment. Thus, Automatic control service for growing environment in the greenhouse have to use the context information as around status information occurred and the status information of crop.

In figure 1, the context information collector collect temperature, humidity, illumination, CO2 concentration, leaf area, number of leaves, number of fruits, size of fruit, color of fruit, etc... to confirm the growing environment data of present crop and the growth step of crop through sensors in real time.

The data collected is passed to the context interpreter. The context reasons the status of crop and growth environment by analyzing the data collected through the situation rule defined such as environment condition of each crop step, associative relation of control elements and the energy efficiency ratio. And then, its results are passed to the service search module. The service search module searches suitable

service based on context information inferred and executable control service rule. The service execution module controls cooling equipment, CO<sub>2</sub> generator, Heater, etc... based on the service inferred in the service search module.

### 3.2 Ontology Model for Automatic Control of the Optimal Growth Environment

In this paper, the proposed service model uses the ontology model for automatic control growth environment of crop in the greenhouse. In other words, the proposed ontology model is used as the common data for natural collaboration between computing resources and to minimize the intervention of human.

The figure 2 is ontology model proposed to provide context-aware control service in this paper.

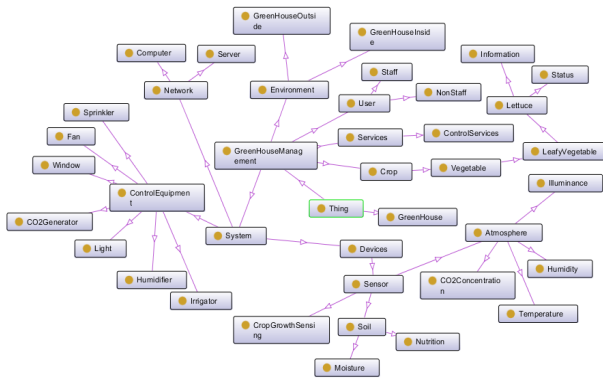


Fig. 2. Ontology Model to provide Context-Aware Control Service

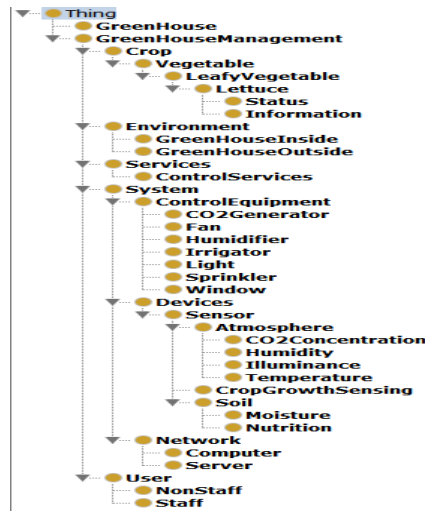


Fig. 3. Ontology Model for auto control

As the figure 3 shows, the top-class of ontology model for automatic control in the greenhouse environment is “GreenHouseManagement”. And it greatly is classified “System”, “Environment”, “Services”, and “User”. “System” class defined network, control equipment and sensor devices used in the greenhouse and “Environment” class defined the influenceable environment factors to the growth environment of crop in exterior and interior of greenhouse. “Services” is class to define the possible service through the greenhouse. But in this paper, we only defined control service. Finally, “User” class defined to assort the authority such as management, control and setting, etc... about greenhouse.

### 3.3 Control Service Process

Figure 4 shows the proposed control service process in this paper. The proposed control system acquires sensing information and knowledge information from the sensors and knowledge DB. And then, it conducts the reasoning by substitute the data to the greenhouse ontology model based on situation rule and service rule in the reasoning engine. Then system selects suitable control service through reasoning rule and control the actuator.

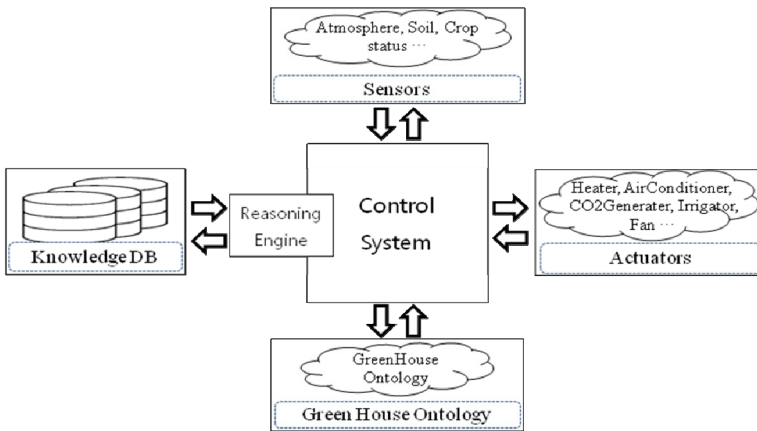


Fig. 4. Control Service Process

## 4 Conclusion

In this paper, we are suggested context-aware control service model based on ontology to complement the demerit of system in the greenhouse that to set simply environment value to provide optimal growth environment to the crop, and to need the interference of human about the particular situation.

The proposed context-aware control service model is easy to extend and to modify as define the factors considered for greenhouse environment control. And it is possible active handling such as suitable process according to the state of the system and automatic control without intervention of human when exception occurred.

In the cultivation under structure, professional agricultural knowledge is required. Therefore, in the future, to increase the efficiency of automatic control needs application of rule including diversity of rule based on expert knowledge.

And, the advanced context-aware control service that can be managed across the crop management needs through mapping of ontology in the field such as cultivation under structure and environment control.

**Acknowledgments.** This work was supported by the Industrial Strategic technology development program, 10040125, Development of the Integrated Environment Control S/W Platform for Constructing an Urbanized Vertical Farm Funded by the Ministry of Knowledge Economy (MKE, Korea).

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0014742).

## References

1. Korea food Security Research Foundation-Causes and solutions of the global food, [http://www.foodsecurity.or.kr/bbs/view.php?bbs\\_id=qnaa07&doc\\_num=57](http://www.foodsecurity.or.kr/bbs/view.php?bbs_id=qnaa07&doc_num=57)
2. Kim, S.-H.: A Study of Ontology-based Context Modeling in the Area of u-Convention. Korea Information Processing Society 28(3), 123–139 (2011)
3. Chen, H., Finin, T., Joshi, A.: Semantic Web in the context broker architecture. In: IEEE International Conference on PERCOM 2004, pp. 277–286. IDEEE Computer Society, Washington (2004)
4. McIlraith, S.: Semantic Enabled Web Services. In: XML-Web Services ONE Conference (June 2002)
5. Ma, Y.B., Kim, J.K., Lee, J.S.: Modeling and Simulation of Ontology-based Path Finding in War-game Simulation. The Korea Society For Simulation 21(1), 9–17 (2012)
6. Lee, B.-M., Kim, J., Kim, J.H., Lee, Y., Kang, U.-G.: A Customized Exercise Service Model based on the Context-Awareness in u-Health Service. Korean Institute of Information Technology 9(2), 141–152 (2011)
7. Kim, J., Park, C., Jung, J., Lee, H., Jung, H., Kim, K.M.H.: Design and Implementation of Knowledge Map Service System based on Ontology. The Korean Institute of Information Scientists and Engineers 30(1), 527–529 (2003)
8. Dey, A.K., Abowd, G.D.: Towards a Better Understanding of Context and Context-Awareness. Technical Report GIT-GVU-99-22, GVU Center, Georgia Institute of Technology (1999)
9. Dey, A.K.: Understanding and Using Context. Personal and Ubiquitous Computing 5(1), 4–7 (2001)
10. Strang, T., Linnhoff-Popien, C.: A context modeling survey. In: Workshop on Advanced Context Modelling, Reasoning and Management, UBICOMP 2004 - The Sixth International Conference on Ubiquitous Computing (2004)

# An OWL-Based Ontology Model for Intelligent Service in Vertical Farm

Saraswathi Sivamani, Nam-Jin Bae, Chang-Sun Shin,  
Jang-Woo Park, and Yong-Yun Cho\*

Information and Communication Engineering, Sunchon National University,  
413 Jungangno, Suncheon, Jeonnam 540-742, Korea  
{saraswathi, bakkepo, csshin, jwpark, yycho}@sunchon.ac.kr

**Abstract.** Recently, the agriculture is undergoing a technical drift in both innovation and technology. As a result, we have vertical farming which helps to increase the rate of food production tremendously, with the appropriate environmental factors in middle of the urban areas. Many researches have been undergone in the agricultural environment, which mainly focused on the smart services in the u-agriculture environment. Ubiquitous computing is one of the major developing technology in this fast evolving pervasive environment. Building a context aware system for the vertical farm is still complex without the shared understanding of the domains which can also be referred as set of entities, relations, functions and services. To resolve this issue, we propose the OWL based ontology model which defines the relationship between the domain factors that helps in both monitoring and controlling services of vertical farm. We have classified the basic concepts for the vertical farm environment which can be extended according to the domain of interest.

**Keywords:** Ubiquitous Computing, Context aware, Ontology, u-Agriculture.

## 1 Introduction

According to the survey, 80% of the world population is expected to live in urban areas by the year 2050[1]. It estimates that new farmland will be needed for cultivation with the present agricultural practice in a year. However, to infinitely increase farmland in accordance with the population increase seems to be impossible. The world might suffer from a shortage of the farmland in the near future, because of the rapid increasing in the population. To overcome this crisis, the existing idea to cultivate small crops indoor is being revamped with expert engineering which is termed as Vertical farming. Vertical farm is a year round crop production which produces the organic crops without any disease or pest. The crops are grown irrelevant of the weather condition and also in a suitable space with the controlled environmental factors such as temperature, luminance, humidity, etc. Currently vertical farm are rapidly evolving in the large scale production of variety of crops in the urban centers [2].

---

\* Corresponding author.

Vertical farming is a fully automated system that consists of sensors and actuators (so-called smart devices) in large scale, also interact with other system without human intervention. The vertical farming as the new agricultural evolution, to realize the technology, we need a ubiquitous computing infrastructure which needs to be aware of the context and provide appropriate data and services. The important factor in realizing this context aware system is the use of a set of common ontologies that support the communication and the relationships. Ontology is a widely accepted tool for the modelling of context information in pervasive computing and also considered to be advantages over the other modelling techniques [3]. Thus the model has been used in many fields which include smart services such as biomedical informatics, library science, architectures and agriculture.

In the Ubiquitous environment, the semantic information of contexts is needed to deal with the complex situations. But it is a known fact that either the sharing or reasoning of the context is very difficult. For the same reason, we use ontology model to fulfil our requirements. In this paper, we propose an OWL based context modelling technique for efficient monitoring and control services of environmental factors in the vertical farming. The proposed ontology focuses on the controlling and monitoring services of the environmental factors. The model can be improvised continuously with needs and requirements. As we have identified the important concepts in vertical farming environment, these concepts can be reused in large requirement without starting from the base.

## 2 Related Work

Smart services attract more attention on the people with its non-human intervention technique. So many researchers have made many researches in the ubiquitous computing field which resulted in the advancement of the computer anxiety [4]. The so-called context aware system is widely used in pervasive computing. The context aware systems helps in the effective usability without the human intervention, only taking the environment context in account. There are many application which are being profited by the context aware system which included varies field such as biomedical, library science, architectures, etc.

The ubiquitous computing took major convergence in the daily life by providing anything to anybody, at anytime and anywhere. As the agriculture being a part of human life, ubiquitous computing is taking its step into the agriculture world. Many studies has been undergoing on u-agriculture which are mainly focused on the automation process for various monitoring and controlling services. The context modelling paves a way for the computer to understand a situation from the real world. There are many context modelling techniques in practices, but according to the survey of context modelling, the ontology model is considered to be the best model. It meets the high demands of ubiquitous computing such as distributed composition, partial validation, richness and quality of information, incompleteness and ambiguity, level of formality and applicability to existing environment [7]. Therefore ontology model was also highly recommended model for vertical farming. As the vertical farming is

in developing stages, a very few attempts are made to develop an ontology model for vertical farming [8]. Therefore, we propose an OWL based Ontology model for the vertical farm by identifying the major concepts for smart environments.

### 3 A Vertical Farm Ontology

The Vertical Farm Ontology is designed for the smart services (monitoring and controlling services) in the vertical farm environment. Figure 1 represents the overview of the upper-level ontology model with the basic entities which is capable of extension according to domain of interest. It represents the semantic relationship between the entities which is more focused on the monitoring and controlling services of vertical farm. The major concepts are classified as Context, Devices, Service, Environment, Network, Location and User.

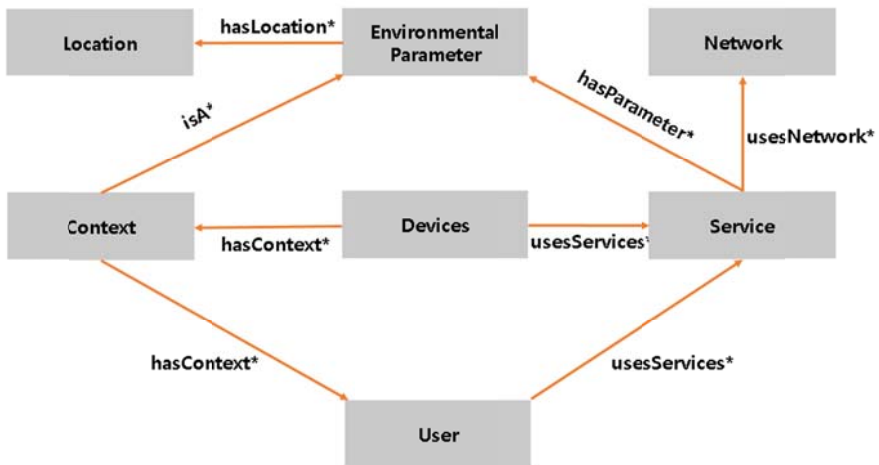


Fig. 1. Overview of Upper-level Ontology for vertical farming

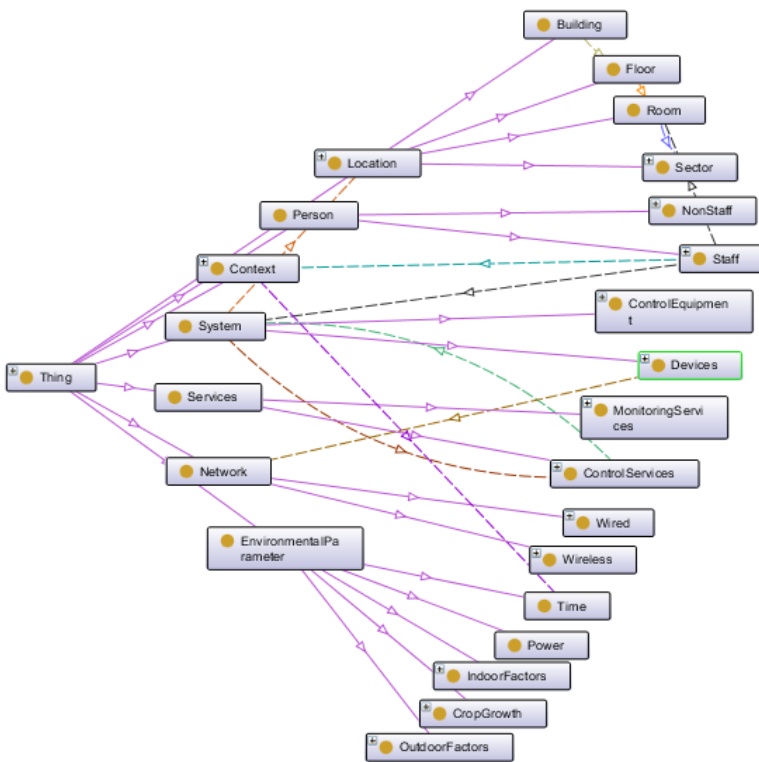
Based on the context information's perspective, the concepts are organized.

- *Context*: The context consists of set of environmental parameters of a single location on scheduled time.
- *Devices*: The devices comprise both the hardware and software devices. The control equipment, sensors, actuators and servers are referred under the Devices.
- *Services*: The Services can be automatic or manual which mainly focus on the monitoring and control services.
- *Environment*: The Environment of the vertical farm are mainly the atmospheric environmental condition such as humidity, temperature, light, CO, etc.
- *Network*: The network specifies the communication protocol which are either wired or wireless communication.



- *Location*: The Location provides the location the crop in the vertical farm environment.
- *User*: The user obtains the environmental entities and takes control over the Services manually.

In this paper, our context ontology is divided into high level ontology (Upper-level) and low-level ontology. The high level ontology captures the basic contextual entities and low level ontology is the domain specific ontologies. The design of the vertical farm ontology majorly concentrates on the services of the vertical farm for any particular crop which can also be classified through the sectors. Figure 2 shows the class representation of vertical farm ontology with classes and the first level sub classes. The sub classes are further divided according to the specific domain interest. For example, one of the sub class IndoorFactor can be expanded as atmosphere, plant and soil conditions. Here the atmosphere contains temperature, humidity, illuminance and carbon-di-oxide. Similarly, the other classes are also subdivided.



**Fig. 2.** High-level Ontology model for vertical farm

The type of properties such as object and data type property are also described in the Figure 2, where the former defines the relationship between instances of two classes and the latter defines the relationship between the instances of classes and the

RDF literals. By defining the legitimate properties, the communication between the systems becomes more reliable even without the human interaction.

## 4 Scenario for Vertical Farm Ontology

In this section, to well illustrate the vertical farming ontology model, we will discuss an experimental scenario.

Here is a farmer who manages a vertical farm firm. In each sector, there are many variety of crops grown in large quantity with its own atmospheric conditions and soil nutrients. The lettuce vegetable is grown in one of the sector in vertical farm which is known as a low-light, low-temperature crop. For this reason, an optimal temperature has to be maintained regardless of time and season.

Scenario A:

- An optimal temperature for the crop during night time is between 10 ~ 15 degree Celsius and 15 ~20 degree Celsius during day time.
- During dawn, if the temperature value exceeds 15 degree Celsius, the state of the temperature changes to “HOT” and thus changes the status of Air-Conditioner to “ON”.
- Similarly, when the sun shines on, the status of the heater is changed to “ON” and the Air-Conditioner status is changed to “OFF” to maintain the daytime temperature.

Scenario B:

- On a rainy day, the status of the outside weather is marked “RAINY”.
- Due to the outside weather (No sunlight), the CO<sub>2</sub> level decreases and the CO<sub>2</sub> generator’s status is changed to “ON”.
- The light is also switched on (Status: ON) as the weather status is “RAINY” to maintain luminance for the photosynthesis.

In the above scenario, all the process is automatic without any human intervention. The semantic interoperability to exchange and share knowledge between the different systems is achieved using the Ontology model. As mentioned in the scenario A, when the temperature varies, the values should be understood between smart plugs and the server which interacts with the knowledge base for the crop’s growth requirements. As we discuss the scenario B, the context information turns to be interrelated. In the scenario B, to maintain photosynthesis, light and the CO<sub>2</sub> generator needs to be turned on, which are interrelated to maintain the steady growth. The Context information based on location can also be considered. The CO<sub>2</sub> generator is turned on (Status: ON), is located in Sector A (isLocatedAt Sector A), in the room 1(isInRoom 1), on the floor 2 (isOnFloor 2). The vertical farming can be more benefited with the help of ontology model as the design focus on the automation process without any human intervention.

## 5 Conclusion and Future Works

In this paper, we have presented a formal context model based OWL ontology to manipulate the context information in the intelligent vertical farm environment. In the intelligent environment, the context model helps in the communication and relationship between the systems and services. The basic concepts proposed can be reused and extended for the agricultural based smart environments. We are widening our studies to bring more refined concepts which will be implemented with the prototype system for the vertical farm environment.

**Acknowledgments.** This work was supported by the Industrial Strategic technology development program, 10040125, Development of the Integrated Environment Control S/W Platform for Constructing an Urbanized Vertical Farm Funded by the Ministry of Knowledge Economy (MKE, Korea).

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0014742).

## References

1. Despommier, D.: Vertical farming, <http://www.eoearth.org/view/article/156849> (retrieved)
2. Despommier, D.: Farming up the city: the rise of urban vertical farms. *Trends in Biotechnology* 31(7), 388–389 (2013)
3. Baldauf, M.: A survey on context-aware systems. *Int. J. Ad Hoc and Ubiquitous Computing* 2(4) (2007)
4. Friedewald, M., Raabe, O.: Ubiquitous computing: An overview of technology impacts. *Telematics and Informatics* 28, 55–65 (2011)
5. Cho, Y., Cho, K., Shin, C., Park, J., Lee, E.: An Agricultural Expert Cloud for a Smart Farm. In: Park, J.J., Leung, V.C.M., Wang, C.-L., Shon, T. (eds.) *Future Information Technology, Application, and Service*. LNEE, vol. 164, pp. 657–662. Springer, Heidelberg (2012)
6. Baek, M., Lee, M., Kim, H., Kim, T., Bae, N., Cho, Y., Park, J., Shin, C.: A Novel Model for Greenhouse Control Architecture. In: Park, J.J.(J.H.), Arabnia, H.R., Kim, C., Shi, W., Gil, J.-M. (eds.) *GPC 2013*. LNCS, vol. 7861, pp. 262–269. Springer, Heidelberg (2013)
7. Strange, T., Linhoff-Popien, C.: A Context Modeling Survey. In: *The Sixth International Conference on Ubiquitous Computing, UbiComp (2004)*
8. Kim, T., Bae, N., Lee, M., Shin, C., Park, J., Cho, Y.: A Study of an Agricultural Ontology Model for an Intelligent Service in a Vertical Farm. *International Journal of Smart Homes* 7(4) (2013)

# Parallel Genetic Algorithm for Solving the Multilayer Survivable Optical Network Design Problem

Huynh Thi Thanh Binh and Nguyen Xuan Tung

School of Information and Communication Technology  
Hanoi University of Science and Technology, Vietnam

**Abstract.** We study the multilayer survivable optical network design problem. Given an undirected graph  $G_1 = (V_1, E_1)$ , a complete undirected and weighted graph  $G_2 = (V_2, E_2, c)$  and a set of customers' demands. The goal is to design connections based on customers' demands with the smallest network cost to protect the network against all failures. This paper focuses on implementing a Parallel Genetic Algorithm (PGA) for solving multilayer optical network, called MSONDP. The experimental results on real world and random instances are reported to show the efficiency of proposed algorithm comparing to the GAMSONDP in term of minimize the network cost and especially running time.

**Keywords:** Survivable Network Design, Genetic Algorithm, Parallel Genetic Algorithm.

## 1 Introduction

Network has brought more and more useful in all of life areas such as economic, culture, military, etc. So, sometimes, only a network failure also can cause serious consequences, particularly about economic. This sets a requirement for the network providers is how to ensure the reliability for their products. It is also the reason for defining Survivable Network Design Problem (SNDP).

SNDP for optical network is becoming essential and gaining much more attention. In this paper, we will solve SNDP for multiplayer optical network which is the latest model of this problem and called the Multilayer Survivable Optical Network Design Problem (MSONDP).

MSONDP problem is defined as the following: Given an undirected and unweighted graph  $G_1 = (V_1, E_1)$  and a complete undirected and weighted graph  $G_2 = (V_2, E_2, c)$  such that  $V_1 \subset V_2$ ,  $E_1 \subset E_2$  and  $c$  is weight of edges in  $G_2$ .  $G_1$  represents the logical layer,  $G_2$  represents the physical layer and each weight  $c$  is cost to connect a corresponding edge in the fact. In this problem, a set of demands  $T$  is required by customers. Each demand  $t_i$  in  $T$  consists of two node-disjoint paths  $L_i^1, L_i^2$  that connect the source node  $o_i$  to the destination node  $d_i$  in  $G_1$ , where  $L_i^1$  is a working path and  $L_i^2$  is a backup path. The goal of the problem is to design connections based on customers' demands with the smallest network cost to protect the network against all

failures. In the other word, it is to find two node-disjoint paths mapping of  $L_i^1, L_i^2$  in graph  $G_2$  for each  $t_i$  such that total cost of all demands is minimal, simultaneously, all nodes in  $L_i^1, L_i^2$ , called required nodes, must be in these two mapping node-disjoint paths in order.

Find an appropriate connection for each demand in customer's set of demands so as to minimize:

$$totalCost = \sum_{i=1}^{|T|} c(t_i) \quad (1)$$

Where  $c(t_i)$  is cost of the  $i$ -th demand,  $|T|$  is the number of customers' demands.

Since the survivability is an essential necessity in any network system, there have many research works about this topic. Many models of problem were researched and solved by specific algorithms. According to [3], with two different approaches, which are exact and approximate algorithms, this problem was solved by many ways.

With exact approach, in 2006, S.Borne et.al studied survivable IP-over-optical network design problem [2]. They gave a 0-1 integer programming formulation for this problem, described some valid inequalities and discussed separation algorithms for these inequalities, simultaneously, introduced some reduction operations. Basing on these, they proposed Branch-and-cut algorithm to solve. However, their algorithm only solves small test sets ( less than 35 nodes and 30 edges in random instances, less than 60 nodes and 102 edges in real instances). With bigger test sets, running time is quite long (about 5 hours) and many of them still have no solution.

After that, in 2009, S. Borne et. al. proposed Branch-and-Cut and Branch-and-Cut- and-Price algorithm based on the path formulation for solving the multilayer capacitated survivable network design problem [9]. However, they only solved with graphs having 10 nodes, 20 edges and the number of demands is 20.

Also following exact approach, Adrian Zymolka defined the cost-efficient design of survivable optical telecommunication networks problem and proposed Branch-and-Price with four branching rules after modeling this problem in integer linear program [8]. His problem was separated into two individually hard sub-problems, one of which is to rout the connection with corresponding dimensioning of capacities and the and the other is to seek for s conflict-free assignment of available wavelengths to the lightpaths using a minimum number of involved wavelength converters. The first problem was solved by three steps algorithm. With the second one, deriving linear program formulation, he proposed an exact Branch-and-Price method to solve. That is an integer linear programming approach for exact solution.

In 2011, MSONDP was proven to be NP-hard by S.Borne et.al [1]. They formulated this problem in terms of 0-1 linear program based on path variables. Then, they discussed the pricing problem and proved that it reduces to a shortest path problem. Using this, they proposed a Branch-and-Price algorithm. However, this is an exact approach for NP-hard problem, so they can only offer solutions for the maximum input is 17 nodes on  $G_1$ , 20 nodes on  $G_2$  and 25 demands.

In 2012, Binh and Ly proposed a genetic algorithm called GAMSONDP to solve MSONDP. This algorithm solved the large test sets that Branch-and-Price [1] did not. However, GAMSONDP exists a problem that it takes so much running time. We

propose parallel genetic algorithm (called PGAMSONDP) for solving this problem to overcome it. Proposed algorithm is expected not only solve larger problem instances, but also reduces running time.

## 2 Proposed Algorithm

### 2.1 Genetic Algorithm

**Individual Representation:** PGAMSONDP uses individual representation in [12]. We create individuals to initialize the initial population by “path-finding” algorithm.

**Crossover Operator:** In PGAMSONDP, we apply two kinds of crossover operator. One of them is “chromosome crossover operator” which was presented in [12]. The other called “modified path crossover operator”. Randomly merge working path of parent 1 and backup path of parent 2.

**Mutation Operator:** We apply three mutation types: delete a node [12], replace a path segment, and chromosome renew.

In “replace a path segment”, we choose a chromosome  $t_i$  in an individual randomly, then reference to the  $i$ -th demand, choose two consecutive nodes  $(n_j, n_{j+1})$  either in  $L_i^1$  or in  $L_i^2$  randomly, use Dijkstra algorithm to find a path from  $n_j$  to  $n_{j+1}$ , replace  $[n_j, n_{j+1}]$  segment in  $t_i$ 's working gene if  $(n_j, n_{j+1})$  are selected in  $L_i^1$  or backup gene if  $(n_j, n_{j+1})$  are selected in  $L_i^2$  by the found path.

The “chromosome renew” is implemented as follows: select an individual randomly, then choose  $k$  ( $0 < k < |I|$ ) chromosomes in this individual such that their total cost is not equal to the individual's. After that, the  $(|I| - k)$  remaining chromosomes are created by above “path-finding” algorithm.

### 2.2 Parallel Genetic Model

Our proposed parallel genetic algorithm is based on island GA. We make a computer as the master and others as slaves. The main idea of our algorithm is that: the master will read input data and send to its slaves. After that, the master will play role as “a store” to keep individuals sent by slaves then send back to them if required. And each slave bases on received information to initialize its own population, implements evolutionary process by itself, then sends its best individuals to the master after each generation. The send – receive process is implemented until the stop condition is satisfied. The following is PGAMSONDP's pseudo code:

```

1 Initialize parallel environment
2 Master m read data input p
3 Slaves ss receive p
4 While(!stopCondition) do
5     receivedGroup <- ss receive individuals from m
6     createGroup <- ss initialize its individuals

```

```

7      evaluate adaptation
8      select
9      crossover operators
10     mutation operators
11     send best individuals to m
12 End while
13 return best individual

```

### 3 Experimental Results

*Problem Instances and Experiment Setup:* We implement PGAMSONDP on both random and real world instances that were presented in [12]. In our algorithm, the population size is 2000 and number of slave is 2, 3, 5, 7. Crossover rate is 40% and mutation rate is 15%. Our system runs 10 times for each instance set.

*Computational Results:* The experiment results show that the min and mean cost found by PGAMSONDP are better than the one found by GAMSONDP. The running time of PGAMSONDP to find the best result is decrease when the number of the slave increase. Especially, PGAMSONDP can find the result on some large problem instances which can not be solved by other known algorithms. The ratio between the computation time and communication time decrease as increasing the number of machines (2, 3, 5, 7).

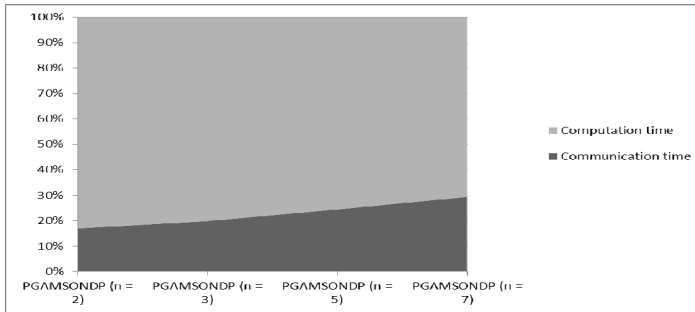
**Table 1.** The results found by PGAMSONDP over 10 running times

Instances		a40_35_2	a40_35_80	a60_55_2	g40_35_2	g40_35_80	g60_55_2
GAMSONDP	best	907	34927	1217	13	364	12
	average	927	35978	1226	13	382	12
	time	814	112798	2604	72	21026	238
PGAMSONDP (n = 2)	best	888	20858	1211	13	331	12
	average	940	21232	1226	13	343	12
	time	51	254	111	35	251	86
PGAMSONDP (n = 3)	best	862	20222	1199	13	327	12
	average	910	21238	1220	13	346	12
	time	39	231	74	28	229	67
PGAMSONDP (n = 5)	best	854	20250	1199	13	313	12
	average	888	20467	1213	13	366	12
	time	31	225	54	22	215	49
PGAMSONDP (n = 7)	best	791	19790	1048	13	310	12
	average	805	20084	1051	13	349	12
	time	31	210	45	20	204	40

Best: The best result found by the algorithm after 10 running time  
Average: The average result found by the algorithm after 10 running time  
Time: The average running time to find best result

**Table 2.** The results found by PGAMSONDP over 10 running times on instances that gamsondp could not solve

Instances		a60_55_1 50	a80_75_4 0	a80_75_ 150	a100_95 _2	a100_95_ 150	g60_55_ 150	g80_75_ 40	g80_75_ 150	g100_95 _150
n=2	Best	111320	14785	79600	1325	102742	748	147	1085	1501
	Avg	115753	14813	79600	1325	102742	774	174	1226	1630
	Time	827	404	1311	260	1950	638	463	1287	1997
n=3	Best	107320	14280	69758	1325	78455	480	160	708	889
	Avg	108753	14280	69758	1345	78455	513	173	736	965
	Time	758	335	1165	177	1692	559	382	1232	1698
n=5	Best	78519	13033	56849	1325	65281	374	151	459	530
	Avg	80582	13033	56849	1325	65281	382	162	469	566
	Time	688	293	1040	126	1483	529	460	1170	1449
n=7	Best	63960	12779	50581	1275	59176	349	150	383	432
	Avg	65312	12779	50581	1275	59176	349	156	392	465
	Time	629	273	936	101	1504	510	357	1243	1562



**Fig. 1.** Average communication time and computation time found by PGAMSONDP  
Problem instance: a6\_4\_2

## 4 Conclusion

In this paper, we proposed the new parallel genetic algorithm for solving MSONDP called PGAMSONDP. We experimented on 8 large random and 8 large real world instances. With each data set, we run 10 times to take the best and average solutions. The results show that our algorithm is effective not only about costs, but also about running time. Besides, our algorithm can give solutions for instances that GAMSONDP could not solve.

In the future, we are defining this problem with the multi-objective and solving it. We also hope that we can solve this problem with larger test sets by these approaches as well as different ones.



**Acknowledgement.** This work was supported by the project funded by the Ministry of Science and Technology, under grant number KC.01.01/10-15. The High Performance Computing center at Hanoi University of Science and Technology provides the facility for the research.

## References

1. Borne, S., Gabrel, V., Mahjoub, R., Taktak, R.: Multilayer Survivable Optical Network Design. In: Pahl, J., Reiniers, T., Voß, S. (eds.) INOC 2011. LNCS, vol. 6701, pp. 170–175. Springer, Heidelberg (2011)
2. Borne, S., Gourdin, E., Liau, B., Mahjoub, A.R.: Design of survivable IP-over-Optical network, pp. 41–73. Springer Science and Business Media (2006)
3. Kerivin, H., Mahjoub, A.R.: Design of survivable networks: A survey. *Networks* 46(1) (2005)
4. Bucsics, T., Raidl, G.: Metaheuristic Approaches for Designing Survivable Fiber-Optic Networks. Institute for Computer Graphics and Algorithms of the Vienna University of Technology (2007)
5. Reeves, C.R., Rowe, J.E.: Genetic algorithms-principles and perspectives: A guide to GA Theory. Kluwer Academic Publishers (2003)
6. Javed, M.S., Thulasiraman, K., Xue, G.(L.): Logical Topology Design for IP-over-WDM networks: A Hybrid Approach for Minimum Protection Capacity. In: ICCCN (2008)
7. <http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/tsp/> (last visited June 2012)
8. Zymolka, A.: Design of Survivable Optical Networks by Mathematical Optimization. Ph.D. Thesis, Mathematik und Naturwissenschaften der Technischen Universität, Berlin (2007)
9. Borne, S., Gourdin, E., Klopfenstein, O., Mahjoub, A.R.: The Multilayer Capacitated Survivable IP Network Design Problem: Valid Inequalities and Branch – and – Cut. In: INOC (2009)
10. Vinodkrishnan, K., Durresi, A., Chandhuk, N., Jain, R., Jagannathan, R., Seetharaman, S.: Survivability in IP over WDM networks, pp. 79–90. IOS Press (2011)
11. Wagner, D., Pferschy, U., Mutzel, P., Raidl, G.R., Bachhiesl, P.: A directed cut model or the design of the last mile in real-world fiber optic networks. In: Fortz, B. (ed.) Spa, Proceedings of the International Network Optimization Conference 2007, Spa, Belgium, pp. 103/1-6 (2007)
12. Binh, H.T.T., Ly, H.D.: Genetic Algorithm for Solving Multilayer Survivable Optical Network Design Problem. *International Journal of Machine Learning and Computing*, 812–816 (2012) ISSN: 2010-3700

# A Context-Aware Collaborative Filtering Algorithm through Identifying Similar Preference Trends in Different Contextual Information

Phung Do<sup>1</sup>, Hiep Le<sup>1</sup>, Vu Thanh Nguyen<sup>1</sup>, and Tran Nam Dung<sup>2</sup>

<sup>1</sup>University of Information Technology, Vietnam National University HoChiMinh City  
{phungdtm,nguyenvt}@uit.edu.vn, lenguyenhaohiep@gmail.com

<sup>2</sup>University of Science, Vietnam National University HoChiMinh City  
trannamdung@yahoo.com

**Abstract.** Three main approaches in Context-Aware Recommender Systems (CARSs) are pre-filtering, post-filtering and contextual modeling. Incorporating contextual information into main process is the different point of contextual modeling from two first approaches. In this paper, we first propose a new context-aware collaborative filtering (CACF) algorithm with contextual modeling approach combined from a clustering technique and matrix factorization method named Similar Trends Identifying (STI). We then compare the proposal with various matrix factorization-based algorithms. Overall, the STI algorithm outperforms some compared algorithms in terms of evaluation metrics and available contextual data sets.

**Keywords:** collaborative filtering, contextual modeling, clustering.

## 1 Introduction

Recommender Systems (RSs) are software tools having techniques to provide suggestions with suitable items for users. However, RSs do not take into account additional information such as time, place, companion and others which can affect preferences of users. This additional information is called context.

Context (or contextual information) is a multifaceted concept and applied widely in different disciplines [1]. In computational environment, the definition “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application” is also suitable for additional information in RSs [2]. In RSs, entity is often a user, an item and the rating from a user over an item. There are two main views of context: representational and interactional [3]. Generally, representational view considers context a predefined set of observable attributes meanwhile interaction view defined context dynamically, it means no enumeration of contextual information is provided beforehand. In this paper, as in the vast majority of RSs with contextual information, the representational view is adopted.

CARSs indicates RSs which incorporate contextual information into recommendation process to model and predicting tastes of users. There are two approaches for CARSs to deal with contextual information and such approaches are classified into two groups: (1) recommendation via context-driven querying and search, and (2) recommendation via contextual preference elicitation and estimation. However, the second approach is the most commonly used in CARSs and it consists of contextual pre-filtering, contextual post-filtering and contextual modeling paradigms [1]. The three techniques above are abstract techniques, to implement them, approaches can be applied consist of content-based filtering, collaborative filtering, hybrid approach and others. Nevertheless, collaborative filtering (CF) is one of most successful techniques [4] that is commonly used in the majority of RSs and it is extended to deal with contextual information called context-aware collaborative filtering (CACF).

CACF can be classified into 3 groups: memory-based, model-based and hybrid in the same way as CF. Prior works with memory-based CACF methods computes the weigh or similarity values between users or items in various ways. These similarity values could be computed by adjusting data including contextual information into 2-dimensional matrix data (*User*×*Item* matrix) in specific values of contextual information and applied similarity CF measures (pre-filtering techniques). Another way is computing the weigh values between pairs (*user*, *item*, *context*) [1] or pairs (*user*, *context*) [5]. Two Weigh and Filter approaches [5] use neighborhoods to compute the probability a user purchases an item in specific contexts with contextual post-filtering technique. Memory-based CACF methods suffer from sparse data (sparsity problems) because these methods base on co-rated items in various contexts. Model-based CACF techniques use data mining or machine learning algorithms to build models from pure data and to make recommendations. Various models using Bayesian networks [6] or extending matrix factorization technique [7] are constructed. As in RSs, model-based methods are better than memory-based ones in addressing the sparsity and scalability problems when the number of users or items tremendously increases. Besides, hybrid methods combine with analyzing semantic using ontology are deployed to attempt in reducing limitations all above methods.

In order to address the sparsity problem in CARSs with a large number of contexts, in this paper, we propose a new model-based CACF algorithm with contextual modeling approach in section 2. Section 3 is experimental setup to compare my proposal with some algorithms. Finally, section 4 is for conclusion and future works.

## 2 The Similar Trends Identifying Algorithm

The Similar Trends Identifying (STI) algorithm is basically based on clustering technique and matrix factorization in RSs. This algorithm with main idea “similar pairs (*item*, *context*) are clustered and converted to new products to reduce contextual dimensions. After that, matrix factorization is applied combined with analyzing the effect of contextual information to predict and recommend potential items to the active users” consists of 5 main steps.

### 2.1 Step 1. Building Item Profiles

Each  $ItemProfile(i,c)$  in the item profiles is a representation vector for the  $i^{th}$  item and context  $c$ , such vector contains ratings for all users pertaining to this item and context.  $ItemProfile(i,c)=(r_{1ic}, r_{2ic}, \dots, r_{uic}, \dots, r_{nic})$ ,  $u=1..n$ ,  $r_{uic}$  is the rating from the  $u^{th}$  user for the  $i^{th}$  item in context  $c$  and  $r_{uic} = 0$  if there is no rating. Notice that,  $c$  is the combination value of contexts. For instance, we have two contexts  $X$  and  $Y$  with numbers of values in  $X$  and  $Y$  are 2 and 3 respectively. Therefore, we have  $2 \times 3 = 6$  combinations of contextual information and with 20 items, there are  $6 \times 20 = 120$  item profiles.

### 2.2 Step 2. Clustering

The main purpose of this step is to identify similar trends in giving the preferences or ratings to items in different contexts. It is handled by applying a clustering technique to obtain  $k$  clusters from item profiles (*step 1*).

There is a number of clustering methods such as k-Means, G-Means, DBScan, Clara and others using Manhattan, Euclidean and other distances to find clusters. However, neighborhoods in RSs are frequently identified through similarity values between users or items, which normally calculated by ratings of co-rated items. Therefore, we present the graph-based clustering technique adapting collaborative filtering.

Item Profiles set ( $D$ ) contains  $N$  numeric vectors. The  $D$  set is considered as an undirected graph  $D=\{V,E\}$ ,  $V$  is a set of  $N$  vertices or nodes corresponding to  $N$  numeric vectors and  $E$  is a set of edges or lines between to vertices. For each vertex  $X$ , building an edge from vertex  $X$  to vertex  $Y$  in case the weigh value between  $X$  and  $Y$  is maximum, where the weigh value is cosine value computed by following formula:

$$weigh(X,Y) = \frac{\sum_{i \in n} X_i Y_i}{\sqrt{\sum_{i \in n} X_i^2} \sqrt{\sum_{i \in n} Y_i^2}} \tag{1}$$

Where  $n$  indicates the length of vectors  $X$  and  $Y$  (or the number of users). It is noticeable that  $weigh(X,Y)$  is computed with values  $X_i$  and  $Y_i$ ,  $i=1..n$  larger than 0 from vector  $X$  and  $Y$  respectively. The value of  $weigh(X,Y)$  indicates the similarity between item profiles  $X$  and  $Y$  which means that users intend to give the similar ratings for the item profiles  $X$  and  $Y$ .

In order to cluster data input  $D$  into  $k$  clusters, one of the graph methods such as Warshall, DFS (Depth First Search) or BFS (Breadth First Search) is applied to find out connected components of undirected graph  $D$ . As the result of this stage, the  $k$  connected components are obtained and it also means  $k$  clusters. In particular, BFS algorithm is used to find  $k$  clusters in this paper. In addition, we also implement G-Means [8] to make comparison with the graph-based clustering technique revealed in the experimental setup section.

### 2.3 Step 3. Building 2-Dimensional Matrix

This step aims to reduce contextual information dimension so that the sparsity problem [4] can be partially tackled. After clustering (*step 2*), a new products set  $P=\{p_1,p_2, \dots,p_j,\dots,p_k\}$  including  $k$  new products corresponding to  $k$  clusters is constructed. The rating value for a new product  $p_j$  for each user is the average rating of this user in the  $j^{\text{th}}$  cluster (notice that we only consider rating values larger than 0 to compute average rating for each user). The matrix  $U \times I \times C$  ( $U$ ser  $\times$  Item  $\times$  Context) is transformed into the 2-dimensional matrix  $U \times P$  ( $U$ ser  $\times$  Product) in this stage. Regarding testing phase, the predictive rating value for the  $i^{\text{th}}$  item in context  $c$  is replaced by the predicting value for a new product  $p_j$  with the  $j^{\text{th}}$  cluster contains  $ItemProfile(i,c)$ .

### 2.4 Step 4. Analyzing the Effect of Contextual Information

The value of  $ContextEffect(c,u)$  indicates the effect value from context  $c$  to the  $u^{\text{th}}$  user. The value of  $ContextEffect(c,u)$  is computed as the average rating value of the  $u^{\text{th}}$  user in context  $c$  (symbolized  $avg(u,c)$ ) subtracts the overall average rating value of the  $u^{\text{th}}$  user (symbolized  $avg(u)$ ).

$$ContextEffect(c,u) = avg(u,c) - avg(u) \quad (2)$$

### 2.5 Step 5. Predicting

In order to predict the rating value for the  $u^{\text{th}}$  user, the  $i^{\text{th}}$  item in context  $c$ , applying first matrix factorization (SVD) [1] technique with 2-dimensional matrix  $U \times P$  from *step 3* with result  $MF(u,p_j)$  depicts the predictive rating value for the  $u^{\text{th}}$  user and the  $j^{\text{th}}$  product  $p_j$  (the  $j^{\text{th}}$  cluster contains  $ItemProfile(i,c)$ ).

Then, the predictive rating value for the  $u^{\text{th}}$  user, the  $i^{\text{th}}$  item in context  $c$  is computed as following expression:

$$\hat{r}_{uic} = MF(u, p_j) + ContextEffect(c,u) \quad (3)$$

The predictive rating values are adjusted in range rating scale  $[min, max]$ . The value of  $\hat{r}_{uic}$  is  $min(max)$  when  $MF(u,p_j)+ContextEffect(c,u)$  smaller (larger) than  $min(max)$ .

## 3 Experimental Setup

We compare the STI with Item Splitting [9] and Context-Aware Matrix Factorization (CAMF) [7] methods. In particular, the Item Splitting method which reduces contextual dimensions and applies matrix factorization to predict unknown ratings is called ISMF. We use different clustering methods in the *step 2* of the STI algorithm, it is the STI algorithm with graph-based clustering technique (STI-GB) (*section 2*) and G-Means algorithm [8] (STI-M). All compared algorithms are based on Matrix Factorization technique (SVD) [1] and these algorithms are configured same parameters.

We use three real world data sets for comparisons and evaluations. The first one is AIST context-aware food preference dataset called Food in [6], this dataset after pre-processing duplicated records contains 5300 ratings from 212 real users over 20 food menus in the real and imaginary level of hungry situations. Therefore, we use these situations along with level of hungry as contextual information. The second one is LDOS-Comoda dataset called Comoda which is used in [10]. Comoda includes 2248 ratings from 82 users in 1225 movies, rating values are judged along with 12 different contextual information. The last one is HMusic collected by our application, in this dataset, there are approximately 2003 ratings from 59 users and 379 songs. Contexts in HMusic contain time stamp when a user listens to track music and the user's mood at that time. All data sets include rating values in range 1–5.

Data sets are partitioned into 5 disjoint sets and we apply 5-fold cross validation with the most common evaluation metrics such as Mean Absolute Error (*MAE*) and Root Mean Square Error (*RMSE*) [1] to estimate algorithms.

**Table 1.** The result of evaluation

<b>Dataset</b>	<b>Metric</b>	<b>ISMF</b>	<b>CAMF</b>	<b>STI-GB</b>	<b>STI-M</b>
<b>Food</b>	MAE	0.876	0.843	0.765	<b>0.719</b>
	RMSE	1.104	1.065	1.027	<b>0.961</b>
<b>Comoda</b>	MAE	0.841	0.874	<b>0.809</b>	0.832
	RMSE	1.067	1.172	<b>1.055</b>	1.064
<b>HMusic</b>	MAE	0.862	0.881	<b>0.784</b>	0.925
	RMSE	1.128	1.174	<b>1.121</b>	1.321

As can be seen in Table 1, the STI-GB outperforms all compared algorithms in all data sets except it achieves *MAE*, *RMSE* higher than the STI-M method only in Food data set. It can be explained that the Food's sparsity is approximately low (about 79.2% compared to nearly 99.99% in Comoda and HMusic) and the number of co-rated items is high that make the STI-M performs well only in this dataset. On the other hand, STI-M dominates ISMF, CAMF in terms of Food and Comoda data sets and better than STI-GB only in Food data set, however, it shows the worst performance under testing with HMusic data set. Moreover, the STI method with G-Means clustering technique is more complicated and more time-consuming than the STI algorithm with graph-based clustering technique. In our experimental setup, graph-based technique with simple concepts namely cosine metric and connected components is easy to figure out and implement comparing to G-Means, a very complicated technique. G-Means using distance metric to identify similar objects sometimes does not fit in finding pairs (*item*, *context*) with the similar trends in rating items in RSs.

To sum up, the STI-GB is the most effective algorithm in predicting unknown rating values under available contextual data sets in this experiment.

## 4 Conclusion and Future Works

In this paper, we propose a novel algorithm (STI) for CACF with contextual modeling approach. Our method clusters similar pairs (*item*, *context*) to convert the multi-dimensional matrix  $User \times Item \times Context$  into the 2-dimensional matrix  $User \times Product$  to reduce sparsity problem. We also compare the STI algorithm with different clustering techniques, which are G-Means algorithm, and the graph-based clustering techniques. The STI algorithm with graph-based clustering technique outperforms compared algorithms in terms of particular data sets and evaluation metrics mentioned.

Due to lack of contextual data sets, we just evaluate algorithms in small-scale data sets. We hope that we will try to run the STI algorithm in large-scale data sets and provide enhancement techniques. Furthermore, we will propose new CACF algorithms and apply them into e-learning recommendation applications with contexts.

**Acknowledgment.** This research is funded by Vietnam National University HoChi-Minh City (VNU-HCM) under grant number 01/CNTT/2013/911VNUHCM-JAIST.

## References

1. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B.: Recommender Systems Handbook. Springer-Verlag New York, Inc., New York (2010)
2. Dey, A.K.: Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 4–7 (2001)
3. Dourish, P.: What we talk about when we talk about context. *Personal Ubiquitous Comput.* 8, 19–30 (2004)
4. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. *Adv. in Artif. Intell.* 2009, 4:2–4:2 (2009)
5. Panniello, U., Gorgoglione, M.: Incorporating context into recommender systems: an empirical comparison of context-based approaches. *Electronic Commerce Research* 12, 1–30 (2012)
6. Ono, C., Takishima, Y., Motomura, Y., Asoh, H.: Context-Aware Preference Model Based on a Study of Difference between Real and Supposed Situation Data. In: Houben, G.-J., McCalla, G., Pianesi, F., Zancanaro, M. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 102–113. Springer, Heidelberg (2009)
7. Baltrunas, L., Ludwig, B., Ricci, F.: Matrix factorization techniques for context aware recommendation. In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 301–304. ACM, New York (2011)
8. Hamerly, G., Elkan, C.: Learning the K in K-Means. In: *Neural Information Processing Systems*, p. 2003. MIT Press (2003)
9. Baltrunas, L., Ricci, F.: Context-based splitting of item ratings in collaborative filtering. In: Proceedings of the Third ACM Conference on Recommender Systems, pp. 245–248. ACM, New York (2009)
10. Odić, A., Tkalčič, M., Tasič, J.F., Košir, A.: Relevant Context in a Movie Recommender System: Users' Opinion vs. Statistical Detection. In: Adomavicius, G. (ed.) Proceedings of the 4th Workshop on Context-Aware Recommender Systems in Conjunction with the 6th ACM Conference on Recommender Systems, RecSys 2012 (2012)

# Survivable Flows Routing in Large Scale Network Design Using Genetic Algorithm

Huynh Thi Thanh Binh and Son Hong Ngo

School of Information and Communication Technology  
Hanoi University of Science and Technology, Vietnam  
{binhht, sonnh}@soict.hut.edu.vn

**Abstract.** We study the survivable network design problem (SNDP) for simultaneous unicast and anycast flows in networks where the link cost follows All Capacities Modular Cost (ACMC) model. Given a network modeled by a connected, undirected graph and a set of flow demands, this problem aims at finding a set of connections with a minimized network cost in order to protect the network against any single failure. In this paper, we propose a new Genetic Algorithm with an efficient encoding to solve the SNDP in networks with ACMC model (A-SNDP). Our encoding scheme is simple and allows large search space. Extensive simulation results on real topology instances show that the proposed algorithm is much more efficient than the Tabu Search and other conventional Genetic Algorithm in terms of minimizing the network cost.

## 1 Introduction

There are many types of connection for data transmission over the Internet. A unicast connection is from one node to another, while an anycast connection is also from one node to another but the destination node has a one or many replicated servers which back up for it. Anycast has in recent years become increasingly popular for adding redundancy to many Internet services [1-3]. In the Internet, any network failure can cause serious consequences. Therefore, the design of survivable networks is a crucial problem. In the survivable network design problem (SNDP) with unicast and anycast flows, the objective is to minimize the network cost to protect network against failures. To guarantee the survivability, we adopt the protection approach [4-9] in that each connection must include a working path and a link-disjoint backup path.

In [4] Gładysz et al. considered the SNDP for networks using ACMC (All Capacities Modular Cost) link cost model (A-SNDP). In the ACMC model, a link has many bandwidth levels, each level has a corresponding cost. Many flow demands can go through the same link thus the link cost is defined as the total of required bandwidth from all demands on that link. The network cost is defined as the total of all link cost.

In this problem, the cost of a link follows All Capacities Modular Cost model [3], that means a link has many bandwidth levels, each level has a corresponding cost. Many flow demands can go through the same link thus the cost of a link is defined as the total of required bandwidth from all demands on that link. The goal is to find a set of connection for all demands such that the network cost (NCost) is minimized:



$$NCost = \sum_i c_i$$

$$\text{where } c_i = C_k, \text{ if } B_{k-1} < \sum_j R_{ij} < B_k$$

here,  $c_i$  is the cost of link  $i$ ;  $B_k$  and  $C_k$  is bandwidth and corresponding cost in level  $k$ ;  $R_{ij}$  is the required bandwidth from the demand  $j$  on the link  $i$ .

The A-SNDP problem is defined as follows. Given a network modeled by an undirected graph where the link cost follows ACMC model and a set of survivable flow demands between node pairs with corresponding bandwidth and type of connection (anycast or unicast), this problem aims at finding a set of connection for all flow demands such that the network cost is minimized. The authors [4] also proposed a heuristic using Tabu Search but their result is still far from optimal approach.

There have been many works on using genetic algorithm (GA) to solve NP-hard problem, particularly in network design problem [8, 12]. In [12], we have developed a genetic algorithm for A-SNDP called CDE-GA that uses Connection Database Encoding for individual representation. However, this encoding method is complex and search space is not diverse enough, thus it limits the performance of GA approach in solving A-SNDP problem.

In this paper, we propose new individual encoding scheme called Complete Connection Encoding (CCE) algorithm for solving A-SNDP. CCE can help to enlarge the search space to find better solution in large network instances. We then design the evolution operators using CCE and simulate our proposed algorithm large instances (Germany50, America). Results obtained in experimentation show that our algorithm are much better than the compared algorithms in terms of minimizing network cost.

## 2 Proposed Algorithm

In the design of a Genetic Algorithm, the encoding is the most important task. There are some methods to encode each individual in a population, such as binary encoding, integer encoding... In this paper, we propose a new encoding mechanism, called Complete Connection Encoding (CCE) to encode individuals in GA. An individual built by CCE is presented as follows: Each individual  $T$  (i.e. a complete solution) is a set of substrings. Each substring  $T_i$ , represents a flow demand  $i$  and has two parts: the working path and the backup path. Illustration of an individual is shown in Fig. 1.

To initialize an individual  $T$ , we create each of its substring  $T_i$  in turn. The working path of  $T_i$  is built by using a path finding algorithm. After that, all the link of this path will be deleted from the graph to find the backup path of  $T_i$ , using the same path finding algorithm. Therefore, to initiate an individual that represents a solution of A-SNDP, we need the time is  $O(|D|n^2)$  where  $n$  is the number of nodes.

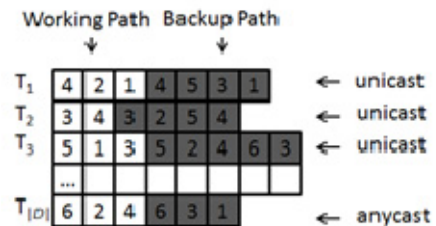


Fig. 1. An example of individual in CCE

For *crossover operator*, we apply two different crossover operators: one-point crossover and path crossover. In one-point crossover, we combine the substrings from  $T_1 \dots T_i$  with  $T'_{i+1} \dots T'_n$  to create the child.

In *path crossover*, we combine the working path of the first parent  $T$  with the backup path of the second parent  $T'$  to create the child  $T_{\text{child}}$ . With the second type of crossover, sometimes, the working path and the corresponding backup path are not link-disjoint anymore. Thus, we have to check the child again and if any substring violates the link-disjoint condition, it will be replaced by the corresponding substring from its parent.

For *mutation operator*: We choose some individuals in the current population randomly. Then, with each selected individual, we choose one substring  $i$  randomly and replace its working path as well as backup path by other couple of link-disjoint paths satisfying the demand  $i$ .

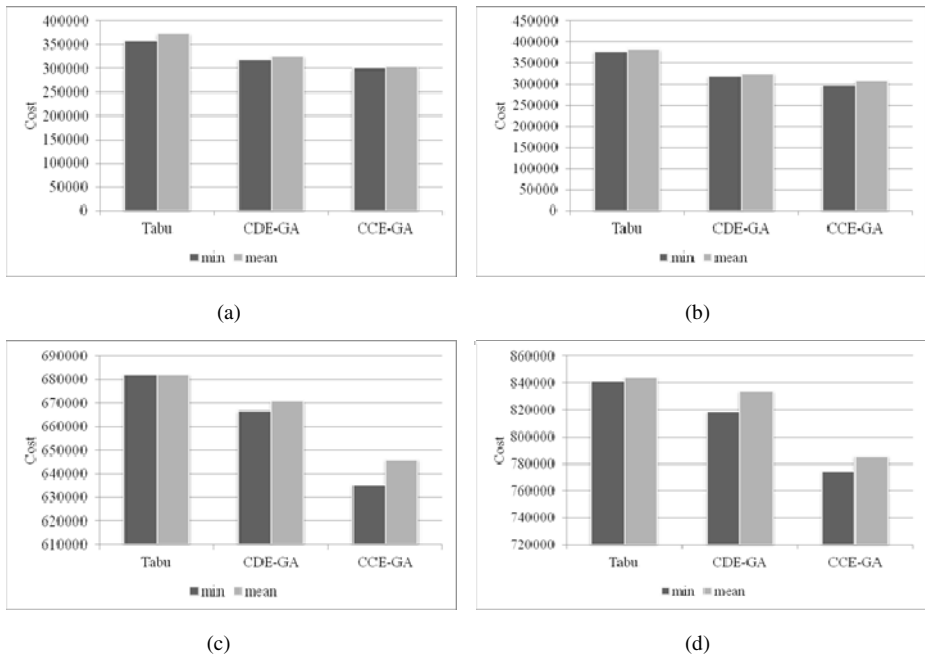
Below is pseudo-code of our proposed algorithm. In line 4, 7 and 14, noGeneration, noCrossover and noMutation stand for the number of generation, the number of crossover operator and the number of mutation operator, respectively.

```

1. Procedure CCE-GA
2. Begin
3.   initPopulation(P) /*Init a population,current = 0*/
4.   While(current < noGeneration) {
5.     Population child =  $\emptyset$ 
6.     h = 0
7.     While(h < noCrossover) {
8.       T1, T2 = SelectParent(P)
9.       T' = Crossover(T1, T2)
10.      child = child  $\cup$  T'
11.      h = h + 1
12.    }
13.    m = 0
14.    While(m < noMutation) {
15.      T = SelectParent(P)
16.      T' = Mutation(T)
17.      child = child  $\cup$  T'
18.      m = m + 1
19.    }
20.    P = P  $\cup$  child
21.    P = Selection(P)
22.    current = current + 1
23.  }
24. End.
```

### 3 Experimental Results

We run our proposed algorithm and compare its performance with the Tabu Search and CDE-GA. Programs are run on a PC with Intel Core 2 Duo U7700, RAM 2GB. For experiments, we used real network topologies (Polska (12 nodes, 36 links, 65 unicast, 12 anycast), Germany17 (17 nodes, 52 links, 119 unicast, 13 anycast), Atlanta (26 nodes, 82 links, 234 unicast, 22 anycast), Germany50 (50 nodes, 176 links, 80 unicast, 20 anycats), TA2(65 node, 216 links, 80 unicats, 20 anycast). All can be downloaded from <http://sndlib.zib.de/home.action> . With each instance, we randomly create 5 test sets. For both CDE-GA and CCE-GA, the number of individual is 300 and the number of generation is 300. In CCE-GA, we set the one-point crossover probability to 17% and the path crossover probability is also 17%. The selection of crossover probability value is tuned after many experiments with the range from 10% to 50%. The mutation rate is set to a small value (3%). Experiment is repeated 10 times for each test set.



**Fig. 2.** Comparison between the best and the average result found by Tabu Search, CDE-GA and CCE-GA on (a) Germany17, (b) Atlanta, (c) Germany50, (d) TA2 networks

Fig. 2 show the computational results of three compared algorithms on five network instances. For each network, we compare the minimum and the average network cost found. It is notable that the minimum and the average result found by CCE-GA are better than the one found by Tabu Search and CDE-GA on almost problem instances.

The average cost found by CCE-GA are 15% and 10% better than Tabu Search and CDE-GA (Polska network), respectively. The best network cost found by CCE-GA are 8% and 4% better than Tabu Search and CDE-GA (TA2 network), while the average network cost found by CCE-GA are better than Tabu Search and CDE-GA about 7% and 6%.

These above results prove the efficiency of GA over Tabu Search for this problem. These results also demonstrate the efficiency of our new encoding scheme because it helps to increase the performance of CCE-GA in compare with CDE-GA. As we have explained in previous section, the new encoding scheme can exploit a larger search space. Therefore it can find a better result than that of CDE-GA.

**Table 1.** The best and average results found on POLSKA(a), on TA2 (b) network after 10 running times

Test #	Cost	Tabu	CDE-GA	CCE-GA
1	Min	13550	12812	12130
	Mean	14777	13923	12653
2	Min	14962	13898	13200
	Mean	15270	14670	13860
3	Min	11226	11484	10792
	Mean	12616	12359	11045
4	Min	14654	13176	14054
	Mean	13691	14318	14690
5	Min	12714	12480	11464
	Mean	12899	13262	11866

(a)

Test #	Cost	Tabu	CDE-GA	CCE-GA
1	Min	894440	850640	848320
	Mean	905968	880571	862527
2	Min	843280	836120	785480
	Mean	850613	837814	785480
3	Min	841280	822120	798640
	Mean	842613	827586	813037
4	Min	858440	836120	806960
	Mean	861756	844252	814589
5	Min	825280	806120	779800
	Mean	827413	810386	789364

(b)

To demonstrate the efficiency of CCE-GA over CDE-GA on large scale networks, we present the detail results on a small network (Polska, Table 1 (a)) and on a larger network (TA2, Table 1 (b)). For Polska network, the average cost found by CCE-GA are better than that found by CDE-GA on 8/10 problem instances (except the test #4) For TA2 network, the average cost found by CCE-GA are always better than that of CDE-GA. This is because the CDE-GA only works well on small network but not for large scale network. We also observe that for all test sets, the minimum cost found by CCE-GA is much better than that of CDE-GA.

**Table 2.** The comarision of the average running time (in second)

	Polska	Germany17	Atlanta	Germany50	TA2
CCE-GA	18	21	95	1951	26216
CDE-GA	13	21	78	941	13232
TabuSearch	1	5	36	666	6514

To compare the tested algorithms in terms of running time, we calculate the average running time of all 10 test sets for each algorithm. Table 2 shows that CCE-GA takes much more running times in compare with Tabu Search and CDE-GA. But if the running time for Tabu Search, CDE-GA is increased, the best result is not improved.

## 4 Conclusion

In this paper, we proposed a new algorithm called CCE-GA for solving A-SNDP. This algorithm uses a new simple encoding scheme called Completely Connection Encoding (CCE) to enlarge the search space to find better solution in large network instances. Experiments are conducted on real network topologies on show that our proposed approach is very efficient in solving A-SNDP. On the big instances, such as Atlanta, Germany50 and TA2 networks, the best and average result found by CCE-GA are much better than CDE-GA and Tabu Search. However, CCE-GA takes much more running times in compare with Tabu Search and CDE-GA to find best result. But if the running time for Tabu Search, CDE-GA is increased, the best result is not improved.

**Acknowledgment.** This work was supported by the project “Meta-heuristic for solving survivable network design problem” funded by the Ministry of Education and Training under grant number B2012-01-28.

## References

1. Johnson, Deering: Reserved IPv6 Subnet Anycast Addresses. RFC 2526 (1999)
2. Ballani, H., Francis, P.: Towards a global IP anycast service. In: SIGCOMM 2005, New York, USA, pp. 301–312 (2005)
3. Walkowiak, K.: Anycast Communication – A New Approach to Survivability of Connection-Oriented Networks. *Comm. in Computer and Info. Science*, 378–389 (2003)
4. Gładysz, J., Walkowiak, K.: Tabu Search Algorithm for Survivable Network Design Problem with Simultaneous Unicast and Anycast Flows. *Intl. Journal of Electronics & Telecom* 56(1), 41–48 (2010)
5. Walkowiak, K.: A Flow Deviation Algorithm for Joint Optimization of Unicast and Anycast Flows in Connection-Oriented Networks. In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2008, Part II. LNCS, vol. 5073, pp. 797–807. Springer, Heidelberg (2008)
6. Walkowiak, K.: A New Function for Optimization of Working Paths in Survivable MPLS Networks. In: Levi, A., Savaş, E., Yenigün, H., Balcısoy, S., Saygın, Y. (eds.) ISCIS 2006. LNCS, vol. 4263, pp. 424–433. Springer, Heidelberg (2006)
7. Grover, W.: Mesh-based Survivable Networks: Options and Strategies for Optical, MPLS, SONET and ATM Networking. Prentice Hall PTR, New Jersey (2004)
8. Nissen, V., Gold, S.: Survivable network design with an evolution strategy. In: Yang, A., Shan, Y., Bui, L.T. (eds.) Success in Evolutionary Computation. SCI, vol. 91, pp. 263–283. Springer, Heidelberg (2008)

9. Sharma, V., Hellstrand, F.: Framework for MPLS-based recovery. RFC 3469 (2003)
10. Vasseur, J., Pickavet, M., Demeester, P.: Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP and MPLS. Morgan Kaufmann, San Francisco (2004)
11. Gladysz, J.: Krzysztof Walkowiak: Optimization of survivable networks with simultaneous unicast and anycast flows. In: ICUMT, Poland, pp. 1–6 (2009)
12. Binh, H.T.T., et al.: Genetic Algorithm for Solving Survivable Network Design with Simultaneous Unicast and Anycast Flows. In: Yin, Z., Pan, L., Fang, X. (eds.) Proceedings of The Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), 2013. AISC, vol. 212, pp. 1237–1247. Springer, Heidelberg (2013)

# Belief Propagation in Bayesian Network

Ji Ryang Chung<sup>1</sup> and Gangman Yi<sup>2,\*</sup>

<sup>1</sup> Samsung Electronics, Suwon, South Korea  
jiryang@gmail.com

<sup>2</sup> Department of Computer Science & Engineering, Gangneung-Wonju National University,  
Wonju, South Korea  
gangman@cs.gwnu.ac.kr

**Abstract.** The current article describes the belief propagation in Bayesian Network. Existing books or online tutorials do not contain explanations with enough detail for a novice to understand the mathematics behind the method. Using a real example of *Water Quality* model, this report will help anyone to understand the belief propagation in Bayesian Network without much knowledge in probability calculus.

**Keywords:** bayesian network, water quality.

## 1 Introduction

Bayesian Network (BN) is a graphical model representing the probabilistic relationships of a set of variables via a directed acyclic graph (DAG). Variables are represented as nodes in BN, and their conditional dependencies are represented in directed edges. A probability function of the variable is assigned in each node. Using the dependency relationship, BN provides diagnostic (bottom-up, "symptoms to cause") and prognostic (top-down, "new information about causes to new beliefs about effects") probabilistic inference. This article describes the belief propagation in Bayesian Network. Existing books or online tutorials do not contain explanations with enough detail for a novice to understand the mathematics behind the method. Using a real example of *Water Quality* model, this will help understanding the belief propagation in Bayesian Network without much knowledge in probability calculus.

## 2 Water Quality Model

The *Water Quality* BN model is made of 8 variables: *Climate*, *Water Storage*, *Management*, *Vegetation*, *Carbol Pools*, *Runoff Erosion*, *Production Traditional*, and finally *Water Quality* as a leaf node. The model depicts how the probabilities and the relationships of different variables can influence upon the quality of water in the given area. The model is illustrated in Fig. 1.

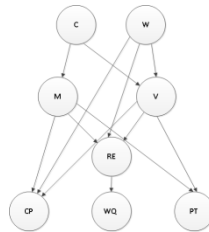
---

\* Corresponding author.

BN provides probabilistic inference so that we can change the probability of a node and see its impact on the water quality (prognosis), or condition the water quality, for example to be high, and see what should be the probabilities of the upper level variables to get the water quality[1]. The inference can be made through the message passing algorithm. Kim and Pearl's message passing algorithm, which is used the most frequently for belief propagation, works only on polytrees, singly-connected networks[2,3]. However, BN usually has a (directly acyclic) graph structure, where two nodes are connected by more than one path. The connection between node W(Water Storage) and R (Runoff Erosion) in Fig. 1 is an example of the multi-path connection, which has both direct (W-R) and indirect (W-VR) connections. Because such cases are not handled by the message passing algorithm, a method called Clustering is developed to transform a graph into a probabilistically equivalent polytree. Clustering algorithm merges multiple nodes to remove the multiple paths between the two nodes.

Junction Tree Clustering algorithm provides systematic method of clustering and adapted as a default inference algorithm in most major BN software packages such as Bayesian-Lab, GeNIe, Hugin, JavaBayes, and Netica Korb and Nicholson[4,5,6]. The algorithm consists of 3 main steps:

1. Moralization: connect all parents and remove arrows
2. Triangulation: add edges to guarantee that every cycle has a length  $< 4$
3. Junction Tree: identify maximal cliques to create supernodes

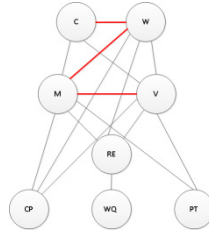


**Fig. 1.** The *Water Quality* Model. Each variable name is abbreviated: C=Climate; W=Water Storage; M=Management; V=Vegetation; CP=Carbol Pools; RE=Runoff Erosion; PT=Production Traditional; WQ=Water Quality. Edges denote the conditional dependencies between variables. Unconnected nodes are conditionally independent. For example, biomass of the vegetation in the area has direct influence on the amount of runoff erosion, but not directly related to the management type.

The first step is named Moralization because it is a process to “marry” the parents of a child node. Fig. 2 shows the result of Moralization of the original *Water Quality* model in Fig. 1. Triangulation is to ensure that the joint probability terms are functions of cliques containing complete sub-graph. Because the moralized *Water Quality* model is already a triangulated graph, another example is used to illustrate this method in Fig. 2. Finally, a Junction Tree with cluster nodes can be created from



identifying maximal cliques in the triangulated graph. Fig. 2 shows the maximal cliques identified from the moralized/triangulated *Water Quality* model, and the resulting Junction Tree is shown in Fig. 5.



**Fig. 2.** Moralized *Water Quality* Model. New connections are added: *C-W*, *W-M*, and *M-V*. *C* and *W* are parents of *V*, *W* and *M* are parents of *CP* and *RE*, and *M* and *V* are parents of *CP*, *RE*, and *PT*.

### 3 Belief Propagation

Now that we have successfully transformed the given *Water Quality* model into a polytree, we are ready to examine how the Kim and Pearl’s message passing algorithm updates beliefs. The total strength of belief is a product of the causal support ( $\pi(X)$ ) from parent and the diagnostic support ( $\lambda(X)$ ) from descendants.

$$BEL(X) = \alpha \lambda(X) \pi(X) \tag{1}$$

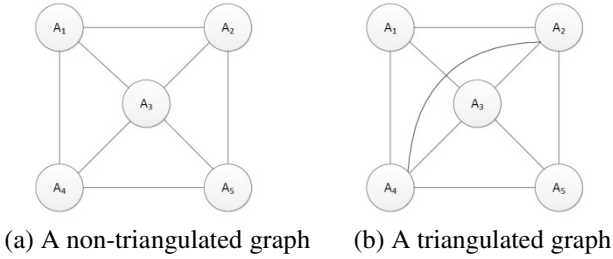
where  $\alpha$  is a normalizing constant that makes  $\sum_X BEL(X) = 1$ . From the original model in Fig. 1,  $P(C)$ ,  $P(W)$ ,  $P(M|C)$ ,  $P(V|C,W)$ ,  $P(CP|M,V,W)$ ,  $P(RE|M,V,W)$ ,  $P(PT|V,W)$ , and  $P(WQ|RE)$  are given. The probability tables given to each variable are in Table 1-8. Note that node probabilities are associated with their conditional dependencies. For the bottom-up scenario, we will fix the *WQ* node in the original graph as an evidence to have *high* value.

#### 3.1 Top-Down Prognosis

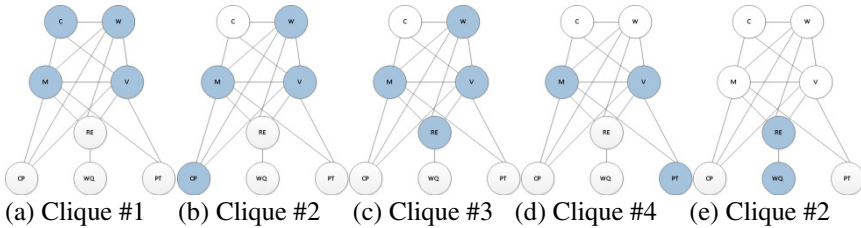
Top-down message passing starts from calculating the belief of the root node  $Z_1$ . From the root ( $Z_1$  in the example), a top-down message is propagated to leaves. Fig. 6 shows two steps that the top-down messages travel. Conventionally,  $\pi_Y(X)$  denotes a top-down message from node *X* to *Y*.

The probability of the initial top-down messages are given by

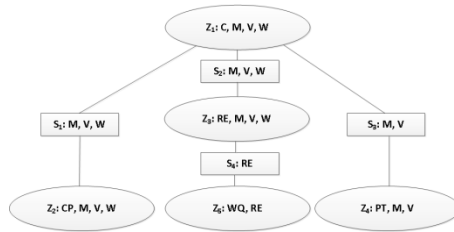
$$\pi(Z_1) = \pi_{Z_2}(Z_1) = \pi_{Z_3}(Z_1) = \pi_{Z_4}(Z_1) = P(Z_1) = P(C,M,V,W) = P(C)P(M|C)P(V|C,W)P(W) \tag{2}$$



**Fig. 3.** Comparison between a non-triangulated and a triangulated graph. In (a), cycle  $A_1$ - $A_2$ - $A_5$ - $A_4$  has length 4. Adapted from Jensen [7].



**Fig. 4.** A total of 5 maximal cliques are found in the moralized/triangulated *Water Quality* model



**Fig. 5.** Junction Tree of the *Water Quality* Model. 5 maximal cliques make 5 nodes in the Junction Tree ( $Z_1$  to  $Z_5$ ). 4 separators consist of the intersection of adjacent nodes ( $S_1$  to  $S_4$ ). The Junction Tree is a singly-connected polytree and there is no two nodes connected by more than one path.

Equation 2 stands for the 108-component vector shown in the  $\pi(Z_i)$  column of Table 9. Because the default bottom-up messages ( $\lambda$ 's) are unit vectors,  $BEL(Z_i) = \pi(Z_i)$ . Before passing  $\pi(Z_i)$  to  $Z_i$ 's children, we can get the marginal probabilities of nodes  $M$  and  $V$  by simply summing up the appropriate rows in Table 9. To take the example of the marginal probability  $P(M)$ ,

$$\begin{aligned}
 P(M = m_1) &= \sum_{id=1}^9 row_{id} + \sum_{id=28}^{36} row_{id} + \sum_{id=55}^{63} row_{id} + \sum_{id=82}^{90} row_{id} = 0.16 \\
 P(M = m_2) &= \sum_{id=10}^{18} row_{id} + \sum_{id=37}^{45} row_{id} + \sum_{id=64}^{72} row_{id} + \sum_{id=91}^{99} row_{id} = 0.58 \\
 P(M = m_3) &= \sum_{id=19}^{27} row_{id} + \sum_{id=46}^{54} row_{id} + \sum_{id=73}^{81} row_{id} + \sum_{id=100}^{108} row_{id} = 0.26
 \end{aligned} \tag{3}$$

Table 10 and 11 show the marginal probabilities of nodes  $M$  and  $V$  respectively. Now, let's consider  $\pi_{Z_1}(Z_1)$  first. The interdependencies between  $Z_1$  and  $Z_2$  are from the shared variables  $Z_1 \cap Z_2$ ,

$$S_1 : P(Z_2|Z_1) = P(Z_2|Z_1 \cap Z_2) = P(CP, M, V, W|M, V, W) = P(CP|M, V, W) \tag{4}$$

As in the case of  $Z_1$ , the top-down message that  $Z_2$  generates is same as the prior probability of  $Z_2$ . Using the Equation 4,

$$\begin{aligned}
 \pi(Z_2) = P(Z_2) &= \sum_{Z_1} P(Z_2|Z_1)P(Z_1) \\
 &= \sum_c P(CP|M, V, W)P(C, M, V, W) = P(CP|M, V, W)P(M, V, W)
 \end{aligned} \tag{5}$$

which means that the p-message from node  $Z_j$  receiving message from  $Z_i$  through a separator  $S_k$  can be calculated by multiplying the interdependency that  $S_k$  represents ( $P(CP|M, V, W)$  in Equation 5) and the joint probability of the variables in the  $S_k$  ( $P(M, V, W)$  in Equation 5). Again, we can get  $P(CP|M, V, W)$  and  $P(M, V, W)$  using Table 5 and Table 9. Table 12 shows the  $\pi(Z_2)$  in the form of 81-component vector. Summing up the rows of the Table 12, the marginal probability of  $CP$  is achieved (Table 13). Top-down messages  $\pi(Z_3)$  and  $\pi(Z_4)$  and the marginal probabilities  $\pi(RE)$  and  $\pi(PT)$  are calculated in the same way. Without redundant explanation, the results are shown in Table 14, 16, 15, and 17.

$$\pi(Z_3) = P(Z_3) = P(RE|M, V, W)P(M, V, W) \tag{6}$$

$$\pi(Z_4) = P(Z_4) = P(PT|M, V)P(M, V) \tag{7}$$

Then, as in Fig. 6(b),  $Z_3$  sends a p-message down to  $Z_5$ .  $\pi(Z_5)$  using Equation 8 is in Table 18 and  $P(WQ)$  calculated by summing up appropriate rows is in Table 19.

$$\pi(Z_5) = P(WQ|RE)P(RE) \tag{8}$$

### 3.2 Bottom-Up Diagnosis

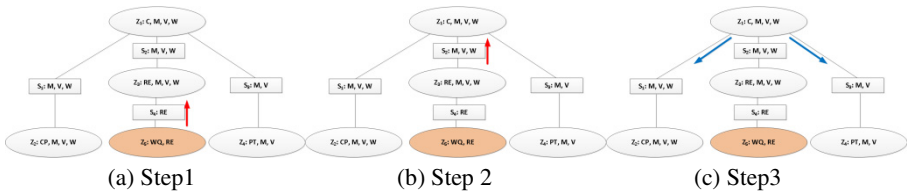
For bottom-up message passing example, we entered an evidence  $WQ = wq_3(\text{high})$ . Starting from where the evidence is entered ( $Z_5$  in the example), a bottom-up message is propagated up to the root. Fig. 7 shows three steps that the bottom-up messages and their by-product top-down messages travel. As shown in Fig. 7(c), bottom-up belief propagation sometimes can produce top-down messages to convey the effect of the evidence to the entire tree. Conventionally,  $\lambda_X(Y)$  denotes a bottom-up message from node  $Y$  to  $X$ .

The initial bottom-up message at  $Z_5$  when the evidence arrives  $\lambda(Z_5)$  is

$$\lambda(Z_5) = \begin{cases} 1 & \text{if } WQ = wq_3 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

and this message prompts  $Z_5$  to generate a corresponding  $\lambda$ -message for  $Z_3$

$$\lambda(Z_3) = \text{Dependency}(S_4) \times \lambda(Z_5) = P(WQ = wq_3 | RE) = \{0.7, 0.2, 0.1\} \quad (10)$$



**Fig. 6.** Bottom-Up Message Propagation. (a) Evidence is entered to node  $Z_4$  to initiate the message to its parent ( $\lambda_{Z_3}(Z_5)$ ). (b) Then,  $Z_3$  sends a message to the root node ( $\lambda_{Z_1}(Z_3)$ ). (c) To deliver the effect of the evidence to the nodes  $Z_2$  and  $Z_4$ ,  $Z_1$  produces top-down messages ( $\pi_{Z_2}(Z_1)$  and  $\pi_{Z_4}(Z_1)$ ).

which in turn to generate new belief of  $Z_3$ . Table 20 shows the updated  $BEL(Z_3)$  (comparing this with Table 14 may be helpful). For the new  $BEL(Z_3)$ , Equation 1 is used with the  $\pi(Z_3)$  given from the above top-down process. Note that the normalizing constant  $\alpha$  in Equation 1 is not 0 this time. With the new  $BEL(Z_3)$ , we get the new marginal probability  $P(RE)$  in Table 21. Next, we get  $\lambda(Z_3)$  with the bottom-up message  $Z_3$  receives ( $\lambda(Z_3) = \lambda_{Z_3}(Z_5) = \{re_1 = 0.7, re_2 = 0.2, re_3 = 0.1\}$ ). Using the same notation as in the Equation 11, we can get  $\lambda_{Z_1}(Z_3)$  as

$$\lambda_{Z_1}(Z_3) = \text{Dependency}(S_2) \times \lambda(Z_3) = \begin{cases} P(RE = re_1 | M, V, W) \\ P(RE = re_2 | M, V, W) \\ P(RE = re_3 | M, V, W) \end{cases} \quad (11)$$

The resulting 27-component vector  $\lambda_{Z_1}(Z_3)$  is in Table 22. As we have  $\lambda(Z_1) = \lambda_{Z_1}(Z_3)$  now,  $BEL(Z_1)$  can be re-calculated using Equation 1. The updated  $BEL(Z_1)$  is in Table 23. By summing up rows in Table 23, new  $P(C)$ ,  $P(M)$ ,  $P(V)$ , and  $P(W)$  are

calculated and shown in Table 24, 25, 26, and 27 respectively. In the last step, top-down messages are sent from  $Z_1$  to  $Z_2$  and  $Z_4$  ( $\pi_{Z_2}(Z_1)$  and  $\pi_{Z_4}(Z_1)$ ).  $P(M, V, W)$  for  $\pi_{Z_2}(Z_1)$  and  $P(M, V)$  for  $\pi_{Z_4}(Z_1)$  can be calculated from Table 23. Updated  $BEL(Z_2)$  and  $BEL(Z_4)$  and new marginal probabilities of  $CP$  and  $PT$  are shown in Table 28, 29, 30, and 31.

## 4 Conclusion

This article presents message propagation in BN using a real world example of *Water Quality* model. Without much mathematical ado, Kim and Pearl’s Clustering method including graph transformation and top-down/bottom-up message propagations are fully explained. We hope this helps fast understanding how belief propagates in BN.

## 5 Tables

$c_1$	0.2	wet
$c_2$	0.5	average
$c_3$	0.25	dry
$c_4$	0.05	extra dry

Table 1: Climate Variable,  $P(C)$

$w_1$	0.05	low
$w_2$	0.7	medium
$w_3$	0.25	high

Table 2: Water Storage Variable,  $P(W)$

	$c_1$	$c_2$	$c_3$	$c_4$	
$m_1$	0.1	0.15	0.2	0.3	minimum
$m_2$	0.6	0.7	0.4	0.2	moderate
$m_3$	0.3	0.15	0.4	0.5	intensive

Table 3: Management Variable,  $P(M|C)$

	$m_1$				$m_2$				$m_3$				
	$c_1$	$c_2$	$c_3$	$c_4$	$c_1$	$c_2$	$c_3$	$c_4$	$c_1$	$c_2$	$c_3$	$c_4$	
$v_1$	0	0.4	0.7	0.8	0	0.1	0.4	0.45	0	0	0.1	0.1	low biomass
$v_2$	0.3	0.3	0.3	0.2	0.2	0.2	0.5	0.4	0.4	0.1	0.4	0.6	medium biomass
$v_3$	0.7	0.3	0	0	0.8	0.4	0.2	0.15	0.9	0.6	0.3	0.4	high biomass

Table 4: Vegetation Variable,  $P(V|C, W)$

	$m_1$								$m_2$								$m_3$												
	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$					
$cp_1$	0.8	0.7	0.6	0.8	0.2	0.1	0.8	0.1	0.1	0.8	0.6	0.3	0.8	0.1	0.05	0.8	0.05	0	0.8	0.5	0.2	0.8	0.1	0.05	0.8	0	0	low	
$cp_2$	0.2	0.25	0.3	0.2	0.3	0.3	0.2	0.3	0.2	0.2	0.2	0.3	0.4	0.2	0.2	0.15	0.2	0.25	0.15	0.2	0.3	0.5	0.2	0.3	0.25	0.2	0.2	0.1	medium
$cp_3$	0	0.05	0.1	0	0.5	0.6	0	0.6	0.7	0	0.1	0.3	0	0.7	0.8	0	0.7	0.85	0	0.2	0.3	0	0.6	0.7	0	0.8	0.9	0	high

Table 5: Carbol Pools Variable,  $P(CP|M, V, W)$

	$m_1$								$m_2$								$m_3$																
	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$	$v_7$	$v_8$	
$rc_1$	0.5	0.3	0.5	0.35	0.35	0.6	0.1	0.4	0.85	0.5	0.4	0.65	0.25	0.45	0.75	0.1	0.5	0.85	0.5	0.5	0.8	0.2	0.55	0.8	0.1	0.6	0.85	0.3	0.1	0.1	low		
$rc_2$	0.3	0.4	0.4	0.25	0.4	0.3	0.2	0.4	0.1	0.35	0.35	0.25	0.3	0.35	0.15	0.25	0.35	0.1	0.35	0.35	0.15	0.3	0.35	0.15	0.3	0.35	0.15	0.3	0.3	0.1	0.05	medium	
$rc_3$	0.2	0.3	0.1	0.4	0.25	0.1	0.7	0.2	0.05	0.15	0.25	0.1	0.45	0.2	0.1	0.65	0.15	0.05	0.15	0.15	0.05	0.5	0.1	0.05	0.6	0.1	0.05	0.1	0.05	0	0	high	

Table 6: Runoff Erosion Variable,  $P(RE|M, V, W)$

	$m_1$				$m_2$				$m_3$				
	$v_1$	$v_2$	$v_3$	$v_4$	$v_1$	$v_2$	$v_3$	$v_4$	$v_1$	$v_2$	$v_3$	$v_4$	
$p_1$	0.7	0.35	0	0.5	0.25	0	0.3	0.15	0	0.15	0	0	unsustainable
$p_2$	0.3	0.45	0.6	0.5	0.5	0.45	0.7	0.5	0.3	0.3	0.5	0.3	sustainable
$p_3$	0	0.2	0.4	0	0.25	0.55	0	0.35	0.7	0	0.35	0.7	maximized

Table 7: Production Traditional Variable,  $P(PT|V, W)$

	$rc_1$	$rc_2$	$rc_3$	
$qv_1$	0.1	0.6	0.7	low
$qv_2$	0.2	0.2	0.2	medium
$qv_3$	0.7	0.2	0.1	high

Table 8: Water Quality Variable,  $P(WQ|RE)$

id	C	M	V	W	$P(C)$	$P(M C)$	$P(V C, W)$	$P(W)$	$\pi(Z_1) = BEL(Z_1)$
1	$c_1$	$m_1$	$v_1$	$w_1$	0.2	0.1	0	0.05	0
2	$c_1$	$m_1$	$v_2$	$w_1$	0.2	0.1	0	0.7	0
3	$c_1$	$m_1$	$v_3$	$w_1$	0.2	0.1	0	0.25	0
4	$c_1$	$m_1$	$v_2$	$w_2$	0.2	0.1	0.3	0.05	0.0003
5	$c_1$	$m_1$	$v_2$	$w_2$	0.2	0.1	0.2	0.7	0.0028
6	$c_1$	$m_1$	$v_2$	$w_2$	0.2	0.1	0.1	0.25	0.0005
7	$c_1$	$m_1$	$v_3$	$w_2$	0.2	0.1	0.2	0.05	0.0007
8	$c_1$	$m_1$	$v_3$	$w_2$	0.2	0.1	0.8	0.7	0.0112
9	$c_1$	$m_1$	$v_3$	$w_2$	0.2	0.1	0.9	0.25	0.0045
10	$c_1$	$m_2$	$v_1$	$w_1$	0.2	0.6	0	0.05	0
11	$c_1$	$m_2$	$v_1$	$w_2$	0.2	0.6	0	0.7	0
12	$c_1$	$m_2$	$v_1$	$w_2$	0.2	0.6	0	0.25	0
13	$c_1$	$m_2$	$v_1$	$w_2$	0.2	0.6	0.3	0.05	0.0018
105	$c_4$	$m_3$	$v_2$	$w_3$	0.05	0.5	0.5	0.25	0.003125
106	$c_4$	$m_3$	$v_2$	$w_3$	0.05	0.5	0	0.05	0
107	$c_4$	$m_3$	$v_2$	$w_3$	0.05	0.5	0.15	0.7	0.002625
108	$c_4$	$m_3$	$v_2$	$w_3$	0.05	0.5	0.4	0.25	0.0015

Table 9:  $BEL(Z_1)$  Table

$m_1$	0.16
$m_2$	0.58
$m_3$	0.26

Table 10: Marginal Probability of  $M$ ,  $P(M)$

$v_1$	0.149
$v_2$	0.4005
$v_3$	0.4505

Table 11: Marginal Probability of  $V$ ,  $P(V)$

$cp_1$	0.189106
$cp_2$	0.21129
$cp_3$	0.599604

Table 13: Marginal Probability of  $CP$ ,  $P(CP)$



id	CP	M	V	W	$P(CP M,V,W)$	$P(M,V,W)$	$newBEL(Z_2)$
1	$c_1$	$m_1$	$v_1$	$w_1$	0.8	0.003506	0.002805
2	$c_1$	$m_1$	$v_1$	$w_2$	0.8	0.017031	0.013699
3	$c_1$	$m_1$	$v_1$	$w_3$	0.8	0.00062	0.000496
4	$c_1$	$m_1$	$v_2$	$w_1$	0.7	0.001576	0.001103
5	$c_1$	$m_1$	$v_2$	$w_2$	0.2	0.035026	0.007305
6	$c_1$	$m_1$	$v_2$	$w_3$	0.1	0.013984	0.001398
7	$c_1$	$m_1$	$v_3$	$w_1$	0.6	0.001701	0.001020
8	$c_1$	$m_1$	$v_3$	$w_2$	0.1	0.042317	0.004232
9	$c_1$	$m_1$	$v_3$	$w_3$	0.1	0.022576	0.002258
10	$c_1$	$m_2$	$v_1$	$w_1$	0.8	0.010042	0.008034
11	$c_1$	$m_2$	$v_1$	$w_2$	0.8	0.033002	0.026402
12	$c_1$	$m_2$	$v_1$	$w_3$	0.8	0.001078	0.000862
13	$c_1$	$m_2$	$v_2$	$w_1$	0.6	0.008687	0.004122
78	$c_3$	$m_3$	$v_3$	$w_3$	0.8	0.028151	0.02252
79	$c_3$	$m_3$	$v_3$	$w_1$	0.3	0.004064	0.001219
80	$c_3$	$m_3$	$v_3$	$w_2$	0.7	0.089758	0.06283
81	$c_3$	$m_3$	$v_3$	$w_3$	0.9	0.045632	0.041069

Table 28: Updated  $BEL(Z_2)$  Table

$c_1$	0.141561
$c_2$	0.206418
$c_3$	0.652021

Table 30: Updated Marginal Probability of CP,  $P(CP)$

id	PT	M	V	$P(PT M,V)$	$P(M,V)$	$newBEL(Z_4)$
1	$p_1$	$m_1$	$v_1$	0.7	0.021137	0.014796
2	$p_1$	$m_1$	$v_1$	0.35	0.050586	0.017705
3	$p_1$	$m_1$	$v_1$	0	0.071594	0
4	$p_1$	$m_1$	$v_2$	0.5	0.044122	0.022061
5	$p_1$	$m_1$	$v_2$	0.25	0.202761	0.05069
6	$p_1$	$m_1$	$v_2$	0	0.341562	0
7	$p_1$	$m_1$	$v_3$	0.3	0.028561	0.008568
8	$p_1$	$m_1$	$v_3$	0.15	0.100224	0.015034
9	$p_1$	$m_1$	$v_3$	0	0.139454	0
10	$p_2$	$m_2$	$v_1$	0.3	0.021137	0.006341
11	$p_2$	$m_2$	$v_1$	0.45	0.050586	0.022764
12	$p_2$	$m_2$	$v_1$	0.6	0.071594	0.042956
13	$p_2$	$m_2$	$v_2$	0.5	0.044122	0.022061
14	$p_2$	$m_2$	$v_2$	0.5	0.202761	0.10138
15	$p_2$	$m_2$	$v_3$	0.45	0.341562	0.153703
16	$p_2$	$m_2$	$v_3$	0.7	0.028561	0.019993
17	$p_2$	$m_2$	$v_3$	0.5	0.100224	0.050112
18	$p_2$	$m_3$	$v_3$	0.3	0.139454	0.041836
19	$p_3$	$m_3$	$v_3$	0	0.022137	0
20	$p_3$	$m_3$	$v_3$	0.2	0.050586	0.010117
21	$p_3$	$m_3$	$v_3$	0.4	0.071594	0.028637
22	$p_3$	$m_3$	$v_3$	0	0.044122	0
23	$p_3$	$m_3$	$v_3$	0.25	0.202761	0.05069
24	$p_3$	$m_3$	$v_3$	0.55	0.341562	0.187859
25	$p_3$	$m_3$	$v_3$	0	0.028561	0
26	$p_3$	$m_3$	$v_3$	0.35	0.100224	0.025078
27	$p_3$	$m_3$	$v_3$	0.7	0.139454	0.097618

Table 29: Updated  $BEL(Z_4)$  Table

$p_1$	0.128854
$p_2$	0.461146
$p_3$	0.41

Table 31: Updated Marginal Probability of PT,  $P(PT)$

## References

- Stephenson, T.A.: An Introduction to Bayesian Network Theory and Usage, IDIAP-RR 00-03 (February 2000)
- Pearl, J.: Belief Networks Revisited. *Artificial Intelligence* 59, 49–56 (1993)
- D’Ambrosio, B.: Inference in Bayesian Networks. *AI Magazine* 20(2), 21–36 (1999)
- Van der Gaag, L.C.: Bayesian Belief Networks: Odds and Ends. *The Computer Journal* 39(2), 97–113 (1996)
- Jensen, F.V., Olesen, K.G., Anderson, S.K.: An Algebra of Bayesian Belief Universes for Knowledge-Based Systems. *Networks* 20, 637–660 (1990)
- Korb, K.B., Nicholson, A.E.: *Bayesian Artificial Intelligence*. Chapman and Hall, Boca Raton (2004)
- Jensen, F.V.: *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. Springer, New York (2001)

# A Method of Search Scope Compaction for Image Indexes

Samuel Sangkon Lee

Computer Science and Engineering, Jeonju University,  
#303 Chonjam-Ro, Wansan-Gu, Jeonju, South Korea, 560-759  
samuel@jj.ac.kr

**Abstract.** It is inferior in generality enough not to be applied in the system wherein various types of multimedia information are stored. In our study, it was intended to improve VP-tree method which is one of the indexing methods for metric space. VP-tree calculates the distance from object which is linked to the leaf node that reaches from root node to final along the matching node within search range during search and then it checks if a particular node is existed within a search range. We propose a compaction method by using triangle inequality at leaf node within search range. Besides, it is proposed to use a triangle inequality at nearest point as a reference point for the query object. Actual performance test for the system using 10,000 video data revealed that search time for the similar videos could be saved by 5~12% as compared with those of conventional methods.

**Keywords:** Multi-dimensional DB, Multimedia DB, Matric Index, VP-Tree, Range Search,  $K$ -Nearest Neighbor Search, Triangle Inequality.

## 1 Introduction

It is a common practice in storing big data for multimedia like documents, images, music, and videos even in the personal computer thanks to the price down of primary or secondary memory devices and their large capacities. Accordingly, a search technique by which only required data by users can be promptly and accurately is highly demanded. To raise search efficiency, it is needed to extract the characteristics from pre-stored data and then to construct index from the characteristics [1]. During actual search, adequate data are extracted from that index. That is why performance of a search system depends on the index construction method.

It is, therefore, important to conduct research for the 'metric space-based index' that performs indexing based on the distance information. As compared with multi-level index preparation based on the characteristic coordinates' value of multidimensional space, a metric space-based indexing method uses only distance information between characteristics once distance axiom is established, consequently indexing is relatively simple. Therefore, it can be applied in the distance calculation method other than Euclidean distance. The metric space-based indexing method, generally is a



hierarchical index tree based on the distance information, thereby can reduce search space during search by partitioning the space (dataset) recursively.

In the present study, to supplement these drawbacks, by improving search algorithm at the leaf node of VP-tree, the reduction of distance calculation frequency is verified. The search range is compacted by applying a triangle inequality<sup>1</sup> at the leaf node of VP-tree. Though Vantage Point was used as a reference in triangle inequality in conventional methods, we focused the fact that analysis range becomes narrowed down as the reference point in the triangle inequality and query object becomes nearer. Therefore, with present search method, by using the nearest neighbor for the query object on the reference point of triangle inequality, the search range was remarkably narrowed and the frequency of distance calculation also could be considerably reduced. However, since the nearest point could not be applied in advance even if the improved method was introduced in the actual system, the compaction technique was realized for the search range using the nearest point by assuming the nearest object to the query object as a virtual nearest point in the search result list in the present study.

As similar researches with the method intended to carry out in the present study, there is AESA (Approximating and Eliminating Search Algorithm) which compacts search range intended to be analyzed using distance list file established in advance [3]. However, the difference between AESA and the method proposed in this study is that while AESA attempts compaction of search range using distance list for all the objects, the method in the present study uses only the objects within leaf node. AESA which deals all objects as subject shows inherent problem of abrupt increases in file reading frequency. While, if VP-tree is used with the method in the present study, only some objects become subjects within leaf node and distance list file is read only when a virtual nearest point (reference point) is updated, ultimately the accessing frequency to the file can be remarkably reduced.

In the following Chapter 2, we explain the set up of VP-tree and its search algorithm, and then compaction method of leaf node is described. In the Chapter 3, the test and evaluation using this improved method are discussed. Lastly in Chapter 4, after concluding the research, future task and prospect of the research are proposed.

## 2 Improvement of VP-Tree

### 2.1 Set-Up Algorithm

In this Chapter, set up algorithm of VP-tree is explained. If we assume that indexing work is executed in data set  $s$  comprised of  $N$  number of data, each node of tree can

---

<sup>1</sup> Triangle inequality is an inequality for three sides of triangle that defines that the sum of length of two sides is larger than that of rest of the side. This inequality is applied in a variety of space. First of all, if vectors of two sides in norm vector space are set as  $x$  and  $y$ , respectively, this inequality can be expressed as below:  $\|x + y\| \leq \|x\| + \|y\|$ . Also, if there are  $x$ ,  $y$ , and  $z$  in a metric space  $M$  and the distance between them is set as  $d$ , below equation is established:  $d(x, y) \leq d(x, z) + d(z, y)$ .

select vantage point<sup>2</sup> (hereinafter referred as 'vp') by random algorithm as presented below. The median value for distance of all the data from  $vp$  selected in root node to  $s$  is set as  $\mu$ . If  $d(p, q)$  is set as distance between point  $p$  and  $q$ , the data set  $S$  is partitioned into two area  $S_1$  and  $S_2$  as below:

$$S_1 = \{s \in S \mid d(s, vp) < \mu\} \text{ and } S_2 = \{s \in S \mid d(s, vp) \geq \mu\}$$

By applying this type of partitioning work in the area  $S_1$  and  $S_2$  recursively, the index is created. All the subsets of  $S_1$  and  $S_2$  are corresponding to one node of VP-tree. In the leaf node, there are some objects stored [3].

#### ■ Set-up Algorithm of VP-tree

- (1) Virtual  $vp$  is randomly selected from data set.
- (2) From virtual  $vp$  till the rest of  $N - 1$  numbers of objects, distances are calculated.
- (3) The mean and variance of distances between these objects are calculated.
- (4) By repeating (1) ~ (3) for several times, the point where variance becomes maximum is decided as  $vp$ .

## 2.2 Search Algorithm

In this Section, algorithms of range search of VP-tree and  $K$ -nearest neighbor search are explained. Range search is a search method in determining set of objects which are present at distance from center of circle till radius using radius of circle between query object and search range. While,  $K$ -nearest neighbor search is a search method by assigning query object and search frequency  $K$  and determines a set of  $K$  cases of object at upper level in the order of nearest distance. In this thesis, test was carried out using  $K$ -nearest neighbor search. Yet, since this search method is based on the range search algorithm, both search methods are described in this section.

Range search calculates distance between leaf object which is linked along the node suitable for search range from root node and then obtains objects existed in the search range. While,  $K$ -nearest neighbor search sets search radius as infinite as an initial value and continuously adds the search results of object along the route in the list. If the number of search becomes over the assigned search number after searching done, the object whose distance is maximum is deleted from search result list so that the number of search in results list doesn't exceed assigned number of search. Also, by repeatedly executing search with maximum distance of list as search range (radius), search is executed by narrowing search radius, and finally search result as much as assigned can be obtained.

---

<sup>2</sup> Literal meaning is as follows: A good position (to watch something); the timing (particularly think over the past).

### 2.3 Compaction of Search Range at Leaf Node

As has described in Section 2.2, in the existing VP-tree, the distance from query object was calculated by accessing to all the objects presented in the leaf nodes during search. However, as a yet another method to improve this method, if triangle inequality is used during search each object inside leaf node, the range of search candidates can be remarkably reduced as below.

In the leaf node, the distance between  $vp$  object and each leaf object is preserved in the distance list. If triangle inequality is used for this distance between  $vp$  object and each object, the calculation frequency for distance can be reduced. If query object is set as  $q$ , the radius in search range is set as  $r$ ,  $vp$  object of leaf node is set as  $v$ , and leaf object linked to leaf node is set as  $o$ , below theorem is established.

- Theorem ① If  $d(v,o) - d(v,q) > r$  is established, leaf object  $o$  is not existed within search range.

**[Proof]** By triangle inequality  $d(v,q) + d(q,o) \geq d(v,o)$ ,  $d(v,o) - d(v,q) > r$  becomes  $d(q,o) > r$ , and  $o$  is not existed within search range. The same is in the case of  $-d(v,o) + d(v,q) > r$  that it becomes  $d(q,o) > r$ . Therefore, above theorem ① can be established.

$d(v,o)$  and  $r$  of theorem ① are an information which are already known during search of leaf node. Since  $d(v,q)$  can determine each leaf node at once, even without calculating the distance from each leaf object, it can be promptly judge if leaf object is existed within search range. Therefore, the frequency of distance calculation and access frequency to leaf object can be reduced. The method that can narrow the candidate of this leaf node (candidate for analysis) is presented. Also, the algorithm of  $K$ -nearest neighbor search is provided as below. In Fig. 1, the area other than slant lines is an area where the equation in theorem ① is established and the distance calculation for the objects present in this area can be omitted. While, the area with slant lines is an area wherein theorem ① is not established, thus the distances of objects presented in this area have to be calculated.

Besides, it is possible to compact search range even when not only  $vp$  of leaf node but also all the  $vp$  object are existed in the path from root node till leaf node are used. In this case, it is required to save distance from all the leaf objects and root nodes linked to leaf node until all the  $vp$  objects existed in the path before saving the distance of leaf node. If triangle inequality is used for the distance calculation between these multiple  $vp$  objects and each leaf object, it is possible to remarkably reduce the distance calculation frequency. If query object is set as  $q$ , radius of search range as  $r$ ,  $k$  number of  $vp$  object existed on the path from root node till leaf node as  $v_i (i = 1, 2, 3, \dots, k)$ , and leaf object linked to leaf node as  $o$  the analysis candidates can be reduced as in Fig. 2. Since this algorithm compares using multiple number of  $vp$  objects, the possibility of narrowing down analysis candidate can be raised as compared to existing methods.

Theorem ① If $d(o_1, o) - d(o_1, q) > r$ established, leaf object $o$ is not existed within search range.
---

<b>[Proof]</b> By triangle inequality $d(o_1, q) + d(q, o) \geq d(o_1, o)$ , $d(o_1, o) - d(o_1, q) > r$ , becomes $d(q, o) > r$ , thus it is clear that $o$ is not existed within search range.
--

### 3 Experimental Results

#### 3.1 Testing Preparation

The test in which improvement method of the present study is actually applied on the VP-tree and implemented in the similar video search was performed. In the test, a computer having specifications with Linux as OS, Intel Core i5-2300 as CPU, and memory 4GB was used. First of all, 10,000 videos were prepared as videos to be registered and characteristics of video were extracted using HSI histogram. HSI Histogram is a color histogram comprised of hue, saturation, and intensity. The four kinds of histogram sizes, namely 12(4×3), 24(8×3), 48(16×3), and 96(32×3) were used. Also, indexing was done with previously mentioned VP-tree for the 10,000 cases of video objects for which extract was completed. The  $vp$  during indexing was calculated as per maximum 100 cases of random data per each time. From around 1,000 videos which were not used during indexing, the distance calculation frequency and average value per one video case of CPU time was calculated using  $K$ -nearest neighbor search.

#### 3.2 Evaluations

The experiment for the  $K$ -nearest neighbor search method was executed as below by using an improvement method for the compaction of each search range. Each case indicated in the graph is the result of test by sequentially using four methods provided below. Here, since there is a research report that it is more effective to use multiple number of  $vp$  rather than compaction of search range using single  $vp$ , VP-tree in later case is adopted in the present test.

- $vp\_all$  : Compaction method of search range using multiple  $vp$ ,
- $vp\_nn$  : Compaction method of search range using the nearest point,
- $vp\_all\_nn$  : Compaction method of search range by combining  $vp\_all$  and  $vp\_nn$ , and
- AESA : Compaction method of search range by AESA

First of all, the test results for the distance calculation frequency at 12 dimension. This figure is the details of test results for the size 12.  $k$  at horizontal axis represents the search frequency, while  $calc\_num$  is the execution number of distance calculation. From the Figure, it is clear that the distance calculation frequency is reduced in the order of  $vp\_all$ ,  $vp\_nn$ ,  $vp\_all\_nn$  and AESA (See curves labeled 1, 2, 3, and 4, respectively).

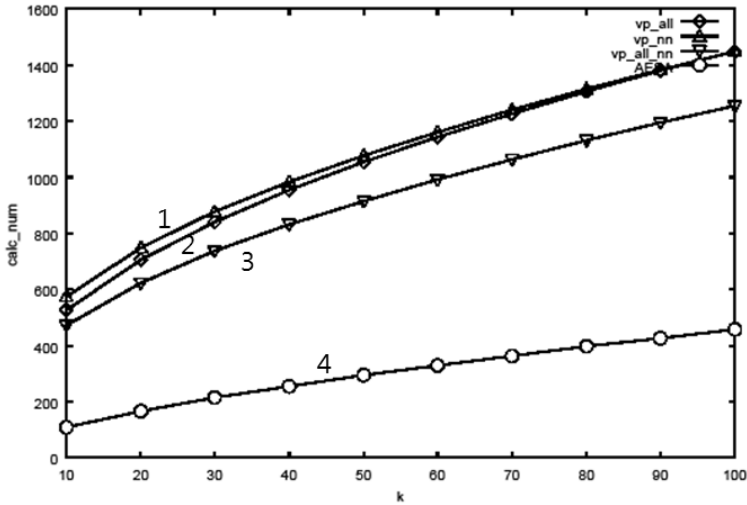


Fig. 1. Distance Calculation Frequency for the 12-dimensional Data

## 4 Conclusions

In this study, it was intended to attribute in reducing distance calculation frequency and in improving search speed by modifying search algorithm of leaf node in the VP-tree. By implementing this improved method, the test for similar video search was executed. It was found that the search speed could be reduced by 5% ~ 10% for similar video search. Besides, VP-tree provided faster search speed and was more useful than in other method. It is to design a search algorithm which can further reduce distance calculation time with less index size in future.

## References

1. Kita, K., Tsuda, K., Shishibori, M.: Information Retrieval Algorithm. Kyoritsu Shuppan, 5–25 (2002) (in Japanese)
2. Ruiz, V.: An Algorithm for Finding Nearest Neighbors in (approximately) Constant Average Time. *Pattern Recognition Letters* 4(3), 145–157 (1986)
3. Shishibori, M., Lee, S.S., Kita, K.: A Method to Improve Metric Index VP-tree for Multimedia Databases. In: *The 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Woolsan, South Korea, pp. 21–26 (2011)

# Efficient Locking Scheme with OPOF on Smart Devices

Hyun-Woo Kim<sup>1</sup>, Anna Kang<sup>1</sup>, Leonard Barolli<sup>2</sup>, and Young-Sik Jeong<sup>1,\*</sup>

<sup>1</sup> Department of Multimedia Engineering, Dongguk University,  
30 Pildongro 1 Gil, Jung-Gu, Seoul, Republic of Korea

<sup>2</sup> Department of Information and Communication Engineering,  
Fukuoka Institute of Technology (FIT)

3-30-1 Wajiro-Higashi, Higashi-Ku, Fukuoka, Japan  
z4538@nate.com, anakang37@gmail.com, barolli@fit.ac.jp,  
ysjeong@dongguk.edu

**Abstract.** Due to the rapid development of touch screens, smart devices have become immensely popular. Different smart devices, such as digital cameras, televisions, door-lock systems, tablet PCs, and smart phones, now have touch screens. Many functions on which people depend are built into such smart devices, and people may now use their smart phones to perform the same tasks they once carried out on desktop PCs. Therefore, touch screens have become popular due to the convenience and ease of use that they offer to users. However, the development of software application programs for popular smart devices focuses on providing diverse functions, therefore creating a risk of personal information leakage. In addition, portability is one of the advantages of smart devices; however, this also means that they may be lost, which is one of their disadvantages. Consequently, different lock functions have been created for information protection. Such lock functions include a different screen being displayed without a security function, a personal identification number (PIN) function, an existing button method, and pattern lock. However, PIN and pattern lock are very vulnerable to shoulder surfing or smudge attack. In this paper, the researcher proposes One Push out Free (OPOF), a function that provides users with an intuitive and easy interface and heightened security for their smart devices. Under OPOF, the pattern that a user input for the lock screen is perceived in different ways. This reduces the risk element for the pattern to be leaked by a third party's unlocking of the locked screen, and when the pattern is incorrectly input, another input mode is provided, heightening security and providing convenience through a simple interface.

**Keywords:** Smart Phone Security, Smart Device Locking System, Touch Screen, Locking System, Security.

## 1 Introduction

The current touch screens are the result of years of development since Samuel C. Hurst's development of an electronic touch interface in 1971. Many application

---

\* Corresponding author.

programs are being developed for smart devices using these touch screens. Smart devices offer the basic functions of schedule management, memo writing, phone number storage, and alarm, as well as e-mail, network games, navigation, augmented reality, and social networking services via the Internet. Such diverse content increases convenience in individuals' lives and enhances companies' task efficiency and cost reduction through enabling a mobile office. As a result, the number of smart device users continues to increase [1, 2, 3, 4, 5, 6]. Therefore, due to their diverse uses, the popularity of smart devices has increased rapidly, and they are now viewed as a necessity. Using a small smart device, most users may conduct tasks once reserved for a desktop PC, without any spatial restrictions, and smart phones, for example, store important information. Although they are small devices, they may store a lot of information, and therefore different lock systems for data security are built into them. The lock system takes various forms, such as drag, motion, face and voice recognition, patterns, and personal identification number (PIN). However, these different lock systems either do not have a security function or the input for unlocking them is inconvenient, and they are very vulnerable to shoulder surfing or smudge attacks, although they are supposed to offer security. Therefore, a new lock system is needed [7, 8, 9, 10, 11, 12].

In this paper, the researcher proposes "One Push Out Free" (OPOF) that provides users with an intuitive interface and offers heightened security. With OPOF, users can recognize pattern input for unlocking of their devices in different ways. As its starting point, OPOF may be applied to the whole touch screen area and by reusing the touch screen area, the strength of the secret pattern is enhanced, thereby providing improved security. The recognition method enables direct or random selection according to the setting, to prevent malicious persons from inferring the secret pattern.

The composition of this paper is as follows. Section 2 deals with locking systems built into smart devices and the attack methods for unlocking touch screen-based smart devices. Section 3 explains the pattern setting method under OPOF and diverse recognition methods. Section 4, 5, and 6 concern the design of OPOF, the embodiment of OPOF, and the overall conclusion and summary and suggestions for future research, respectively.

## **2 Related Works**

This paper first examines the screen locking function currently built into touch screen-based smart devices. Then it discusses malicious users' common attack methods aimed at inferring the secret pattern of a locked screen.

### **2.1 Previously Developed Locking Systems**

Common locking systems in smart devices include drag, motion, face and voice recognition, pattern, PIN, and password. In general, these locking systems are used to automatically turn off the screen and reactivate the screen using a power button after the screen has been turned off. Table 1 shows the names of the locking systems built into smart devices and explains their functions.

**Table 1.** Diverse Locking Systems Built into Smart Devices and Their Functions

<b>Embedded Locking Systems</b>	<b>Lock Functions</b>
Drag	Drag is a common locking system. It is basically a screen, defined by the user, with the purpose of hiding the content on the device. To unlock the locked screen, the user drags a certain area with a random starting point on the touch screen, or presses a certain button displayed and then drags to a certain area. Since only a certain area is dragged, anyone can unlock the screen, including malicious persons; therefore, it is a very vulnerable security option.
Motion	The recognition of motion depends on the smart device, and it is not frequently used. In order to unlock the locked screen, the smart device is tilted while the user touches the touch screen. It is convenient because it has a simple interface, like drag, but it offers low security because anyone can unlock the locked screen.
Face Recognition	This security setting demands that the user be at a certain distance and present the precise focal point of the face. Having to similarly display the facial expression set by the user in order to unlock the locked screen is also inconvenient, and inconsistency may occur according to the recognition distance. In addition, since other persons with a similar face may unlock the locked screen, face recognition does not offer high security.
Face and Voice Recognition	Face and voice recognition is a method that adds voice recognition to the face recognition method in order to remove the risk of other similar looking persons unlocking the locked screen—a disadvantage of the face recognition method. However, when the user speaks to the smart device to activate the voice recognition, his/her voice may be easily exposed to others, and if someone else can mimic the tone or intonation of the voice, this makes it possible for that person to unlock the locked screen. Therefore, this security method is very vulnerable. In addition, if there are problems with recognition, the user must try again, possibly numerous times, which is inconvenient.



**Table 1.** (continued)

Pattern	Pattern is a common locking system, similar to drag. It is composed of a $3 \times 3$ grid and access is granted by dragging each point. The number of recognized patterns is limited, and the security strength depends on the locking pattern set by the user. Screen locking through a pattern provides an easy interface and security.
PIN	In order to unlock the locked screen, the user can input a PIN. The PIN is composed of a combination of numbers, and the weaker the relationship between the numbers and the user, the higher the security. However, since the PIN it is made up of only numbers, its security strength is weaker than that of a password, and it is not easy to swiftly unlock the screen.
Password	Password is the most commonly used screen locking method. It can be composed of numbers, letters, and special characters, or a combination of these. It offers very strong security, but inputting a password to unlock the locked screen also poses a risk of leakage through screen exposure. In addition, a complex combination inclined toward high security may result in the user forgetting the password or slow down the unlocking process.

## 2.2 Representative Methods of Attack against Screen Locking

Even though many screen locking systems exist, the security that they offer is fundamentally vulnerable. This is due to the exposure of the touch screen input required to unlock the locked screen.

There are two common methods of attack to unlock a locked screen using a touch screen: shoulder surfing and smudge.

A shoulder surfing attack is simple and may allow the attacker to steal information easily. It is an attack method whereby a malicious person watches or records with a hidden camera when a user inputs a secret pattern to unlock a locked touch screen. Smart devices with inbuilt pattern, PIN, and password unlocking functions are very vulnerable to this attack method. In the case of PIN, since the numbers are aligned, it has even stronger vulnerability than the other screen unlocking methods. Although pattern is expected to offer strong security, due to the large space for movement, the area for input is limited to a  $3 \times 3$  grid, and since the secret pattern points are all accessible through dragging, this method does not offer high security.

A smudge attack literally utilizes a smudge. When a user inputs a secret pattern to unlock the locked touch screen, a fingerprint or trace remains on the touch screen. Smudge is a method to find out the secret pattern through the generated trace. Unless

the user is meticulous about maintaining the cleanliness of the touch screen, which he/she uses frequently, it can be exploited by a malicious person to launch an attack.

### 3 One Push Out Free

OPOF sets and unlocks the locking function using directionality. As a method for perceiving directions, a total of eight directions—north, northwest, northeast, south, southwest, southeast, west, and east—are provided. The starting point is possible on any part of the touch screen. The perception of directionality is made at the moment when the user's finger leaves the circular area generated at the starting point. Since only directionality is perceived, the touch screen may be reused, enabling the setting of diverse patterns. In addition, when perception occurs outside the circular area, another area may become the starting point, and perception is enabled by repeatedly entering and exiting the circular area. The diversity of secret patterns generated and the diverse inputs of the same secret pattern when inputting a pattern aimed at unlocking the screen may heighten security against malicious persons.

OPOF provides systematic random directionality in order to complement the user's vulnerability to shoulder surfing and smudge attacks. For random directionality, when a user inputs all set patterns, the directionality generated in the system should be input. For the directionality generated in the system, the current time is applied as a seed value to prevent overlapping. The number of directionalities generated in the system is provided so that the user or the system can determine the maximal length. Therefore, it will be difficult for shoulder surfing to determine the secret pattern that the user has set. Moreover, it is difficult for a smudge attack to judge whether the trace is a pattern set by the user or a systematic pattern resulting from the systematic directionality provided by OPOF, thereby heightening security.

### 4 Design of OPOF

The OPOF design proposed in this paper is largely composed of a user interface that receives input from the user; a point path trace that pursues the input; a direction (D)-manager that manages the input directionality; an intelligent direction pattern (IDP)-manager that manages the systematic directionality; a database(DB)-manager that stores and manages the input directionality; a lock (L)-service that provides services for screen locking; a coordinate converter that converts the input directionality so as to show it to the user; and an activity to show the OPOF setting and locked screen to the user. Fig 1 shows a structural map of the overall functions of OPOF.

The user interface consists of a direction and mode. The direction perceives the user's movements north, northeast, northwest, south, southeast, southwest, west, and east. The perceived directionality is sent to the IDP-manager and D-manager. Two kinds of modes are provided so that the user may select whether to lock the screen or unlock the locked screen.

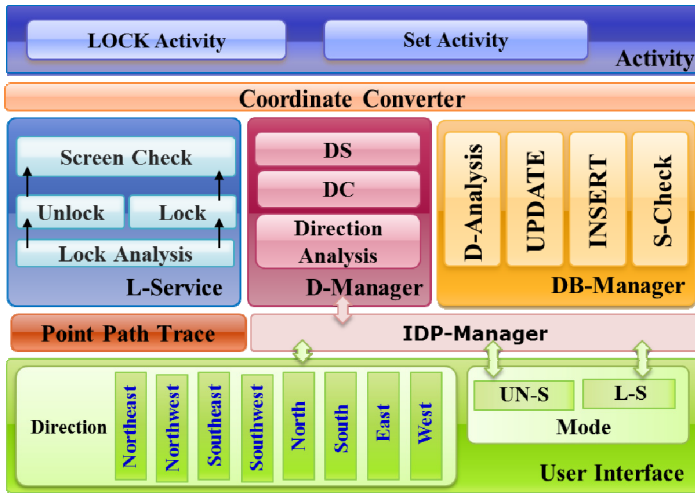


Fig. 1. Architecture of OPOF

Point path trace plays the role of calculating the distance from the point the user pressed to the point where the user moved in the user interface. This calculated distance is sent to a coordinate converter and may be viewed on a real-time basis.

The D-manager comprises direction analysis to analyze the directionality delivered from user interface, direction check to investigate whether the two inputs are the same in the case of setting a direction pattern, and direction save to send an index on directionality according to whether it is the same or not.

The IDP-manager operates according to whether the user set or not. It stores the data produced from the point path trace and compares the distances moved to unlock the locking and directionality and the direction pattern set by the user. Using the compared data, in order to reduce attacks by malicious persons, it demands that the user input a random direction pattern. The random direction pattern is induced to input directionality with low frequency in the user's input to unlock the locked screen. This may heighten the security strength against smudge attacks.

The DB-manager consists of D-analysis that analyzes the data conveyed and stores them in SQLite, update that aims at revising a directional pattern previously set, insert that aims at inputting a new directional pattern when it is executed for the first time, and s-check that checks whether the stored direction pattern exists and allows the user to select the mode for the unlocking of the screen.

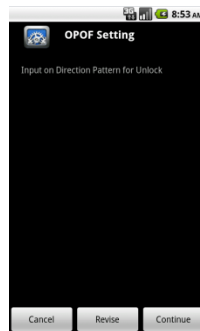
L-service comprises lock analysis that analyzes whether screen locking has been executed, unlock and lock that are executed according to the lock analysis, and screen check that is intended to perform a test of the movement direction when screen locking is performed. In screen check, when the input directionality differs from the stored directionality, the input directionality is initialized. The user may determine the number of initializations, and they are used as the number of attempts to unlock the screen.

Coordinate converter plays the part of processing the user's directionality input and sending it so that it may be used in activity.

Activity comprises set activity, through which the user sets the direction pattern to be used for screen locking, and lock activity, which is executed when pressing the power key or when the screen is turned off, in case the user sets and executes screen locking.

## 5 Implementation of OPOF

When OPOF is executed for the first time, there is no information about the secret pattern using directionality; therefore, the user must input a directional pattern whose locking will be set. Fig 2 shows a screen where OPOF has been executed for the first time; the screen consists of a touch area for inputting the pattern and a cancel button for canceling the pattern, a revise button to revise when a directional pattern has been wrongly input, and a continue button to perform the next progression after inputting the directionality.



**Fig. 2.** Initial Execution Screen of OPOF

Fig 3 displays a screen that has perceived the directionality of the user who touched it. When each direction—north, northeast, east, southeast, south, southwest, west, and northwest—is perceived, it is shown to the user as an icon at the selected starting point.

Like Fig 3, Fig 4 shows a screen when a total of 10 directions—north, northeast, east, southeast, south, southwest, west, southwest, west, and northwest—are input through the perception of directionality. By showing the input directionality to the user, he/she may verify it, and when the user wishes to revise it, he/she may do so using the Revise button at the bottom of the screen.

Fig 5 displays a screen demanding that the user input a directional pattern when he/she sets the locking of the screen. The left side of the figure is a screen without any touch input and the right side is a screen showing that any location may be used as the starting point for the input of directionality.

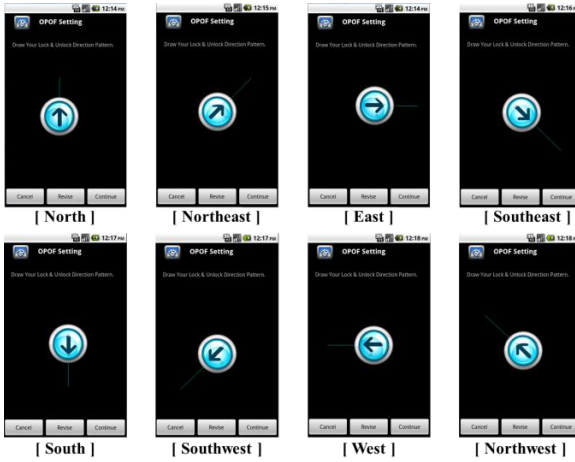


Fig. 3. Screen Displaying the Perception of Directionality



Fig. 4. Screen When 10 directionalities are input by the user

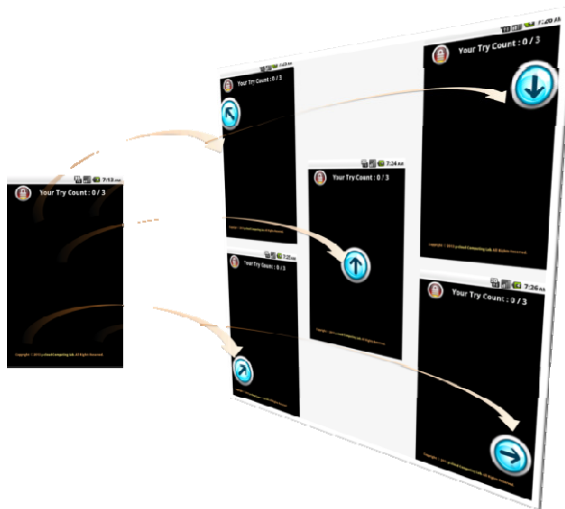


Fig. 5. Initial Screen (Left) and Input Screen (Right) When Screen Locking Is Set

## 6 Conclusion

Touch screen-based smart devices are utilized by many different people, due to their portability and many inbuilt functions, and as the storage capacity of these devices increases, they can be used to store a lot of data related to tasks, as well as personal data. As a result, they offer increased convenience. However, the number of malicious persons and the risk of losing data have increased. In particular, if a device's touch screen is frequently exposed, the device may become a target for malicious persons.

In this paper, the researcher proposed OPOF in order to address the vulnerability of touch screens to shoulder surfing and smudge attacks. OPOF perceives directionality and therefore the touch screen may be reused, enabling the setting of screen lock with many directional patterns. When the screen is locked, OPOF provides several methods for unlocking it that differ from the previous method of inputting the same pattern; therefore, it is difficult for a malicious person to infer the pattern. In addition, in order to increase security against smudge attacks, systematic random directions were provided. Through systematic random directions, the inference of traces becomes all the more difficult, and the inconvenience of having to polish the touch screen every time is resolved, allowing the user to use the device with ease.

In the future, in order to derive systematic random directions more intelligently, the researcher intends to converge OPOF with an algorithm for the neural network. In addition, the researcher aims to study a middleware that may transmit the location of the device to a contact number set by the user. This middleware will be installable in all smart devices and other digital devices that enable mail and Internet access, so that the user may use it conveniently.

**Acknowledgment.** This research supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-4007) supervised by the NIPA (National IT Industry Promotion Agency).

## References

- [1] Alberts, C.J., Dorofee, A.J.: *Managing Information Security Risks: The OCTAVE Approach*. Addison-Wesley Professional (July 2002)
- [2] Mulliner, C., Vigna, G., Dagon, D., Lee, W.: Using Labeling to Prevent Cross-Service Attacks Against Smart Phones. In: Büschkes, R., Laskov, P. (eds.) DIMVA 2006. LNCS, vol. 4064, pp. 91–108. Springer, Heidelberg (2006)
- [3] Carey, M.: *Enterprise Risk Management: How to Jumpstart Your Implementation Efforts*. International Risk Management Institute (February 2005)
- [4] Shtykh, R.Y., Jin, Q.: A Human-Centric integrated approach to web information search and sharing. *Human-centric Computing and Information Sciences* 1(2) (November 2011)
- [5] Obaidat, M.S., Zarai, F.: Novel algorithm for Secured Mobility and IP Traceability for WLAN Networks. *Journal of Convergence* 3(2), 1–8 (2012)
- [6] Singh, S.K., Sabharwal, S., Gupta, J.P.: A Novel Approach for Deriving Test Scenarios and Test Cases from Events. *J. Inf. Process. Syst.* 8(2), 213–240 (2012)

- [7] Park, M.: The Evolution of the Mobile Phones with Touchscreen and the Prospect of Future: Focused on the SRI-Tech. Master's Thesis, Incheon University (2011)
- [8] Jaegal, B.: Trend of Smartphone Market and Mobile OS. Samsung Wireless Division (June 2010)
- [9] Gong, Y.: Implications and Agreement of Smartphone. Korea Information Society Development Institute 22(4), 480 (2010)
- [10] Kim, C.-S., Yoon, S.-B., Lee, M.-K.: Shoulder-Surfing Resistant Password Input Method for Mobile Environment. Journal of Korea Institute of Information Security and Cryptology 20(3), 93–104 (2010)
- [11] DeAlvarez, A.M.: A Framework for Password Selection. In: UNIX Security Workshop II (1998)
- [12] ITU-T, Security Aspects of Mobile Phones, T09 SG17 100407 TD PLEN 1012 (April 2010)

# A Requirement Analysis of Awareness-Based Vessel Traffic Service System for Maritime Safety\*

Byunggil Lee<sup>2</sup>, Namje Park<sup>1,\*\*</sup>, and Juyoung Kim<sup>2</sup>

<sup>1</sup> Department of Computer Education, Teachers College, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
namjepark@jejunu.ac.kr

<sup>2</sup> Electronics and Telecommunications Research Institute (ETRI),  
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Korea  
{ap424, bglee}@etri.re.kr

**Abstract.** Combining next u-Navigation concept with the next-generation maritime safety technology such as the situation recognition technology, the risk-based navigation support management technology, etc., the next-generation VTS(Vessel Traffic Service) technology should be developed. This means the necessity of developing the technology differentiated from the existing VTS in operation. Consequently, this paper defines the requirements of the situation-recognition-based next-generation VTS to materialize the e-Navigation, and proposes the followings: the situation-recognition processing structure support based on this processing.

**Keywords:** VTS, Vessel Traffic Service, e-Navigation, Maritime.

## 1 Introduction

The Vessel Traffic System (VTS) field that is about maritime safety has mostly relied on overseas technology, unlike the shipbuilding industry that has recently retained the leader's position in the global market as a traditional industry. The VTS technology in the maritime safety field consists of maritime IT technology and has desperately required the grafting with the up-to-date IT technology.

In the maritime field, the concept of "e-Navigation" for the grafting of the electronic information technology was introduced into Europe, and the popularity of this concept has been rising internationally in recent two to three years. The e-Navigation promoted by IMO is about collecting/integrating/expressing/analyzing the marine data between ships and the land in harmony through the electronic method with the purpose of marine safety/security and marine environment protection through the improvement of the sailing-related services.

---

\* This work was supported by ETRI through Maritime Safety & Maritime Traffic Management R&D Program of the MLTM/KIMST. (No.D10902411H360000110, Development of u-VTS for Maritime Safety).

\*\* Corresponding author.



VTS plays the following key roles with the purpose of materializing the e-Navigation in the marine environment: collecting/integrating/analyzing the various data related to marine traffic control and then providing the data to the applicable ships. The VTS Committee (a professional IALA group) has been actively discussing VTS' role, VTS service, etc., in order to establish the new concept suitable for the e-Navigation environment.

Internationally, there is a trend that VTS has been evolving to Vessel Traffic Management (VTM) recently and VTS' overall concept has expanded as the framework of the methods and services to enhance the followings: the safety in the sailable waters, the efficiency in security and shipping, and the marine environment protection. In other words, this service architecture has been changing into a new service type, not only in the maritime traffic safety and business service, but also in the maritime computing environment, and foretells the change in the existing VTS and its concept.

Through combining this e-Navigation concept with the next-generation maritime safety technology such as the situation recognition technology, the risk-based navigation support management technology, etc., the next-generation VTS technology should be developed. This means the necessity of developing the technology differentiated from the existing VTS in operation. Consequently, this paper defines the requirements of the situation-recognition-based next-generation VTS to materialize the e-Navigation, and proposes the followings: the situation-recognition processing structure support based on this processing.

## 2 Vessel Traffic System

In general, the elements composing VTS are shown in Figure 1. VTS is the system in which the followings are connected to one another: the VTS Center on the land, the base station site on which various sensors (sensing devices such as CCTV, Radar, DF, MET, etc.) and AIS are installed, the Control Center that actually operates VTS, and it is a complicated system consisting of various types of telecommunications networks that connect ships, satellites, and sensing devices.

As for the VTS-related study overseas, through the Framework Programme (FP) project, they promoted various studies whose objects include the next-generation technology of VTS, Vessel Traffic Management & Information System (VTMIS), and Port Control Management Service PCS). As the MarNIS project to be implemented between 2012 and 2020, these study is marked by providing the VTM and Search And Rescue (SAR) services through collecting various information such as the ship's dynamic/static data and waters climate/geography/environment by means of various media and processing the data safely and efficiently. Besides, in the MarNIS project, they have been conducting the aids-to-navigation study (including the marine mobile communication network technology) for the enhanced multimedia telecommunication. Especially, they applied the enhanced controlling function, multimedia telecommunications function, etc., and have been conducting the follow-up studies and developments continuously for the actual service implementation and the international standardization.

In the past, VTS had a narrow meaning or a classical meaning that VTS' purposes were mainly to control the applicable area through monitoring and controlling vessel traffic in the radar signal control area. However, VTM that has been promoted by IALA recently has no limit to the control area and no limit to competent offices for the expansion of information sharing. In other words, not only competent offices have expanded to various sensors such as AIS, satellites, etc., but also the concept of information sharing has expanded to the provision of aids-to-navigation and various information.

In the existing VTS, low-cost telecommunication system in the vest waters does not exist and remote telecommunication technology except for satellites is not available. Thus the aids-to-navigation function in the ship's maritime situation is weak and seldom used. Besides, the risk management is done at the level that the alarms sounds for the anchored ships are generated on the basis of the degree of risk that is calculated based on the simple distances and speeds so that the risk management is seldom used in the service, which is a weak point. As the technology for safe ship control, various algorithms that calculate the degree of ship collision risk and apply it to sailing have been studied continuously. For example, a risk management method that calculates the degree of collision risk at the ship's location on the basis of the fuzzy theory is available, but the accuracy of the moving direction according to the ship's intention and the predicted location is most important.

However, in general, the existing sailing patterns or the preferred routes might be available to a ship-navigation officer, and these cannot be obtained through a simple mining for the existing sailing data. It is because the situation on the sea is always dynamic and keeps changing. Consequently, it is desirable to analyze the contexts of the situation-specific preferred routes, establish the profile data of the detailed historic tracks, and predict the degree of the collision risk by combining this data with that about the situation on the sea. The control system requires an efficient system structure because it should be able to recognize the current situation on the sea by analyzing the data about numerous contexts and to transfer the applicable situation to the controller rapidly and intuitively.

Besides, as the control system also requires the situation analysis function to find a dangerous situation through the analysis of the situation on the sea, the visualized screen for intuitive recognition, and the function to create and transfer data for processing the collision-avoiding event, in this thesis, the author proposes the situation-recognition structure that satisfy these requirements.

### **3 Requirement Analysis the Proposed Aids-to-Navigation Structure**

The aids-to-navigation function of the current VST is used, relying on the controller's voice; most items including a collision risk, a burdened region, etc. are operated, relying on the controller's experiences; and the characteristics of the automatic aids-to-navigation support by the system or those of the aids-to-controller's decision-making are not considered. Besides, the existing VST does not use the detailed

information of the past traffic routes as a technique to control at the present time. And the aids-to-navigation telecommunication system is seldom used as the low-speed telecommunication structure in AIS' binary transmission mode. However, recently, the international issues including the aids-to-navigation technology (e-Navigation) show the trend of changing VTS to the system grafting the latest IT technology.

This paper, with the purpose of resolving these issues, intends to design the system to aid the controller's decision-making through recognizing new situations on the sea, predicting routes based on the recognition, and guessing the degree of risk, and also to design the sailing support structure through the new VHF telecommunication network.

First, the paper intends to analyze the aids-to-navigation requirements for the next-generation VTS and to design the VTS structure adopting the analysis result. VTS has the following functional requirements to provide the intelligent controlling service with the next-generation situation-recognition-based aids-to-navigation characteristics that is different from the existing VTS:

(1) Multi- VTS sensor data collection and various situation recognition data collection. VTS should have the function to collect the followings: the radar data for the situation-recognition data collection, the data about ships and sailors through the interworking with AIS and PortMIS, CCTV data, and the data collected through other sensor networks.

(2) The analysis of the degree of the maritime collision/stranding risk and the avoidance of them. VTS should have the function to calculate the degree of the risk of stranding or the collision between ships in real time through the ship navigation data and to perform a collision in consideration of all route conditions.

(3) Ship/waters-specific risk management and the automatic safe sailing data creation VTS should have the function to create the data about the risk management in consideration of the applicable waters-specific Dynamic Under keep Clearance (DUKU) through using the ship navigation data and the data about the application waters-specific safe sailing and to transmit the data automatically.

(4) The optimal sea route search and ship berth/unberth management. VTS should provide the navigation data through predicting the followings, provide the navigation data by exploring the carrying vessel's optimal route, and provide the data about ship entrance/clearance or ship berth/unberth: the high possibility of collision due to the complicated traffic situation and the ship entrance/clearance in bad weather, the piloting and moving situation of a hazard material.

(5) The customized controlling in consideration of the port/ship characteristics. VTS should consider the port characteristics such as narrow channel, shoaling zone, etc., and provide the necessary controlling function according to the characteristics of the ship type (cargo ship, passenger ship, dangerous cargo vessel, etc.) and the types of main ships entering/clearing each port.

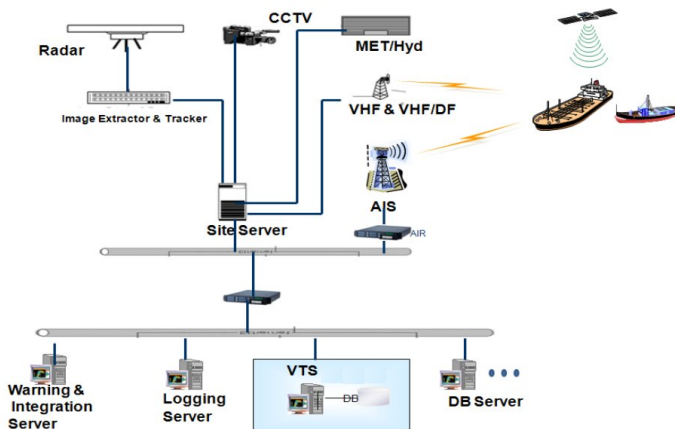
(6) Modeling and simulation-based 3D monitoring. VTS should provide the 3-dimensional monitoring of 3-dimensional ports based on the electronic navigation chart through using modeling and simulation-based 3D virtual reality technology.

(7) The countermeasure against oil pollution and the early warning of the accident response. VTS should be able to understand and predict the pollution degree in case of an accident for an early response towards oil pollution accidents, and provide an early alarming to relevant institutes in case of an accident.

(8) Support of situation-specific controller's decision-making. VTS should introduce/provide the various algorithms and systems to support the decision-making of sailing data support system according to the situation on the sea.

(9) Marine device-based maritime situation recognition data collection and analysis. VTS should be able to directly collect the maritime situation data such as bad weather through interworking with the various structures on the sea and the telecommunication/sensor devices on ships, to collect various situational data through radio/wired networks, and to analyze the collected data.

(10) Situation-specific extraction of the sailor preference through the data-mining method. During the vessel traffic control, VTS should be able to perform the data-mining according to the situation for the existing routes and then to use it for route prediction according to the real-time situation on the sea.



**Fig. 1.** System Architecture of VTS

This u-VTS has been designed based on the model described in the above with the configuration in Figure 1. Overall, it consists of 25 individual systems and includes the followings: the multi-VTS sensor network and terminal (consisting of radar, AIS, CCTV, other sensor networks), the sensor data processing system (7~10, performing target extraction and tracing), the data integration and controlling sophistication system (fusing sensing data and aids to controller's decision-making), the control data management system, operation and management system, and interworking (I/F)G/W. In other words, developing the mobile object situation recognition technology in the ubiquitous intelligent space requires the sophistication of situation deduction function and the sailor preference management function. Besides, the uVTS should classify/log the various sailing history that the sailor of the target (object, ship) in details, and

provide the detailed route prediction reflecting the optimal preference through the data-mining method.

## 4 Conclusion

According to the recent worldwide trend, the VTS and the maritime computing environment have been rapidly changing to adopt the e-navigation-based intelligent broad control service architecture through the active fusion with the rapidly-developing IT technology. Besides, the maritime mobile radio telecommunications technology is expected to become the basic infrastructure for providing the service of the type desired by all users in the maritime environment where users are mobile.

The paper focused on proposing the system design and processing structure that are based on the situation recognition technology for the next-generation VTS, so in the next-step study, it is expected to present the result of the specific study about the performance analysis and service platform regarding the application of this technology.

## References

1. Laxhammar, E., Falkman, G., Sviestins, E.: Anomaly detection in sea traffic – a comparison of the Gaussian Mixture Model and the Kernel Density Estimator. In: ISI 2009 (2009)
2. Kawaguchi, A., et al.: Towards the development of intelligent navigation support systems for group shipping and global marine traffic control. IET Intell. Transp. Syst. 3(3), 257–267 (2009)
3. Lee, B.-G., Han, J.-W., Jo, H.-S.: Design of situation awareness and aids to navigation structure of VTS for maritime safety. The Journal of Korea Information and Communications Society 35(7) (2010)
4. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
5. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010, Part II. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
6. Park, N.: Secure UHF/HF Dual-band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
7. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. International Journal of Ad Hoc and Ubiquitous Computing 8(4), 205–219 (2011)
8. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
9. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. The Journal of Korea Information and Communications Society 35(5), 785–793 (2010)

# Analysis of Open Interface Function in USN Service Middleware System<sup>\*</sup>

Yilip Kim<sup>1</sup>, Dong-Hwan Park<sup>2</sup>, Hyo-Chan Bang<sup>2</sup>, and Namje Park<sup>1,\*\*</sup>

<sup>1</sup> Major in Computer Education, Faculty of Science Education,  
Graduate School, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
{yilipkim, namjepark}@jejunu.ac.kr

<sup>2</sup> Electronics and Telecommunications Research Institute (ETRI),  
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Korea  
{dhpark, bangs}@etri.re.kr

**Abstract.** Rather than the USN applications specialized for individual sensor networks, therefore, this paper proposes an open application interface of USN Service Middleware Platform that provides common functions for the sensor data processing of sensor networks and applications in order for the development of application services for which broadband USN application services based on the various type of sensor networks and links of services are required. Using the application interface, efficiency in developing applications, integrating application services and maintaining application services can be improved a lot.

**Keywords:** Open USN Service, Semantic, Platform Model, USN, WSN.

## 1 Introduction

In the Ubiquitous environments, various kinds of heterogeneous sensor networks are deployed. The common interface of sensor networks is a standard to remove the dependency of application services on the heterogeneity of such various sensor networks. The sensing data that are acquired and analyzed continuously or periodically are processed and transmitted according to the requirements of diverse application services. USN Service Middleware Platform is a common platform that provides common interfaces to reduce the dependency on the multiple sensor networks for efficient collection, management and supply of sensor data by the common processing function for sensor network abstraction, high level of intelligent analysis and provision of sensor data, and so on.

---

<sup>\*</sup> This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A4A01 013587).

<sup>\*\*</sup> Corresponding author.

Rather than the USN applications specialized for individual sensor networks, therefore, this paper proposes an open application interface of USN Service Middleware Platform that provides common functions for the sensor data processing of sensor networks and applications in order for the development of application services for which broadband USN application services based on the various type of sensor networks and links of services are required. Using the application interface, efficiency in developing applications, integrating application services and maintaining application services can be improved a lot.

## **2 Open USN Service Middleware Platform Model**

The reference model of USN Service Middleware Platform proposes roles, types and functions of the platform for USN application service and defines the platform architecture to provide such functions.

USN Service Middleware Platform is linked to the multiple heterogeneous sensor networks by the sensor network common interface standard as shown in the figure. The platform abstracts sensor networks to cope with the heterogeneity and commonality of the heterogeneous sensor network. The platform monitors the state of sensor network or provides the control commands via the abstraction layer of sensor network. Especially it acquires the sensor data periodically and continuously by the sensor data collection command to the sensor network or transmits the sensor data from the sensor network to the platform periodically and continuously. In order to carrying out such sensor network abstraction function, the key information of sensor network is displayed, managed and used according to the USN metadata model standard. Especially for the external application service, it provides the function to generate, retrieve and manage the metadata by the USN metadata directory service interface standard.

## **3 Function Requirements of USN Service Middleware Platform**

USN Service Middleware Platform should satisfy the following requirements in order to provide the function to acquire and analyze sensor network monitoring information or sensor data periodically and continuously in the links of multiple heterogeneous sensor networks for USN application services.

### **3.1 User Authentication and Authorization**

USN Service Middleware Platform should provide authentication for the multiple USN application services registered in the platform when they request for their access to the platform and authorization to allow the requested service functions to access to certain sensor networks. The detailed requirements as follows.

- 1) USN application service should be able to be registered as the user that is uniquely identified in the platform.

- 2) Each user can select one between ID based method and authentication certification based method and the platform can perform authentication in the method set by the user.
- 3) In case of authentication failure, proper error messages should be issued.
- 4) Authority for the user to access to the registered platform service can be set by user in the platform.
- 5) In case of using the platform service for which authentication is declined, proper error messages should be issued.
- 6) Authority for the user to access to the registered sensor network can be set by user in the platform.
- 7) The platform should be able to perform appropriate authentication for the legal use in case that the user uses the sensor network via the service function.
- 8) In case of using the sensor network for which authentication is declined, proper error messages should be issued.
- 9) The platform should be able to handle the release request of the USN application service for access to the platform.

### **3.2 Platform Session Management**

USN Service Middleware Platform should provide the session management function to generate and remove the application sessions for the application services after their getting authorization from the platform and the function to generate and remove the service sessions for the service functions of platform. The detailed requirements are as follows.

- 1) The platform should be able to generate new application sessions for the authorized application services.
- 2) The platform should be able to provide unique identifier to the application session of authorized application service.
- 3) The platform should be able to maintain the application session continuously and to remove it by the explicit request including the application session identifier.
- 4) The platform should be able to generate service session by platform service for an effective application session.
- 5) The platform should be able to provide unique session identifier to the generated service session.
- 6) The platform should be able to remove the service session by the session identifier for the service session.
- 7) The platform should remove all service sessions related to the application session when the application session is removed.

### **3.3 Service Message/Notification**

USN Service Middleware Platform provides the processed results by the response to the requests for service function processing. But it cannot response to such requests of application service instantly because sensor network state and sensor data from the sensor network are collected and processed continuously or periodically. Therefore,



the processed results should be provided asynchronously against the requests of application service in two types: message and notification. Message is made when the platform performs and makes the data processing results for the data collected from the sensor network. Notification is made when abnormal states or events are detected by imposing certain conditions on the sensor network, the platform and the results of service functions. The detailed requirements are as follows.

- 1) The platform should be able to provide the processed results by the instant response to the service function request of the authorized application service.
- 2) The platform should be able to provide asynchronous messages continuously or periodically in case that the results of service function request of the authorized application service are generated continuously or periodically.
- 3) The platform should be able to provide notifications periodically or continuously when certain types of changes and detections accorded with conditions occur in the sensor networks, platform and service functions that the authorized application service requests.
- 4) The application service should register the message destination to receive messages or notifications from the platform and such destination information should contain receiving method (JMS, JDBC and RMI), receiving address (URL or IP address and port information), receiving information (parameters according to the receiving methods), etc.
- 5) The application service can request notification by designating the notification type registered in the platform. Of the recognized types of notifications, the platform should be able to provide the notification only requested by the application service.
- 6) The platform should be able to generate the unique identifier of message or notification according to the receiving destination registration of the application service and to provide it to the application service.
- 7) The platform should be able to send a message of a notification for the registration of message or notification destination of the application service.
- 8) The platform should be able to cancel the registration of destination explicitly by the directory or notification destination identifier based on the requests of application services.
- 9) The platform should be able to delete the registration of message or notification destination according to the deletion of application sessions or service sessions.

### **3.4 Sensor Network Monitoring and Control**

USN Service Middleware Platform should be able to monitor the changes of state information of the multiple sensor networks linked to the platform and to control the operation of sensor networks by detecting the requests of application services or abnormality of sensor networks. The detailed requirements are as follows.

- 1) The platform should be able to acquire the state information of sensor networks continuously by performing the sensor network monitoring operation after the sensor networks are normally linked to the platform.

- 2) The platform should be able to update the information registered on USN metadata when the state information of sensor networks are changed and to provide the application service requesting sensor network notification with the notification of state change of sensor network.
- 3) The application service should be able to control the authorized sensor networks by the commands of start, suspension, stop, etc. of monitoring operation.
- 4) The application service should be able to change the monitoring cycle for the authorized sensor network.

### 3.5 Sensor Data Acquisition/Processing

USN Service Middleware Platform should be able to acquire sensor data continuously from the multiple sensor networks linked to the platform, to provide the application service with the data and to provide the application service with processed results of the acquired sensor data requested by the application service. The detailed requirements are as follows.

- 1) The platform should be able to handle the sensor data query requests of multiple application services simultaneously to acquire sensor data from the multiple registered sensor network including sensor or actuator.
- 2) The platform should be able to handle the event query requests by providing the results to the application services in case that certain given conditions are satisfied.
- 3) The platform should be able to handle the continuous query requests by providing the sensor data acquired and processed according to the given time period and cycle.
- 4) The platform should be able to handle the stream query requests by providing the sensor data acquired and processed continuously according to the time window and moving slice.
- 5) The query process requests of application services should be possible by the request interface including the command lines having the similar grammar of SQL or by the request interface based on parameters by stag.
- 6) In case of query process requests of application services, the sensor network identifier should be designated as the unique identifier following the USN metadata standard or in the alias name, and the sensor types of sensors should be designated in the sensor type name.
- 7) In case of query process requests of application services, each query should include the processing of aggregate functions of max, min, avg, count, etc.
- H. In case of query process requests of application services, conditional equation should be included to limit the sensor values.
- 8) In case of query process requests of application services, the process of logical sensor networks should be supported to bind more than two sensor networks logically to be a single identifier.
- 9) The platform should be able to provide query identifier that is uniquely recognized in the query requests.

- 10) The platform should be able to control the query processes, that are performed continuously or periodically by the application services, by suspending, resuming and stopping them.

## 4 Conclusion

The USN service middleware platform is a common platform that provides not only interconnection between services and multiple heterogeneous sensor networks that conform to the sensor network common interface standard but also functions of continuous data gathering and processing. By categorizing the necessary functions and interface provided to USN service by USN service middleware platform, this paper defines a open application programming interfaces, which can be used in implementing various USN application services and service middleware platform.

## References

1. Lee, J.: Open APIs of USN Service Middleware Platform. Telecommunications Technology Association(TTA) Standard, TTAK.KO-06.0284 (2012)
2. Park, D.-H., Kim, S.-J., Bang, H.-C.: Semantic Open USN Service Platform Architecture. In: ICTC 2012, pp. 12–15 (2012)
3. Kim, M., Lee, J.W., Lee, Y.J., Ryou, J.-C.: COSMOS: A Middleware for Integrated Data Processing over Heterogeneous Sensor Networks. ETRI Journal 30(5), 696–706 (2008)
4. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
5. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
6. Park, N.: Secure UHF/HF Dual-band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
7. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. International Journal of Ad Hoc and Ubiquitous Computing 8(4), 205–219 (2011)
8. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
9. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. The Journal of Korea Information and Communications Society 35(5), 785–793 (2010)

# Security Controls Based on K-ISMS in Cloud Computing Service<sup>\*</sup>

Jongho Mun<sup>1</sup>, Youngman Jung<sup>1</sup>, Jeeyeon Kim<sup>1</sup>, Youngsook Lee<sup>2</sup>,  
Kidam Park<sup>3</sup>, and Dongho Won<sup>1,\*\*</sup>

<sup>1</sup> College of Information and Communication Engineering, Sungkyunkwan University, Korea  
{jhmoon, ymjung, dhwon}@security.re.kr, jykimpro@daum.net

<sup>2</sup> Department of Cyber Investigation Police, Howon University, Korea  
ysooklee@howan.ac.kr

<sup>3</sup> Wins Technet Co. Ltd., Korea  
kdpark@wins21.co.kr

**Abstract.** Cloud computing service can be provided by diverse forms such as XaaS, MSP and others. The most common examples of XaaS are Software as a Service, Infrastructure as a Service, and Platform as a Service. A MSP(Management Service Provider) is a type of IT service companies that provides network, server and specialized applications to end user and organization. Also cloud computing service offers many benefits of reducing cost by sharing storage resource, but is vulnerable to threats. Therefore the issue of information security should be solved to mitigate the threat in cloud computing environment. The cloud computing service providers had better establish the ISMS (Information Security Management System) to manage risks systematically. In this paper, we propose the security controls for cloud computing service which are based on K-ISMS through a comparative analysis of domestic and foreign cloud computing service guidelines. These security controls will be useful for organization's ISMS.

**Keywords:** Cloud Computing, Cloud Computing Service, Security Controls, K-ISMS, Virtualization.

## 1 Introduction

At first, John McCarthy who is computer scientist of the USA, presented the concept of cloud computing as "Computer is used as a public facility", in 1960s. Since then, many organizations have defined cloud computing. In 2008, Gartner [1] defined cloud computing as an alternative delivery and acquisition model for IT-related services. In 2011, NIST (National Institute of Standards and Technology) [2] defined cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage,

---

<sup>\*</sup> This research was supported by the Wins Technet Co. Ltd. in 2013.

<sup>\*\*</sup> Corresponding author.

applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Recently, information area of our lives is increasing gradually, with the development of information and communication technology. As a result, the amount of information to be processed is increasing. Many companies and organizations introduce the cloud computing because of the cost savings of processing much information, efficiencies, flexibility, innovation, and market opportunities. According to SERI (Samsung Economic Research Institute) [3]'s findings, the market size of cloud computing will be \$200 billion. According to Gartner [1], however, since cloud-computing environment are externally provided and shared, organizations need to evaluate risk in areas such as data integrity and privacy and need to understand issues in areas such as e-discovery, compliance, and audit reporting.

In 2011, TechAmerica Foundation's Commission of the Leadership Opportunity in U.S. Deployment of the Cloud (Cloud2 Commission) released the "Cloud First, Cloud Fast: Recommendations for Innovation, Leadership and Job Creation [4]". According to this report, the cloud computing service providers collaborate with the NIST, relevant associations and standards bodies to assess and evolve current best practices and standards, to strengthen cloud security metrics, and to facilitate information sharing. For the reason, some organizations have defined specific certification standards such as FedRAMP (Federal Risk and Authorization Management Program), CCM (Cloud Controls Matrix) and so on. Also, many organizations provides security controls for cloud computing service.

Cloud computing service offers many benefits of reducing cost by sharing storage resource, but is vulnerable to threats. Therefore the issue of information security should be solved to mitigate the threat in cloud computing environment. The cloud computing service providers had better establish the ISMS to manage risks systematically. In this paper, we propose the security controls for cloud computing service which are based on K-ISMS through a comparative analysis of domestic and foreign cloud computing service guidelines. These security controls will be useful for organization's ISMS.

The rest of the paper is organized as follows: Section 2 analyzes the security risks and the certificate standards of cloud computing service. Section 3 compares the standards introduced in the previous section. Section 4 proposes a new security controls for cloud computing based on K-ISMS. Finally, we conclude this paper in Section 5.

## 2 Related Work

In this section, we analyze security risks of cloud computing service, and investigate the guidelines of cloud computing service.

### 2.1 Security Risks of Cloud Computing Service

Recently, Google, Amazon.com, Sun, and other companies that provide storage services and virtual services are emerging. However, in the middle of February 2010, the

Amazon network host service S3 (Simple Storage Service) was broken down for 4 hours. This made customers think about the security of cloud computing service again. For this reason, the organization must establish the security measures for threat in order to build a cloud computing environment or take advantage of cloud services.

Cloud computing assembles all the computing resources and manages them automatically through virtualization software. The key characteristics of cloud computing service include on-demand outsourcing service, virtualization, information consignment, resource sharing, and variety of terminal device. Because of the characteristics of cloud computing, it has the security risks such as Table 1.

**Table 1.** Security risks of cloud computing

<b>Risks</b>	<b>Description</b>
Vulnerability of virtualization	Cloud services have the vulnerability of virtualization if the services use existing virtualization techniques.
Information leakage due to consignment	It is possible that information is leaked by internal user or malicious user in cloud services because user's information is stored in remote cloud server.
Service failure due to resource sharing and centralization	Users in cloud services share physical resources provided by service provider. Because most resource are concentrated on the server, Services of all users are stopped if some problems occur.
Information leakage due to diversity of devices	Cloud services have the security vulnerability of all devices such PC, smart phone, smart TV, etc.
The difficulty of applying security in distributed processing	It is difficult to do data encryption, user authentication, and access control because large amounts of data are stored across multiple servers.
Legal and regulatory issues	Cloud services have various environmental characteristics unlike the traditional way. Therefore the problem may occur if the existing laws is applied.

Gartner [5] says, cloud computing has “unique attributes that require risk assessment in areas such as data integrity, recovery, and privacy, and an evaluation of legal issues such as e-discovery, regulatory compliance, and auditing”. Therefore, if information security evaluation criteria is established for cloud computing service, and the cloud service provider and service is evaluated, then the customers will have confidence in cloud computing service.

## 2.2 Cloud Service Information Security Guide

In 2011, Korea Communications Commission (KCC) and Korea Internet Security Agency (KISA) issued a guide for information security of cloud service [6]. This guide consists of the characteristics and vulnerabilities of cloud service, information security controls for cloud service provider, and information security controls for cloud service customer. Also, it provides a check list for cloud service provider and customer as appendix. Table 2 shows the check list for cloud service provider.

**Table 2.** A check list of the cloud service information security guide for cloud service provider

Control Area	Control Name	Number of controls	
Cloud service security policy	Information Security Policy	2	
Security organization operation and personnel security	Security Organization Operation	6	
	Personnel Security	5	
Asset classification and control	Asset Protection	3	
	Asset Classification	2	
Cloud service incident management procedure	Security Issue Report	2	
	Incident Response	3	
Service Sustainability	Service Sustainability	3	
Compliance	Compliance	7	
Physical security	Security Control	5	
	Equipment Protection	6	
Communication and Operation	Operation Management	2	
	System Management	1	
	Malicious Code Response	2	
	Backup	1	
	Network Security Management	1	
	Audit	6	
	Security Requirements	1	
	Information System Security	1	
	Applied System Security	3	
	Crpytography	2	
	System file security	3	
	Development Management	5	
	Vulnerability Management	1	
	Technical Security	System Security	1
		Information Security	2
Security Test		1	
Access Management		1	
Total		77	

### 2.3 FedRAMP (Federal Risk and Authorization Management Program)

FedRAMP [7] is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services. The FedRAMP includes a subset of NIST Special Publication 800-53 Revision 3 security controls specifically selected to provide protection in cloud environments. And a subset has been defined for the FIPS 199 low categorization and the FIPS 199 moderate categorization. In 2012, the GSA (General Services Administration) released the “FedRAMP Baseline Security Controls Ver. 1.0” that federal agencies and cloud service providers must implement within a cloud computing environment to satisfy FedRAMP requirements. Moreover, the FedRAMP is in the

process of updating the FedRAMP security control baseline based on the recently released 800-53 Revision 4, Security and Privacy Controls for Federal Information Systems and Organizations.

## 2.4 Gartner Report

In 2008, the technology analyst firm Gartner [5] issued a report about seven cloud-computing security risks. According to Gartner, customers must demand transparency, avoiding vendors that refuse to provide detailed information on security program. Here are seven of the specific security issues Gartner says customers should raise with vendors before selecting a cloud vendor.

**Privileged User Access.** Get as much information as you can about the people who manage your data. Ask providers to supply specific information on the hiring and oversight of privileged administrators, and the controls over their access.

**Regulatory Compliance.** Customers are ultimately responsible for the security and integrity of their own data, even when it is held by a service provider. Traditional service providers are subjected to external audits and security certifications. Cloud computing providers who refuse to undergo this scrutiny are signaling that customers can only use them for the most trivial functions.

**Data Location.** When you use the cloud, you probably won't know exactly where your data is hosted. In fact, you might not even know what country it will be stored in. Ask providers if they will commit to storing and processing data in specific jurisdictions, and whether they will make a contractual commitment to obey local privacy requirements on behalf of their customers.

**Data Segregation.** Data in the cloud is typically in a shared environment alongside data from other customers. The cloud provider should evidence that encryption schemes were designed and tested by experienced specialists.

**Recovery.** Even if you don't know where your data is, a cloud provider should tell you what will happen to your data and service in case of a disaster. Any offering that does replicate the data and application infrastructure across multiple sites is vulnerable to a total failure. Ask your provider if it has "the ability to do a complete restoration, and how long it will take."

**Investigative Support.** Investigating inappropriate or illegal activity may be impossible in cloud computing. Cloud services are especially difficult to investigate, because logging and data for multiple customers may be co-located and may also be spread across an ever-changing set of hosts and data centers. If you cannot get a contractual commitment to support specific forms of investigation, along with evidence that the vendor has already successfully supported such activities, then your only safe assumption is that investigation and discovery requests will be impossible.



**Long-term Viability.** Ideally, your cloud computing provider will never go broke or get acquired and swallowed up by a larger company. But you must be sure your data will remain available even after such an event. "Ask potential providers how you would get your data back and if it would be in a format that you could import into a replacement application.

## 2.5 ENISA (European Network and Information Security Agency)'s Cloud Computing Risk Assessment Guide

ENISA [8] is an EU agency created to advance the functioning of the internal market. In 2009, the ENISA came out with the Cloud Computing Risk Assessment guide. This guide is an in-depth and independent analysis that outlines some of the information security benefits, key security risks of cloud computing and recommendations for information security. The ENISA classifies risks into four categories: policy and organizational, technical, legal, and not specific to the cloud. Also, it classified cloud-specific vulnerabilities and assets and information assurance requirements are given as follows:

1. Personnel security
2. Supply-chain assurance
3. Operational security
4. Identity and access management
5. Asset management
6. Data and Services Portability
7. Business Continuity Management
8. Physical security
9. Environmental controls
10. Legal requirements
11. Legal recommendations
12. Legal recommendations to the European Commission

## 2.6 CSA-CCM (Cloud Security Alliance Cloud Controls Matrix)

CSA-CCM [9] specifically proposed to provide fundamental security principles to guide cloud vendors and to support prospective cloud customers in assessing the overall security risk of a cloud provider. This guide provides a controls framework that gives comprehensive understanding of security concepts and standards that are aligned to the Cloud Security Alliance guidance in 14 domains. The foundations of the Cloud Security Alliance Controls Matrix rest on its customized relationship to other industry-accepted security standards, regulations, and controls frameworks such as the FedRAMP, ISO 27001/27002, ISACA COBIT, PCI, NIST, Jericho Forum and NERC CIP and will augment or provide internal control direction for service organization control reports attestations provided by cloud providers. The recently version

of the CCM is 1.4 and the CCM consists of the 11 domains which are given as follows:

1. Compliance(CO)
2. Data Governance(DG)
3. Facility Security(FS)
4. Human Resources(HR)
5. Information Security(IS)
6. Legal(LG)
7. Operations Management(OP)
8. Risk Management(RI)
9. Release Management(RM)
10. Resiliency(RS)
11. Security Architecture(SA)

## 2.7 CSC (Cloud Service Certification)

CSC [10] is cloud service certification program in Korea. In 2012, the Korea Cloud Service Association enforced the CSC. The CSC proposed to provide cloud service certification apropos of cloud service provider. The screening criterion of the CSC consists of seven controls; availability, expandability, performance, data management, security, service sustainability, and service support. Table 3 shows specification of the CSC.

**Table 3.** Screening criterion of the CSC

Control Area	Control Specification
Availability	The agency shall apply for various actions constantly in order to provide cloud computing services.
Expandability	Service provider for cloud computing should have policies to providing the expanded resources to meet the needs.
Performance	Service provider for cloud computing sholud maintain the performance to ensure the quality of service.
Data Management	Service provider for cloud computing should have policies securely to store user's data.
Security	Management system to effectively implement the organization's security should be established.
Service Sustainability	Service provider for cloud computing should have human-based and material-based resouces.that users can believe.
Service Support	Service provider for cloud computing sholud have support systems that users can be satisfied with.

### 3 Comparison and Analysis

#### 3.1 Comparison and Analysis of Cloud Computing Service Certification Standards

In this paper, we compare and analyze between Cloud Service Information Security Guide and other cloud computing service certification standards. Table 4 shows the result of the comparison.

**Table 4.** Comparison analysis of cloud computing service certification standards

Cloud Service Information Security Guide (Number of detailed controls)	A percentage of mapping				
	FedRAMP	Gartner	CSA-CCM	ENISA	CSC
Information security policy(5)	100.0	0.0	100.0	0.0	100.0
Security Organization's operation(13)	53.8	0.0	61.5	0.0	53.8
Personnel security(9)	100.0	0.0	100.0	100.0	100.0
Asset protection(8)	100.0	37.5	100.0	100.0	100.0
Asset classification(5)	100.0	0.0	100.0	100.0	100.0
Security issue report(7)	0.0	0.0	0.0	0.0	100.0
Incident response(9)	77.8	44.4	77.8	0.0	100.0
Service Sustainability(10)	70.0	10.0	70.0	80.0	100.0
Compliance(31)	22.6	3.2	22.6	22.6	0.0
Security control(21)	57.1	0.0	57.1	100.0	100.0
Equipment protection(17)	11.8	0.0	11.8	100.0	100.0
Operation management(12)	58.3	0.0	58.3	91.7	0.0
System management(10)	60.0	0.0	60.0	0.0	100.0
Malicious code response(4)	0.0	0.0	0.0	0.0	0.0
Control of mobile code(3)	33.3	0.0	100.0	0.0	0.0
Backup(3)	100.0	0.0	100.0	0.0	100.0
Network security management(4)	100.0	0.0	100.0	0.0	0.0
Audit(11)	54.5	0.0	54.5	0.0	0.0
Security requirements(3)	100.0	0.0	100.0	0.0	0.0
Information system security(5)	0.0	0.0	0.0	0.0	0.0
Applied system security(8)	25.0	0.0	25.0	0.0	0.0
Cryptography(6)	100.0	16.7	100.0	0.0	0.0
System file security(11)	54.5	0.0	36.4	0.0	63.6
Development management(16)	62.5	0.0	62.5	43.8	0.0
Vulnerability Management(4)	100.0	0.0	100.0	0.0	0.0
System security(3)	33.3	0.0	33.3	100.0	0.0
Information security(12)	75.0	0.0	75.0	100.0	0.0
Security test(5)	0.0	0.0	0.0	0.0	0.0
Access management(3)	100.0	66.7	100.0	66.7	66.7
Total(258)	53.5	5.4	55.0	43.4	47.3

### 3.2 Additional Security Controls

According to mapping result, the Cloud Service Information Security Guide contains almost all requirement and criteria of other certification standards about cloud computing service. Also, it is more detailed. On the basis of mapping result, we obtain additional security controls such as Table 5 which is not redundancy.

**Table 5.** Additional Security Controls

Control Area	Control Specification	ID	Standard
Session Rock	Service provider prevents further access to the system by initiating a session rock after defined time period of inactivity or upon receiving a request from a customer.	AC-11	FedRAMP
Wireless Access	Provider monitors for unauthorized wireless connections to the information system, including scanning for unauthorized wireless access points, and takes appropriate action if an unauthorized connection is discovered.	AC-18(2)	
Non-Repudiation	Cloud service provider must implement a secure digital signatures.	AU-10(5)	
Virtualization Techniques	Provider employs virtualization techniques to present information system components as other types of components, or components with differing configurations.	SC-30	
Acquisitions	Requires, if no U.S. Government Protection Profile exists for a specific technology type but a commercially provided information technology product relies on cryptographic functionality to enforce its security policy, then the cryptographic module is FIPS-validated.	SA-4(7)	
Data Location	Customer must know exactly where his data is hosted.	-	Gartner
Investigative Support	Provider must support the some investigation such as e-discovery and forensic.	-	
Long-term Viability	Even if cloud service is merged, provider must ensure that the availability of customer data is sustained.	-	
Management Program	The Information Security Management Program should be developed, documented, approved, and implemented.	IS-01	CSA-CCM
Industry Knowledge/Benchmarking	Industry security knowledge and benchmarking through networking, specialist security forums, and professional associations shall be maintained.	IS-12	

**Table 5.** (continued)

Workspace	<i>Policies and procedures shall be established for clearing visible documents containing sensitive data when a workspace is unattended and enforcement of workstation session logout for a period of inactivity.</i>	IS-17	
eCommerce Transactions	Electronic commerce (e-commerce) related data traversing public networks shall be appropriately classified and protected from fraudulent activity, unauthorized disclosure or modification in such a manner to prevent contract dispute and compromise of data.	IS-28	
Quality Testing	A program for the systematic monitoring and evaluation to ensure that standards of quality are being met shall be established for all software developed by the organization.	RM-03	
Data and Service Portability	Procedures, APIs and API interfaces for exporting data should be implemented in a standard format.	-	ENISA
Availability	The agency shall apply for various actions constantly in order to provide cloud computing services.	-	CSC
Expandability	Service provider for cloud computing should have policies to providing the expanded resources to meet the needs.	-	
Virtualization Security	Provider should establish measures to prevent vulnerabilities in virtualized environments.	-	
Performance	Service provider for cloud computing should maintain the performance to ensure the quality of service.	-	

## 4 Proposed Security Controls Based on K-ISMS

### 4.1 K-ISMS (Korea-Information Security Management System)

K-ISMS [11] is an information security management system which is introduced a domestic standard. The K-ISMS is based on the BS7799 and classifies certification

criteria into two categories; information security management process and information security measure. In 2013, the certification criteria have been revised to reflect the information security trend. A number of controls of revised certification criteria is 104 and has the security controls such as Table 6.

**Table 6.** Certification Criteria of K-ISMS

<b>Classification</b>	<b>Control Area</b>	<b>Number of controls</b>
Information Security Management Process	Information Security Policy Establishment and Fixing the Scope	2
	Senior Responsibility and Organization Construction	2
	Risk Management	3
	Information Security Measure Implementation	2
	Follow-up Management	3
	Subtotal	12
Information Security Measure	Information Security Policy	6
	Information Security Organization	4
	Outsider Security	3
	Information Asset Classification	3
	Information Security Training	4
	Personnel Security	5
	Physical Security	9
	System Development Security	10
	Cryptography Control	2
	Access Control	14
	Operation Security	22
	Incident Management	7
IT Disaster Recovery	3	
Subtotal	92	
Total		104

#### 4.2 Proposed Security Controls Based on K-ISMS

In the section 3, we obtain additional security controls such as Table 5 which is not redundancy. On the basis of the security controls, we propose a new security controls based on K-ISMS such as Table 7. Our proposed security controls classifies certification criteria into two parts; core criteria and additional criteria. The cloud service provider certified K-ISMS can minimize unnecessary certification assessment work by considering additional criteria.

**Table 7.** Proposed Security Controls based on K-ISMS

Core Criteria	1. Information Security Policy, 2. Information Security Organization 3. Outsider Security, 4. Information Asset Classification, 5. Information Security Training, 6. Personnel Security, 7. Physical Security, 8. System Development Security, 9. Cryptography Control		
	10. Access Control	Additional Controls	
		Wireless Access	Provider monitors for unauthorized wireless connections to the information system and takes appropriate action if an unauthorized connection is discovered.
		Non-Repudiation	Cloud service provider must implement a secure digital signatures.
	11. Operation Security	Additional Controls	
		Management Program	The Information Security Management Program should be developed, documented, approved, and implemented.
		Industry Knowledge/ Benchmarking	Industry security knowledge and benchmarking through networking, specialist security forums, and professional associations shall be maintained.
		Workspace	Policies and procedures shall be established for clearing visible documents containing sensitive data when a workspace is unattended and enforcement of workstation session logout for a period of inactivity.
		Acquisitions	Requires, if no U.S. Government Protection Profile exists for a specific technology type but a commercially provided information technology product relies on cryptographic functionality to enforce its security policy, then the cryptographic module is FIPS-validated.

**Table 7. (continued)**

		eCommerce Transactions	Electronic commerce (e-commerce) related data traversing public networks shall be appropriately classified and protected from fraudulent activity, unauthorized disclosure or modification in such a manner to prevent contract dispute and compromise of data.
	12. Incident Management	Additional Controls	
		Investigative Support	Even if cloud service is merged, provider must ensure that the availability of customer data is sustained.
		Data Location	Customer must know exactly where his data is hosted.
	13. IT Disaster Recovery		
Additional Criteria	14. Virtualization Security	Virtualization Techniques	Provider employs virtualization techniques to present information system components as other types of components, or components with differing configurations.
		Virtualization Security	Provider should establish measures to prevent vulnerabilities in virtualized environments.
	15. Quality Management	Expandability	Service provider for cloud computing should have policies to providing the expanded resources to meet the needs.
		Performance	Service provider for cloud computing should maintain the performance to ensure the quality of service.
		Availability	The agency shall apply for various actions constantly in order to provide cloud computing services.
		Quality Testing	A program for the systematic monitoring and evaluation to ensure that standards of quality are being met shall be established for all software developed by the organization.



## 5 Conclusion

Our main concern is to discuss the security controls used to protect cloud users. In this paper, we analyze security risks of cloud computing service, and investigate the security controls of many guidelines for protecting cloud computing service. Also, we compare Cloud Service Information Security Guide with the other cloud computing service security guidelines, and draw the additional security controls. Finally, we remove the repetitious security controls in many guidelines and then propose the security controls for cloud computing service which are based on K-ISMS and consist of core criteria and additional criteria. It will expect that cloud computing service providers use our security controls to establish their ISMS.

## References

1. Daryl, C.P., Thomas, J.B., Tom, A., David, W.C., David, M.S.: Cloud Computing: Defining and Describing an Emerging Phenomenon. Gartner (2008)
2. Peter, M., Timothy, G.: The NIST Definition of Cloud Computing. NIST SP 800-145 (2011)
3. Samsung Economic Research Institute – SERI (2012), <http://www.seri.org>
4. TechAmerica Foundation.: Cloud First, Cloud Fast: Recommendation for Innovation, Leadership and Job Creation, Cloud<sup>2</sup> Commission Report (2011)
5. Gartner.: Seven cloud-computing security risks (2008), <http://www.infoworld.com/d/security-central/gartner-seven-cloud-computing-security-risks-853?page=0,0>
6. Cloud Service Information Security Guide (2012), <http://www.kisa.or.kr>
7. The Federal Risk and Authorization Management Program, FedRAMP (2013), <http://www.fedramp.gov>
8. European Network and Information Security Agency, ENISA (2013), <http://www.enisa.europa.eu>
9. Cloud Security Alliance (CSA) Cloud Control Matrix, CCM (2013), <http://www.cloudsecurityalliance.org>
10. Cloud Service Certification, CSC (2012), <http://www.excellent-cloud.or.kr>
11. Korea Information Security Management System, K-ISMS (2013), <http://isms.kisa.or.kr>

# Implementation of Smart Grid Educational Application

Yeonghae Ko and Namje Park\*

Major in Elementary Computer Education, Department of Primary Education,  
Graduate School of Education, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
{smakor, namjepark}@jejunu.ac.kr

**Abstract.** This paper looks through several educational cases applying STEAM using SMART GRID. Then it is followed by the presentation of an educational application for making up the deficiency of examined cases. The educational application is able to calculate and analyze wattage by using the QR codes, which leads to easy solution and self-initiative work for elementary school students regarding their developing stage and interests. The application will be verified its effectiveness and implented augmented reality in the next stage. It is also expected to have beneficial effect on the educational field.

**Keywords:** STEAM, Smart Grid, QR Code, Elementary School.

## 1 Introduction

Recently, Korea experienced the start of the ubiquitous era in which we can use internet anytime anywhere as smart devices have been supplied rapidly. According to the analytic data published by Strategy Analytics in 2013, Korea's Smartphone supply rate, that means the Smartphone supply amount against population, is No. 1 in the world. In is confirmed that Korea's Smartphone supply rate of 67.6% is over four time of the world average supply rate of 14.8% and much higher than No. 2 Norway (55%), No. 3 Hong Kong (54.9%), No. 4. Singapore (53.1%), and No. 5 Australia (50.2%). However, at the same time, Korea has been experiencing the downsides of widespread use of the Smartphone. The Smartphone addiction rate in Korea of 11.1% means that more than 1 out of 10 people are addicted to the Smartphone. Especially, students' addiction to the Smartphone is higher than the average of the entire users, which has become a serious issue. This resulted from lack of basis of using smart devices for learning.

In this situation, Korea has been making efforts to build the infrastructure for the educational quality enhancement in line with the rapidly-changing information and telecommunications environment. It is necessary to establish the application of the newly-appearing mobile environment to the education, to create the content suitable for the rapidly-changing educational information and telecommunications situation, and provide the service.

---

\* Corresponding author.

The education program using smart devices provides the possibility of supplementing the weak points that can occur in the education field. In the STEAM program using Smart Grid, students worked on the task to measure the power consumption of electric home appliances but had a great difficulty in completing the task or failed to do it as the necessary task time was too long and the necessary information to complete the task was insufficient. To resolve this issue, this paper designed and implemented the educational application that can supplement the weak points of the Smart Grid education for elementary school students.

## **2 Theoretical Background**

### **2.1 Smart Grid and QR-Based STEAM Education**

Smart Grid means the next-generation intelligent power network that uses energy efficiently through the intelligent power transmission and distribution based on the grafting of IT technology to the existing power network. As Smart Grid is closely related to our lives and it is new technology with IT applied, it can attract students' interest and requires fused thinking for the issue resolution procedure, so it is suitable for STEAM education.

As a kind of two-dimensional code, the QR code is the bar code of a new concept that can be perceived more easily than the existing bar codes and can record a large amount of data. The QR code's patent is open so that anybody can use it and vast information can be stored, so it is emerging as the standard of two-dimensional codes. Android is one type of Smartphone OSs, and various Android applications are being developed as the Smartphone supply rate has been rapidly increasing recently. If Android is used for an educational purpose, users can download applications on their Smartphone without an additional device and use educational content easily. Especially, due to the free development, Android is used much for educational applications development.

### **2.2 The Relationship between Smart Grid STEAM Program and the Educational Application**

The STEAM education program using Smart Grid received students' interest and enabled students to learn about science continuously. However, they had a great difficulty with the task to measure/calculate/arrange power. First of all, to measure the power consumption during a certain period brings the difficulty of repeating measurements at a certain interval for longer than one day and that of finding the content about standby power or power consumption from electric products. The educational application is the tool that can supplement this weak point. Learners can create a virtual learning environment similar to the real life through an application, freely operate tools in it, and complete a task in the self-directed way.

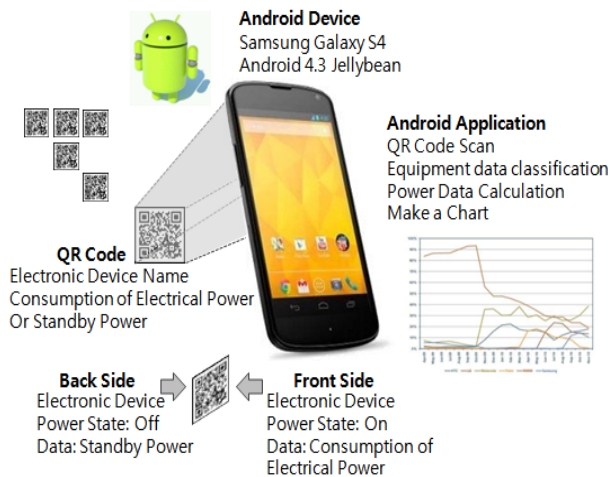
### 3 Educational Application Development Using Android and the QR Code

#### 3.1 Development Environment

In this paper, the educational application was developed on Microsoft Windows 7 OS through Java Eclipse Kepler and Android SDK 4.0.3 (API 15). The verification following the development was done on Android 4.3 Jellybean through Samsung Galaxy S4 and Galaxy Note 2.

#### 3.2 Application Configuration

On the application developed in this study, Android Smartphone fetches the QR Scanner to enter a value into according to the user's selection. The QR Scanner perceives the consistent pre-defined rules in the input images from Smartphone's camera and searches for the data in the code. Here, the code contains the information such as various electronic devices' names, the current power status, power consumption, or standby power.



**Fig. 1.** Concept of Proposed Education Application

As the user operates it, the application reads the data contained in the QR Code one by one and saves the values in the database. The user can complete the pattern of using electronic devices freely while repeating the inputting and operation of the QR code in turns. If all inputting is completed, the application analyzes the power consumption for each hour and each electronic device, and then displays the result in a graph on the screen. Besides, the application is configured such that it automatically analyzes the electricity fee for the power consumption, the fee for the actually used power, and fee for the electricity lost due to the standby power, so that the user can easily obtain the analyzed data without additional complicated calculation.

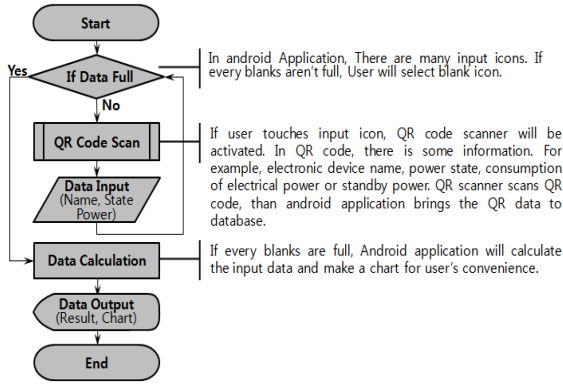


Fig. 2. Block diagram of Proposed Education Application

### 3.3 User Interface Configuration

The application developed in this study is to help students complete their tasks easily, so it is developed to have the simplest operation method and interface as possible. It is developed such that students can scan codes and enter data with simple touches.



Fig. 3. Interface of Proposed Education Application

## 4 Conclusion

It is effective to complete a task in the real life, but to complete a task in the real life accompanies many difficulties. In general, if the task completion time takes long, the task accuracy degrades and a student’s concentration on the task also degrades. Besides, if the necessary information to complete a task is insufficient, the case that a student fails to complete the task occurs more frequently. What can supplement this weak point is IT. The education program using IT devices and software can bring the

real life into the virtual education field and resolve the issue with task completion in the real life simultaneously. Especially, nowadays, due to the widespread use of smart devices, the merit of the smart devices-using education has been standing out more.

The application for the Smart Grid education proposed in this study supplements the weak points of the general Smart Grid education and can increase the education effectiveness. It is possible to reduce the task load by measuring the daily power consumption status in a short time through the application, and the self-directive learning and research are possible as the learner can adjust the electronic devices' status actively. Besides, the application has the following merits: the possibility of checking immediately on the application the electronic device-specific standby power or power consumption that is difficult to check in the real life, the possibility of comparing power consumptions and checking the power consumption data in the visible graph.

Like this, the application for the Smart Grid education that uses the QR code developed in this study has the merit that elementary school students can use it easily. However, it is necessary to do more investigation for the analysis of the effectiveness of using the developed application in the actual learning situation. Besides, the application proposed in this study has limits that it is difficult to check the information about the electronic devices that measure power consumption. To supplement this weak point, the study plans to supplement the application by seasoning the augmented reality in future such that it is possible to visibly display the type and current status of the electronic devices to measure when the QR code is recognized. Second, the application of the developed STEAM education program to the education field resulted in the great change in students' attitude towards engineering and technology. Especially, the change in the career perception about engineers showed a meaningful difference before and after the test, which confirmed that the Rub Goldberg's Inventions creation STEAM education had a positive effect on the selection of the engineering/technology-related jobs. It was also noticed that the students' satisfaction with the program and their interest in it were high and that the frequency of the characteristics of Investigative (I) related to engineers in Holland's career interest type inspection increased.

The STEAM actively promoted by the Ministry of Education, Science and Technology has not been extended deeply into the school fields, yet. Besides, we cannot say that teacher's understanding of and will about STEAM education are active. Currently, many studies about program development are in progress and the efforts to expand the impact of the program application to schools have been made. Thus teachers also should make efforts to realize the true STEAM through the study and efforts of STEAM education.

## References

1. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)

2. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
3. Park, N.: Secure UHF/HF Dual-band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
4. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. *International Journal of Ad Hoc and Ubiquitous Computing* 8(4), 205–219 (2011)
5. Park, N., Kim, Y.: Harmful Adult Multimedia Contents Filtering Method in Mobile RFID Service Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS (LNAI), vol. 6422, pp. 193–202. Springer, Heidelberg (2010)
6. Park, N., Song, Y.: AONT Encryption Based Application Data Management in Mobile RFID Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS (LNAI), vol. 6422, pp. 142–152. Springer, Heidelberg (2010)
7. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
8. Park, N.: Secure Data Access Control Scheme Using Type-Based Re-encryption in Cloud Environment. In: Katarzyniak, R., Chiu, T.-F., Hong, C.-F., Nguyen, N.T. (eds.) *Semantic Methods. SCI*, vol. 381, pp. 319–327. Springer, Heidelberg (2011)
9. Park, N., Song, Y.: Secure RFID Application Data Management Using All-Or-Nothing Transform Encryption. In: Pandurangan, G., Anil Kumar, V.S., Ming, G., Liu, Y., Li, Y. (eds.) WASA 2010. LNCS, vol. 6221, pp. 245–252. Springer, Heidelberg (2010)
10. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. *The Journal of Korea Information and Communications Society* 35(5), 785–793 (2010)
11. Kim, Y., Park, N.: Development and Application of STEAM Teaching Model Based on the Rube Goldberg’s Invention. In: Yeo, S.-S., Pan, Y., Lee, Y.S., Chang, H.B. (eds.) *Computer Science and its Applications. LNEE*, vol. 203, pp. 693–698. Springer, Heidelberg (2012)
12. Park, N., Cho, S., Kim, B., Lee, B., Won, D.: Security Enhancement of User Authentication Scheme Using IVEF in Vessel Traffic Service System. In: Yeo, S.-S., Pan, Y., Lee, Y.S., Chang, H.B. (eds.) *Computer Science and its Applications. LNEE*, vol. 203, pp. 699–705. Springer, Heidelberg (2012)
13. Kim, K., Kim, B.-D., Lee, B., Park, N.: Design and Implementation of IVEF Protocol Using Wireless Communication on Android Mobile Platform. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H. (eds.) *SecTech/CA/CES<sup>3</sup> 2012*, vol. 339, pp. 94–100. Springer, Heidelberg (2012)
14. Ko, Y., An, J., Park, N.: Development of Computer, Math, Art Convergence Education Lesson Plans Based on Smart Grid Technology. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H. (eds.) *SecTech/CA/CES<sup>3</sup> 2012*. CCIS, vol. 339, pp. 109–114. Springer, Heidelberg (2012)
15. Kim, Y., Park, N.: The Effect of STEAM Education on Elementary School Student’s Creativity Improvement. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H. (eds.) *SecTech/CA/CES<sup>3</sup> 2012*. CCIS, vol. 339, pp. 115–121. Springer, Heidelberg (2012)

# Development of Open Service Interface's Instructional Design Model in USN Middleware Platform Environment\*

Jeongyeun Kim<sup>1</sup>, Dong-Hwan Park<sup>2</sup>, Hyo-Chan Bang<sup>2</sup>, and Namjpe Park<sup>1,\*\*</sup>

<sup>1</sup> Major in Computer Education, Faculty of Science Education,  
Graduate School, Jeju National University,

61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
namjpepark@jejunu.ac.kr, inarasam@naver.com

<sup>2</sup> Electronics and Telecommunications Research Institute (ETRI),  
218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700, Korea  
{dhpark, bangs}@etri.re.kr

**Abstract.** This paper contains interface definition provided by the USN service middleware platform, which intends to eliminate the sensor network dependency of USN services and to provide common functions in gathering and processing sensor data continuously. First, we categorize the functions and application interface types provided by platform. And we define interfaces for user authentication and authorization, service session management, message and notification processing, sensor network monitoring and control, and sensor data query processing.

**Keywords:** Open USN Service, Semantic, Platform Model, USN, WSN.

## 1 Introduction

USN Service Middleware Platform is linked to the multiple heterogeneous sensor networks by the sensor network common interface standard as shown in the figure. The platform abstracts sensor networks to cope with the heterogeneity and commonality of the heterogeneous sensor network. The platform monitors the state of sensor network or provides the control commands via the abstraction layer of sensor network. Especially it acquires the sensor data periodically and continuously by the sensor data collection command to the sensor network or transmits the sensor data from the sensor network to the platform periodically and continuously. In order to carrying out such sensor network abstraction function, the key information of sensor network is displayed, managed and used according to the USN metadata model standard. Especially for the external application service, it provides the function to generate, retrieve and manage the metadata by the USN metadata directory service interface standard.

---

\* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2013S1A5A8026151).

\*\* Corresponding author.



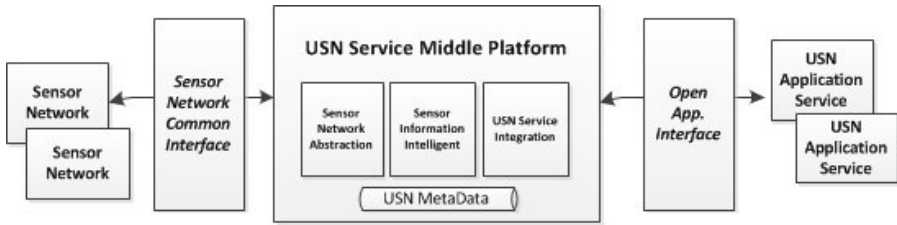


Fig. 1. USN Service Middleware Platform Architecture

This paper contains interface definition provided by the USN service middleware platform, which intends to eliminate the sensor network dependency of USN services and to provide common functions in gathering and processing sensor data continuously. First, we categorize the functions and application interface types provided by platform. And we define interfaces for user authentication and authorization, service session management, message and notification processing, sensor network monitoring and control, and sensor data query processing

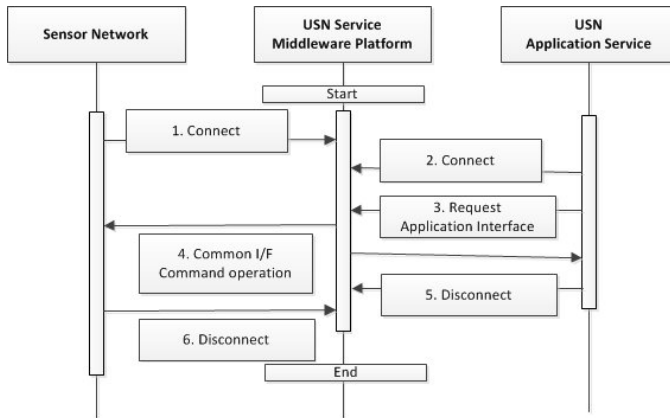
## 2 Service Flow of USN Service Middleware Platform

### 2.1 Flow of Platform Operation

The flow of platform operation is shown in the figure 2 above. The platform enters the operational state by the initial actuation and then starts being linked to the sensor networks and the application services. Multiple sensor networks are linked by the connection requests in the operational state of platform and then the application services are also linked to the platform in its operational state. The platform linked with the sensor networks processes the commands continuously according to the common interface standard on the service requests of application services and the application service performs its actions by the application interface with the platform. The connections of applications and sensor networks can happen. Finally the platform stops its function as a platform in the finish stage. (Figure 2.)

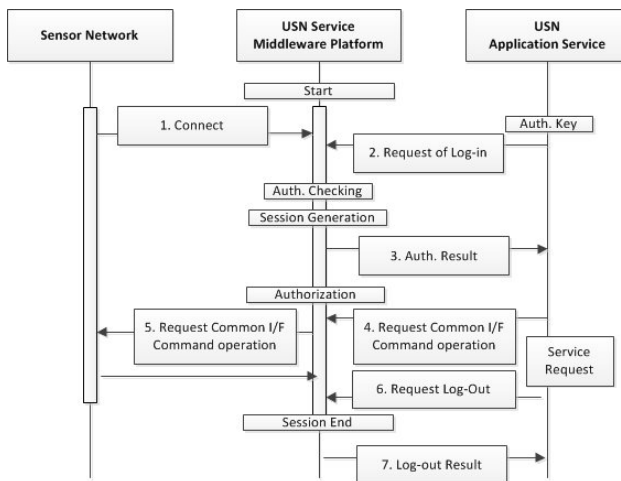
### 2.2 Flow of User Authentication/Authorization

The authentication during the request of application service and the flow of authentication by the service request are shown in the figure above. After the platform is activated and linked with the sensor networks, the application service should request authentication first to get the platform service. The authentication is made when the application service sends the generated credential to the platform and the platform interprets the credential and checks if the application service is a registered user or not. The credential should be treated confidentially and can be generated selectively in one of th methods of ID/Password based authentication and Certificate based authentication. The platform generates application sessions when the authentication is confirmed and sends the Session ID, the result of authentication, to the application service.



**Fig. 2.** Platform Operating Flow

After the process of authentication, the application service sends requests using the service interface provided by the platform. These requests make the platform carry out the authorization process that decides whether the services or resources provided by the platform can be used, and the results performed in the sensor networks are provided to the application services. The basic services provided by the platform are sensor network control, sensor network monitoring, sensor data query process, etc. and there might be the other types of expanded services. The resources provided by the platform mean the sensor networks registered on and linked to the platform. The platform should maintain the information on permission of services and resources to each user for the purpose of authorization process. The application service is disconnected by sending log-out request to the platform. (Figure 3)



**Fig. 3.** Authorization process of Application Service

### 2.3 Flow of Service Session Management

USN Service Middleware Platform makes groups of functions provided to the applications services and defines the groups as the platform service. In other words, application interfaces provided by abstracting the sensor network functions, multiple number of application interfaces provided for monitoring of sensor networks and query process interfaces for acquisition and analysis of sensor data should be used together with the related interface in each interface group. The processed results according to each service request can be transmitted to the application services in diverse ways depending on the characteristics of each service.

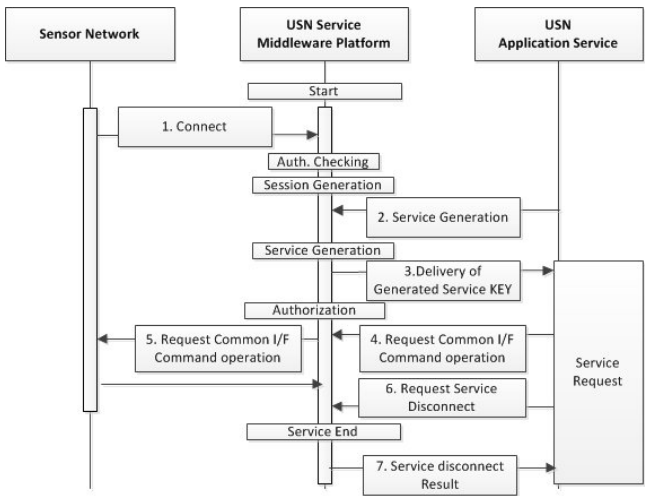


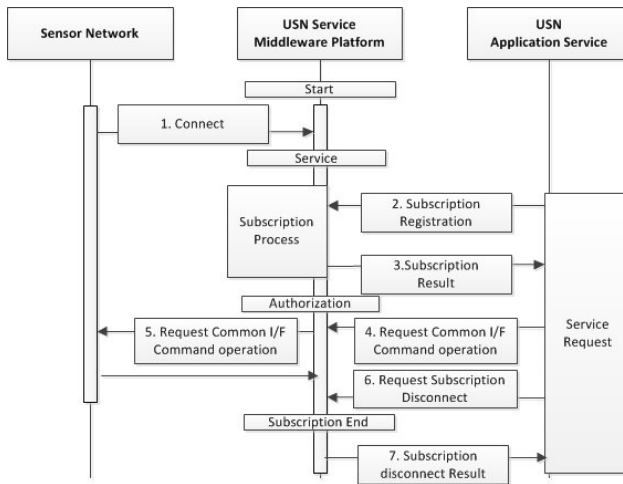
Fig. 4. Flow of Platform Service Session Management

The process flow of service sessions that are generated and removed in the platform for the application services are shown in the figure above. First, the sensor networks are connected to the platform. The platform generate application sessions after authentication for the application service, and the application service performs requests for generation of service session (S2, S3) prior to the request for platform service. The application service can generate multiple number of service sessions for the services provided by the platform. The application service make requests (S4) via the service application interface within the generated service session. Finally, the application service should request the removal of service sessions prior to the requests for the removal of application sessions and the platform should remove the remaining sessions for the service on getting the requests of removal from the application service.

### 2.4 Flow of Message and Notification

The service results for the service requests of application are defined as Message and Notification. The message is not the return value of processed results that is

instantly transmitted in response to the service requests but the results type that is generated to be sent to the application service after taking process continuously on the requests. In other words, it is not the results transmitted synchronously to the requests of application service but the results type generated by the process in the platform and transmitted asynchronously. The notification is the result type that is pushed to the service when events occur to the platform or state changes are detected. The notification are different from the message in the point that notification type and notification contents (reasons) are included and transmitted.



**Fig. 5.** Flow of Message and Alert Process

The transmission flow of message and notification to the application service is shown in the figure above. First, the sensor network is linked (S1) to the platform. The application can register the subscription URI by requesting subscription of message or notification to each service session. Once the subscription URI is registered, the platform transmits messages and notifications continuously whenever they are generated as the processed results for the service requested by the application. The application service can change and remove the subscription URI. Once the application service removes the subscription URI, the platform does not send the processed results to the subscription URI and discards them.

## 2.5 Flow of Sensor Network Monitoring and Control

The flows of sensor network monitoring and control are shown in the figure above. The sensor networks linked to the platform transmit the monitoring information periodically via the sensor network common interface. The platform updates the monitoring information of the multiple sensor networks linked to the platform dynamically and maintains them to be used in the related application or platform. USN application service gets the monitoring state information from the platform in

notification forms. When there are changes of the monitoring state information, the platform sends notification to the notification subscription URI of the registered service if necessary. The application service uses application interface to transmit the control requests like change of sensor network monitoring cycle, transmission method, On/Off of network elements, etc. and the platform gives control commands the sensor networks and transmit the results to the service.

## 2.6 Query Process of Sensor Data and Actuator

The sensor data query is a request that the application service sends to get the sensor data acquired from the sensor networks in the forms that each application desires. The flow of sensor data query process is shown in the figure above. First the sensor network is linked to the platform. USN application service requests query process in the query statement expressed based on the query type and query syntax that USN Service Middleware Platform supports. The platform provides the identifier on the query process request to the service and then transmits the results of query process for the sensor data acquired from the sensor network to the service. The application service controls the query by sending control requests like suspend, resume and stop for the consecutively performed query such as continuity query or stream query. USN Service Middleware Platform should support the following types of sensor data query.

## 3 Conclusion

Enabling various USN applications through the standard application interface to get access to and use the sensor networks constructed in a variety of service domains and service Intellectualization elements, the standard removes the dependency of the USN application on specific sensor networks. In addition, it is expected to enable the link and integration between various USN applications in diverse service domains. This, in turn, will entail more active development of USN applications, greatly enhance the efficiency of their use, and eventually lead to the notable growth of the USN industry.

## References

1. Lee, J.: Open APIs of USN Service Middleware Platform. Telecommunications Technology Association(TTA) Standard, TTA.KO-06.0284 (2012)
2. Park, D.-H., Kim, S.-J., Bang, H.-C.: Semantic Open USN Service Platform Architecture. In: ICTC 2012, pp. 12–15 (2012)
3. Kim, M., Lee, J.W., Lee, Y.J., Ryou, J.-C.: COSMOS: A Middleware for Integrated Data Processing over Heterogeneous Sensor Networks. *ETRI Journal* 30(5), 696–706 (2008)
4. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) *APWeb Workshops 2006*. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
5. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. *International Journal of Ad Hoc and Ubiquitous Computing* 8(4), 205–219 (2011)

# The Analysis of Case Result and Satisfaction of Digital Textbooks for Elementary School Students

Jeongyeun Kim and Namje Park\*

Major in Computer Education, Faculty of Science Education,  
Graduate School, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
namjepark@jejunu.ac.kr, inarasam@naver.com

**Abstract.** This paper examines the effectiveness of using digital textbooks in an elementary school on Jeju Island that has been using digital textbooks for 5 years, and analyzes the result of the survey about the satisfaction of the students and the teachers. The effectiveness of using digital textbooks in other cities is similar to this study result. To review the effectiveness of digital textbooks, the paper studied the self-directed learning ability and the issue resolution ability and then made a comparison before and after the use of digital textbooks. It is desired that the government will reflect the study result in configuring the digital textbook platform and that teachers and students will acquire some ideas while using digital textbooks so that the trial and errors in using digital textbooks completely can be minimized.

**Keywords:** Digital Textbook, Elementary School, Textbook, Smart Education.

## 1 Introduction

As the ubiquitous environment has been introduced through the development of the state-of-the-art information and communications technology and the information and telecommunications network, the society has been changing rapidly and is flooded with the tremendous amount of new technical knowledge and technology daily. Following this stream of times, companies started looking for people with different talent; the men of talent focusing on 3R capacities to the men of talent with creativity, issue resolution ability, communication ability, and cooperation ability.

In the educational field, the development of an information-based society has brought a groundbreaking change to the uniform and standardized educational method of the traditional industrial society. Following the grafting of IT technology to education, the change to education adopting digital technology (such as the appearance of cyber learning and digitized textbooks) was requested. In the highly informatized society with the high-speed telecommunications network developed, there is a request for the innovation of the education environment and the education paradigm change to the intelligent customized teaching/learning system so that a new

---

\* Corresponding author.

knowledge circulation structure through the cooperation for the enhancement of the 21 century learner's capacity can be revealed.

As a response, the government announced the Smart Education Policy Promotion Strategy in 2011. Smart Education was introduced as the intelligent customized education that is more than simple utilization of ICT devices and accompanying the change in education content, educational methods and evaluation, education environment, etc. As the individually customized learning system based on the characteristics of the next-generation learners who will be familiar with the digital environment from birth, Smart Education, which emphasizes mutual communication and feeling the same through the enhancement of the interactions between students, between teachers and students, and between teachers, is regarded as the driving force of the educational system innovation.

The core of Smart Education converges to digital textbooks and the clouding service. The government already started a digital textbook-based research school focusing on the 6th grade mathematics in 2006 and expanded the study range nationwide. In 2014, the government plans to introduce digital textbooks for the junior high school classes such as society, science, and English; and perform a test operation for elementary schools. In 2015, the government expects to supply digital textbooks completely in 2015. As the futuristic textbooks that have or are related to the functions including textbooks, references, question banks, dictionaries, notebooks, multimedia element data, digital textbooks are expected to be developed through open sources so that they can be used on all terminals such as PC, Smartphone, Smartpad, etc. These digital textbooks are expected to play the following roles: elevating the next-generation's learning efficiency, reducing the educational gap caused by different locations and different incomes, and reducing the dependency on private education through the normalization of public education.

In 2014, Korea will introduce digital textbooks for junior high school subjects such as society, science, and English while starting a test operation for elementary schools. The key word of what the Korean government's digital education policy intends to promote is self-directed learning. However, for this digital textbook policy to bear the intended fruit, it is necessary to share the result of using digital textbooks with teachers and students.

This paper examines the effectiveness of using digital textbooks in an elementary school on Jeju Island that has been using digital textbooks for 5 years, and analyzes the result of the survey about the satisfaction of the students and the teachers. The effectiveness of using digital textbooks in other cities is similar to this study result. To review the effectiveness of digital textbooks, the paper studied the self-directed learning ability and the issue resolution ability and then made a comparison before and after the use of digital textbooks.

It is desired that the government will reflect the study result in configuring the digital textbook platform and that teachers and students will acquire some ideas while using digital textbooks so that the trial and errors in using digital textbooks completely can be minimized.

## **2 Digital Textbook Operation Study Result**

### **2.1 Students' Satisfaction about Teaching/Learning Activities**

To check the satisfaction about teaching/learning activities from the teachers who used digital textbooks, the teachers in charge of classes (the 4th grade ~ the 6th grade) were interviewed, and paper textbooks and digital textbooks were compared during the interviews, regarding 10 items (including variety, convenience, students' concentration degree). The result was that the teachers in charge of the 4th grade classes expressed their opinions that the digital textbooks provided more useful and a vast amount of information than paper textbooks from the following two viewpoints: material variety for supporting teaching and learning, the interaction among students. On the contrary, the teachers in charge of the 5th and 6th grade classes expressed their opinions that digital textbooks were helpful in provoking the students' learning motivation and interest, but that not much difference in content and materials between two types of textbooks was found. Besides, all the teachers in charge of the classes from the 4th grade to the 6th grade were negative towards the question of whether digital textbooks were more convenient to teach with than paper textbooks. Especially, they pointed out that the delay in supplying the new digital textbooks due to the revised education courses resulted in the lower utilization of them to the class and that the content were insufficient.

### **2.2 An Analysis and Review of the Effectiveness of Using Digital Textbooks**

#### **1) Self-directed Learning Ability**

The result of comparing/analyzing students' self-directed learning ability before and after the use of digital textbooks was as follows for the dependent variables (as self-efficacy, meta-cognition, information search and task resolution, internal motivation, and self-reflection): For the 4th grade students, changes to self-efficacy was meaningful with the figure of  $p < .05$  while for the 5th and the 6th grade students, all dependent variables showed no statistically meaningful difference.

#### **2) Issue Resolution Ability**

Regarding the dependent variables of issue resolution ability (issue clarification, cause analysis, alternative development, plan execution, performance evaluation), the result of comparing/analyze students' issue resolution ability before and after the use of digital textbooks shows that both the 4th grade students and the 6th grade students showed no statistically meaningful difference in all dependent variables. However, the 5th grade students showed some meaningful difference in issue clarification ( $p < .05$ ) and cause analysis and an alternative development ( $p < .01$ ).

#### **3) Health Issue Analysis**

The VDT syndrome/internet addiction/sight tests were performed twice (before and after the use of digital textbook) for the students from the 4th grade to the 6th grade with the purpose of investigating the impact of digital textbook utilization to the health issue.



### **A. VDT Syndrome**

Changes to dependent variables of VDT syndrome (eye symptoms, psychological symptoms, overall body symptoms, musculoskeletal symptoms) were examined before and after the use of digital textbooks. The students took a self-diagnosis test, and t-test was done for corresponding samples. The results showed that the dependent variables did not show statistically significant differences before and after the use of digital textbooks. The results were the same for the 4th grade students in their first year of using digital textbooks, and 5th grade students and 6th grade students in their second year.

### **B. Internet Addiction Inspection Result**

Both pre- and post-test were conducted using a self-diagnosis questionnaire for internet addiction provided by the National Information Society Agency of Korea. The comparison of the results showed that, in the pre-test, 86.7% were classified as normal users, 8.0% fell under a potentially risky group, and 5.3% under a highly risky group. In the post-test, 88.0% were classified as normal users, 6.7% fell under a highly risky group, and 5.3% under a potentially risky group. That is, the share of highly risky group rose by 1.4%p, while that of a potentially risky group fell by 2.7%p, and that of normal users raised 1.3%p. It suggested that the use of digital textbooks did not have direct impact on inducing the Internet addiction. However, among the 4th grade students who were in the first year of using digital textbooks, 26.0% were classified as either a highly risky or a potentially risky group..

### **C. Eyesight Examination Result**

To inspect the students' eyesight, pre- and post-test were conducted in April 2010 and October 2011 for the students from the 4th grade to 6th grade. The analysis results showed that the left and the right eyesight each improved by 0.08 and 0.01 for the 4th grade students, declined by 0.03 and 0.05 for the 5th grade students, and for the 6th grade students, the left eyesight deteriorated by 0.09 and the right eyesight improved by 0.02. Overall, the eyesight fell by 0.01 on average.

## **2.3 Digital Textbook Content Analysis Result**

The analysis of digital textbook content for the students from the 4th grade to 6th grade showed the following results: First, errors included word spacing, word layout, resolution of words and figures, paragraph layout, fixed size of materials, and failure to play video clips. Second, digital textbooks for the 4th grade students contained rich multimedia materials, but they were hard to install due to voluminous sizes. Third, digital textbooks for higher grades had much less multimedia materials that they did not significantly vary from paper textbooks.

## **2.4 Digital Textbook Utilization Daily Log Creation Analysis**

Teachers' teaching logs and the students' learning logs were analyzed and compared to draw opinions on the suitability of digital textbooks as educational material.

### **1) The Analysis of Teachers' Daily Log Result**

Teachers' daily log contained the following opinions. First, digital textbook content for the 4th grade students were useful for the class and for stimulating students' interest as they had materials and content that enhanced teaching. However, multimedia materials need to be redesigned in accordance with the teaching and learning plans. Second, digital textbook content for the 5th grade students and the 6th grade students hardly contained multimedia materials that are applicable in the class that the teachers and students had to seek relevant materials on their own, which is inconvenient. Third, it takes substantial time to connect to cyber home tutoring, and it is impossible to selectively choose necessary parts such as additional materials or in-depth materials. Fourth, in case of science class, digital textbooks provided answers in advance, before students had a chance to make observations or participate in experiments that its use might dampen the students' investigative ability or power of thinking. Fifth, the screen size is too small and the resolution of figures is too low that it is hard to concentrate. Sixth, it would be desirable if a greater variety of materials and resources can be offered in addition to those already on paper textbooks.

### **2) The Analysis of Students' Daily Log Result**

Students' daily log contained the following opinions. First, digital textbook content for the 4th grade students contained a wealth of information and learning materials, which enable in-depth study, supplementary self-learning, and they are more stimulating and enticing compared to textbook materials. Second, for the 5th grade students and 6th grade students, digital textbook content hardly contained video clips or multimedia materials that they were hardly different from paper textbooks. Third, some pointed out that sometimes it was hard to log on, and tablet PCs occasionally malfunctioned. Fourth, some criticized that digital textbook content for the 5th grade students and 6th grade students did not have sufficient materials compared to those used in 2010.

## **2.5 Digital Textbook Utilization Checklist Analysis Result**

The teachers wrote up digital textbook utilization checklist at the end of each month, and the checklists were analyzed in terms of infrastructure, class operation, ratio of utilizing digital textbooks in class, and overall management of the schools in research. In all aspects, the utilization of digital textbooks declined between the first year and the second year. This is mainly because provisioning of content was delayed due to revised curricula for the 5th grade students and 6th grade students, so that digital textbook was used for a short period of time. Another reason is that both quality and organization of digital content varied little from paper textbooks. Also, the old personal tablet PCs often malfunctioned.

## **3 Conclusion**

Next-generation students live and grow up in a digital culture from a very early age, surrounded by smart devices. They have an excellent ability to adapt to new

information and communications technologies, unlike the older generation, sensitive to digital learning materials in the areas of their interest, and they have outstanding learning abilities. The ubiquitous learning environment is already firmly in place to enable learning experiences unrestricted by time and space, using information devices. In 2014, digital textbooks will be introduced for junior high school classes including society, science and English, and will be tested in grammar schools, which raises widespread concerns. How do the present-day students who are familiar with digital environment perceive paper textbooks? Many studies on the utilization of digital textbooks have addressed the issue. Teachers can add interesting teaching materials or customize their styles, which can enhance competitiveness of the future, in addition to default materials. This would bring quality of digital textbooks to a level field with other smart educational devices. When the key of smart education is customized self-directed learning, optimizing the use of digital textbooks based on the analysis of their effectiveness and the factors that influence students' social skills would contribute to efficient learning experiences.

## References

1. Kosan Elementary School: Simulation and Application for Vacational of Digital textbooks (2010)
2. Handong Elementary School: Web-based Digital textbooks that help to cultivate the ability of self-directed Learning (2012)
3. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
4. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. Communications in Computer and Information Science, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
5. Park, N.: Secure UHF/HF Dual-Band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
6. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. *International Journal of Ad Hoc and Ubiquitous Computing* 8(4), 205–219 (2011)
7. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
8. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. *The Journal of Korea Information and Communications Society* 35(5), 785–793 (2010)

# Cryptanalysis of Encrypted Remote User Authentication Scheme by Using Smart Card\*

Jongho Mun<sup>1</sup>, Jiye Kim<sup>1</sup>, Woongryul Jeon<sup>1</sup>, Youngsook Lee<sup>2</sup>, and Dongho Won<sup>1,\*\*</sup>

<sup>1</sup> College of Information and Communication Engineering,  
Sungkyunkwan University,

300 Cheoncheon-dong, Jangan-gu, Suwon-si, Gyeonggi-do, 440-746, Korea  
{jhmoon, jykim, wrjeon, dhwon}@security.re.kr

<sup>2</sup> Department of Cyber Investigation Police, Howon University,  
727 Weolha-li, Impi-Myeon, Gunsan-si, Jeonrabuk-do, 573-718, Korea  
ysooklee@howan.ac.kr

**Abstract.** Remote user authentication scheme is one of the most convenient authentication schemes to deal with secret data over insecure channels. In 2012, Yassin et al. proposed encrypted remote user authentication scheme by using smart card. They claimed that their scheme is secure against various attacks. In this paper, we demonstrate that their scheme is insecure and vulnerable to outsider attack, smart card stolen attack, offline password guessing attack, and masquerade attack.

**Keywords:** Smart card, Remote user authentication, Security.

## 1 Introduction

Smart card-based authentication schemes are becoming day by day more popular. In the view of fact that several remote user authentication schemes using smart card [1][2][3][4][5][6][7] have been proposed. In 2004, Das et al. [8] proposed a dynamic identity based remote user authentication scheme using smart cards. However, their scheme is vulnerable to various attacks. In 2009, Wang et al. [9] presented a more secure dynamic ID-based remote user authentication scheme and demonstrated the weakness of Das et al.'s scheme such as impersonate attack and lack mutual authentication. However, Wang et al.'s scheme suffers from malicious attacks and has some feasible security risks.

Recently, Yassin et al. [10] demonstrated that Wang et al.'s scheme is vulnerable to password guessing attack, DOS attack and server impersonate attack and proposed an enhancement of Wang et al.'s scheme. However, in this paper, we find that Yassin et al.'s scheme is vulnerable to outsider attack and smart card stolen attack.

---

\* This research was funded by the MSIP(Ministry of Science, ICT&Future Planning), Korea in the ICT R&D Program 2013.

\*\* Corresponding author.

The rest of the paper is organized as follows: Section 2 briefly reviews Yassin et al.’s authentication scheme. Section 3 describes the weaknesses of Yassin et al.’s scheme. Finally, we conclude this paper in Section 4.

## 2 Review in Yassin et al.’s Scheme

This section reviews an encrypted remote user authentication scheme by using smart card proposed by Yassin et al. [10]. Yassin et al.’s scheme consists of four phase; registration phase, login phase, authentication phase and password change phase. The notations used in this scheme are summarized as Table 1.

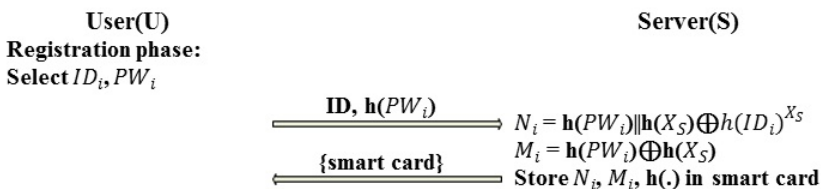
**Table 1.** Notations used in this paper

Notation	Description
U	A user
S	A remote server
ID, PW, SC	U’s identity, password, and smart card
$\oplus$	The bitwise XOR operation
$\parallel$	String concatenation
$X_S$	S’s secret key, which is kept secret and only known by S
$h(\cdot)$	A collision resistant one-way hash function

### 2.1 Registration Phase

In this phase, the user  $U_i$  initially registers with the remote server S as follows:

1.  $U_i \Rightarrow S : \{ID_i, h(PW_i)\}$ . User  $U_i$  sends his selected identity  $ID_i$  and hashed password  $PW_i$  to the remote server S over a secure channel.
2. S computes  $N_i = h(PW_i) \parallel h(X_S) \oplus h(ID_i)^{X_S}$ ,  $M_i = h(PW_i) \oplus h(X_S)$ , where  $X_S$  is a secret key kept by S in private.
3.  $S \Rightarrow U_i : \{SC\}$ . S stores the secure information  $\{h(\cdot), N_i, M_i\}$  into a new smart card SC and sends a smart card to user  $U_i$  over a secure channel.



**Fig. 1.** Registration phase

### 2.2 Login Phase

When a user  $U_i$  wants to login S,  $U_i$  inserts his smart card into the card reader and inputs his password  $PW_i$ . The smart card fulfills the following steps:

1. Compute  $h(X_S) = M_i \oplus h(PW_i)$  and  $Z' = h(PW_i) \parallel h(X_S)$ .
2. Generate a random number  $r_i$ . Compute  $K_i = h(r_i \oplus Z')$ ,  $C_i = K_i \oplus (Z' \oplus N_i)^{r_i}$ , and  $f_i = h(ID_i)^{r_i}$ .
3. Calculate  $CID_i = Z' \oplus h(T \oplus r_i)$ , there T is the current time stamp of the input device.
4. Encrypt  $(r_i, T, N_i, CID_i)$  by using  $K_i$ .
5. SC  $\rightarrow$  S :  $\{C_i, f_i, E_{K_i}(r_i, T, N_i, CID_i)\}$ . Smart card sends login request message M to the remote server.

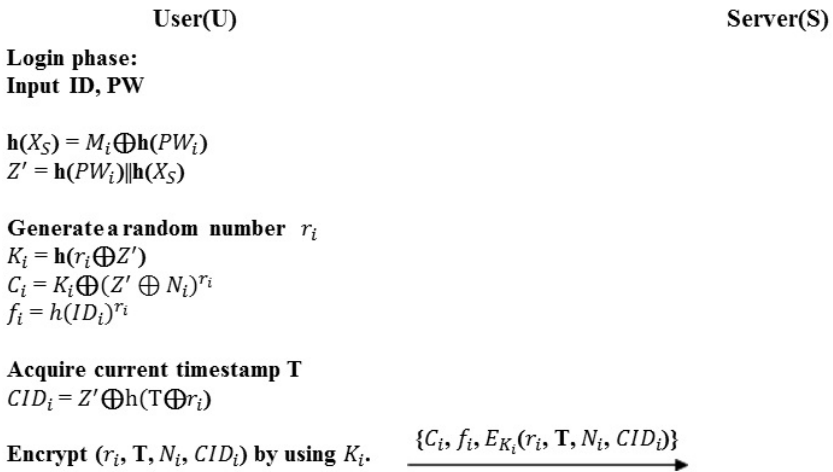


Fig. 2. Login phase

### 2.3 Authentication Phase

After receiving login request message at time  $T'$ , S performs the following computations:

1. S computes  $K_i = C_i \oplus f_i^{X_S}$ , and decrypts  $E_{K_i}(r_i, T, N_i, CID_i)$ .
2. S checks the freshness of time stamp T. If  $T' - T \leq \Delta T$  contains, S persists the next step. Otherwise, S rejects the session.
3. S compute  $Z'' = CID_i \oplus h(T \oplus r_i) = h(PW_i) \parallel h(X_S)$ , checks whether  $(N_i \oplus Z'')^{r_i}$  is equal to  $f_i^{X_S}$ . If so, S accepts the user  $U_i$ 's login request.
4. S  $\rightarrow U_i$  :  $\{E_{K_i}(a', T')\}$  S computes  $a' = h(Z'' \parallel r_i \parallel T')$  and sends message  $M' = E_{K_i}(a', T')$  to  $U_i$ .

- When  $U_i$  receives the message  $M' = E_{K_i}(a', T')$  at time  $T''$ ,  $U_i$  checks whether  $T'' - T \leq \Delta T$ . If not hold,  $U_i$  overthrows the message  $M'$  and terminate this phase. Otherwise  $U_i$  decrypts message  $M'$  by using  $K_i$ , computes  $a = h(Z' || r_i || T')$ , and compares  $a$  with  $a'$ . If so,  $U_i$  decides that the remote server S is authenticated.

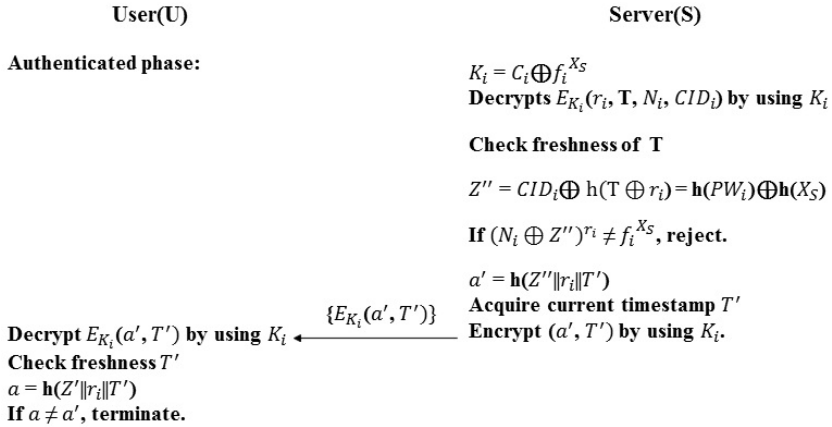


Fig. 3. Authentication phase

## 2.4 Password Change Phase

When  $U_i$  wants to change his password from  $PW_i$  to  $PW_i^n$ ,  $U_i$  implors this phase. The password change phase needs to pass the following steps:

- User  $U_i$  must have executed the above login and authentication phase. The server S authenticates his old password  $PW_i$ .
- After the successful mutual authentication,  $U_i$  inserts his new password  $PW_i^n$ . Then, smart card computes  $h(X_S) = M_i \oplus h(PW_i)$ ,  $N_i^n = N_i \oplus h(PW_i) || h(X_S) \oplus h(PW_i^n) || h(X_S)$  and replaces the old  $N_i$  with the new  $N_i^n$ .

## 3 Security Flaws in Yassin et al.'s Scheme

In this section, we present that Yassin et al.'s scheme is insecure and vulnerable to outsider attack, smart card stolen attack, offline password guessing attack and masquerade attack.

### 3.1 Outsider Attack

Any adversary  $U_a$  who is the legal user and owns a smart card, can get information  $(h(\cdot), N_a, M_a)$ , then he compute:  $h(X_S) = M_a \oplus h(PW_a)$ . Thus an adversary can get

$h(X_S)$  which same for each legal user and is very sensitive information, the hash value of secret key of the server. Furthermore, an adversary can computes secret key  $X_S$  of remote server.

1. The adversary calculates  $h(X_S) = M_a \oplus h(PW_a)$ , and then  $h(ID_a)^{X_S} = N_a \oplus h(PW_a) \parallel h(X_S)$ .
2. Assume that  $R_0 = ID_a$ , and  $i=1$ .
3. Calculate  $R_i = h(R_{i-1})$ . If  $R_i = h(ID_a)^{X_S}$  then  $i=X_S$  else  $i=i+1$
4. Repeat 3.

### 3.2 Smart Card Stolen and Offline Password Guessing Attack

Smart card stolen attack means an adversary who possessed with smart card performs any operation which the smart card and obtains any information. If an adversary  $U_a$  steals the SC of legitimate user  $U_c$  and obtains the parameters  $N_c$  and  $M_c$ , then he can easily computes out the hash value of the password of the real user  $U_c$  by computing  $M_c \oplus h(X_S)$ . Now, an adversary performs an off-line password guessing to get the current password of the user.

1. The adversary calculates  $h(PW_c) = M_c \oplus h(X_S)$ .
2. The adversary selects a random password  $PW_c^*$ , calculates  $h(PW_c^*)$  and compares it with  $h(PW_c)$ . If so, the adversary infers that  $PW_c^*$  is user  $U_c$ 's password. Otherwise the adversary selects another password nominee and performs the same processes, until he locates the valid password.

### 3.3 User Masquerade Attack

A legal but malicious user  $U_a$  can get the value  $h(X_S)$  from his own card, which is same for each user and can get the value  $h(PW_c)$  from legitimate user  $U_c$ 's smart card. If he knows the identity  $ID_c$  of user  $U_c$ , he can easily masquerades as  $U_c$  to login and access the remote server because he can computes  $Z'$ .

### 3.4 Server Masquerade Attack

An outsider adversary  $U_a$  can easily masquerades as to remote server because he knows secret key  $X_S$  of remote server. If an adversary intercepts login request message  $\{C_i, f_i, E_{K_i}(r_i, T, N_i, CID_i)\}$  that the user  $U_i$  sends to the server S. An adversary uses his knowledge of  $X_S$  and computes  $K_i = C_i \oplus f_i^{X_S}$ . Then an adversary decrypts  $E_{K_i}(r_i, T, N_i, CID_i)$  and can computes  $a'$ . Thus he can easily masquerades as server S.

## 4 Conclusion

In this paper, we have presented a cryptanalysis of Yassin et al.'s scheme. We indicate that Yassin et al.'s scheme is vulnerable to outsider attack, smart card stolen



attack, off-line guessing attack, user masquerade attack and server masquerade attack. Finally, our further research direction ought to propose a secure user authentication scheme which can solve these problems.

## References

1. Chien, H., Chen, C.: A remote authentication scheme preserving user anonymity. In: Proc. Advanced Information Networking and Applications, vol. 2, pp. 245–248 (2005)
2. Lee, Y., Nam, J., Won, D.: Security enhancement of a remote user authentication scheme using smart cards. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4277, pp. 508–516. Springer, Heidelberg (2006)
3. Nam, J., Kim, S., Park, S., Won, D.: Security analysis of a nonce-based user authentication scheme using smart cards. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 90(1), 299–302 (2007)
4. Yi, W., Kim, S., Won, D.: Smart Card Based AKE Protocol Using Biometric Information in Pervasive Computing Environments. In: Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2009, Part II. LNCS, vol. 5593, pp. 182–190. Springer, Heidelberg (2009)
5. Jin, Q., Lee, K., Won, D.: Cryptanalysis of a two-factor user authentication scheme over insecure channels. In: ISA 2012 (2012)
6. He, D., Wu, S.: Security flaws in smart card based authentication scheme for multi server environment. *Wireless Personal Communications*, 0929–6212 (2012)
7. Son, K., Han, D., Won, D.: A Privacy-Protecting Authentication Scheme for Roaming Services with Smart Cards. *IEICE Transactions on Communications* 95(5), 1819–1821 (2012)
8. Das, M.L., Saxena, A., Gulati, V.P.: A dynamic ID-based remote user authentication scheme. *IEEE Transactions on Consumer Electronics* 50(2), 629–631 (2004)
9. Wang, Y.Y., Liu, J.Y., Xiao, F.X., Dan, J.: A more efficient and secure dynamic ID-based remote user authentication scheme. *Computer Communications* 4(32), 583–585 (2009)
10. Yassin, A.A., Jin, H., Ibrahim, A., Zou, D.: Encrypted Remote User Authentication Scheme by Using Smart Card. In: Wang, F.L., Lei, J., Gong, Z., Luo, X. (eds.) WISM 2012. LNCS, vol. 7529, pp. 314–323. Springer, Heidelberg (2012)

# The Development of Convergent STEAM Program Focused on Rube Goldberg for Improvement of Engineer Career Awareness of Elementary School Students

Yilip Kim and Namje Park\*

Major in Computer Education, Faculty of Science Education,  
Graduate School, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
{yilipkim, namjepark}@jejunu.ac.kr

**Abstract.** This paper will focus on the development of the engineering /technology-related STEAM program for elementary students on the basis of the study results that the technology-based STEAM education had a positive effect on the improvement of junior high school students' attitude towards technology and that STEAM activities improved elementary school students' scientific research ability and helped them have a positive attitude towards science. Furthermore, the change will be checked in the elementary school students' preference to engineering/technology-related jobs by applying the engineering technology-related STEAM program to the students actually, analyze the degree of satisfaction from and that of the interest in the STEAM program, and also check the change in the student's career interest type frequency based on Holland's career characteristics type theory.

**Keywords:** STEAM, Rube Goldberg, Career, Elementary School.

## 1 Introduction

The new education paradigm that many countries including the US recently pay attention to is fusion education. To be in line with this flow of times, Korea is conducting STEAM education, which is the STEM education with Arts added. What actually triggered the introduction of STEAM education to Korea was to train those with the talent of science technology and to prevent the national loss caused by the students who avoided studying natural science. However, many students fail to consider their aptitude/interest/future enough in selecting their career and job, and especially, preference for technology and engineering-related career is noticeably lower than that for the careers related to liberal art and social sciences. In one sentence, the objective of the STEAM education in Korea is to cultivate the talented students of science technology who will lead Korea in future by engrave the positive perception of science technology and engineering into students' mind and leading them to have interest in science and technology.

---

\* Corresponding author.

This paper will focus on the development of the engineering/technology-related STEAM program for elementary students on the basis of the study results that the technology-based STEAM education had a positive effect on the improvement of junior high school students' attitude towards technology and that STEAM activities improved elementary school students' scientific research ability and helped them have a positive attitude towards science. Furthermore, the change will be checked in the elementary school students' preference to engineering/technology-related jobs by applying the engineering technology-related STEAM program to the students actually, analyze the degree of satisfaction from and that of the interest in the STEAM program, and also check the change in the student's career interest type frequency based on Holland's career characteristics type theory.

## **2 Theoretical Background**

### **2.1 STEAM Education**

STEAM is to train the men of talents with the STEAM literacy that enables the creative and comprehensive issue resolution through increasing the interest in and the understanding of the fused knowledge/course/instinct in the various fields related to science technology on the basis of the experience about the contents fused in various fields through creative designs and emotional touches.

#### 1) Creative design

It is the comprehensive course in which learners find the optimal scheme and resolve an issue through the revelation of creativity, efficiency, economic feasibility, aesthetical characteristics, etc., with the purpose of creating the products such as knowledge, products, works, etc. in the given situation.

#### 2) Emotional touch

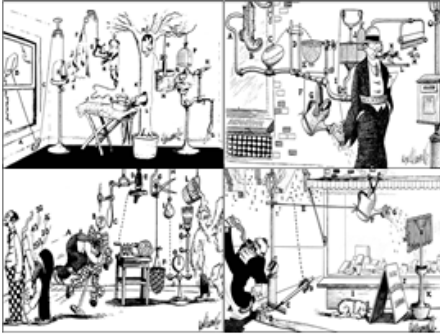
This includes various activities that help learners feel a positive emotion about learning and experience the pleasure of achievement and the value of failure.

#### 3) Content integration

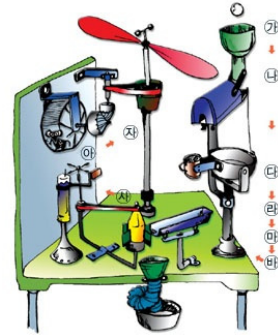
It means the dynamic integration of more than one teaching contents.

### **2.2 Rube Goldberg's Invention**

Rube Goldberg's Invention means "the most complicated machine to resolve the simplest task." This name came from the act that Rube Goldberg drew a cartoon about this idea. For example, this invention includes the machine that cleans lips and the device that prevents one's manager from noticing his/her late arrival at work. The highly accurate physical laws are applied to each step of these inventions. However, the result is in vain or poor, compared to the effort. Rube Goldberg defined this observation as "human beings' remarkable ability of putting the most effort to obtain the least result."



**Fig. 1.** Rube Goldberg's Invention Image Cartoons



**Fig. 2.** The Device that Spins a Propeller Using a Mouse

As it is very interesting to create Rube Goldberg's inventions, it helps students think outside of the box during issue resolution and exert creativity in the actual life.

### 2.3 Holland's Career characteristics Type Theory

Holland's career selection theory has been an important clue to understand career interest, career selection, and career satisfaction since the 1950s; and it has the theoretical background of the followings: the initial type study of an individual's interest and characteristics, the individual environment theory. The study of investigating the differences among individuals based on career aptitude has been on the rise recently, and especially, the many studies of Holland's career aptitude have been conducted. Holland (1985, 1977) classified human beings' characteristics into 6 types: Realistic, Investigative Artistic, Social, Enterprising, Conventional. According to the theory, the closer an individual is to one of these types, the more noticeably he or she displays the characteristics and behavior of the specific type.

## 3 Education Program Materials Development

Through the application of student's 12 sections rollercoaster creation program to the actual school field, the necessary parts were supplemented, and the teaching materials were developed in consideration of the effective teaching/learning method for both students and teachers. The teaching materials consist of total 20 sections, and these sections are arranged according to the standard tools of STEAM education: situation presentation, creative design, emotional experience. Besides, the content of teaching materials is designed such that students form a team for a project class themselves and perform the tasks in each section.

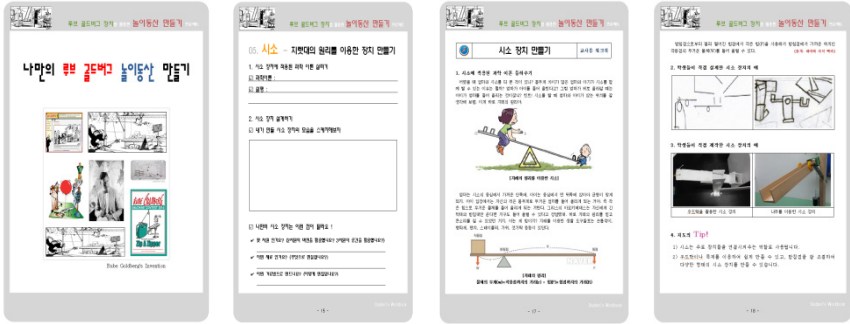


Fig. 3. The Text Materials for the Project to Make an Amusement Park Utilizing Rube Goldberg’s Inventions

3.1 The Field Application Object

The study objects were some 5thgradeand6thgradestudentsat○○ Elementary School (total 40) as shown below.

Table 1. The field application object

Classification	Male	Female
The number of students	28	12
Total	40	

3.2 Study Design

“One-Group Pretest-Posttest Design” of Fraenkel & Wallen (1996) was used as the test design model to investigate how the application of the STEAM education program using Rube Goldberg’s Inventions affects students’ selection of the engineering/technology-related career.

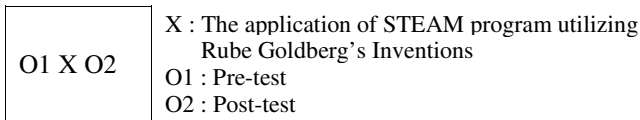


Fig. 4. A Test Design for the Study

As the design to deduct the cause and effect through investigating the change in dependant variables following the program application, “One-Group Pretest-Posttest Design” is mainly used to investigate the program’s effect.

3.3 Field Application Result and Interpretation

A. The analysis of the result from the participation into the Rube Goldberg’s Inventions Creation Competition in Korea

The change in students' attitude towards the engineering/technology through the STEAM program using Rube Goldberg's Inventions was confirmed in the actual competition. Students obtained an excellent result at 2012 Jeju-si Science Festival Goldberg's inventions Creation Competition, 2012 Korean Students' Creativity Championship Competition Challenge 3. Goldberg Invention Creation, and 2012 National Gwacheon Science Museum Center Goldberg's Invention Creation Competition.

The students who participated in the competition in which they applied the STEAM program using Goldberg's Inventions directly experienced engineering, technology, and basic science theories through practices. Above all, the students could be absorbed into the Goldberg's Invention creation activities with interest and could change their perception of engineering and technology through the participations in competitions.

B. An analysis of the field application result

- 1) An analysis of the effect of the program application on the change in students' preference of the engineering/technology-related career

**Table 2.** The field application result

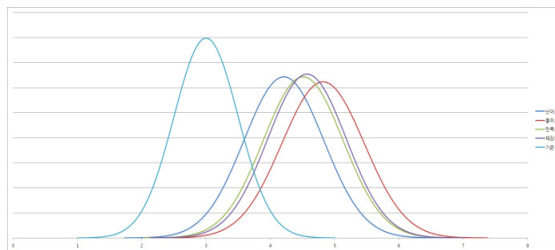
Question	Classification	N	M	SD	t	P
All questions	Pre-test	40	2.84	.98	-1.852	.045*
	Post-test		3.05	.90		

\*p<.05, \*\*p<.01, \*\*\*p<.001

To investigate the effect of the Rube Goldberg's Inventions creation STEAM program on the change in students' preference of engineering/technology-related career, matching sample t-verification was conducted. The result showed a statistically meaningful difference ( $t=-1.852, p<.05$ ).

- 2) An analysis of students' satisfaction and the class effectiveness following the program application

The average degree of difficulty of the entire classes was 4.21, which means that most students understood classes easily. The average degree of the interest in classes was 4.66, which means that most students participated in the class with high interest. The average degrees of satisfaction with the class (Question 9) and that of understanding the class (Question 7) were 4.57 and 4.62, respectively, both of which are high figures.



**Fig. 5.** Students' satisfaction and the class effectiveness following the program application

### 3) Holland's career interest type analysis

The change in the frequency of Investigative (I) students shown through the comparison before and after the participation in Rube Goldberg's Invention creation STEAM education program is described below.

The result is meaningful in the sense that the Rube Goldberg's Inventions creation STEAM education program helped students have confidence on science and technology and become the Investigative (I) type in career and job selection.

## 4 Conclusion

This paper was conducted to develop the STEAM education program and teaching materials with the purpose of inducing the positive change in elementary school students' attitude towards engineering/technology and the change in their perception about engineers regarding their job and career selection.

First, the Rube Goldberg's Inventions creation is the optimal activity for STEAM. It gave participants the chance to discover and develop their aptitude and talent as an engineer through the course that included the presentation of interesting situations, the creative design to resolve an issue alone, and the emotional experience of feeling cooperation and achievement during the creation process.

Second, the application of the developed STEAM education program to the education field resulted in the great change in students' attitude towards engineering and technology. Especially, the change in the career perception about engineers showed a meaningful difference before and after the test, which confirmed that the Rub Goldberg's Inventions creation STEAM education had a positive effect on the selection of the engineering/technology-related jobs. It was also noticed that the students' satisfaction with the program and their interest in it were high and that the frequency of the characteristics of Investigative (I) related to engineers in Holland's career interest type inspection increased.

The STEAM actively promoted by the Ministry of Education, Science and Technology has not been extended deeply into the school fields, yet. Besides, we cannot say that teacher's understanding of and will about STEAM education are active. Currently, many studies about program development are in progress and the efforts to expand the impact of the program application to schools have been made. Thus teachers also should make efforts to realize the true STEAM through the study and efforts of STEAM education.

## References

1. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
2. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. *The Journal of Korea Information and Communications Society* 35(5), 785–793 (2010)

# Key Management Scheme Using Dynamic Identity-Based Broadcast Encryption for Social Network Services\*

Youngman Jung, Yoonho Nam, Jiye Kim, Woongryul Jeon,  
Hanwook Lee, and Dongho Won\*\*

College of Information and Communication Engineering, Sungkyunkwan University,  
300 Cheoncheon-dong, Jangan-gu, Suwon-si, Gyeonggi-do, 440-746, Korea  
{ymjung, yhnam, jykim, wrjeon, hwlee, dhwon}@security.re.kr

**Abstract.** Recently, Social Network Service market is getting bigger and bigger. Then there are many security threats by malicious users. In addition, because sensitive data is concentrated on the central server, privacy can be exposed to SNS provider as well as malicious users. To overcome this problem, many previous researches suggest decentralized systems for SNS. In these systems, sensitive data may not be stored in central server. When a user transmits a message, the server does not interfere with the process. Thus, the user who transmits a message needs way to manage the keys that are used for message encryption scheme. In this paper, we suggest the efficient key management scheme using Dynamic Identity-Based Broadcast Encryption. Using this scheme, it is possible to communicate securely between users in decentralized social network.

**Keywords:** Social Network Services, Decentralized, Dynamic Identity-Based Broadcast Encryption.

## 1 Introduction

Recently, due to the spread of mobile devices such as smart phone, Social Network Services' market is getting bigger and bigger. According to eMarketer's survey, the number of SNS users seems to surpass the 10 million users worldwide[1].

According as many users use SNS, privacy exposure has become a problem. To overcome this problem, each service provides various solutions. Nevertheless, privacy exposure is still a problem, because sensitive data generated by users stores in central server.

To overcome this problem, many previous researches suggest decentralized systems for SNS[3][4][5][6][7][8][9]. Unlike the centralized systems, the systems have the concept of peer-to-peer network. Thus, the communication between the users is carried out without the central provider.

---

\*This research was funded by the MSIP(Ministry of Science, ICT&Future Planning), Korea in the ICT R&D Program 2013.

\*\* Corresponding author.



Because the decentralized systems do not have the central provider, the users in the systems should have the security systems without the central provider. Above all, the systems need the encryption schemes for transmitting the users' information and the key management for encryption.

In this paper, we suggest the key management for encryption in decentralized systems. The system does not have the central server. For the efficient key management in decentralized system, we use the concept of the dynamic identity-based broadcast encryption scheme proposed by Jiang[10][11].

This paper is organized as follow. Section 2 reviews Identity-Based Broadcast Encryption scheme and Jiang's Dynamic Identity-Based Broadcast Encryption scheme. Section 3 explains security requirements for SNS. Section 4 proposes the key management scheme using Dynamic Identity-Based Broadcast Encryption scheme for SNS. Section 5 analyzes the scheme. Section 6 concludes.

## 2 Preliminaries

### 2.1 Identity-Based Encryption[12]

Identity-Based Encryption is a type of public key Encryption. The encryption is suggested by Shamir in 1984. In this cryptosystem, the public and identifiable information such email, phone number is used as a public key, and the private key generator generates the corresponding private keys.

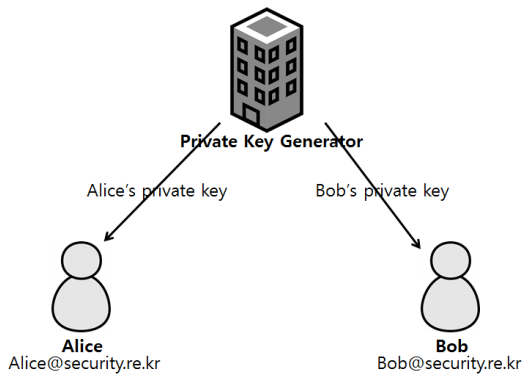


Fig. 1. Identity-Based Cryptosystem

### 2.2 Identity-Based Broadcast Encryption[10]

The concept of Identity-Based Broadcast Encryption was introduced by Delerablee. In Identity-Based Broadcast Encryption schemes, the authority generates users' private key using master secret key and each user's identity, and the user transmits an encrypted message through a broadcast channel. Then valid receiver can decrypt the message using his own private key.

### 2.3 Jiang, H's DIBBE scheme[11]

Identity-Based Broadcast Encryption has some problems. First, the maximum number of user should be predetermined. Second, each receiver must know all of the receivers. However, it is natural that each receiver knows only his own information. To solve these problems, the concept of Dynamic Identity-Based Broadcast Encryption was introduced by Jiang. This scheme consists of four algorithms: Setup, Extract, Encrypt, and Decrypt.

#### 1. Setup( $\lambda$ )

Given the security parameter  $\lambda$ , a bilinear map group system  $B = (p, G_1, G_2, G_T, e(\cdot, \cdot))$  is constructed such that  $|p| = \lambda$ . Also, two generators  $g \in G_1$  and  $h \in G_2$  are randomly selected as well as a secret value  $\gamma \in Z_p^*$ . Choose a cryptographic hash function:  $H: \{0,1\}^* \rightarrow Z_p^*$ .  $B$  and  $H$  constitute system public parameters. The master secret key is defined as  $MSK=(g,\gamma)$ . The public key  $PK=(v,h)$  is where  $v = e(g,h)$ .

#### 2. Extract( $MSK, ID$ )

Given  $MSK=(g,\gamma)$  and the identity  $ID$ , it outputs  $sk = \frac{1}{g^{\gamma \cdot h(ID)}}$ .

#### 3. Encrypt( $S, MSK, PK$ )

Assume for notational simplicity that  $S = \{ID_j\}_{j=1}^S$ . Given  $PK=(v,h)$ , the broadcaster randomly picks  $k, r \leftarrow Z_p^*$  and computes  $Hdr = (T_1, T_2, C_1, C_2)$  and  $K$  where

$$T_1 = r \cdot \gamma^S \text{ mod } p, \quad T_2 = \prod_{i=1}^S (H(ID_i)) \text{ mod } p, \\ C_1 = g^{-k/r}, \quad C_2 = h^{k \cdot \gamma \cdot T_1 \cdot T_2 / r}, \quad K = v^{k/r}.$$

Encrypt outputs  $(Hdr, K)$ . (Then  $K$  is used to encrypt the message)

#### 4. Decrypt( $S, ID_i, sk_{ID_i}, Hdr, PK$ )

In order to retrieve the message encryption key  $K$  encapsulated in the header  $Hdr = (T_1, T_2, C_1, C_2)$ , user with identity  $ID_i$  and the corresponding private key  $sk = \frac{1}{g^{\gamma \cdot h(ID_i)}}$  computes

$$K = \left( e(C_1, h^{p_{i,S}(\gamma)}) \cdot e(sk_{ID_i}, C_2) \right)^{\frac{H(ID_i)}{T_2}} \\ \text{with } p_{i,S}(\gamma) = \frac{(T_1 - 1) \cdot T_2}{H(ID_i)} \text{ mod } p$$

## 3 Security Requirement

According as many users use SNS, privacy exposure has become a problem. In this section, we define the target of preserving privacy, adversary, and message authentication. Then we introduce some security requirements in SNS.

Many people use SNS in order to communicate with each other. Sometimes, a user sends messages to other users, and the user receives messages from others. Users in SNS can communicate with each other through transmitting the messages. If unspecified user can read messages sent by the user, privacy exposure would occur. Therefore, we regard messages as a target of preserving privacy, and it is assumed that users that are not included in receiver group are possible adversaries. If they can read a message, sensitive information can be stolen unintentionally. Therefore, every message transmitted by users can be encrypted. Also, in the process of transmission, messages can be modified by malicious users. Therefore, encrypted message should be attached with a MAC. A key that is used for MAC is derived from the key for the encryption. Then a user who received a message validates MAC. Through this process, the user can be convinced that the message is sent by authorized user.

We need some security requirements for communicating securely. We explain security requirements for SNS in Table 1[1][2][13][14][15][16][17][18][19][20].

**Table 1.** Security Requirements for SNS

Security Requirement	Explanation
Access Control	SNS should prevent leakage of personal information and unintended spread of information. Thus, SNS should support access control about users' contents.
Confidentiality	SNS should provide encryption in order to maintain the confidentiality of the content transmitted between users. Also, SNS shall ensure the confidentiality of the information that is stored on a server to cope with intentional information leakage of internal managers.
Availability	The user who wants to communicate with other users can access from every device.
Integrity	SNS should provide the integrity of the information generated by authorized users.
Privacy	SNS should protect users' privacy from any others both internal and external malicious users.
Forward Secrecy	The user who wants to transmit messages to other users adds them to receiver group. If involved users in the group are removed, then they cannot read the messages transmitted by the user.
Backward Secrecy	If the user adds some users to receiver group at some point, involved users in the group cannot read messages written by the user in the past.

## 4 Key Management Scheme Using DIBBE

The users who want to receive a message sent by user A should send his own ID and public key to user A. Using these values, user A can send a message to the user who sends the values.

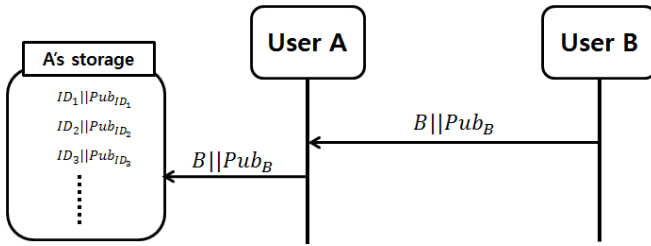


Fig. 2. Joinable Phase

Detailed key management scheme consists of four steps.

1. Setup

For sending a message, user A generates his own master secret key and PK.

2. Extract

User A who wants to send a message is provided with receivers' identity and public key from his own storage. Using master secret key and receivers' identity, he generates receivers' private key  $sk_{ID}$ . And he transmits the private key that is encrypted using receivers' public key, respectively.

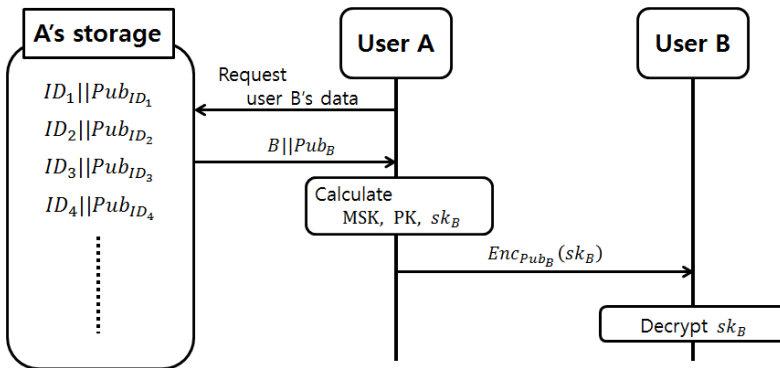


Fig. 3. Setup & Extract steps

3. Encrypt

User A generates  $k$  and  $r$  randomly, and he calculates  $Hdr$  and  $K$  using  $k$ ,  $r$ , and receivers' identity. Then he broadcasts  $Hdr$  and  $PK$ .

4. Decrypt

User B who is included in receiver group can calculate the key  $K$  that can decrypt the messages transmitted by user A. But, User C who is not included in receiver group cannot calculate the key.

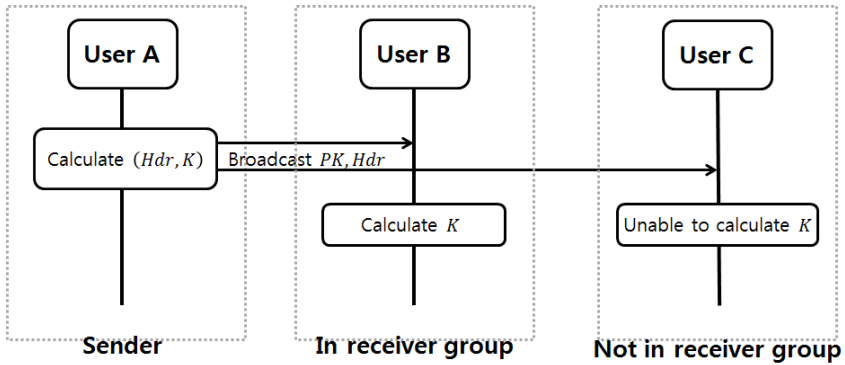


Fig. 4. Encrypt & Decrypt steps

If user A wants to add new receiver, then he performs Extract step. Also, if user A wants to remove a receiver, then he recalculates Hdr using other  $k$  and  $r$ . And he broadcasts Hdr.

Through these steps, user A can share the key  $K$  with users in receiver group.

## 5 Security Analysis

The suggested key management scheme is to manage keys for encryption in decentralized systems for SNS. In these systems, there is no SNS provider. Because of this, the systems are safe from internal attackers.

Despite this advantage, the user in the systems should have the security systems without the central provider. In this paper, we provide the security systems using DIBBE. The users who want to send a message broadcast the message to all users. If the user who receives the message is included in receiver group, then the user can read the message. Otherwise, the user cannot read the message.

If the user who transmits the message wants to add new receiver or to remove a receiver, then the user can simply modify the receiver group.

Although the decentralized systems do not have the SNS provider, the systems can have the security system using various key management schemes. We suggest the key management scheme using DIBBE suggested by Jiang, H. Using this scheme, the user who wants to transmit a message can transmit the message securely.

- Access Control

The user who transmits messages chooses identity of users who receive the messages from the user's storage. Because the user encrypts the messages using the key made of receivers' identity stored in the user's storage, Users who receive the messages can decrypt the encrypted messages using their own private key. Therefore, Users who can decrypt the encrypted messages are regarded that the users have accessible permission.

- Confidentiality

Every message transmitted by users is encrypted using a proper key. The key used to encrypt is the key made of receivers' identity. Thus, to decrypt the encrypted message, receivers should use the private key corresponded to their identity.

- Availability

The user who wants to transmit a message can always encrypt the message using receivers' identity stored in the user's storage. Also, receivers can always decrypt the encrypted message because the message is encrypted using the key made of their own identity.

- Integrity

In our scheme, every encrypted message should be attached with a MAC. It is assumed that MAC is composed of cryptographically secure hash function. Because it is hard to forge the MAC, integrity of messages transmitted by users can be checked by validating the MAC.

- Privacy

A user who transmits a message should encrypt the message using receivers' identity. Receiver can decrypt the encrypted message using his own private key. Thus, users who are not included in receiver group cannot read the encrypted message.

- Forward Secrecy & Backward Secrecy

In our key management scheme using DIBBE, the user who wants to transmit a message encrypts the message using receivers' identity. Thus, in the process of encryption, users who will receive the message are determined. In figure 5, user A included in receiver group can read messages broadcasted by Sender, but user B who is not included in receiver group cannot read the messages.

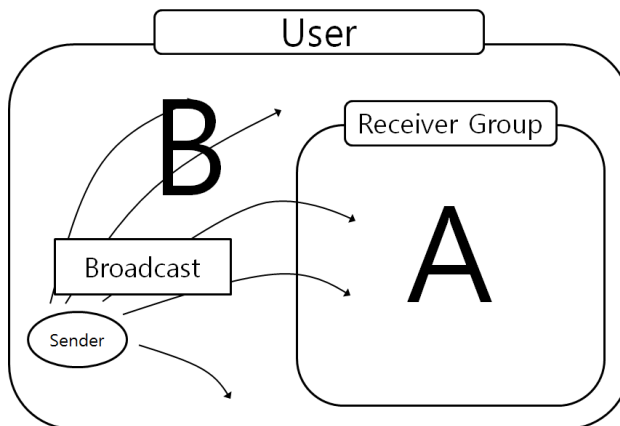


Fig. 5. Receiver Group

If the user who sends a message wants to add new user to receiver group, the user makes new user's private key in extract step. Then the user sends the key to new user and the user encrypts a message using new user's identity. Thus, new receiver can read the encrypted message. However, the receiver cannot read the message that was encrypted previously. Therefore, our scheme supports backward secrecy. If the user don't want to share his information with a receiver anymore, he will not use the receiver's identity in encrypt step. Then the receiver cannot read the message that is encrypted subsequently. Therefore, our scheme supports forward secrecy.

## 6 Conclusion

According as many users use SNS, various security threats have occurred. For this reason, SNS provides various solutions. But, the SNS provided by SNS providers does not solve the problem of privacy exposure by SNS providers. To overcome this problem, many previous researches suggest decentralized systems for SNS. We suggested the key management for encryption in decentralized systems. For the efficient key management, we used the concept of the dynamic identity-based broadcast encryption scheme proposed by Jiang. Using this method, the user who transmits a message may not store sensitive information in the central server.

## References

1. Jeong, H., Won, D.: A method for the protection of the message in social network service(SNS) using access control and hash-chain. *Journal of The Korea Institute of Information Security & Cryptology* 1(23), 81–88 (2013)
2. Lee, C., Jung, Y., Jung, J.J., Won, D.: Dynamic User Reliability Evaluation Scheme for Social Network Service. *Journal of The Korea Institute of Information Security & Cryptology* 2(23), 157–168 (2013)
3. Yeung, C.A., Licaardi, L., Lu, K., Seneviratne, O., Lee, T.B.: Decentralization: The future of online social networking. In: *Proc. Int. Joint Conf. W3C Workshop* (2009)
4. Datta, A., Buchegger, S., Vu, L.H., Strufe, T., Rzađca, K.: Decentralized online social networks. In: *Handbook of Social Network Technologies and Applications*, pp. 349–378 (2010)
5. Cutillo, L.A., Molva, R., Strufe, T.: Privacy preserving social networking through decentralization. *Wireless On-Demand Network Systems and Services*, 145–152 (2009)
6. Cutillo, L.A., Molva, R.: Safebook: a privacy-preserving online social network leveraging on real-life trust. *IEEE Communications Magazine* 47(12), 94–101 (2009)
7. Aiello, L.M., Ruffo, G.: Secure and flexible framework for decentralized social network services. In: *IEEE PERCOM*, pp. 594–599 (2010)
8. Buchegger, S., Schiöberg, D., Vu, L.-H., Datta, A.: PeerSoN: p2p social networking. In: *Proc. 2nd ACM EuroSys Workshop on Social Network Systems*, pp. 46–52 (2009)
9. Aiello, L., Ruffo, G.: LotusNet: tunable privacy for distributed online social network services. *Computer Communications* 1(35), 75–88 (2012)

10. Delerablée, C.: Identity-based broadcast encryption with constant size ciphertexts and private keys. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 200–215. Springer, Heidelberg (2007)
11. Jiang, H., Xu, Q., Shang, J.: An efficient dynamic identity-based broadcast encryption scheme. In: 2010 Second International Symposium on Data, Privacy and E-Commerce (ISDPE), pp. 27–32 (2010)
12. Boneh, D., Franklin, M.: Identity-Based Encryption from the Weil Pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001)
13. Ajami, R., Ramadan, N., Mohamed, N., Al-Jaroodi, J.: Security Challenges and Approaches in Online Social Networks: A Survey. *International Journal of Computer Science and Network Security* 11(8) (August 2011)
14. Masoumzadeh, A., Joshi, J.: Ontology-based access control for social network systems. *IJIPSI* 1(1), 59–78 (2011)
15. Beato, F., Kohlweiss, M., Wouters, K.: Scramble! Your Social Network Data. In: Fischer-Hübner, S., Hopper, N. (eds.) PETS 2011. LNCS, vol. 6794, pp. 211–225. Springer, Heidelberg (2011)
16. Zilpelwar, R.A., Bedi, R.K., Wadhai, V.M.: An Overview of Privacy and Security in SNS. *International Journal of P2P Network Trends and Technology* 2(1) (2012)
17. Lawler, J.P.: A Study of the Perceptions of Students on Privacy and Security on Social Networking Sites(SNS) on the Internet. *Journal of Information Systems Applied Research* 3(12) (June 2010)
18. Lawler, J.P., Molluzzo, J.C., Doshi, V.: An Expanded Study of Net Generation Perceptions on Privacy and Security on Social Networking Sites(SNS). *Information Systems Education Journal* (February 2012)
19. Nam, J., Park, M., Han, S., Paik, J., Won, D.: Scalable Group Key Exchange for Securing Distributed Operating Systems. *Journal of Information Science and Engineering* 28(5), 829–857 (2012)
20. Choi, D., Jeong, H., Won, D., Kim, S.: Hybrid Key Management Architecture for Robust SCADA Systems. *Journal of Information Science and Engineering*, 281–298 (March 2013)



# Protecting Mobile Devices from Adversarial User by Fine-Grained Analysis of User Behavior

Yonggon Kim, Ohmin Kwon, Sunwoo Kim, Byungjin Jeong, and Hyunsoo Yoon

Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea

**Abstract.** In recent years, mobile devices have become the mainstream medium for maintaining and processing a variety of information, including personal or sensitive information. To prevent a leakage of information to unauthorized users, a user authentication scheme that is appropriate for mobile environment is surely needed. In the meantime, it is important to guarantee sufficient usability of the mechanism so that it does not affect the user experience of the mobile device in a significant way. To pursue both usability and security perspectives of mobile devices, we propose a novel way of user authentication where fine-grained user behaviors are exploited to prohibit unauthorized access from an adversarial user. We examine behavioral patterns of daily usage of mobile devices in very fine-grained and implicit fashion. As a result, we suggest and evaluate the overall system exploiting user behaviors to separate valid user and adversarial user.

**Keywords:** Authentication, Mobile Device, Usable Security.

## 1 Introduction

In recent years, there has been dramatically increasing demand on mobile devices. Due to large improvement of their portability and computing performance, benefits of mobile devices are now comparable to traditional computing devices. These days, mobile devices like smartphone and tablet are hugely participated in various domains of our daily life including social relationships, financial management, and even business work. As the number of important tasks dealt by mobile devices increases, the proper security system needs to be developed so that users can use their devices safely and the attacks can be segregated.

One of the most important security feature needed is an authentication mechanism. Although multiple user authentication mechanisms have been proposed and implemented, a careful adjustment of classical user authentication mechanism is needed for a smartphone<sup>1</sup> because of its smaller screen and limitation of input devices. The most widely used user authentications on a smartphone are PIN-based and pattern-based passcode systems. It is very intuitive to build such

---

<sup>1</sup> Although we will use the term mobile device and smartphone interchangeably in many context of the paper, our results are mainly focused on smartphone.

systems for developers and to manipulate for users. However, with such authentication, mobile device only supports two states, either locked all the time or no lock at all. Thus, only two access control models which are either “too hard” or “too soft” exist for common mobile device users[1]. As a result, despite of the simplicity of the authentication mechanisms, 36% of users use their smartphones without enabling any authentication [2]. Also, 15% of total passcodes used are ten most popular passcodes of which the top three are 1234, 0000, and 2580 [3]. A device that has proper (not too hard and not too soft) authentication system would allow users to use lock only when it is necessary. In this paper, we investigate what is the necessary condition to use lock system by modeling the situation in which the device is vulnerable to adversarial users.

Due to the importance of user convenience in mobile devices and the limitation of input methods, authentication based on implicit data processing has been researched in broad ways. Since there exists a variety of sensors on mobile devices, prior works have exploited various kinds of user contexts including face image, voice sound, and touch screen gestures. In particular, the movement of user measured by accelerometer sensors is easily obtained in mobile frameworks like Android and iOS. Also, they contain unique features which are different for each user so that in the prior work[4], information of device movement collected by accelerometer and orientation sensors are exploited to suggest user verification mechanism. In this work, we propose a way to use the combination of movement information and timing result of device events such as screen on/off. As a result, we offer a novel mechanism to detect adversarial users using very simple user context data and heuristics. Furthermore, to avoid any significant degradation of user experience, we propose *implicit authentication* that operates in a user transparent manner. Since we make use of the user context data which are easily obtainable and have negligible noise, our authentication mechanism does not require any explicit interaction with device users.

## 2 Related Works

*Persistent authentication* used in static devices such as Desktop is not suitable for mobile environments that have high chances of information leakage due to device losses. To prevent the information leakage in mobile environments, *transient authentication* is recommended. But currently used methods for transient authentication require user’s frequent participations whenever mobile devices are activated, which could significantly degrade their usability. Anthony et al.[5] proposed a method using wearable token for transient authentication which considers both usability and security. In this work, it checks whether the valid user is close by using short range communication between mobile device and wearable token. If he/she is not close, it protects stored data by using cryptographic controls for preventing information leakage.

Transient authentication needs user’s frequent participations for security. It is also necessary to use additional devices if usability is needed to be considered. To alleviate this problem, implicit authentication is proposed, which satisfies

**Table 1.** A criterion of determining adversarial user activity

	Screen On		
Safe State	User		
	Slow Screen On	Fast Screen On	Direct Screen On
Danger State <sup>2</sup>	Adversary	User	User

both security and usability without explicit participations of users or additional devices. It uses multiple efficient sensors equipped in mobile devices. They can recognize an event happened by a current user and check whether it is happened by the valid user or not. Then, it authenticates his valid identity. From this way, implicit authentication can provide high usability since it does not require user's frequent participations. The events used for implicit authentication are classified into location-based[6], biometric-based[7], and behavior-based[8]. For correctness of recognition in the machine learning process, a large number of training data are needed, which are collected from the valid user for a long time.

A decision based on only whether the current user is valid or not provides just all-or-nothing access. Thus, it causes inconvenience to users who use shared mobile devices or utilize diverse applications which have different sensitive-degree information. For these cases, *multi-level authentication* is proposed by using access control[9] or authentication score[10] based on measurement of how much interactions with the valid user are happened. Depending on the decision of the access control scheme or the authentication score, leveled-access is allowed to the current user.

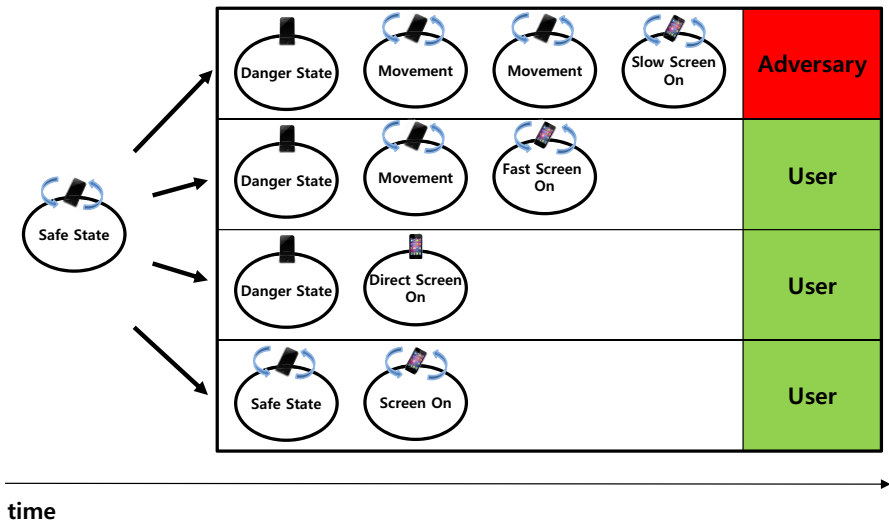
### 3 Methodology

One of the most important purposes of mobile user authentication is to prevent the information leakage when the device has been lost and revealed to anonymous users. To propose user authentication suitable for such objective, we suggest a novel way to recognize dangerous situation (and adversarial user) by analyzing simple user context data.

As mentioned earlier, we manipulate two kinds of device events, which are screen on/off event and movement event of the mobile devices. These events can be easily manipulated by using broadcast receiver in the android framework. As shown in Table 1, we classify states of mobile device into two groups: danger state and safe state. When there exists no movement of the device for consecutive 30 seconds and no screen on event during those 30 seconds, the state of device is set as danger state. Intuitively, danger state implies a possibility of the situation

<sup>2</sup> When there's no movement of device for consecutive 30 seconds with screen off.

that a user unintentionally left the device in solitary and moved to different place. Safe state implies that the device never has been into consecutive 30 seconds of “sleep” after last valid authentication. In the safe state, the activity of screen on is always confirmed to the valid activity of the correct user. It is a quite reasonable decision since the safe state implies that the device has been continuously in motion after valid authentication. Thus, the possibility of device exposure to an unauthorized user during the safe state is quite low. However, since a condition to enter the danger state is somewhat simple, we devise an additional condition to distinguish an adversarial action and a normal action by analyzing and classifying the activity of screen on event. In dangerous state, the activities of screen on are classified into three groups. *Direct screen on* means that a user turns on the screen without any movement. We consider it as a valid user behavior since for unauthorized user, it is unlikely to turn on the screen without grabbing or touching the device itself. *Fast screen on* implies that there exists movement of device before the event of screen on. Also “fast” means that a time interval from the movement to screen on event is relatively short. On the other hand, *slow screen on* is almost same as *fast screen on* but with relatively long interval. We have a conjecture that screen on events of unauthorized user have a large portion in the class of *slow screen on*. Also, a relatively small portion of valid user’s screen on events will be *slow screen on*.



**Fig. 1.** High level description of proposed authentication. It gives four scenario of distinguishing valid user and adversarial user.

Based on above conjectures, we propose a novel user authentication that is very simple but shows an effective capability in perceiving unauthorized access to smartphone when it has unintentionally been left behind. Overall description of the mechanism is shown in Figure 1. Our authentication classifies users

whenever the screen on event has occurred. Only if a *slow screen on* event is activated, the authentication scheme treats it as unauthorized action and requires additional authentication like passcode-based systems. However, we observe that there exists unexpectedly *slow screen on* for a valid user, because it is impossible to distinguish the real purpose of movement of the device. For example, after 30 seconds of unmoved period, if a user walks with device for a minute and tries to turn on a screen, then our authentication scheme classifies such situation as a *very slow screen on* event. Fortunately, the important observation is that the unauthorized users always turn on the screen as soon as they have found the device, except the situation when they try to give the device to whom has authority, such as an officer. Therefore, we treat such *very slow screen on* event as a valid action from valid user. We formally define *very slow screen on* event as a screen on event which takes more than 30 seconds from initial move to screen on.

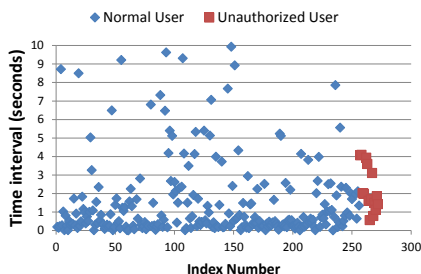
### 4 Experiments and Results

To evaluate our authentication scheme, we implement it on the Android framework and make experiments for 21 different users. We make two experiments which consist of a valid user experiment and an unauthorized user experiment to analyze our scheme in two different perspectives. 6 and 15 users have participated in our first experiment and second experiment, respectively. In a former experiment, we setup our application to 6 participants' smartphone and classify every screen on event. The result is shown in Table 2.

For latter experiment, in order to model a situation of authorized access, we leave a smartphone in various places like cafe or public road, and observe the behavior of users who have a try to access the smartphone. As shown in Table 2, we obtain a result which show the meaningful difference compared to the result of the first experiment. An average value of time interval from initial movement to screen on is much higher for unauthorized access.

**Table 2.** Comparison of experimental results

	Valid User	Unauthorized User
Safe Screen On	410	N/A
Direct Screen On	181	0
Slow or Fast Screen On	256 (2,114)	15 (3,146)
Very Slow Screen On	59	0
Total	906	15



**Fig. 2.** Scatter plot of time interval from initial movement to screen on event. X-axis is for index number of experimental datas.

Every time-interval instance except *very slow screen on* is plotted in Figure 2. For simplicity, we omit the results which are longer than 10 seconds. As shown in Figure 2, many instances of the normal user are plotted below 1 second while unauthorized users' instances are focused near 2 and 4 seconds.

To distinguish *slow screen on* events and *fast screen on* events, we need to choose a threshold value. Our authentication scheme shows 11% of false rejection rates and 13% of false acceptance rates with threshold value set as 1(second).

## 5 Conclusion

In this work, we have suggested a novel user authentication scheme which exploits two simple smartphone sensor data, i.e, movement of device and screen on/off event. By implicit authentication, users are unaware of its existence and operation mechanism so that it significantly improves the usability of previous passcode based authentications. We have focused on the difference between normal usage of the smartphone device and adversarial access, by which we have suggested a very simple way to protect the device from adversarial users. As a result, with appropriate threshold value setting, we have captured 13 of 15 adversarial users, while only 11% of false rejection of valid access has occurred.

**Acknowledgments.** This work was supported by Samsung Electronics and the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MEST) (No. 2012-0005390).

## References

1. Hayashi, E., Riva, O., Strauss, K., Brush, A.J.B., Schechter, S.: Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. In: Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS 2012, pp. 2:1–2:11. ACM, New York (2012)
2. Siciliano, R.: More Than 30% of People Dont Password Protect Their Mobile Devices (2013), <http://blogs.mcafee.com/consumer/unprotected-mobile-devices>
3. De Luca, A., et al.: Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 987–996. ACM, New York (2012)
4. Conti, M., Zachia-Zlatea, I., Crispo, B.: Mind how you answer me!: transparently authenticating the user of a smartphone when answering or placing a call. In: Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2011, pp. 249–259. ACM, New York (2011)
5. Nicholson, A., Corner, M., Noble, B.: Mobile device security using transient authentication. IEEE Transactions on Mobile Computing 5(11), 1489–1502 (2006)
6. Seifert, J., et al.: Treasurephone: Context-sensitive user data protection on mobile phones. In: Floréen, P., Krüger, A., Spasojevic, M. (eds.) Pervasive 2010. LNCS, vol. 6030, pp. 130–137. Springer, Heidelberg (2010)

7. Trewin, S., et al.: Biometric authentication on a mobile device: a study of user effort, error and task disruption. In: Proceedings of the 28th Annual Computer Security Applications Conference, ACSAC 2012, pp. 159–168. ACM, New York (2012)
8. Shi, E., et al.: Implicit authentication through learning user behavior. In: Burmester, M., et al. (eds.) ISC 2010. LNCS, vol. 6531, pp. 99–113. Springer, Heidelberg (2011)
9. Ni, X., Yang, Z., Bai, X., Champion, A., Xuan, D.: Diffuser: Differentiated user access control on smartphones. In: IEEE 6th International Conference on Mobile Adhoc and Sensor Systems, MASS 2009, pp. 1012–1017 (2009)
10. Riva, O., Qin, C., Strauss, K., Lymberopoulos, D.: Progressive authentication: deciding when to authenticate on mobile phones. In: Proceedings of the 21st USENIX Conference on Security Symposium, Security 2012, p. 15. USENIX Association, Berkeley (2012)

# Design of an RDFizer for Online Social Network Services

Junsik Hwang, Hyosook Jung, Sujin Yoo and Seongbin Park\*

Korea University, Seoul, Korea

{js.seth.h, est0718, mynameislydia, hyperspace}@korea.ac.kr

**Abstract.** Recently, the number of online Social Networking Services (SNSs) such as Facebook or Twitter has been increasing and many users have various activities such as talk, image sharing and, making recommendations, etc. In addition, they can write their own profiles that contain a lot of personal information and values. If we can apply Semantic Web technology such as Resource Description Framework (RDF) to the information existing on SNSs, it can help analyzing valuable information in a machine-understandable way. In this paper, we propose an RDF Schema that defines social activities on SNSs. Then, we implemented a system by which data existing on SNSs can be converted into an RDF document using the proposed RDF schema.

**Keywords:** Semantic Web, Social Network Service, Semantic SNS.

## 1 Introduction

The number of data produced in social networks has recently increased and the types of data format are different among the social networks. In this situation, some services are created for the purpose of using SNS with efficiency. For instance, social comment services combine questions and answers with SNSs for better customer services. There are services that can spread an URL which contains information, pictures, video or online stores as a way of social marketing. In fact, SNSs tend to get connected each other more and more and it becomes complex owing to interoperability based on Open APIs. To solve this complexity, a service emerges that manages distributed bookmarks over SNSs. And a social network federation system that tries to integrate multiple SNS has been introduced [1][2].

In this paper, we present a system that transforms the data on SNSs which are written in various formats into an RDF document. First, we designed an RDF Schema where social activities can be defined. Then, we developed a system that converts data on SNSs into an RDF document (i.e., RDFizer [3]). By using our system, users do not have to spend much time on formatting data existing in different formats. In addition, both experts and end-users can easily convert data about social activities on SNSs into an RDF document.

This paper is structured as follows. Section 2 describes related works to our research. Section 3 explains the proposed RDF Schema as well as the system in detail. Section 4 concludes the paper.

---

\* Corresponding author.



## 2 Related Works

SNSs have priceless data. In social search, Aardvark, data from twitter have been extracted [4]. In addition, some efforts have been made to apply Semantic Web technology to data of SNSs [1][2][5]. These researches tend to exploit well-known RDF Schema such as FOAF [6] and SIOC [7]. However, since SIOC focuses on describing “online communities”, it is not a proper schema to describe the structure of SNS [8].

FlickrWrapper [9] is using Flickr and DBPedia and it is some kind of meta search which merges results into an RDF file. FlickrWrapper searches pictures over Flickr[10] by just keyword and user activities are not considered at all. It is the purpose of our research to make a semantic environment of SNSs by converting SNS data to RDF documents using an RDF Schema and an RDFizer. Precisely, it includes user activities that consist of who, when, where, what etc.

In general, the main data of SNS is who, when, with whom and what did. The “who” is the account of the SNS, “when” is the time whose article is written, “what did” presents the action of publishing the text. “Who” is divided by the target conversation as the author of answered question or the public when text has not a specific audience. The main difference between our system called Semantic SNS and existing studies of SNSs with Semantic Web technology lies in the fact that we focus on social activities and try to integrate SNSs. Studies related to the characteristics of SNS[11][12][13] show that SNS have distinct features on their use. It is a tool for social communications and contributes sharing many contents even used in marketing. User’s experience over SNS occurs in a short time and is very intensive event.

## 3 Proposed System

In this section, we explain the RDF schema (RDFS) that we propose. Then we describe the structure of the proposed system as well as how it can be used. A proposed RDFS adopts well-known RDFSs such as Dublin Core [14], FOAF and SIOC. Dublin core is used for describing a common web resource, and FOAF is used for the profile of a user. The proposed RDFS defines vocabularies for external link (ss:externalLink) and message receiver (ss:receiver). An external link usually has a form of URL in text, message receiver is provided in different way in each SNS site. In Twitter, special tag is used for marking user id as listener. In the case of Facebook’s comment, it is inferred from who is writer of replied text. Commonly these messages have short lengths and replied each other, then a bunch of text composes a dialog (ss:dialog). Semantic SNS considers these features of SNS data. Figure 1 shows the structure of the proposed RDF Schema.

As an RDFizer, the main functionality of Semantic SNS is to generate an RDF document that contains data which comes from an SNS. In this research, social activities, written on SNSs as user’s action or its result are mainly data which is a form of text, image, reply, recommender etc. Most SNSs maintain these data and Open APIs provide these data with internal identifier. In case of Twitter and Facebook, we can

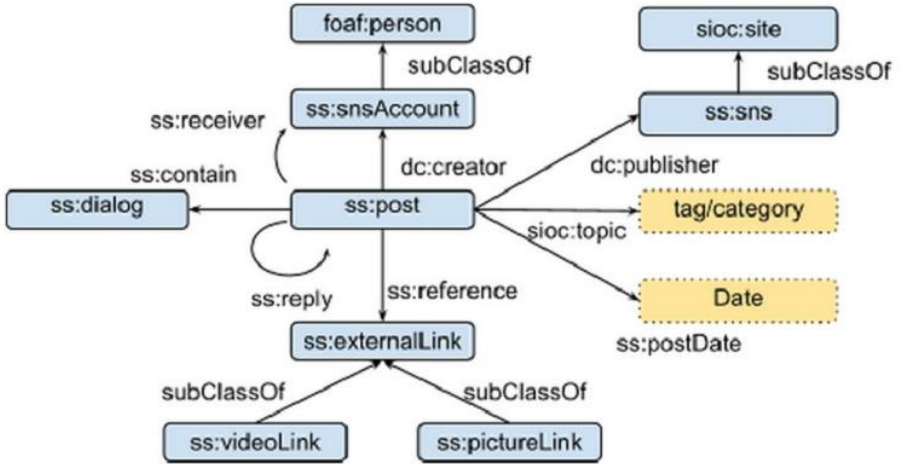


Fig. 1. Structure of the proposed RDF Schema

make a URI of data from these identifiers hosted by Twitter or Facebook. For instance, Twitter’s URI has the following form, ‘https://twitter#!/SCREENNAME/status/IDENTIFIER OF STATUS’, where SCREEN NAME is account of user and IDENTIFIER OF STATUS is contained in Open API’s results.

As an RDFizer, Semantic SNS reuses these URIs and attaches properties to them. Consequently, a user can get an RDF document with one’s social activities. Semantic SNS needs proper authorization for accessing. Both Facebook and Twitter, provide authorization processes through OAuth [15]. After Semantic SNS has user’s permission, it accesses data through Open APIs[16][17] which are provided by both SNSs, respectively. In the current version, our system tries to have authorization of Facebook and Twitter one at a time. So, a user must have accounts on both systems.

Figure 2 shows the components of Semantic SNS. We used ‘node.js’[18] as a platform which works like a web server application. And a MVC Web framework, ‘Express.js’[19], is used for handling http request and response. And it needs OAuth, which means that the system manages some requests of HTTP on difference servers (exactly, Semantic SNS, Facebook and Twitter) over connectless context. When it gets authorization, it has to format data from SNSs. The Web Request Handler is built for processing http requests. The OAuth Manager deals with authorization states and manages OAuth sequence. In current version, OAuth Manager handles Facebook and Twitter and these functionalities are supported by facebook-js[20]and twitter-js[21]. The RDFizer collects data and sets its format.

In order to generate an RDF document from data on SNSs, we have to deal with different data formats of SNSs. Facebook proposes concept of Open Graph[22] as a data scheme, and its data is queried by FQL (Facebook Query Language) [23]. For instance, if a user wants to query one’s facebook’s status history, the user just sends a request with a query sentence like ‘SELECT uid, status id, time, source, message

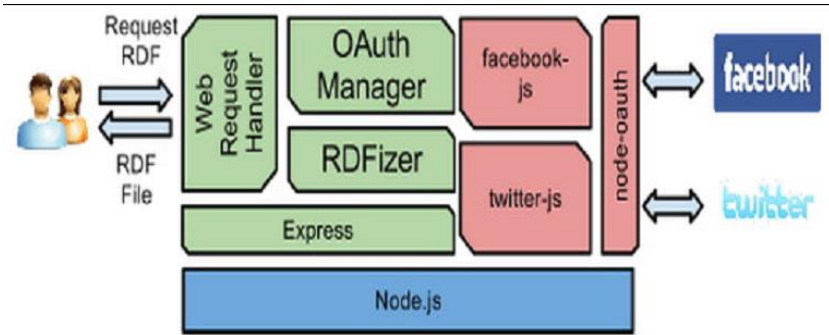


Fig. 2. Semantic SNS architecture

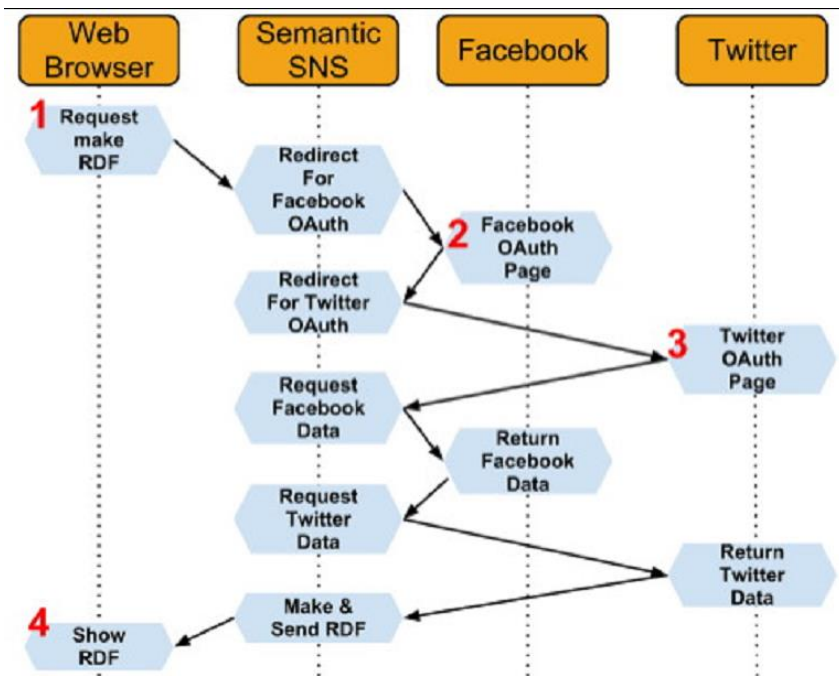


Fig. 3. Activity diagram of Semantic SNS

FROM status WHERE uid= me()'. Twitter affords Twitter REST API[24], it consists of urls for getting Twitter's data. In Semantic SNS, the RDFizer needs tweets written by a user, which are obtainable from timeline. So we use an API, 'statuses/home timeline', designed for getting online user's tweets. Figure 3 is an activity diagram of the proposed system.

In this activity diagram, the starting point is a user's request which is regarded an order for making an RDF document. Clicking the "start" button in the web page lets the system try to get authorization for accessing SNSs. It means that a web page

redirects to Facebook OAuth page. At that moment, the user will see a Facebook's page with question about permission. If the user agrees with accessing of Semantic SNS, then the system goes to the next step (i.e., starting with Twitter). Like Facebook, the system needs a permission to get data from Twitter. So Twitter's OAuth page will be shown. After user's agreements, Semantic SNS starts working. It starts to gather user's data contained in social activities, through Open API of each SNS. It sequentially demands data of Facebook and Twitter. After data collecting, Semantic SNS makes an RDF document and sends it to user's web browser. Then, the user will see an RDF document that contains information about social activities.

## 4 Conclusion

In this paper, we propose an RDF Schema that defines social activities existing on SNSs. We also implemented a system that converts data on SNSs into RDF documents using the proposed RDF Schema. A recent survey shows that the percentage of member overlaps between social network sites is relatively high [25]. This means that we might need to consider different SNSs for a single user in order to exploit social activities of the user. This aspect has been reflected in our system from the start of our research. Therefore the proposed system can help integrating data which come from different SNSs in a machine readable format (i.e., RDF document). Currently, we are working on ways by which the proposed RDF Schema can be extended. We plan to extend our system so that it can deal with social activities data from different SNSs than Facebook and Twitter.

## References

1. Zhou, B., Wu, C.: Semantic Model for Social Networking Federation. *Advances in Information Sciences & Service Sciences* 3(11), 1 (2011)
2. Chao, W., Guo, Y., Zhou, B.: Social networking federation: A position paper. *Computers & Electrical Engineering* 38(2), 306–329 (2012)
3. RDFizers (2008), <http://simile.mit.edu/wiki/RDFizers>
4. Horowitz, D., Kamvar, S.D.: The Anatomy of a Large-Scale Social Search Engine. In: *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, USA, April 26-30*, pp. 431–440 (2010)
5. Rowe, M., Ciravegna, F.: Ciravegna, Getting to Me Exporting Semantic Social Network Information from Facebook, *Social Data on the Web Workshop*. In: *Proceedings of the ISWC 2008 Workshop on Social Data on the Web (SDoW 2008)*, Karlsruhe, Germany (October 27, 2008)
6. FOAF (2009), <http://www.foaf-project.org/>
7. SIOC, <http://sioc-project.org/> (October 24, 2006)
8. Berrueta, D., et al.: SIOC Specification (March 25, 2010), <http://rdfs.org/sioc/spec/>
9. Flickr Wrapper (2007), <http://www4.wiwiss.fu-berlin.de/flickrwrapp/>
10. Flickr, <http://www.flickr.com/>

11. Schaefer, C.: Motivations and Usage Patterns on Social Network Sites. In: 16th European Conference on Information Systems, ECIS 2008, Paper 143, Galway, Ireland (2008)
12. Hargittai, E.: Whose Space? Differences Among Users and Non-Users of Social Network-Sites. *Journal of Computer-Mediated Communication* 13(1), article 14, 276–297 (2007)
13. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: *Proceedings of the 19th International World Wide Web (WWW) Conference*, Raleigh, USA, April 26-30, pp. 591–600 (2010)
14. Dublin Core, <http://dublincore.org/>
15. Hammer-Lahav, E.: Oauth community, OAuth (September 5, 2007), <http://oauth.net/>
16. Twitter developers (January 24, 2011), <http://dev.twitter.com/>
17. Facebook developers (August 2006), <http://developers.facebook.com/>
18. Node.js (2009), <http://nodejs.org/>
19. Express, <http://expressjs.com/>
20. Facebook-js, <https://github.com/masyllum/facebook-js>
21. Twitter-js, <https://github.com/masyllum/twitter-js>
22. Facebook open graph, <http://developers.facebook.com/docs/opengraph/>
23. Facebook query language (FQL), <http://developers.facebook.com/docs/reference/fql/>
24. Twitter REST API, <https://dev.twitter.com/docs/api>
25. Patriquin, A.: Connecting the Social Graph: Member Overlap at OpenSocial and Facebook (2007), <http://blog.compete.com/2007/11/12/connecting-the-social-graph-member-overlap-at-opensocial-and-facebook/>

# A High Performance and Bandwidth Efficient IDMA Scheme with Large Receiver MIMO Technologies

Pararajasingam Niroopan<sup>\*</sup>, Kasun Bandara, and Yeon-ho Chung

Department of Information and Communications Engineering,  
Pukyong National University, Busan, Korea  
{niroopan86,kassae6}@gmail.com, yhchung@pknu.ac.kr

**Abstract.** Interleave Division Multiple access (IDMA) is a new multiple access scheme, and Multiple-Input Multiple-Output (MIMO) is an emerging technology in the wireless broadband communications. In this paper, we investigate the performance of the IDMA scheme with Large Receiver MIMO (LR-MIMO) with different antenna configurations for the multiuser detection. The bit error rate (BER) performance of the proposed system is analyzed for different channel conditions. Numerical results show that the LR-MIMO-IDMA scheme provides superior performance with the aid of MIMO detection techniques over frequency selective fading channel. Also, this scheme further enhances bandwidth efficiency by reducing the repetition length of the IDMA scheme.

**Keywords:** IDMA, Interleaver, MMIMO, Multiple detection.

## 1 Introduction

A new multiple access scheme called Interleave Division Multiple Access (IDMA), which is a special case of CDMA, was recently proposed as a spread spectrum multiple access scheme [1], [2]. In an IDMA scheme, users are distinguished by different chip-level interleaving methods instead of by different signatures as in a conventional CDMA scheme. A special benefit of IDMA is that it allows a very simple chip-by-chip iterative multiuser detection strategy and per user cost of this algorithm is independent of the number of users. Moreover, the advantages of combining forward error correction (FEC) code and spreading lead to improved cellular performance and also increase the throughput.

However, the performance of the IDMA scheme suffers over frequency selective fading channel, although it gives better performance over flat fading channel. Thereby, the IDMA scheme requires additional compensation methods to combat the multipath fading effect. Li Ping et al.[3] proposed the rake Gaussian (RG) approach to the conventional IDMA scheme for a multipath channel.

Multiple-Input Multiple-Output (MIMO) system has proved in the recent past to provide very high capacity without any increase in the transmission bandwidth and power [4]-[6]. The information theoretic capacity of these MIMO channels was

shown to grow linearly with smaller numbers of transmitter and receiver antennas in the scatters-rich environment. Therefore, combining this large receiver MIMO (LR-MIMO) architecture with the IDMA scheme can result in superior performance with bandwidth efficiency, spatial multiplexing and ISI reduction in high data rate transmission.

In this paper, we focus on an improvement of the conventional IDMA scheme with the aid of LR-MIMO detection technique. In Section 2, we present a brief introduction of transceiver structure in the IDMA scheme. Then, the LR-MIMO-IDMA design is described in Section 3. In Section 4, numerical simulation results are presented and Section 5 concludes with remarks.

## 2 IDMA Scheme

The transmitter and receiver structures of an IDMA scheme with  $K$ -simultaneous users are shown in Fig.1. At the transmitter, the block size of  $N$ -length information bits from each user- $k$  is denoted as  $d_k = [d_k(0), \dots, d_k(N-1)]^T$ ,  $k=1, 2, \dots, K$ . The data sequence is encoded using a convolutional code into  $b_k = [b_k(0), \dots, b_k(N_C-1)]^T$ . That is, the code rate is defined as  $R_1 = N/N_C$ . Then each bit of  $b_k$  is again encoded using a low rate code such as a spread encoder with a rate of  $R_2 = 1/S_k$ , where  $S_k$  is a spreading factor. Thus, the overall code rate is  $R_1 R_2$ , which produces a chip signal. The second encoder output is fed into the user specific interleaver ( $\pi_1, \pi_2, \dots, \pi_K$ ) for user separation, which generates  $x_k(j)$ ,  $j=1, 2, \dots, J$ , where  $J$  is the user frame length. The resultant signal is then transmitted through the multiple access channel. In the receiver, the received signal is given by

$$r(j) = \sum_{k=1}^K h_k x_k(j) + n(j), \quad j=1, 2, \dots, J \quad (1)$$

where  $h_k$  is the channel gain for user- $k$ ,  $x_k$  is the corresponding transmitted signal and  $n$  is the additive white Gaussian noise (AWGN) process with zero mean and variance,  $\sigma^2 = N_0/2$ . It is assumed that the channel coefficients  $\{h_k\}$  are known a priori at the receiver.

This received signal is passed to a multi-user detection (MUD) receiver that consists of an elementary signal estimator (ESE) and  $K$  a posteriori probability (APP) decoders (DECs), one for each user. The ESE performs chip-by-chip detection to roughly remove the interference among users. The outputs of the ESE and DECs are extrinsic log-likelihood ratios (LLRs) about  $\{x_k\}$  defined as

$$e(x_k(j)) = \log \left[ \frac{p(y | x_k(j) = +1)}{p(y | x_k(j) = -1)} \right] \quad (2)$$

Those LLRs are further distinguished by  $e_{ESE}(x_k(j))$  and  $e_{DEC}(x_k(j))$ , depending on whether they are generated by the ESE or DECs. For the ESE section,  $y$  in (2) denotes the received channel output while for the DECs,  $y$  in (2) is formed by the deinterleaved version of the outputs of the ESE block. These results are then combined using a turbo-type iterative process for a pre-defined number of iterations. Finally the DECs produce hard decisions on information bits for each user.

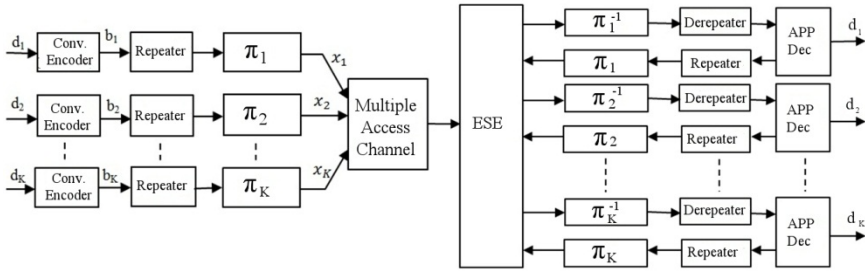


Fig. 1. The transmitter and receiver structure of the IDMA scheme

### 3 Large Receiver MIMO-IDMA Scheme (LR-MIMO-IDMA)

The proposed LR-MIMO-IDMA transmitter extends the IDMA transmitter with the multiuser. As shown in Fig.2, the information bits from each user data is coded by convolutional code and a repetition code. This coded data are interleaved by a user-specific interleaver. Next, the data are encoded using orthogonal space-time block code (OSTBC). Finally, the resultant signal is transmitted through the frequency selected fading channel.

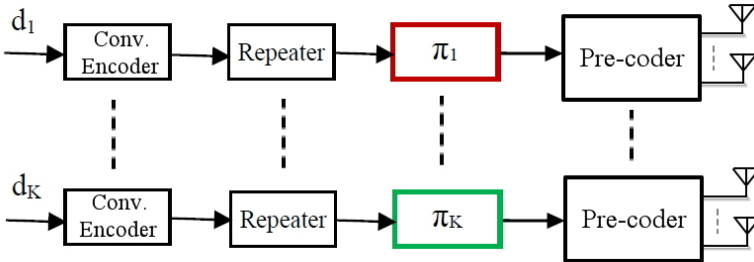


Fig. 2. The LR-MIMO-IDMA transmitter

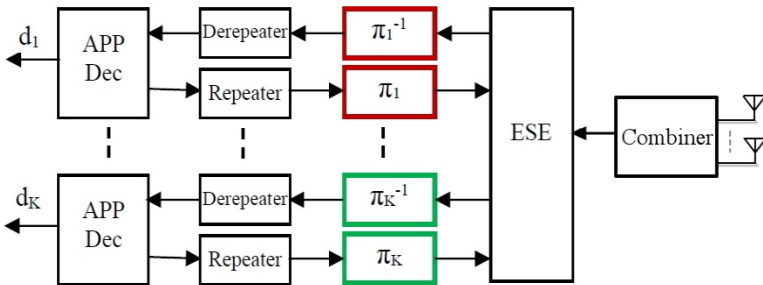


Fig. 3. The LR-MIMO-IDMA receiver



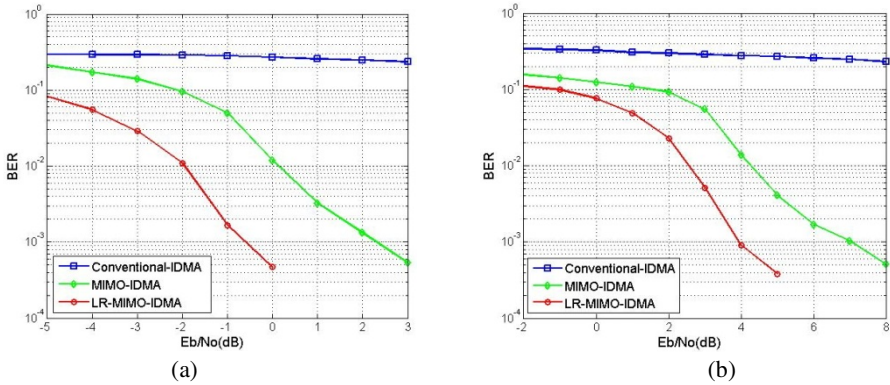
The receiver structure of the LR-MIMO-IDMA is shown in Fig.3. In the receiver, the OSTBC combiner combines the received signal from all receive antennas with the channel estimates in order to extract the soft information of the symbols encoded by the OSTBC encoder. For multi-user separation, the output data of OSTBC combiner is fed into iterative turbo-type process in the same way as the conventional IDMA receiver. Using the ESE and APP decoders, multi-user data are separated.

### 4 Simulation Results

We analyze the bit-error rate (BER) performances of the LR-MIMO-IDMA scheme, MIMO-IDMA scheme and the conventional IDMA scheme over frequency selective fading channel. It is assumed that all users in the system use the same energy level. The parameters used in the simulation are shown in Table 1.

**Table 1.** Parameters used in the simulation over frequency selective fading channel

Parameters	Specifications
No. of users (K)	5, 10
Tx. Antennas	3
Rx. Antennas	3, 8
Data length	1024 bits
Encoder	Convolutional code (23, 35) <sub>8</sub>
Repetition (S)	4, 8
Modulation	QPSK
Iteration	10
Interleaver	Random interleaver



**Fig. 4.** Performance comparison with S=4 in the Indoor Office (a) K = 5 (b) K = 10

The channel model to be used for simulation is an indoor office with 6 paths [8]. The delay spread of the channel is found to be 100 ns. The repetition length is 4. Fig.4 shows the BER performance of LR-MIMO-IDMA, MIMO-IDMA and conventional IDMA scheme with 5 and 10 simultaneous users, respectively. It is shown that the conventional IDMA scheme is significantly affected by frequency selective fading channel. Compared with the MIMO-IDMA scheme, LR-MIMO-IDMA scheme gives better BER performance in the indoor multipath environments.

We further make an evaluation over an outdoor channel model that is urban/suburban low-rise with 6 paths [8]. The delay spread of the channel is found to be 750 ns. Fig.5 shows the BER performance of the systems over the outdoor multipath environment with 5 and 10 simultaneous users. The LR-MIMO-IDMA scheme gives the significant improvement over the conventional IDMA scheme and the MIMO-IDMA scheme.

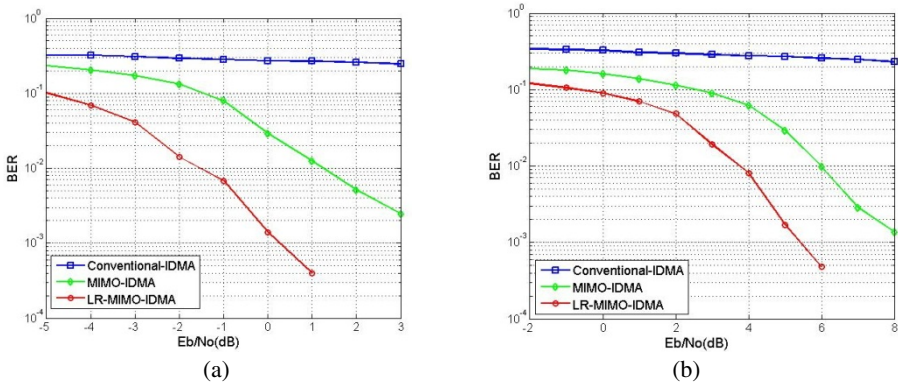


Fig. 5. Performance comparison with S=4 in the Outdoor Urban/Suburban Low-Rise (a) K = 5 (b) K = 10

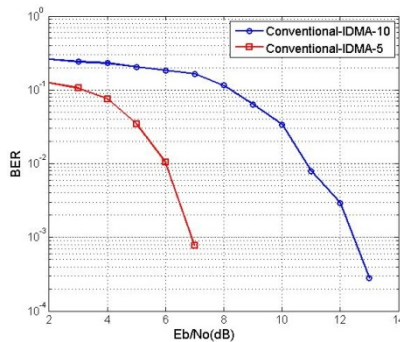


Fig. 6. Performance of the conventional IDMA system with K=5, 10 and S=8

It is clear that the conventional IDMA fails to provide acceptable performance over frequency selective fading channel. To compensate this effect, the receiver of the conventional IDMA system requires equalizer or rake receiver. However, the LR-MIMO-IDMA system uses the orthogonal space time block coding with perfect channel estimation in the receiver. This lends itself to the improvement for the reliability of data transmission.

Further, a comparison has been made in terms of bandwidth efficiency. We have conducted a performance evaluation of the conventional IDMA scheme with a repetition length of 8 over flat fading channel. Fig.6 shows the performance for 5 and 10 simultaneous users. Even with the repetition length of 8 (double the repetition length) and also a more friendly channel model of flat fading, the conventional IDMA scheme shows poorer BER performance compared with the LR-MIMO-IDMA scheme. Therefore, the proposed LR-MIMO-IDMA scheme outperforms the conventional IDMA as well as MIMO-IDMA and also offers reduced length of the repetition code, thereby significantly enhancing bandwidth efficiency.

## 5 Conclusion

To address the poor performance of the IDMA over frequency selective fading channel, we have proposed the LR-MIMO-IDMA for significant performance improvement under various channel models. The performance evaluation shows that the LR-MIMO-IDMA scheme outperforms the conventional IDMA and MIMO-IDMA under the indoor and outdoor environment. Moreover, the proposed system greatly enhances bandwidth efficiency by reducing the repetition length of the IDMA scheme. Therefore, the proposed LR-MIMO-IDMA can be considered as a good candidate for future high rate high capacity multiuser mobile radio systems.

**Acknowledgement.** This work was supported by a Research Grant of Pukyong National University (2013 year).

## References

1. Ping, L., Liu, L., Wu, K.Y., Leung, W.K.: Interleave-division multiple-access. *IEEE Transactions on Wireless Communications* 5, 938–947 (2006)
2. Lee, J.S., Miller, L.E.: *CDMA system engineering handbook*. Artech House Publishers, Boston (1998)
3. Ping, L., Liu, L., Wu, K.Y., Leung, W.K.: A simple approach to near-optimal multiuser detection: interleave-division multiple-access. In: *Wireless Communications and Networking*, vol. 1, pp. 391–396. IEEE (2003)
4. Zheng, L., Tse, D.N.C.: Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels. *IEEE Transactions on Information Theory* 49(5), 1073–1096 (2003)
5. Yu, Q., Lambotaran, S.: Iterative (Turbo) Estimation and Detection Techniques for Frequency Selective Channels with Multiple Frequency Offsets in MIMO System. In: *IEEE 65th Vehicular Technology Conference*, pp. 2015–2018. IEEE (2007)

6. Vakilian, V., Frigon, J., Roy, S.: Performance Evaluation of Reconfigurable MIMO Systems in Spatially Correlated Frequency-Selective Fading Channels. In: IEEE Vehicular Technology Conference, pp. 1–5. IEEE (2012)
7. Rusek, F., Persson, D., Lau, B.K., Larsson, E.G., Marzetta, T.L., Edfors, O., Tufvesson, F.: Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays. IEEE Signal Processing Magazine 30(1), 40–60 (2013)
8. Technical Report of RF Channel Characterization and System Deployment Modeling JTC(AIR) (1994)

# Security Considerations for Smart Phone Smishing Attacks

Anna Kang<sup>1</sup>, Jae Dong Lee<sup>1</sup>, Won Min Kang<sup>2</sup>,  
Leonard Barolli<sup>3</sup>, and Jong Hyuk Park<sup>2,\*</sup>

<sup>1</sup> Department of Multimedia Engineering, Dongguk University, Korea

<sup>2</sup> Department of Computer Science and Engineering,

Seoul National University of Science and Technology, Korea

<sup>3</sup> Department of Information and Communication Engineering,

Fukuoka Institute of Technology, Japan

anakang37@gmail.com,

{jdllee731, wkaghdsk0, jhpark1}@seoultech.ac.kr,

barolli@fit.ac.jp

**Abstract.** Recently, phishing, a new type of crime, has increased due to the expansion of the use of smart phones, and financial losses have been increasingly reported. As damage cases, which have become sophisticated through the use of smart phones, crimes occurring in relation to smart phones are becoming a concern. In this paper, we discuss crimes such as phishing, voice phishing, and smishing, which can occur in a smart phone environment. In particular, we discuss the issue of smishing attacks, as well as the problems and preventive measures of these attacks.

**Keywords:** Smart phone security, Phishing, SMS, Smishing attack.

## 1 Introduction

Since the 2000s, information and communication related crimes have increased due to the spread of mobile phones. There has been a recent increase in crimes using smart phones, which are widely used, and a crime called phishing, which has become more organized and intelligent, has brought a great amount of financial damage to a large number of people, and has even led to the suicide of its victims, creating a serious social issue [1].

Due to technological advancements, banking transaction modes have changed greatly. Rather than visiting a physical bank directly, more and more customers are now using computers or mobile phones, such as smart phones, to access banking services. However, transactions in a smart phone environment are inherently vulnerable to financial fraud because the identification process is performed remotely. In practice, it is difficult for a party other than the account holder to make a transaction offline, whereas online transactions can be completed by anyone who has

---

\* Corresponding author.

the necessary digital certificate. Pharming or smishing, which has become an issue in Korea recently, exploits the difficulty in identifying account holders during electronic financial transactions [2].

This paper is organized as follows: In chapter 2, we will discuss related works about phishing and smishing attacks that have occurred recently. In chapter 3, we discuss security considerations against smishing attacks and a detailed operation process for the measure in a smart phone environment. Finally, in chapter 4, we finish with a conclusion.

## **2 Related Works**

### **2.1 Smishing Attacks**

Phishing is a cyber-fraud crime that obtains personal certificate numbers, credit card numbers, or bank account numbers illegally by masquerading as a trustworthy entity, such as a web site or email from a financial institution, in electronic communication. Phishing can lead to financial damages, and its attack method changes continuously so that it is difficult to counterattack [3, 4].

Recently, phishing has been combined with Trojans, worms, and viruses to take advantage of the vulnerability of computers, making it much more sophisticated. Trojans are received via an attachment file in an email using social engineering techniques and run on a user's computer to collect the user's IDs or passwords or to transfer computer monitoring information to other sites. Recently, techniques have been created to make a path to a phishing site when a user connects to a preferred site. This is done after a user host file is modified to re-set up a transfer path using a worm and virus, and the ID and password of a user who was connected to a trusted site is collected and transferred to an attacker [5, 6].

There is a new type of phishing technique that steals the personal information of a user using the short message service (SMS) of a mobile phone. Smishing (SMS + phishing), which was named by McAfee, an Internet security company, occurs when a web site link is sent via text messaging to a smart phone user. When the user connects to the web site, a Trojan virus is inserted on the smart phone, allowing it to be controlled by a third party.

Fig. 1 shows the process of smishing, which is an important issue due not only to its more advanced and intelligent spy features, but also its more systematic and intelligent type of spread compared to existing methods, which strengthen the survival of the malicious application.

### **2.2 Types of Smishing Attacks**

In this paper, the types of smishing attacks can be considered as two categories:

The first type is when an attacker sends an SMS text message that includes information regarding purchase details, exchanges, refunds, or cancellations to deceive the user; it also includes the phone number of the attacker. Then, the smart

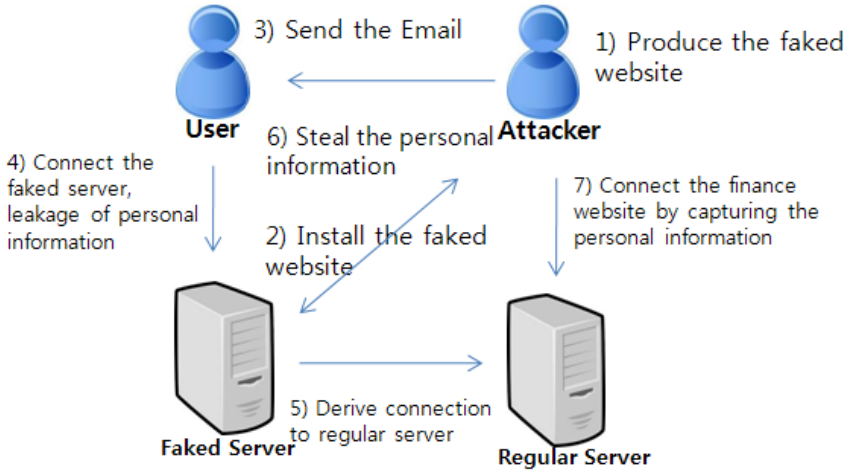


Fig. 1. Phishing generation flow diagram

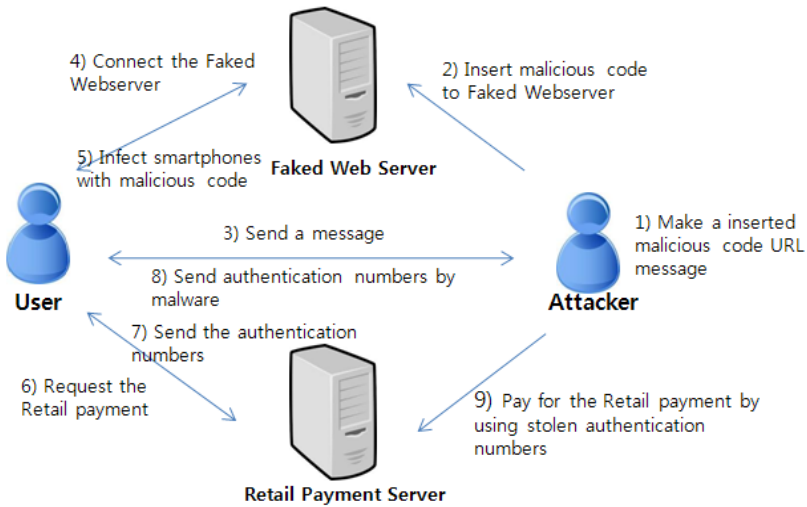


Fig. 2. A basic flow of smishing attack

phone user calls the number in the text message and the attacker requests the user’s personal information or authorization code required for online micropayment. The personal information or payment information of a victim can be stolen using a phone conversation between victim and attacker [7].

The second type is when an attacker inserts a malicious code on his/her masqueraded web site and sends a deceptive text message to a user. Once the user checks the message and connects to the web site with the malicious code, the user’s smart phone is infected by that malicious code. When a user makes a micropayment, the malicious code intercepts the authorization code received from the authorized

server; this number is then sent to the attacker. Then, the attacker can make a micropayment using the verified authorization code via the user's smart phone. This malicious code plays a role in intercepting authorization codes transmitted from a micropayment system. An attacker sends a deceptive SMS text message that attracts a user's interest without raising any suspicions. The text message includes a URL in which a malicious code is running. After the user clicks the URL, the user's smart phone is infected by the malicious code [8]. Fig. 2 shows a basic flow of smishing attacks.

### 3 Security Considerations for Smishing Attacks on Smart Phones

In this chapter, we discuss security considerations for smishing attacks on smart phones. First, we discuss URL validation tests that use smart phone applications as well as the management of downloaded applications using the Smishing box. In addition, we explain the operation process in detail as a response to the aforementioned smishing attacks.

#### 3.1 URL Validation Test Using Smart Phone Applications

Smart phones are highly vulnerable to smishing attacks because a malicious code can be inserted on a user's smart phone so that the user's key log or personal information can be exposed. The URL validation test using a mobile application can inspect the validity of the URL included in the message and will inform the user of the validation results after the user clicks a URL in a text message. Due to its attack method, smishing is not a type of direct hacking on a user's smart phone, in that it does not insert malicious codes or programs; thus, this test performs a comparison and inspection of the URL received in a text message using a smart phone application. Fig. 3 shows the operation process of the URL validation test application.

The operation process for the URL validation test is performed with the following five steps:

- STEP 1** An attacker sends a text message that includes a URL to a user.
- STEP 2** The application checks whether the text message includes a URL, and if a URL is not included, it moves the message to a message history box.
- STEP 3** The application checks whether the text message includes a URL that is clicked by the user, and if it is not clicked, it moves the message to a message history box.
- STEP 4** If the user clicks the URL, the application inspects the URL by comparing it with authorized URLs stored in a database.
- STEP 5** The application determines the validity of the compared URL. If the URL is not validated, the application deletes the URL and sends an alert message to a user.



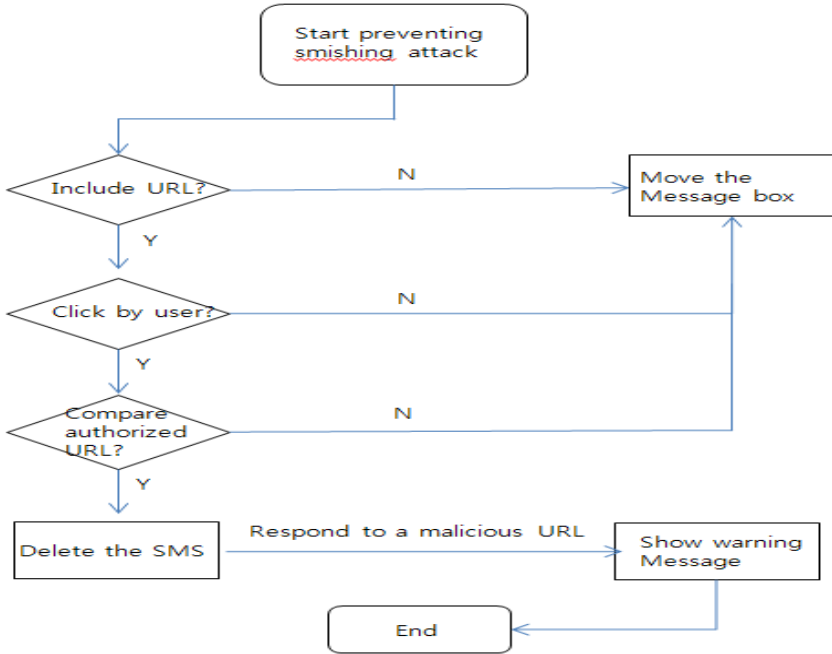


Fig. 3. Operation process of the URL validation test application

### 3.2 Management of Downloaded Applications Using the Smishing Box

The smishing box manages downloaded applications on a smart phone by creating a small box. That is, the downloaded applications exist within the smishing box and all actions or data that occur in the box do not affect the smart phone in terms of data security. Fig. 4 shows the operation process of the smishing box.

The operation process of the smishing box is performed with the following five steps.

- STEP 1** An attacker sends a text message that includes a URL to a user.
- STEP 2** The application checks whether the text message includes a URL, and if a URL is not included, it moves the message to a message history box.
- STEP 3** The application checks whether the text message includes a URL that is clicked by the user, and if it is not clicked, it moves the message to a message history box.
- STEP 4** If the user clicks the URL and a download begins and completes, the completed downloaded application is moved to the smishing box.
- STEP 5** The smishing box allows the installation of only applications registered in the Android market or App store and removes any unauthorized applications.

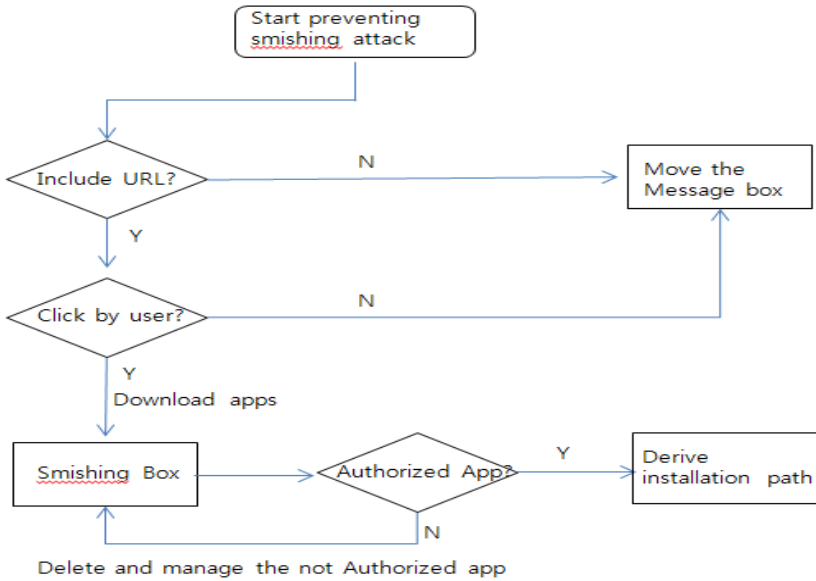


Fig. 4. Operation process of the smishing box

Since the smishing box is run as an independent application, even in cases where installation of applications is done automatically, personal information, such as a contact list or personal albums, will not be violated.

#### 4 Conclusion

As individual payment means by transactions in cyberspace have increased, crimes such as phishing, pharming, and smishing have also increased. In this paper, we discussed security considerations for smart phone smishing attacks.

In particular, we discussed smishing attacks and provided measures to prevent these types of attacks, and explained the operation processes in detail.

In the future, additional studies are required to analyze Korea-specific smishing attack types and their problems as well as measures that use the Smishing Box and the effective use of resources in smart phones.

**Acknowledgments.** This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center)) support program (NIPA-2013-H0301-13-4007) supervised by the NIPA(National IT Industry Promotion Agency). This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0024052).

## References

1. Kang, J.-Y., Yoon, J., Kim, Y.: Phishing/Pharming Examples and Countermeasure Analysis. The Korean Institute of Information Scientists and Engineers 12(2), 171–180 (2013)
2. Kim, K.-Y., Kang, D.-H.: Smartphone security technology in an open mobile environment. Korea Institute of Information Security and Cryptology 19(5), 21–28 (2009)
3. Kim, J., Maeng, Y., Nyang, D., Lee, K.: Cognitive Approach to Anti-Phishing and Anti-Pharming. Korea Institute of Information Security and Cryptology 19(1) (2009)
4. Kim, T.-H., Lee, J.-H., Lee, D.-H.: Study on Mobile OTP(One Time Password) Mechanism based PKI for Preventing Phishing Attacks and Improving Availability. Korea Institute of Information Security and Cryptology 21(1), 15–26 (2011)
5. Kim, S.-R.: Phishing of the dual criminality. IT and Legal Research 4, 251–290 (2010)
6. Ryu, G.-A., Kim, K.-I., Jeon, W., Lee, Y.: An Anti-Phishing Method using the Client's IP Address. Korean Society for Internet Information 13(2), 21–22 (2012)
7. Kang, S., Kim, S.: Study on the personal Information Retrieval of Smartphone Messenger Service. Korea Institute of Information Security and Cryptology 23(1), 97–107 (2013)
8. KISA Internet Incidents Analysis Division, Code Analysis Team, Spreading smishing malicious apps will become intelligent Increasingly, KISA (March 2013)

# Technology Venture Startup Invigoration Strategy for Building Infrastructures for the Business Startup Ecosystem

Hye-Sun Kim<sup>1</sup>, Yunho Lee<sup>2,\*</sup>, and Hyoung-Ro Kim<sup>1</sup>

<sup>1</sup> Dept. of Business Incubation Agency, Induk University,  
12 Choansan-ro, Nowon-gu Seoul, 139-749, Korea

<sup>2</sup> Dept. of Social Welfare, Kyung Hee Cyber University,  
1 Hoegi-Dong, dongdaemun-Gu, Seoul, 130-701, Korea  
daisyhsun@hanmail.net, anne6@khcu.ac.kr,  
hrkim@induk.ac.kr

**Abstract.** The people who wish to establish a business should have excellent relevant knowledge, technology, and entrepreneurship skills, and should be able to address the issue of job creation accompanied by employment. Startup companies accounted for about 6.6% of the country's gross domestic product (GDP) in 2009 and 9.8% of the total employment nationwide, according to Korea Institute of Startup & Entrepreneurship Development, and are highly likely to create job opportunities. Technology venture startups apply the research outcomes from enterprises, universities, research institutes, etc. to real businesses and can provide new products through technology innovations, can create new markets, and can vitalize regional economies. As such, it is necessary to conduct continuous studies, pursue policy development, and make proposals on the technology venture startup ecosystem.

**Keywords:** Startup, startup ecosystem, technology venture, startup environment, technology venture invigoration strategy.

## 1 Introduction

The Ministry of Science, ICT, & Future Planning, newly established by the South Korean government, announced that key to achieving a creative economy is business startups, and emphasized that the key subjects for realizing a creative economy are the construction of a business startup ecosystem where even students, housewives, etc. could establish a business. The ministry added that a virtuous circulation structure established by creating new jobs through business startups and rearing businesses through the enterprise welfare expansion policy will become a foothold for national economic growth. In this regard, the government is interested in Israel, an advanced country in terms of cultivating venture companies, and shows great concern about the policy of Technion Israel Institute of Technology, the cradle of venture companies

---

\* Corresponding author.

[1, 2]. Israel boasts the establishment of some 5,000 venture companies over the last 15 years, over 100 of which have been listed on the NASDAQ. Moreover, the USA has 30-fold more medicine universities as Israel does, but Israel has the biggest number of business startups in the bio healthcare sector (about 40%) in the whole world.

Business startups, therefore, are very important as they can serve as the foundation and driving force of national competitiveness for creating a new, 21<sup>st</sup>-century economy. In an environment where the industrial paradigms are rapidly changing, the utilization of external innovation through business ventures is considered integral and can contribute to addressing the unemployment problems of the highly educated as it can contribute to economic growth and high-quality job creation.

Therefore, in this paper, for continuous policy development and proposals on the technology venture startup ecosystem, the related policies are investigated in Chapter 2, theoretical considerations are studied in Chapter 3, and a conclusion is formulated in Chapter 4 through strategy proposals for utilizing technology venture startups.

## 2 Related Policies

The Ministry of Science, ICT, & Future Planning recently announced that universities can provide education for business startups; can expand the technology-holding companies that provide tailored support to business founders; can utilize business startups by collecting ideas, cultivating talented individuals, commercializing technologies presented by universities, linking the industry-academy-research sectors, creating new large-scale markets and demands, and reinforcing protection measures and financial support for intellectual properties; and can support the growth of venture and small- and medium-sized companies. In addition, the ministry plans to install an “infinite imagination room,” which collects various ideas from the citizens and links them with researches and business startups at two places in all cities, districts, and boroughs by the end of 2017.

The ministry maintains its thrust of reinforcing thorough education and system maintenance in the primary and middle schools for technology venture startups. In addition, the ministry has a plan to cultivate a high-quality workforce in strategic ICT sectors, such as software (SW) and big data, and to develop special workforces to meet the private demands by nominating/supporting 40 information communication academies led by the private sector by 2017.

Small & Medium Business Administration has declared that it intends to build a merger and acquisition (M&A) network that mediates merging and acquisitions among small- and medium-sized companies, and is promoting a plan to extend the benefits for small- and medium-sized companies from three to ten years. In addition, the administration is studying the introduction of a cloud funding system by revising the Support for Small and Medium Enterprise Establishment Act.

In particular, the culture and intellectual knowledge service industry sector expects utilization effects in funding and investments. Besides, through the introduction of Israel’s venture investment system described in the beginning of this paper, the

administration is making progress in establishing a system that will link the government with and will support matching-type research and development (R&D) funds for companies with venture capital investments as establishment and operation funds.

### 3 Theoretical Considerations

#### 3.1 Business Startup Ecosystem

The term *ecosystem* refers to a combination of organisms interacting with one another and its surrounding inorganic environments that affect such organisms. The business startup ecosystem in technology venture establishment is used with the concept of a virtuous circulation ecosystem, in which a virtuous cycle is created through associations with various organizations and institutions in the industrial, academic, and research fields that enable business startups, growth and accumulation, and interactions of knowledge, talented individuals, and funds flow that support such associations. The startup ecosystem of exceptional technology ventures can be largely divided into a period of satisfaction of the ecological conditions, an ecosystem formation period, and an ecosystem establishment period [2].

**Table 1.** Ecological Stages of a Technology Venture Startup

Ecological Stage	Conditions	Induction of a Business Startup
Satisfaction Period	<ul style="list-style-type: none"> <li>✓ Accumulating a certain size of technology/manpower (necessary condition)</li> <li>✓ Maintaining a support system for a technology venture starter (sufficient condition)</li> </ul>	<ul style="list-style-type: none"> <li>✓ Helping an expected business starter with technologies at companies, universities, institutions, etc. to become a venture starter</li> </ul>
Formation Period	<ul style="list-style-type: none"> <li>✓ Converting to a large-scale technology venture startup and initiating growth through a combination of the necessary and sufficient conditions</li> </ul>	<ul style="list-style-type: none"> <li>✓ Increasing technology venture startups by supplying funds, manpower, etc.</li> <li>✓ Supporting information, facilities, etc. to nurture technology venture companies</li> <li>✓ Continuously inducing technology venture startups through public relations, etc.</li> </ul>
Establishment Period	<ul style="list-style-type: none"> <li>✓ Building a new high-tech industry through the accumulation of exceptional technology companies</li> </ul>	<ul style="list-style-type: none"> <li>✓ A large-scale technology venture startup is completed.</li> <li>✓ A connective cooperation system with successful companies is established, and a virtuous circulation ecosystem of technology-based venture startups is built.</li> </ul>

Source: Shin Chang-Ho and Kim Mook-Han (2012), *A Plan for Building a Virtuous Circulation Ecosystem of Technology Venture Startups in Seoul*

### 3.2 Technology Venture Startup

#### (1) Definition

The term *technology venture startup* refers to the establishment of a company that creates innovative technologies and knowledge, and is differently defined by institutions as the common standard for defining a technology venture group is practically ambiguous [1, 6]. Korea Technology Finance Corporation (KTFC) and Korea Institute of Startup & Entrepreneurship Development (KISED) have selected the cutting-edge high-technology business line as the key business line for technology venture startups, define the establishment of a business by the relevant industry as the real technology venture startup, and investigate the trends of technology venture business establishment using venture companies and other companies that have obtained technology innovation certification as an InnoBiz company as a surrogate variable of technology venture startups.

#### (2) Scope

The scope of the technology venture startup includes the establishment of a business involving not only cutting-edge, high-tech technologies but also knowledge as the motive power in a broad sense. KTFC cites the manufacturing, specialized service, and knowledge culture business types as the companies that conduct production and product/service sales activities with new technologies or ideas, and that can become middle-standing enterprises characterized by high-risk, high-revenue, and high-speed growth.

**Table 2.** Scope of the Technology Venture Startup

Classification	Business Line
Cutting-Edge Technology Business	Computer and office appliances, electronic parts and components, image/sound/communication equipment, medical/precision/optical appliances and clocks, aircraft and space shuttle parts, pharmaceutical products
High-Tech Business	Chemical compounds and products, other machineries and equipment, electricity conversion equipment, automobiles and trailers, other transportation equipment

Source: Korea Institute of Startup & Entrepreneurship Development (KISED)

Korea Research Institute for Vocational Education & Training (KRIVET, 2011) limits the knowledge technology venture startup to the contents and software of the IT fusion sector, manufacturing fusion sector, and knowledge-based service business in the middle classification of Korean Standard Industrial Classification Method, and excludes the livelihood-type business and the traditional manufacturing sector. KISED (2011) classifies five cutting-edge technology businesses and five high-tech businesses in accordance with the technological levels and regards them as the scope of the technology venture setup [1].

## 4 Conclusion: Proposed Invigoration Strategy for Technology Venture Startup

Recently, the government expressed concerns about the construction of a virtuous circulation structure in the venture business startup ecosystem, and announced that it would pour its efforts into improving the systems and building a startup ecosystem so that businesses can be started without government support in the long term.

In an effort to invigorate technology venture startups by building infrastructures for the business startup ecosystem investigated earlier, this paper proposes the following:

(1) that systematic technology startup education be offered for nurturing and developing the manpower needed to build the startup infrastructures

The designing of a business-startup-friendly education system will help improve the interested parties' understanding of business startups. Education on business startups can be offered in primary, middle, and high school, under the slogan "Business startup is a part of education."

There is a need to enable the recognition of business establishments and professions from a young age, to reinforce the education on entrepreneurship, and to consider a plan for R&BD participating companies to obligatorily complete the business startup education.

(2) that the operation of the government-driven business incubator be converted and extended to the private-centered operation system

Through consignments, the incubating system should provide a wide range of support for selecting business startup items, research and development targets, manufacturing trial products, commercializing, developing sales routes, management, managing human resources and others, building infrastructures, and inducing business startups specialized by region.

(3) that the importance of intellectual property rights be recognized, and that education be reinforced through the related business practitioners, including business founders, potential business founders, experts, universities, etc.

Lawsuits not only on leakages of cutting-edge technologies and infringements of others' rights through the globalized free trade agreements (FTAs), such as the recent legal battle on patents and designs between Apple and Samsung, but also on violations of the trademarks and designs of small and medium-sized companies, individuals, etc. are on the rise day by day. As more and more people are suffering in such cases, many of whom do not know their relevant rights, education on how to respond to such situations needs to be offered. As South Korea, however, restricts the individual possession of intellectual property rights to university professors, few research institutes are established. Through the Office of University-Industry Cooperation's revision of the provisions for the across-the-board possession of intellectual property rights, the system can be improved so that the researchers' intellectual property rights can be upheld.

To achieve such, there is a need to create a globalized competition landscape by maintaining the intellectual property rights system and establishing acquisition strategies for the intellectual property rights of the expected business founders;



construct environments where people can receive compensation for their various ideas, technologies, and designs; and build a sustainable business startup culture by fostering the trade and investment environments where ideas can be sold and bought.

(4) that the support funds be converted from loan-centered to investment-centered to systematically reduce the risk factors of credit rating deterioration if the business fails. In addition, the investment mentoring system should be introduced, and the mentors should be continuously educated and managed.

Further, the clouding funding scheme promoted by the government should be settled so that the investments can be vitalized by the private sectors by expanding the tax deduction benefits for the angel investments, or extending the period. If the investments were made by the mother companies or by large companies, the tax deductions should be increased to return the benefits to the investing companies. In this case, an investment mentoring system should be introduced to successfully lead business startups through systematic mentoring by a group of experts rather than having such startups solve all their problems by themselves, and the mentors should manage experts through DB management and continuous education. Innopoli in Finland introduced specialized mentoring management by experts, and as a result, 90% of its venture companies have survived.

(5) that M&As be vitalized through the acquisition of venture companies by large companies

The large ICT companies in the USA are particularly aggressive in the acquisition of venture companies, and a virtuous circulation system is naturally formed in the order of business startup-growth-sale-reestablishment. In South Korea, however, large companies are inactive in terms of venture company acquisition, and a social environment where the initial founders of venture companies are not well treated exists. Therefore, encouragement is necessary to establish the growth-sale-reestablishment system. A numerical investigation showed that Google has taken over 69 companies from 2006 to 2011, Microsoft 52 companies, and Samsung only 17 companies.

(6) that a cultural maintenance scheme or system be provided to turn failures into assets so as to exhibit creativity for the reinforcement of the venture ecosystem

The personal guaranty responsibility of the representative director should be relaxed for the exploration of another venture by turning a failure experience into a medicine rather than a poison. Further, through such system, the exit barriers should be eased so that such representative directors could retire from the business while minimizing the losses if failure is expected.

(7) that global competitiveness be reinforced as the domestic market is limited in size.

The growth of domestic companies is restricted due to the limitations of the domestic market. Due to their lack of knowledge about export/import, experience, etc., however, companies often fail without attempting to penetrate overseas markets. To overcome such problems, export/import experts should be trained to lend practical assistance to such companies through appropriate human resource expansion in international trade, the social standing of such experts should be upgraded, and the recognition of many talented individuals (i.e., first-class citizens) as global trading

experts should be expanded. Of course, as the cases where companies return to the domestic market after achieving success in the USA or European markets are currently increasing, such is considered definitely possible if the language barrier will be removed.

(8) that a business startup space hub be established for use as a business startup space, where the expected business starter simply brings a notebook for convenience, by utilizing spaces such as a library of a local autonomous entity. The space should be arranged to sufficiently play the role of a preparation space before participating in the incubating system, and a sustainable support system should be introduced to induce a virtuous circulation of business startups by providing a cooperative space, networking opportunities, etc. to people who subscribed at the site or online homepage. The space should include a concept of small investment by the government and/or the private sectors, and should be used as a cooperative, incubating, and multi-purpose space. It will be better, for instance, if additional points can be given to users who borrow business-startup-related books at this place.

## References

1. Shin, C.-H., Kim, M.-H.: A Plan for Building a Virtuous Circulation Ecosystem of Technology Venture Startup in Seoul. The Seoul Institute (2013)
2. Yang, H.-B., Park, J.-B.: Construction of Business Startup Ecosystem for Youth. Korea Institute of Engineering and Technology (2011)
3. Kang, H.-H.: To Build 'Busan Good for Business' by Upgrading Business Startup Ecosystem. Busan Development Institute (2011)
4. Lee, Y.-J.: An Approach for Building Infrastructures for Knowledge Management. Graduate School of Information & Science, Sungsil University (Master) 2 (2006)
5. Lee, S.-G., Kim, Y.-S.: Utilization of Technology Venture Startup through Industry-Academia Cooperation: Around University Technology Business. Industry-Academia Support Center Affiliated with Seoul Development Institute (2010)
6. The Small and Medium Business Administration, Supplement Projects for Utilization of Technology Venture Startup. The Small and Medium Business Administration (Outside Edition) (2008)
7. Daeduk Research & Development Area, Step Up for Utilization of Technology Venture Startup, The Ministry of Science & Technology (2006)
8. Kim, W.-M., Lee, J.-H.: The Role of Universities for Utilization of Technology Concentrated Business Startup. Journal of the Korean Management Association 18(5), Serial 52, 2079–2106 (2005)
9. The Ministry of Science & Technology, Supports for Utilization of Business Startup Incubation of Gwangju Science & Technology Institute (2001)
10. Park, J.-H., et al.: Status of Business Startup Policy and Utilization Plan for Business Startup for Youth at Business Startup Ecosystem View. Korea Association of Business Education 10 (2012)
11. Kim, G.-S.: Implications from Policy Discussion Data: Construction Plan for Venture Ecosystem for Rearing Up the Small and Medium Venture Companies. Office of a Member of Parliament Kim, G.-S. (2012)

12. Park, Y.-R., Jeon, S.-J., Joo, J.-W.: Structural Problems and Improvement Plan of ICT Venture Ecosystem. Gwacheon Information Society Development Institute (2012)
13. Han, J.-H., Shin, J.-G.: Development Process and Implications of Venture Ecosystem in South Korea. Korea Industrial Technology Foundation
14. Rae, D.: Mid-Career Entrepreneurial Learning, pp. 562–574. The Derbyshire Business School, University of Derby, Derby (2005)
15. GERA, The Global Entrepreneurship Monitor, Global Report (2011)

# Smart-Contents Visualization of Vehicle Big Data Using Vehicle Navigation Status Information

Hae-Jong Joo<sup>1</sup>, Suck-Joo Hong<sup>2,\*</sup>, and Dong-Su Park<sup>3</sup>

<sup>1</sup>Dept. of LINC, Dongguk University, 82-1 Pil-dong 2-ga, Jung-gu, Seoul, 100-272, Korea

<sup>2</sup>Dept. of CIEP, Mokwon University, Doanbookro 88, Seo-ki, Daejeon, 302-729, Korea

<sup>3</sup>NOVOSYS, 5 Fl., LIFE Tool Center, 614-33 Guro 2-dong, Guro-gu, Seoul 152-865, Korea

hjjoo@dongguk.edu, sjhong@mokwonu.ac.kr,  
toward21c@empas.com

**Abstract.** The recent technological advances have engrafted information technology (IT) on vehicles, making such vehicles more convenient and safer, and enabling them to meet the consumers' needs. As traffic congestion due to the increased number of vehicles has become a social issue, however, various other connected vehicle technologies for traffic and safety services using the data from the surrounding vehicles are being developed, standard industrial technologies connecting vehicles' infotainment system with mobile devices are under way, and many automobile and IT companies are participating in the promotion of their standardization. This paper discusses an effective method of managing and applying vehicle navigation information by deducing a method of collecting vehicle navigation information and the analysis factors of such information. If such analysis factors and collection technologies of vehicle navigation information manage to accommodate and rapidly apply the customers' demands using precise, highly reliable, and effective visualization methods, they can serve as important factors in creating company profits and in dominating the market in advance. In addition, the transmission of vehicle navigation information in real time is likely to provide an effective way of managing the problem occurrence factors and of instilling effective driving habits, etc. for safe driving.

**Keywords:** Vehicle Navigation Status Information, Vehicle Big Data, Smart-Contents Visualization.

## 1 Introduction

The technological advances and standardization in the automobile sector that started in earnest in 2004 have gone far, aiming to provide external information and multimedia environments to the vehicle space, and their actual services have been focused on the provision of traffic information, road guides, and various life convenience information.

---

\* Corresponding author.

The recent technological advances have engrafted various cutting-edge equipment and information technology (IT) on vehicles, and various sensors and state-of-the-art IT parts and components are largely being attached to improve the convenience and safety of vehicles. Accordingly, the percentage of IT parts and components in vehicle is likely to climb from 25% in 2010 to 40% in 2015 [1].

The trends of the telematics technological advances and services not only in South Korea but in the whole world are heading towards safety and driving convenience rather than infotainment. As the concerns about the telematics services for vehicle diagnosis and management, vehicle convenience equipment control, etc. are rising, studies on and the commercial-service development of a terminal service platform are under way.

Many services that display basic vehicle data for the driver after collecting the internal vehicle data from an external equipment (smartphone, pad, telematics terminal, etc.) via Bluetooth have been introduced, using the Onboard Diagnostic System II (OBD-II) ports originally developed for vehicle emission control. In this study, the OBD-II ports were used to collect vehicle information and other data, and the data were transmitted to the information center using wireless communication, depending on the data type. Service directions on fuel saving and vehicle failure prediction are then suggested in this paper based on the results of the analysis of the data regarding each driver's driving habits and the travel distances, traffic information, etc.

## **2 Vehicle Navigation Status Information Factors**

In this study, the collection of vehicle navigation information using the OBD-II standard was considered to collect and save vehicle navigation information, and vehicles supported by a standard other than OBD-II were excluded.

### **2.1 OBD-II Standard**

The Society of Automotive Engineers (SAE) in the USA established a standard for the plug connectors that process diagnostic test signals and the onboard diagnosis program OBD in 1988. The OBD standard was thereafter developed under the names OBD-1.5 and OBD-II through supplements [2].

### **2.2 OBD-II Protocol**

All OBD-II-supported vehicles largely use three or five different detailed standard signal methods, such as VPW-PWM (SAE-J1850), CAN communication (ISO 15765, SAE-J2234), or the ISO methods (ISO 1941-2, ISO 14230-4) [3]. The signal methods differ by vehicle manufacturer, and different signal methods are used by vehicle model. As such, the OBD-II interface should be taken into account considering the three different methods cited above.

The vehicle navigation information can be monitored and saved through real-time communication with the electronic control unit (ECU) using the OBD-II standard

protocol, and if vehicle failure occurs, the failure details are identified through the standardized five-digit failure diagnostic code.

### 1) OBD PIDs (Onboard Diagnostics Parameter IDs)

OBD PID is the code used for demanding information from the vehicle for the failure diagnosis. The latest OBD-II standard, SAE J1979, stipulates ten different modes, as follows [2]:

- 0x01. Show current data
- 0x02. Show freeze frame data
- 0x03. Show stored Diagnostic Trouble Codes
- 0x04. Clear Diagnostic Trouble Codes and stored values
- 0x05. Test results, oxygen sensor monitoring (non CAN only)
- 0x06. Test results, other component/system monitoring (Test results, oxygen sensor monitoring for CAN only)
- 0x07. Show pending Diagnostic Trouble Codes (detected during current or last driving cycle)
- 0x08. Control operation of on-board component/system
- 0x09. Request vehicle information
- 0x0A. Permanent DTC's (Cleared DTC's)

In this paper, the mode 01 data are collected and analyzed for monitoring the vehicle driving status. The user data collected and used at mode 01 are provided in the following types:

- Vehicle status inspection
  - Analyzing the codes generated from the vehicle's ECU about problems with various parts and components of the vehicle
- Vehicle navigation record management and statistical information
  - Vehicle navigation starting/finishing time, navigation time, travel distance, average speed, highest speed, number of rapid accelerations, fuel injection time, etc.
  - Vehicle navigation records by weekday, week, month, and year
- Real-time vehicle information display
  - Providing real-time information on the vehicle speed, dashboard, coolant temperature, battery voltage, etc., as with the real vehicle dashboard screen
  - Providing various sensor data based on OBD
  - Recording and managing the real-time location information
- Information related to the driver's driving habits
  - Statistical data about the average mileage, numbers of sudden starting and braking, idle speed time, number of excessive RPM, congestion time, and excessive-speed time

## 3 Big-Data Analysis Model Using Vehicle Information

### 3.1 Construction of a Model for Vehicle-Big-Data Analysis

The construction of a model for analyzing vehicle big data can largely be divided into the construction of vehicle data collection environments and of a vehicle information center where all the data are to be collected and analyzed.

1) *Vehicle Data Collection Environments*

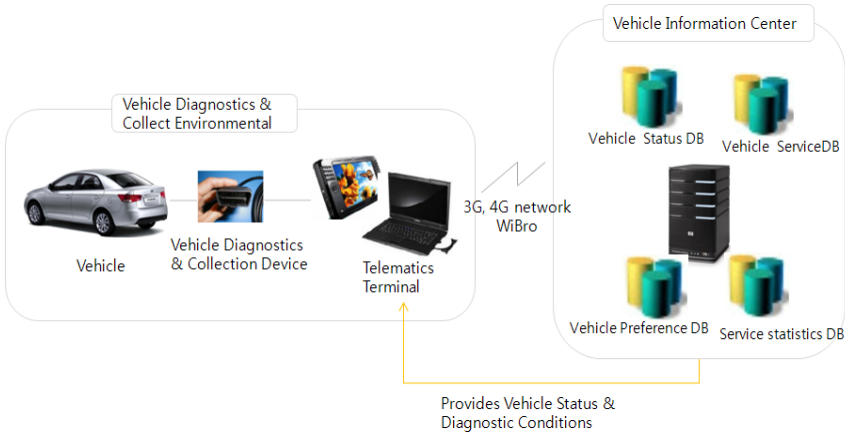
The environments consist of the OBD-II interface that collects data from the vehicle ECU, and of the smart appliances and smart appliance applications by which the user confirms the collected data and transmits them to the collection center.

2) *Vehicle Information Center*

The center carries out the functions of saving and management after analyzing the data transmitted from the vehicle diagnostic and collection equipment, for the user to receive his/her own information and to be provided with driving information and convenience service. The types of received data are divided into the 1<sup>st</sup> data (original data) and the 2<sup>nd</sup> data (original-data fabrications). The 2<sup>nd</sup> data are transmitted to the collection center after being converted into the vehicle information collection environments.

3) *Schematic Diagram of the Big-Data Analysis System for Vehicle Use*

The system to be constructed will collect the navigation information from a driver's vehicle, will conduct vehicle diagnosis, will provide information to the driver, and will transmit the collected data to the vehicle integrated information center for utilization.



**Fig. 1.** Schematic Diagram of the Big-Data Analysis System for Vehicle Use

4) *Big-Data Analytical Factor Decision and Analysis*

The vehicle navigation information analysis system selects the items to be managed regularly and to be provided to the driver for driving safety and convenience. In addition, the navigation status, driving tendency of each driver, etc. are analyzed and subdivided for use as basic data for supplying a suitable type of service to each driver.

In this paper, such data are largely divided into two kinds in the aspects of safety and vehicle management, as follows:

- Vehicle navigation and status management
  - Vehicle status: Selecting the items directly related to the navigation, like managing the vehicle error details, managing the real-time error occurrence and others, and checking the speed, RPM, disability code, and sensor status

- Vehicle navigation: Selecting the items necessary for the navigation, including the navigation time, average speed, travel distance, and movement path
- Driving tendency and statistical management
  - Driving tendency: Selecting the items related to the driver’s driving tendency, like the sudden acceleration, sudden braking, and number of navigations
  - Navigation statistics: Providing the driver with the details of the vehicle navigation by day, week, month, and year

#### 4 Utilization of Big-Data Analysis Results Using Vehicle Information

If the vehicle navigation and status data is received, analyzed, and provided to the driver and/or to the users who need them, not only the driver’s own vehicle navigation management but also convenient service in association with various services will most likely be available. In addition, if the driver’s location and navigation speed data is provided, the service that will enable the other drivers who are travelling around such area to easily check the traffic situation will be available.

Fig. 2 shows the implicated correlation of various convenience service connections through the big-data analysis of the analysis items.

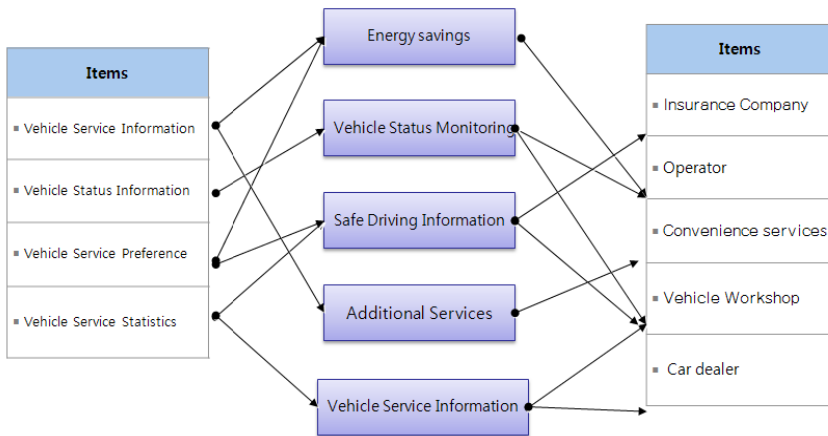


Fig. 2. Utilization of Big-Data Analysis Results Using Vehicle Information

#### 5 Conclusion

The recent fusion of the automobile industry and information technology (IT) has introduced new technologies to satisfy the drivers’ demands related to being able to drive more conveniently and safely, and vehicles are being transformed into intelligent-type vehicles equipped with various sensors and network equipment. In addition, the



environments are rapidly changing, allowing people to use their smartphones to check and enjoy various types of services and convenience functions using vehicle navigation, status, and location information, among others.

In this paper, a model for providing and developing the services currently needed by drivers is proposed based on the monitoring and analysis of drivers' navigation details, driving habits, navigation status, etc. using the collected vehicle status and information. The results of this study can be applied to more effectively collect a large quantity of driver information so that drivers can save on costs, and can be considered a cornerstone for companies' establishment of a platform where they can develop optimal services and continuously manage their customers by investigating the driver's tendencies, etc.

## References

1. Jung, K.-H.: Smart Car and Future Social Changes. KT Economic Business Research Institute.
2. E/E Diagnostic Test Modes, SAE standard J (1979)
3. Jang, S.-T.: A Development of Real-Time Monitoring System at Vehicle Based on OBD-II. Korea University of Technology and Education (August 2009)

# A Model for Analyzing the Effectiveness of Smart Mobile Communication Quality Measurement

Bong-Hwa Hong<sup>1</sup>, Hae-Jong Joo<sup>2,\*</sup>, and Sang-Soo Kim<sup>3</sup>

<sup>1</sup>Dept. of Information and Telecommunication, Kyung Hee Cyber University,  
1 Hoegi-Dong, dongdaemun-Gu, Seoul,

<sup>2</sup>Dept. of LINC, Dongguk University, #710, 82-1 Pil-dong 2-ga,  
Jung-gu, Seoul, 100-272, Korea

<sup>3</sup>Contents Vision Corp., #613, 82-1 Pil-dong 2-ga, Jung-gu, Seoul, 100-272, Korea  
bhhong@khcu.ac.kr, hjjoo@dongguk.edu, cqsky@paran.com

**Abstract.** The recent developments with regard to smart media require objective analyses of the effectiveness of the quality evaluations of the 4G and super-high-speed Internet; as such, it is necessary to establish a standardized model for analyzing the effectiveness of such quality evaluations. The studies related to the performance analysis of the information technology and system have focused on the identification of the correlation among various factors basically related to the performance of such technology and system. Such studies are ultimately relevant to the IS Success Model that can be used to successfully achieve the intended targets of the information system. This paper proposes an alternative list of performance indicators for the analysis of the effectiveness of the quality evaluation of the information technology and system, as well as a reference model for the procedure, definition, etc. for deducing the key performance indicators through the types and cases of common performance indicators applicable to the performance analysis from the past relevant studies. In addition, the effectiveness analysis model is reestablished and applied to systematically conduct effectiveness analysis of the quality evaluation business of smart media based on the literature and preceding case searches. The model of the evaluation system from the analysis perspective is as follows: plan (1<sup>st</sup>)→ do (progress: 2<sup>nd</sup>)→ see (result: 3<sup>rd</sup>).

**Keywords:** Quality Measurement Model, Effective Analysis Model, Smart Mobile Communication.

## 1 Introduction

This paper applies the types and cases of common performance indicators applicable to performance analysis from the past relevant studies to the procedure, definition, etc. for deducing an alternative list of performance indicators for quality evaluation effectiveness analysis, and the key performance indicators. The studies related to the

---

\* Corresponding author.

performance analysis of the information technology and system have focused on the identification of the basic correlations among various factors related to performance. Such studies are ultimately relevant to the IS Success Model, which can be used to successfully achieve the intended targets of the information system[1].

This paper aims to promote quality evaluation effectiveness analysis by quality evaluation business type (4G, super-high-speed Internet) by determining the reliability and validity of the deduced effects, by promoting smart mobile communication quality evaluation effectiveness analysis in a systematic and scientific way. Accordingly, this paper maintains the consistency and objectivity of the smart mobile communication business through the deduction of the common results, provides a foundation for establishing a standard model for the effectiveness measurement, and paves the way for the establishment of a generalized development stage system (degree of normality) through the effectiveness analysis of the quality evaluation business. To achieve this, this paper explains the related study “A Study on the Performance Analysis Model of the Information System” in chapter 2, proposes a definition and systematic diagram of the analysis of the effectiveness of the smart mobile communications quality evaluation, and presents the model indicator selections in chapter 3. Lastly, a conclusion is drawn.

## **2 Related Studies**

### **2.1 Initial Model of Delone and Mclean (1992), and Its Limits**

The factors affecting the success of the information system consist of the performance factors in six areas: system quality, information quality, system usage, user’s satisfaction level, individual performance, and organizational performance[1]. The performance measurement of the information system is applied using various performance measurement indicators with various properties, such as the properties of the system itself, the information property generated from the system, the users’ properties, and the organization’s properties.

The information system success factors consist of the relations affecting the stages quality--> use--> effects, as shown in <Table 1>; the factors affecting the quality of the information system consist of the system and information quality; the use of the system is subdivided into the information usage and the degree of satisfaction of the user; and the final effects are divided into the individual performance and the organizational performance affected by the individual performance.

As the initial model of Delone and Mclean (1992) presented in <Table 1> considers the influence factors of the information system itself, it exposes several limits. First, it omits the variable pertaining to the organizational situation where the information system is used; second, the “use” of the information system can be replaced with the “usefulness” of the system rather than simply its usability; third, despite the fact that the information system function itself contains the service factors that satisfy an organization’s IT demands, such factors are not considered in the performance measurements of most models. Therefore, given that the role of the information department increases with time, the “service quality” area is appropriately added to the role of the information system department.

Delone and Mclean (2002) partially modified their models to accommodate the criticisms of other studies after 1992. “Service quality” was added to “system quality” and “information quality,” “use intention” was additionally introduced to the “system use” area, and the two were integrated into “pure effects” while considering that the benefits of separating the “individual influence” from the “organizational influence” were actually insignificant. The “pure effects” are characterized by the model that includes the “system use” and the reflux character re-affecting the “users’ satisfaction.”

**Table 1.** Six performance factors of Delone and Mclean

Stage	Quality ⇨	Use ⇨	Effects
Area	System Quality	Information Usage	Individual Performance
<b>Performance Factors</b>	<ul style="list-style-type: none"> <li>. Data precision</li> <li>. Data recency</li> <li>. Data contents</li> <li>. Use availability</li> <li>. Humane factors</li> <li>. Approach convenience</li> <li>. Reality of the requirements</li> <li>. Usefulness of the function</li> <li>. System precision</li> <li>. System flexibility</li> <li>. System accuracy</li> <li>. System integration</li> <li>. System efficiency</li> <li>. Resource availability</li> <li>. Response time</li> </ul>	<ul style="list-style-type: none"> <li>. Use amount/time</li> <li>. No. of questions</li> <li>. Connecting time, use functions</li> <li>. Used record volume</li> <li>. Use frequency</li> <li>. Volume of output report</li> <li>. Regular use volume</li> <li>. User type</li> <li>. Direct/indirect user</li> <li>. User/non-user</li> <li>. 1<sup>st</sup>/2<sup>nd</sup> user</li> <li>. Use characteristics</li> <li>. Use for intended purpose</li> <li>. Proper use</li> <li>. Used information types</li> <li>. User level</li> <li>. Repeated use, voluntary use</li> <li>. Motivation to use</li> </ul>	<ul style="list-style-type: none"> <li>. Interpreting the level of information</li> <li>. Learning level</li> <li>. Precise interpretation</li> <li>. Information recognition</li> <li>. Information recall</li> <li>. Confirming problems</li> <li>. Effectiveness of decision making</li> <li>. Quality of decision making</li> <li>. Improved decision making</li> <li>. Precision and time of decision making</li> <li>. Certainty and participation of decision making</li> <li>. Improvement of individual productivity</li> <li>. Change of decision making</li> <li>. Inducing management activity</li> <li>. Project outcome</li> <li>. Plan quality</li> <li>. Individual influence degree</li> <li>. Individual evaluation of the information system</li> </ul>
<b>Area</b>	<b>Information Quality</b>	<b>Users' Satisfaction Level</b>	<b>Organizational Performance</b>
<b>Performance Factors</b>	<ul style="list-style-type: none"> <li>. Importance, suitability, and availability</li> <li>. Usability and interpretation possibility</li> <li>. Simplicity, form, and contents</li> <li>. Accuracy</li> <li>. Precision level</li> <li>. Sufficiency</li> <li>. Completeness</li> <li>. Reliability</li> <li>. Recency</li> <li>. Proper timing</li> <li>. Peculiarity</li> <li>. Comparability</li> <li>. Degree of quantification</li> <li>. Removal of distortion</li> </ul>	<ul style="list-style-type: none"> <li>. Satisfaction in a particular field</li> <li>. Overall satisfaction level</li> <li>. Single-item measurement</li> <li>. Multiple-item measurement</li> <li>. Information satisfaction level</li> <li>. Difference between the demanded and output information</li> <li>. Software satisfaction degree</li> <li>. Decision making satisfaction level</li> </ul>	<ul style="list-style-type: none"> <li>. Portfolio of application program</li> <li>. Scope of application program</li> <li>. No. of key application programs</li> <li>. Reduction of operational cost, decreasing no. of staffs</li> <li>. Increase of total productivity</li> <li>. Increase of revenue and sales</li> <li>. Increase of market share and profit</li> <li>. ROI and ROA</li> <li>. Net profit/costs</li> <li>. Cost/benefit ratio</li> <li>. Share price</li> <li>. Increase of work amount and product quality</li> <li>. Contribution level to the object achievement</li> <li>. Effectiveness of service</li> </ul>

**2.2 Model of Myers, Kappelman, and Prybutok (1997)**

Myers, Kappelman, and Prybutok (1997) reconstructed the evaluation model by adopting the contingency theoretical approach that included the service quality, added the work groups, and considered the organization and/or external environment at the same time by expanding the model of Delone and Mclean (1992) and the model of Pitt, Watson, and Kavan (1995).

They insisted that the “service quality” was considered in the information system proposed by them in that more IT demands should be satisfied, and they added the work group as an intermediate medium between the individuals and organization considering the organizational environment where the team role was emphasized[1].

### 2.3 Model of Parker and Benson(1998)

Parker and Benson (1998) presented a method that evaluated the economic influence on the companies affected by the information technology and system through the concept of 「information economics」 [1]. 「Information economics」 is a method that is based on the value chain of the company by expanding the profits to the concept of value, divides the value into the business area actually creating the values and the information technology area supporting such, and selects the information technology that is suitable for attaining the vision and implementing the strategy of the company through the connection and reconstruction of the costs and values created among such business and information technology areas. In addition, 「information economics」 is not limited to calculating the benefits in monetary terms, unlike the existing cost benefit analyses, but investigates the correlation in the economic aspect, considering the generated performance as a “value.”

## 3 A Model for Analyzing the Effectiveness of the Smart Mobile Communication Quality Evaluation

### 3.1 Conceptual Definition

The effectiveness analysis model is reestablished and applied based on the literature and preceding case searches to systematically analyze the effectiveness of the quality evaluation of the smart mobile communications (4G and LTE) businesses. The evaluation system is as follows (from the analysis perspective): plan (1<sup>st</sup>)→ do (progress: 2<sup>nd</sup>)→ see (result: 3<sup>rd</sup>)[2-5].

**Table 2.** Model Concepts of Quality Evaluation Effectiveness Analysis

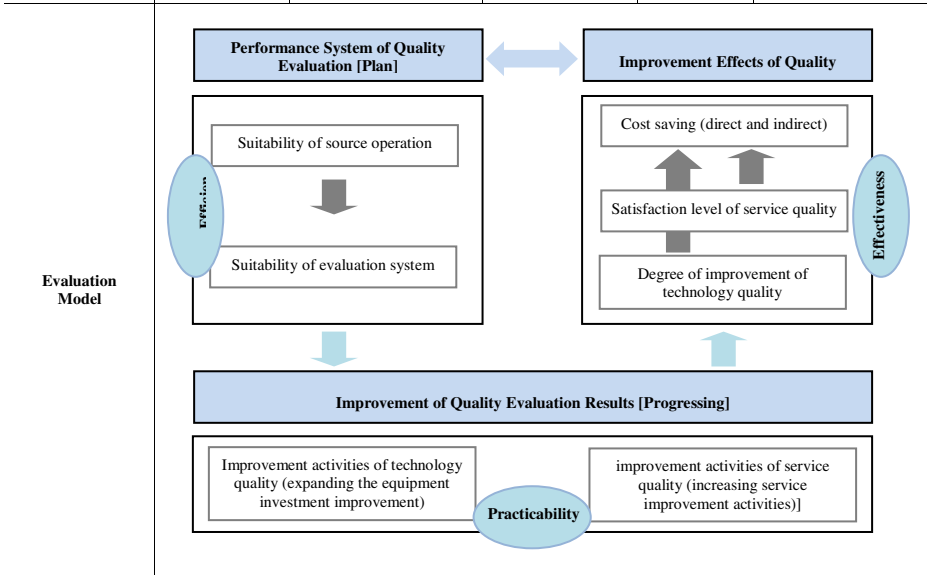
		Operant Definition
	<b>Plan</b> (1 <sup>st</sup> )	Formulating plans for funding, manpower, promotion, etc. for the establishment of a quality evaluation business, deducing the problems to be addressed and validating the reliabilities by evaluating their suitability to the overall business progress ( <b>Effectiveness Measurement</b> )
	<b>Do</b> (2 <sup>nd</sup> )	Evaluating the direct activities, including the responses of the businesspeople/users/ government, investment, etc., and their influences on the results of the quality evaluation business ( <b>Practicability Measurement</b> )
	<b>See</b> (3 <sup>rd</sup> )	Evaluating the quality improvement, satisfaction level, cost reduction, etc. that occur as the results of the response level to and activities of the businesspeople /users/government in relation to the quality evaluation business ( <b>Effectiveness Measurement</b> )

### 3.2 Systematic Diagram of the Models

The system was established to separately enable effectiveness measurement by deducing the evaluation areas, evaluation purposes, and subject institutions by business promotion stage; was designed to enable the effectiveness measurement of the overall process of the business promotion; and consists of a virtuous circulation structure that can be continuously improved by applying the results of each stage to the next business plan.

**Table 3.** Systematic Diagram of the Quality Evaluation Effectiveness Analysis Model

Business Stage	Plan (1 <sup>st</sup> )	Do (2 <sup>nd</sup> )	See (3 <sup>rd</sup> )		
Evaluation Area	Business operation	Improvement activities	Operational effects		
Evaluation Purpose	Efficiency	Practicability	Effectiveness		
Creation Stage of Evaluation Effectiveness	Input	Improvement (process)	Effects (outcome)		
			Quality	Satisfaction level	Cost saving
Subject Institution	Government (policy institution)	(Businesspeople, users, government)	Businesspeople	Users (clients)	Businesspeople, users, government



### 3.3 Selection of Quality Evaluation Effectiveness Analysis Indicators

To deduce the logical and reasonable indicators for effectiveness analysis, a literary search was conducted based on the preceding studies on the similar cases; the indicators were selected considering the evaluation areas and factors by business stage; and an indicator pool was completed based on such selection. After deciding

the selection standard for five indicators using the SMART technique[6,7] from the indicator pool consisting of the preceding study cases, the initial draft for the indicators was prepared, and finally, the indicators were confirmed through the verification of the advisory committee consisting of experts in the relevant fields.

The selection of the optimized indicators cannot be achieved in a short time, but this is possible through continuous development. As such, it is necessary to have a procedure for filtering out the suitable indicators from the alternative indicator list composed by various studies, to measure the performance appropriate to the strategy by applying the changing strategy. A business like that involving quality evaluation considers the selection of indicators and the system among the indicators important because it has many ripple effects through the 2<sup>nd</sup> and 3<sup>rd</sup> stages because the effects occur directly. To satisfy such conditions, the standard for the indicator selection was used and applied with the SMART technique. The term “SMART” applies the specific level (Specific), measurability (Measurable), practicability (Action-oriented), relation (Relevant), and timeliness (proper Timing) as the indicator selection standards, and plays the role of showing which part is considered important in developing the indicators.

## 4 Conclusion

In this study, a common effectiveness measurement method was induced by quality evaluation business type (4G, super-high-speed Internet) by determining the reliability and validity of the deduced effects, for the promotion of smart mobile communication quality evaluation effectiveness analysis in a systematic and scientific way. Therefore, this paper maintains the consistency and objectivity of the smart mobile communication business through the deduction of the common results, provides a foundation for establishing a standard model for effectiveness measurement, and paves the way for the establishment of a generalized development stage system (degree of normality) through the effectiveness analysis of the quality evaluation business. In addition, in this study, models for evaluating all the effects on the overall quality evaluation business were developed and designed to provide a comprehensive view based on the procedures, so that the evaluations can include the effects occurring in the promotion of the quality evaluation business.

This paper defines the effects of quality evaluation and evaluates them by promotion stage, compares such effects by stage to deduce the excellent or inadequate stages, and proposes that the results thereof be applied by the government in its future business promotions.

## References

1. NIA V-RIR-09133, Analysis of effectiveness in 3G and high-speed internet quality test (December 2009)
2. Takahashi, A., Hands, D., Barriac, V.: Standardization Activities in the ITU for a QoE Assessment of IPTV. *IEEE Communications Magazine* 46(2), 78–84 (2008)

3. ITU-R Rec. BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R (December 2002)
4. ITU-T Rec. P.910, Subjective video quality assessment methods for multimedia applications, ITU-T (September 1999)
5. Shumeli, R., Hadar, O., Huber, R., Maltz, M., Huber, M.: Effects of an Encoding Scheme on Perceived Video Quality Transmitted Over Lossy Internet Protocol Networks. *IEEE Tr. Broadcasting* 54(3), 628–640 (2008)
6. Kim, H., Choi, S.: A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service. In: *IEEE ICACT 2010* (February 2010)
7. Dai, Q., Lehnert, R.: Impact of Packet Loss on the Perceived Video Quality. *IEEE INTERNET 2010* (September 2010)
8. ITU-T J.143 User Requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television Series J: Transmission of Television, Sound Programme and Other Multimedia Signals Measurement of the Quality of Service (May 2000)
9. Yamagishi, K., Hayashi, T.: Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services. *IEEE ICC 2008* (May 2008)
10. Hurricane WAN Emulation & Network Simulation, PacketStorm Communications, Inc., <http://www.packetstorm.com/>



# Characteristics Analysis and Library Development for Common Lamps by Using PSPICE Modeling

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2,\*</sup>, Ho-Bin Song<sup>3</sup>, and Ok-Hwan Kim<sup>1</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Budaе-dong, Cheonan-si, Chungnam, 330-717, Korea  
{yckim59, owkim}@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & electric Engineering, Daewon University College,  
316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

<sup>3</sup> Contents Vision CO. #613, 82-1, Dongguk University Chungmuro Media Center,  
Pildong 2-ga Jung-go, Seoul, 100-272, Korea  
songhobin@daum.net

**Abstract.** Recently, common lamps have been used in a variety of fields, including simple illumination science, and their domain of use is increasingly widening. In this study, we configure a library through modeling and verify its accuracy through simulations for widely used and representative lamps such as CCFL, fluorescent lamps, and HID lamps. On the basis of our experiments, we also perform a lamp simulation using PSPICE, which allows us to take advantage of the lamp library easily.

**Keywords:** CCFL, HID lamp, Fluorescent lamp, PSPICE.

## 1 Introduction

In the modern world, the common lamp is one of the most important and indispensable items and is used in various fields in addition to simple lighting.

However, the development and the application of various circuits using lamps require simulations that are currently performed with an alternative and simplified model, which is essentially a serial connection of loads such as resistors and inductors. Such an alternative simulation methodology can provide rough characteristics but has clear limitation in accurately verifying the functionality. If more detailed and precise lamp modeling is available, an accurate circuit analysis using lamps can be carried out accordingly.

To address such challenges, in this work, we have modeled widely used lamps and built a library. We have used PSPICE, a general-purpose and easy-to-use simulation tool. We have modeled a CCFL, a fluorescent lamp, and HID lamp, which have become popular recently and are now used widely, and applied them to our simulation setup, which has led to good functional characteristics as a result. With the output from this study, we expect that a user can obtain more accurate lamp simulation results with considerably less effort.[1]

---

\* Corresponding author.

## 2 Characteristics of Lamps and Modeling

### 2.1 CCFL Modeling

A cold cathode fluorescent lamp (CCFL) is suitable for miniaturization because of its simple structure. By changing the fluorescent material type and mixture ratio, we can obtain an arbitrary fluorescent color. Further, the CCFL has longer lifetime expectancy than a hot cathode fluorescent lamp and requires only a simple lighting circuit. Further, the CCFL shows a negative resistance characteristic, requires a ballaster to limit the negative current, and has resistor-like characteristics under a stable condition.[2]

Basically, the CCFL has four important parameters: firing potential for lighting, sustained voltage, frequency, and pipe current. The brightness of the CCFL depends not on the operating frequency but on the lamp current.

The effective voltage and current characteristics of the CCFL in this study have negative impedance at the sensitization level, and a lamp can represents two sensitization characteristics, as expressed in equation (1).

$$V_{\text{rms}} = 60.966 + 110.45 \cdot e^{-1.9404 \times I_{\text{rms}}} - 48.578 \times e^{-60.182 \times I_{\text{rms}}} \quad (1)$$

where the first term denotes the default value of the voltage-current curve, the second term shows the negative impedance nature, and the last term indicates the positive impedance nature.

The parameters specified in equation (1) can be derived by computing the least square root and the equivalent impedance to a resistor-like lamp at high frequency, as expressed in equation (2).

$$R_{\text{LAMP}} = \frac{V_{\text{rms}}}{I_{\text{rms}}} \quad (2)$$

From the above, we can compute the momentary voltage of a lamp  $V_t$  by using the momentary current of a lamp  $I_t$ , as shown in equation (3).

$$V_t = \frac{60.966 + 110.45 \cdot e^{-1.9404 \times I_{\text{rms}}} - 48.578 \times e^{-60.182 \times I_{\text{rms}}}}{I_{\text{rms}}} \times I_t \quad (3)$$

On the basis of equation (3), we modeled the CCFL. The momentary voltage of a lamp  $V_t$  can be described as a function of the momentary current  $I_t$  and the effective current  $I_{\text{rms}}$ . The effective value of the lamp current in a PSPICE model can be computed using an RC integration circuit, where the current source  $I_s$  is defined as the square of the momentary current of a lamp  $I_t$ , and the output voltage  $V_A$  can be obtained by integrating over the period  $T$ , as in equation (4).

$$V_A = \int_0^T I_t^2 \frac{dt}{T} = I_{\text{rms}}^2 \quad (4)$$

Therefore, the square root of  $V_A$  is the effective value of the current  $I_t$ . By using such an RC integration circuit to compute the effective current of a lamp, the lamp model equation in equation (3) can be transformed into the PSPICE model in Fig. 1.

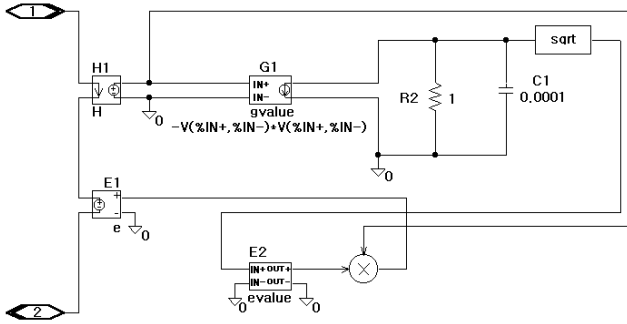


Fig. 1. PSPICE model for CCFL

## 2.2 Fluorescent Light Modeling

As a type of a low-pressure gas discharge lamp, fluorescent light can be best described by the flow of current through the carrier gas. Gas discharge or arc plasma occurs in a lamp tube. The discharge length consists of several regions between the positive column and the anode drop, which play an important role in the overall lamp behavior. Therefore, we model fluorescent light by focusing mainly on these two regions.

The Cassie equation where conductance  $G$  is a dependent variable is expressed in equation (5), which describes the arc behavior in the case of a high current.[3]

$$G = \frac{v \cdot i}{E_0^2} - \theta \frac{dG}{dt} \tag{5}$$

where  $E_0$  is the arc voltage,  $\theta$  is the arc time ratio,  $v$  is the lamp voltage, and  $i$  is the lamp current.

The Mayr equation is given as equation (6); it describes the arc behavior in the case of a low current.

$$G = \frac{i^2}{P_0} - \theta \frac{dG}{dt} \tag{6}$$

where  $P_0$  is the power loss.

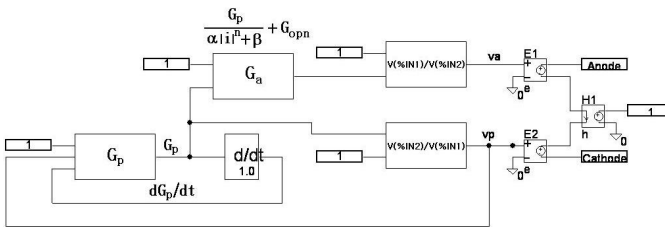


Fig. 2. PSPICE model for fluorescent light

The Mayr and Cassie equations are effective in the cases of 0 or low current and high voltage ranges, respectively. The positive column can be simulated using equation (5) and (6).

Fig. 3 shows the anode voltage and the lamp current of a fluorescent light.

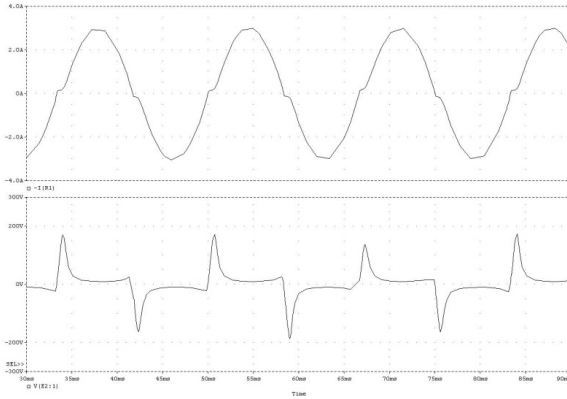


Fig. 3. PSPICE model for fluorescent light

### 2.3 HID Lamp Modeling

Our PSPICE model for a high-intensity discharge lamp (HID) can be built using physical properties, lamp parameters, and universal constants.[4]

The increase amount in arc heating  $dQ$  can be described as follows:

$$dQ = (P_{in} - P_{out})dt \tag{7}$$

where  $P_{in}$  is the input power to a lamp,  $P_{out}$  is the output power from the lamp, and  $dt$  is the time difference.

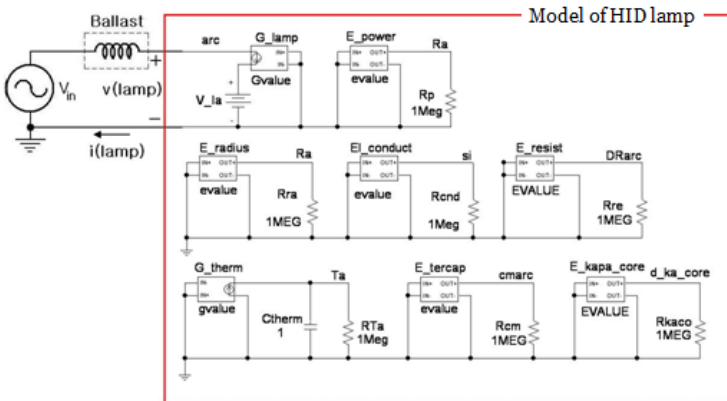


Fig. 4. PSPICE model for HID lamp

After applying interaction equations to equation (7), we can obtain the differential equation given in equation (8).

$$\frac{dT}{dt} = \frac{\frac{V_{\text{lamp}} I_{\text{lamp}}}{L_g} - 2\pi R_{\text{el}}[\varepsilon(T) + k(T - T_{\text{amb}})]}{\pi [c_{\text{arc}} \rho_{\text{arc}} R_{\text{el}}^2 + k_{\text{Tred}} c_{\text{per}} \rho_{\text{per}} (R_{\text{tube}}^2 - R_{\text{el}}^2)]} k\omega \quad (8)$$

Once we mathematically describe equation (8) in PSPICE, we can build the HID lamp model shown in Fig. 4. Fig. 5 respectively show the waveforms of the HID lamp obtained from our PSPICE simulation model at 60 Hz.

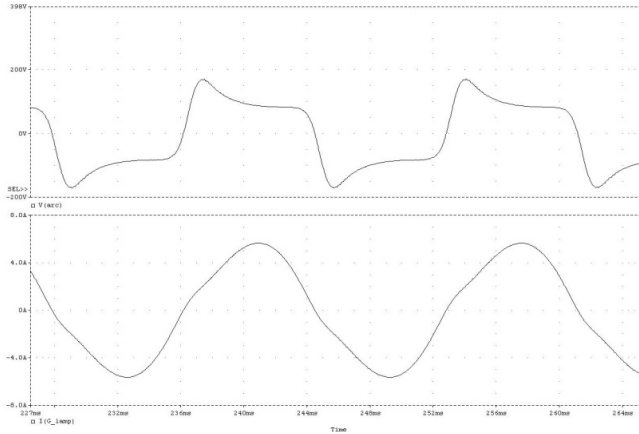


Fig. 5. Voltage and current curve for the HID lamp at 60[Hz]

### 3 Conclusion

In this study, we built an easy-to-use platform with PSPICE-based modeling and library for CCFLs, fluorescent lamps, and HID lamp. Further, we performed simulations to verify the correctness of our models and library. We hope that our library can allow more accurate simulations for lamps using only the basic parameters.

### References

1. Zaitso, T., Shigehisa, T., Shoyama, M., Ninomiya, T.: Piezoelectric Transformer Converter with PWM Control. In: IEEE APEC 1996 Proc., pp. 279–283 (March 1996)
2. Hwang, L.H., Yoo, J.H., Xiao, H., Zhi, Y., Kim, H.S., Oh, H.S., Cho, M.T., Choi, G.S.: A study on the modeling of piezoelectric transformer for CCFL using a PSPICE. In: 7th International Conference on Power Electronics, ICPE 2007, pp. 122–126 (October 2007)
3. Perdigao, M., Saraiva, E.S.: MATLAB-SIMULINK fluorescent-lamp models. In: 39th International, UPEC 2004, vol. 2, pp. 855–859 (September 2004)
4. Paul, K.C., Takemura, T., Hiramoto, T., Yoshioka, M., Igarashi, T.: Three-Dimensional Modeling of an HID Lamp Operated by a Direct Current. In: ICOPS 2005, p. 191 (June 2005)

# A Study on Brake Torque for Traction Motors by Using the Electric Brake

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2\*</sup>, and Ho-Bin Song<sup>3</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Buda-dong, Cheonan-si, Chungnam, 330-717, Korea  
yckim59@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College,  
316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

<sup>3</sup> Contents Vision CO. #613, 82-1, Dongguk University Chungmuro Media Center,  
Pildong 2-ga Jung-go, Seoul, 100-272, Korea  
songhobin@daum.net

**Abstract.** In this paper, a scaled model propulsion system was designed and tested for a electrical brake until stopping of the vehicles. A brake torque control method at the moment of vehicle stop was proposed accordingly. The test results for the electric brake were drawn while controlling the motor at low speed until the stop of the electric locomotives. Also, the speed detection method was used by implementing an observer that estimates the position and speed of rotor by using a resolver. Power converter was constructed as a converter-inverter system. Further, an improved brake method that uses only an electric brake till motor stop is proposed by comparing those in the blending brake that uses a air brake while reducing brake torque at vehicle stop.

**Keywords:** Electric brake, Traction motor, Brake torque, Converter-inverter.

## 1 Introduction

It has been reported that brake technology can pursue the effective energy use by expanding the use of a regenerative brake. Brake force is secured with technology development that expands a regenerative brake not only in the electric brake till stopping the motor but also in the high speed region. Also, brake force was reported to be increased and ride comfort is excellent by the new technology compared with the conventional.[1][2]

A pure electric brake possibly makes the mechanical brake device reduced by reducing air brake use. Moreover, it is superior in the maintenance and environmental aspects by reducing wear and tear of the brake shoe and lining as well as consequently generating dust.[3][4][5]

---

\* Corresponding author.

In this paper, a scaled model propulsion system was designed and tested for a electrical brake until stopping of the vehicles. A brake torque control method at the moment of vehicle stop was proposed accordingly. The test results for the electric brake were drawn while controlling the motor at low speed until the stop of the electric locomotives. Also, the speed detection method was used by implementing an observer that estimates the position and speed of rotor by using a resolver. Power converter was constructed as a converter-inverter system. Further, an improved brake method that uses only an electric brake till motor stop is proposed by comparing those in the blending brake that uses a air brake while reducing brake torque at vehicle stop.

## 2 Electric Brake Method Till the Stop of Electric Locomotive

The brake of an electric locomotive uses both an air brake and an electric brake. M car uses the air brake as a supplement for the shortage of brake force according to the electric brake specification. Since electric braking force is less in high speed region, speed reduction is executed with a blending brake which uses an air brake along with an electric brake, while at constant torque, speed reduction is executed with only an electric brake.[6] A vehicle is stopped by changing the electric brake into the air brake at the speed of around 6~7[km/h]. Stopping is executed by using the method as shown in Fig. 1(b). The proposed ideal brake process when electric brake is used to stop the vehicle is presented in Fig. 1(a).

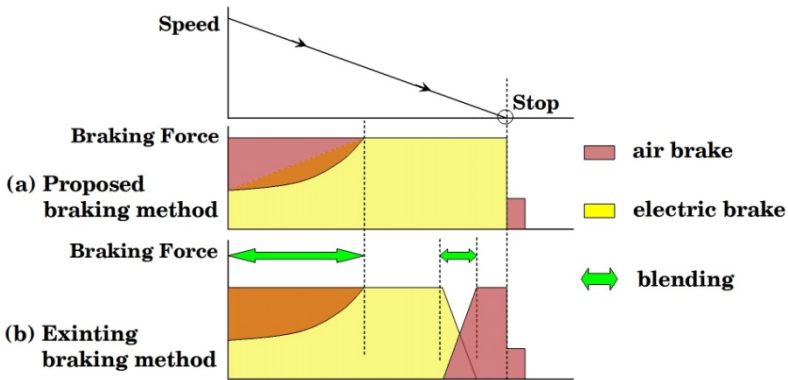


Fig. 1. Mode of electric brake until stop

For the torque reduction method in Fig. 2, section (1) is a zone wherein the vehicle is decelerated as brake torque  $T$ , section (2) is a reduction starting zone of torque at  $\omega_0$ , and section (3) is a control zone of brake torque.[7]

In section (3) of Fig. 2, when brake torque is made proportional to the vehicle speed, brake torque becomes equation (1).

$$T = k\omega \tag{1}$$

Under deceleration condition, there exists brake torque for section (1) of Fig. 2 and  $\omega_0$  which is decided by proportional factor  $k$  in equation (1). Besides, since the condition for torque becomes zero in the torque control by Equation (1) is speed zero, it satisfies the condition for the stop moment.

Rotational speed and brake torque in the brake section can be expressed as equation (2).

$$\frac{d\omega}{dt} = -\frac{T}{J} \tag{2}$$

From equation (1) and equation (2), the rotational speed in section (3) becomes equation (3).

$$\omega = \omega_0 e^{-\frac{k}{T}t} \tag{3}$$

Reduction starting speed of brake torque  $\omega_0$  and deceleration time constant are decided by proportional factor  $k$ . When this proportional factor is applied to actual vehicles, it is a factor that needs to be controlled by considering ride comfort, etc.

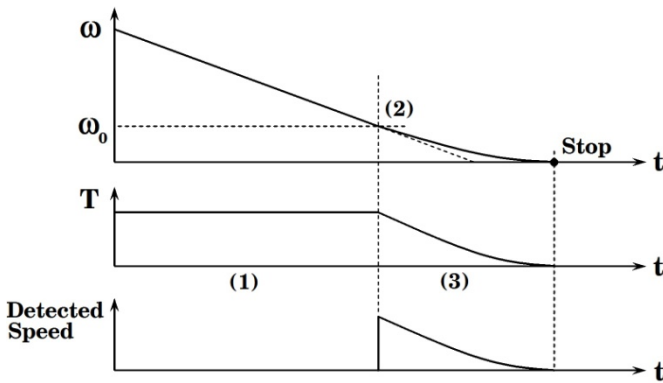


Fig. 2. Torque control at the moment of vehicle stop

### 3 Experimental of Motor Brake and Its Driving

When a vehicle is stopped by an electric brake, speed measurement at low speed and torque control are key functional elements. Especially, the precision in the speed measurement is extremely important to stop the motor smoothly. Therefore, position and speed of rotor were measured using a resolver in this study. By generating brake torque that is proportional to speed at low speed rather than generating suspension torque, travelling prevention after stop the rotor was achieved.

Fig. 3 is the measured wave form that stops rotor with torque proportional to the speed by detecting the speed that starts torque reduction when regenerative brake was executed during forward-reverse operation of motor.



In Fig. 3,  $i_q^*$  is the set value of current to the torque.  $\omega$  the magnified waveform estimated in the speed detector as low speed region. The recording time was limited as 17.5[rpm] for the forward-reverse direction for instrumentation. The third waveform in the Figure 3 is the detected speed waveform to control torque that is proportional to the speed. The measurement was performed with the same magnification as  $\omega$ .  $i_a$  is the measurement of line current of motor.

The test equipment was adjusted as torque was varied depending on steps in the startup and brakes. The control was made to be switched at the point where torque and set torque becomes same by Equation (1) at the moment of stop while the regenerative brake was executed. A close examination of  $i_q^*$  in Fig. 3 gives an idea that current set value doesn't follow step changes at the motor stop. It is clear that torque control was switched by control program. In addition, by controlling torque

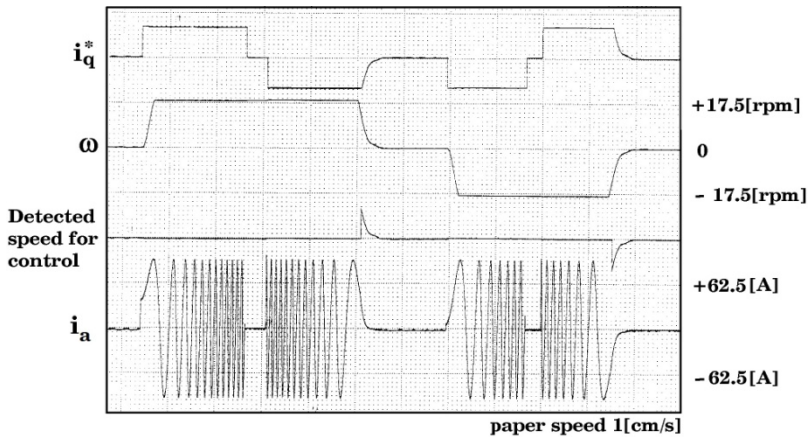


Fig. 3. Brake waveform of motor

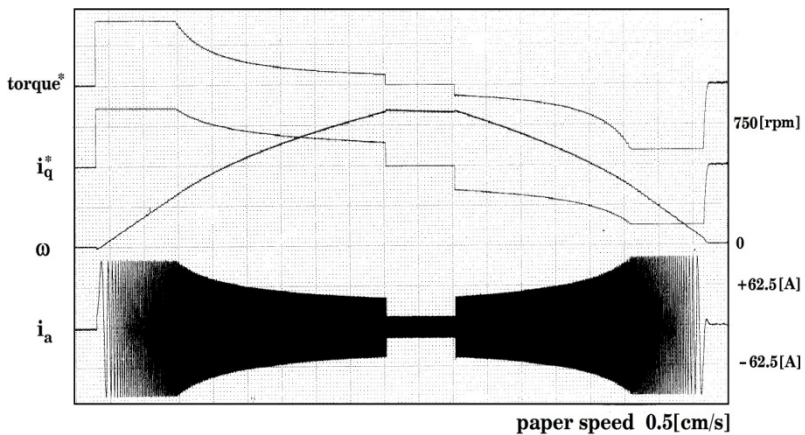


Fig. 4. Drive experiment for the inertial load

without having off on the power converter even after the motor stops, suspension torque proportional to the speed could be maintained and stable movement could be observed. In Fig. 3, it can be seen that conversion of speed-torque control is executed at near 10[rpm].

The vehicle operation was performed by a drive motor and stop it by a regenerative brake. In Fig. 4, the speed which startup the characteristic drive was set at 360[rpm] and torque was set zero at near 800[rpm]. After starting the regenerative brake, the motor was stopped with torque that is proportional to the speed. Torque was set as step changes same as previous experiments as shown in Fig. 4. The smooth driving at stop of motor was thereby observed.

## 4 Conclusion

A scaled model propulsion system was designed for the electric brake until stopping the vehicle. With experimental results, a control method of brake torque at the moment of motor stop was proposed. With test equipment, drive and brake tests were repeated in this study. Since the control method of brake force that was proportional to the speed at the moment of motor stop was implemented in the test, soft drive effect could be achieved at the stop only with the electric brake.

Besides, by carrying out an air brake movement at the same time of reducing brake torque at the stop, environmental perspective and energy use aspect and performance of vehicle could be maximized with the improved brake method that uses only an electric brake until vehicle stop as compared with those in the blending brake. The additional benefits of reducing vehicle weight and cost reduction for maintenance as well as ride comfort, energy efficiency and noise reduction could be expected.

## References

1. Kim, Y.-C., Song, H.-B., Cho, M.-T., Lee, C.-S., Kim, O.-H., Park, S.-Y.: A Study on the Improved Stability of Inverter through History Management of Semiconductor Elements for Power Supply. In: Kim, T.-h., Ramos, C., Kim, H.-k., Kiumi, A., Mohammed, S., Ślęzak, D. (eds.) ASEA/DRBC 2012. CCIS, vol. 340, pp. 155–162. Springer, Heidelberg (2012)
2. Kraiem, H., Flah, A., Hamed, M.B., Sbita, L.: High Performances Induction Motor Drive Based on Fuzzy Logic Control. *IJCA* 5(1), 1–12 (2012)
3. Kim, Y.-C., Song, H.-B., Cho, M.-T., Moon, S.-H.: A Study on Direct Vector Control System for Induction Motor Speed Control. In: Park, J.J(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. LNEE, vol. 181, pp. 599–612. Springer, Heidelberg (2012)
4. Fuchs, E.F., Masoum, M.A.S.: Analyses and Designs Related to Renewable Energy Systems. Springer Science+Business Media, LLC 2011, 557–687 (2011), doi:10.1007/978-1-4419-7979-7\_12
5. Aravindh Kumar, B., Saranya, G., Selvakumar, R., Swetha Shree, R., Saranya, M., Sumesh, E.P.: Fault detection in Induction motor using WPT and Multiple SVM. *IJCA* 3(2), 9–20 (2010)
6. Treerong, J.: Fault Detection of Electric Motors Based on Frequency and Time-Frequency Analysis using Extended DFT. *IJCA* 4(1), 49–58 (2010)
7. Pal, S., Tripathy, N.S.: Remote Position Control System of Stepper Motor Using DTMF Technology. *IJCA* 4(2), 35–42 (2010)

# A Virtual Cluster Scheme Technology for Efficient Wireless Sensor Networks

Seongsso Cho<sup>1</sup>, Bhanu Shrestha<sup>1</sup>, Young-gi Kim<sup>2</sup>, and Bong-Hwa Hong<sup>2</sup>

<sup>1</sup> Department of Electronic Engineering, Kwangwoon University,  
20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea

<sup>2</sup> Department of Information Communication, Kyunghee Cyber University, Korea  
{css,bnu}@kw.ac.kr, smileknock@gmail.com, bhhong@khcu.ac.kr

**Abstract.** Once nodes are deployed in the wireless sensor network, as it is impossible to replace batteries, the amount of transmittable information depends on how to use the limited energy efficiently for longer network life. Virtual Cluster Routing (VCR) builds the efficient virtual cluster considering node compactness, selects the closest node and manages the routing table to reduce overhead significantly by referencing routing table information within the virtual cluster without communicating with other clusters via the head, and has a lower transmission delay and higher survival time than the routing scheme.

**Keywords:** Sensor Network, MANET, LEACH, Sensor Node, Sink Node.

## 1 Introduction

The Wireless Sensor Network (WSN) consists of small sensor nodes including microcontrollers, wireless transmitters, and sensing modules. And node information is transmitted to the data collecting node, sink node mostly via multi-hop communication. At first, for the military purpose, the WSN was developed to monitor and reconnoiter areas inaccessible to people by deploying many sensor nodes. However, its usage expanded to environmental monitoring, building risk analysis, patient monitoring, medical service and others to collect various information [1]. In the sensor network, modeling the Mobile Ad-Hoc Network (MANET) environment, without fixed bases such as Access Points (AP), relatively many sensor nodes are deployed in the wide sensor field to create various dynamic topologies and have nodes autonomous and independent from each other [2-3]. What should be considered here is how to use the limited energy resource of the sensor node efficiently. The sensor node is powered by the battery, which cannot be replaced or recharged due to its operational environment. When assessing the WSN's performance, energy efficiency, accuracy of detected data and service quality are considered. Among them, the key item is energy efficiency. As the sensor consumes energy over time, energy efficiency decides how long the sensor can operate before energy runs dry [4]. In the sensor network, the existing routing scheme can be divided into plane-based one and

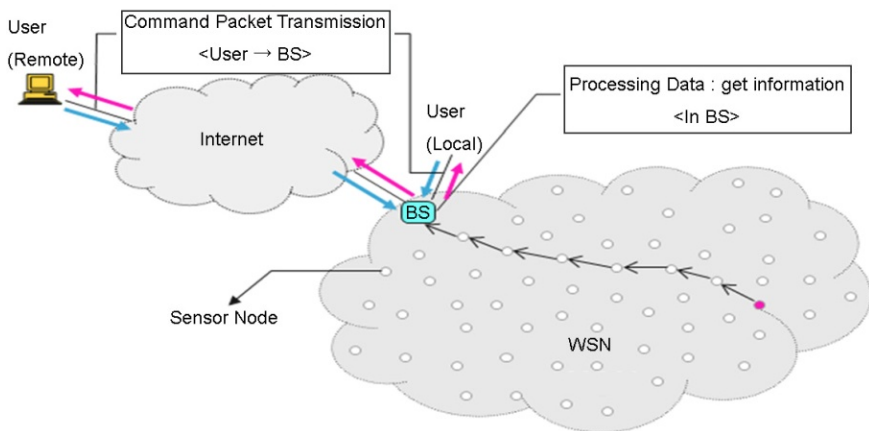
hierarchy-based one depending on the node configuration. While nodes exchange information with each other at the same level in the plane-based routing scheme, many clusters are built to create the node hierarchy for data transmission in the hierarchy-based routing scheme [5-6].

If the plane-based routing scheme is applied to the large-scale sensor network, routing table management and increased routing messages increase routing overhead. To address this problem, as a hierarchy-based routing scheme, Low Energy Adaptive Clustering Hierarchy (LEACH) has been suggested [7-8]. In case of LEACH, due to inefficient clusters and increased dependence on cluster heads, routing overhead increases. In this study, to overcome these problems, combining the plane-based routing scheme with the hierarchy-based one, Virtual Cluster Routing (VCR) is suggested. VCR builds efficient VCR considering node compactness. The virtual cluster communicates with other clusters not only by heads but also by referencing the routing table within the virtual cluster to select the nearest node. Also, as only routing information within the cluster is maintained, overhead for routing table management can be reduced significantly.

## 2 Related Research

### 2.1 Overview of the Sensor Network Routing Protocol

Figure 1, shows the sensor network where each node reads information through its sensor [9].



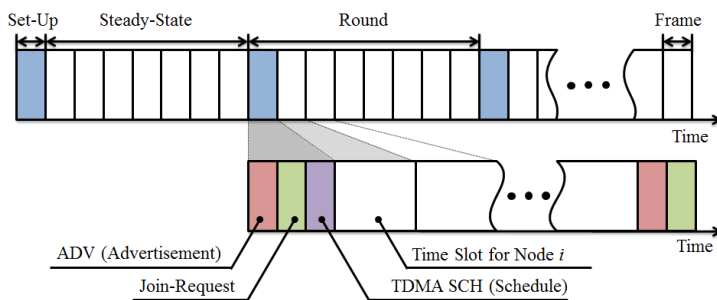
**Fig. 1.** Wireless sensor network architecture

Depending on the network type, the protocol can be divided into the plane routing one and the cluster-based hierarchical routing one. In the plane routing protocol, the network is considered as one area, all nodes equally participate in routing. The cluster-based protocol partitions the network into several clusters and classifies nodes into the hierarchy by its role [4]. Data collected by lower nodes are transmitted to upper nodes, and then the upper nodes

combine them to send to the BS. Well-known protocols are LEACH, LEACH-Centralized (LEACH-C), and TEEN [10-11]. Also, based on network’s operating modes and target applications, it can be classified into proactive one and reactive one. In the proactive network, nodes in the field run only during their cycle to sense and collect data, and then send it to their upper nodes. For periodic data monitoring, LEACH and LEACH-C are suitable. On the other hand, in the reactive network, all nodes in the field sense data sequentially, react to a data change immediately, and then send changed data to their upper node directly. This is suitable for time-critical applications. The most well-known example is TEEN.

## 2.2 LEACH Protocol

The WSN using LEACH consists of multiple clusters. Each protocol, there are the cluster head (CH) which controls all sensor nodes in the cluster, fuses data from sensor nodes, and then send it to BS; and non-CH which collects and sends data to its CH. Especially, as the CH should fuse data from non-CHs, and then sends it to the remotely-located BS, much energy is consumed. So, to have all nodes play a role as CH evenly, whenever the round begins, the CH is selected from all nodes according to specified probability [10].



**Fig. 2.** Timeline showing operation of LEACH

As shown in Figure 2, operation and configuration of the LEACH protocol consist of rounds. In each round, by starting up with set-up, where the head is selected to form a cluster, and then the steady state where data is transmitted from non-CH to CH, and then CH to BS [9-10]. During the unassigned slot time, the node switches to the sleep mode to save energy. When one round ends, a new round begins, a new CH is selected, and then the above procedure is repeated. Even when previously-collected data is the same as the one sensed currently, data is transmitted to the CH. In other words, as unnecessary data is sent among member nodes, consuming energy. At the same time, as the CH is selected according to probability and the cluster is built based on the selected CH’s location, the cluster can be built in the unfavorable geologic structure.

## 2.3 Characteristics of the Sensor Network

In the sensor network, there are processors which process detectable and collected information, small sensor nodes which send such information, and sink nodes which collect

and send such information to the outside [12]. Different from existing networks, the sensor network is basically designed to automatically collect remote information, and is widely used in scientific, medical, military and commercial applications.

The sensor network differs significantly in its application, control and configuration. In the traditional network, Quality of Service (QoS) should be guaranteed, and the configuration, routing and mobility control of mobile nodes for high bandwidth use is important. However, in the sensor network, as many small sensors are running in the environment where people cannot access easily and power cannot be resupplied, energy control for sensor nodes is very important. Furthermore, compared to traditional wireless environments such as Ad Hoc, several hundreds to tens of thousands nodes are compacted to create a sensor network. Therefore, routing many nodes can create routing overhead. This issue should be addressed in the large-scale sensor network.

### 3 Virtual Cluster Routing (VCR)

In The virtual cluster exchange routing information only with its closest node to build a network. At this point, the sensor node in the virtual cluster can send data to different cluster nodes or sink nodes without going through the virtual cluster head. To build a virtual cluster, based on node compactness, virtual cluster heads are selected. And then, based on selected heads, multiple virtual clusters are built. For partition nodes not included in the virtual cluster, the virtual cluster is built. Also, to exchange data among virtual clusters, their level is set.

To select the virtual cluster head, all sensor nodes send the ADV message to themselves and nearby nodes. At this point, the ACK message for the ADV message is not sent. Based on the sum of ADV messages from surrounding nodes, adjacent node information is determined. If a node receives messages whose sum is bigger than the standard value based on the node compactness, that node is selected as the cluster head. If two of neighboring nodes are selected as the cluster head, the node with a bigger ADV message sum will be selected as the virtual cluster head.

To minimize the overhead in building the virtual cluster, the virtual cluster head is selected only once. At this point, the standard value for head selection is cut in half. More nodes can be involved in selecting the virtual node head. The formula for selecting the cluster head is shown in Formula (1).  $\square$

$$\text{Head Selection Standard} = [(N * \pi R^2) / 2A] \quad (1)$$

When the cluster head is selected in the LEACH scheme, according to Formula (2), each node calculates its possibility to become the head cluster.  $C_i(t)$  is an indicator function. During  $r \bmod(N/k)$ , if the node was the cluster head, the indicator function's value is 0, and if not, it is 1. In other words, if a node was the head during  $r \bmod(N/k)$  even once, it cannot be selected as a head again.

$$P_i(t) = \begin{cases} \frac{k}{N - k(r \bmod \frac{N}{k})} & : C_i(t) = 1 \\ 0 & : C_i(t) = 0 \end{cases} \quad (2)$$

In Formula (2),  $i$  is node's indicator,  $t$  is time,  $N$  is number of nodes,  $k$  is number of clusters, and  $r$  is round. During a certain round, as the head is selected from nodes which have not been a head before, the number of rounds increases, resulting in a simple increase in  $P_i(t)$ . This pattern repeats at the cycle of  $N/k$  to have all nodes selected as a head node having equal probability.

## 4 Test and Performance Evaluation

### 4.1 Test Environment

In As shown in Table 1, network size, number of nodes, transmission range and number of heads were selected for various sensor network topologies to conduct tests.

**Table 1.** Test Environments

Topology	Network Size	No. of Nodes	Transmission Range	No. of Heads
A	500m*500m	100	100m	6
B	1000m*1000m	200	200m	12
C	2000m*2000m	300	300m	18
D	4000m*4000m	400	400m	25
E	5000m*5000m	500	500m	31

### 4.2 Survival Time Test

To evaluate the energy efficiency of VCR in the sensor network, the survival time was compared with the existing routing scheme. In this test, the initial energy of 100J was given to Topologies A and E, and then CBR data was transmitted by every 0.5 seconds. Table 2 shows network survival times. As shown here, Topology A has a long survival time in the order of AODV, DSDV, LEACH, and VCR.

**Table 2.** Comparison of Network Survival Times

Routing Scheme	Topology A	Topology E
AODV	1249	199
DSDV	1392	213
LEACH	1434	229
VCR	1521	253

VCR survived 22%, 9% and 6% longer than AODV, DSDV, and LEACH respectively. In the AODV scheme, as route searching messages increased rapidly whenever a new route was set, energy consumption increased proportionally to route searching messages. In the DSDV scheme, as 100 routing tables were maintained to send data, this led to overhead, resulting in high energy consumption. In the LEACH scheme, as the overhead caused by routing tables and routing messages was reduced, energy was less consumed than AODV and DSDV. However, as clusters were formed

every round, energy was consumed more than the proposed VCR scheme. Also, in the test with Topology E where there was a significant increase in the number of nodes, transmission range and network size, VCR's survival time was longer than AODV, DSDV and LEACH by 27%, 19% and 11% respectively.

## 5 Conclusion

In the sensor network, the plane-based routing scheme increases overhead due to routing table management and increased routing messages. To address this problem, the hierarchy-based routing scheme is suggested and has lower overhead by managing the routing table with multiple clusters. However, inefficient cluster formation and dependency on the cluster head selected every round leads to cluster overhead. In the VCR scheme, considering the compactness of sensor nodes existing in the network, the virtual cluster head is selected, resulting in higher efficiency. Also, as the virtual cluster is formed only in the highly compacted area, cluster overhead can be reduced. Moreover, as the virtual cluster level is set, data is sent only within the virtual cluster where the node belongs by referring to the routing table. This can reduce overhead caused by routing messages and routing tables. The test confirmed the proposed VCR scheme had the lower transmission delay in the large-scale sensor network with increased network size, transmission range and number of nodes.

## References

1. Ian, F., Akyildiz, S.W., Sankarsubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. *IEEE Communications Magazine*, 102–114 (August 2002)
2. Anna, H.: *Wireless Sensor Network Designs*. John Wiley & Sons, Ltd. (2003)
3. Paolo, S.: Topology control in wireless ad hoc and sensor networks. *ACM Computing Surveys (CSUR)* 37(2) (2005)
4. Heinzelman, W.B.: Application-specific protocol architectures for wireless networks, Ph.D. dissertation, Mass. Inst. Technol., Cambridge (2000)
5. Jiang, M., Li, J., Tay, Y.: Cluster Based Routing Protocol. IETF Draft (August 1999)
6. Chatterjee, M., Das, S.K., Turgut, D.: An On-demand Weighted Clustering Algorithm (WCA) for Ad-hoc Networks. In: *Global Telecommunications Conference, GLOBECOM 2000*, vol. 3, pp. 1697–1701. IEEE (2000)
7. Luo, H., Liu, Y., Das, S.K.: Routing correlated data in wireless sensor networks: A survey. *IEEE Netw.* 21, 40–47 (2007)
8. Zhang, Y., Gulliver, T.A.: Quality of Service for Ad-hoc On-demand Distance Vector Routing. *IEEE Conference Publications* 3, 192–196 (2005)
9. Cho, S., Shrestha, B., La, K.-H., Hong, B., Lee, J.: An Energy-Efficient Cluster-Based Routing in Wireless Sensor Networks. In: Kim, T.-H., et al. (eds.) *FGCN 2011, Part I. CCIS*, vol. 265, pp. 15–22. Springer, Heidelberg (2011)
10. Koutsonikola, D., Das, S., Hu, Y.C., Stojmenovic, I.: Hierarchical Geographic Multicast Routing for Wireless Sensor Networks. *Wireless Netw.* 16, 449–466 (2010)
11. Dimokas, N., Katsaros, D., Manolopoulos, Y.: Energy-efficient distributed clustering in wireless sensor networks. *Journal of Parallel and Distributed Computing* 70, 371–383 (2010)



# Implementation of Wireless Electronic Acupuncture System

You-Sik Hong<sup>1</sup>, Bong-Hwa Hong<sup>2,\*</sup>, and Baek-Ki Kim<sup>3</sup>

<sup>1</sup>Dept. of Computer Science, Sangji University, Korea

<sup>2</sup>Dept. of Digital Media Engineering, Kyunghee Cyber University, Korea

<sup>3</sup>Dept. of Information & Telecommunication Eng.,

Gangneung-Wonju National University, Korea

h5674korea@gmail.com, bhhong@khcu.ac.kr, bkkim@gwnu.ac.kr

**Abstract.** In this paper we proposed diagnosis system in which we could use simultaneously pulse and tongue diagnosis data, and measurement data of oxygen saturation using technology of bluetooth. And also we developed adaptive wireless acupuncture system by using pulse diagnosis system to adjust strength and time of acupuncture and several acupuncture points of patients for whom intellectual fuzzy technology is applied. With this acupuncture system we can obtain optimal acupuncture time and use at any where and any time easily by input our physical condition to smart phone and in the web. We implemented smart wireless electronic acupuncture system to get acupuncture easily using intelligent diagnosis system.

**Keywords:** fuzzy rules, diagnosis, wireless, acupuncture.

## 1 Introduction

More than 60 percent of the electronic acu punctures are developed in the country using low frequency and the rest is developed using instantaneous electro stimulation. Existing low-frequency therapeutic apparatuses are simple frequency generator(16~32Hz) which attaches the electrodes to patient's diseased area. Patient cannot be treated effectively because it does not provide detailed frequency(three decimal places) but uncertain frequency. Furthermore, it can't find acupuncture points because it has no consideration for patient's sex, age, weight, illness, etc. And it causes problem that children and elderly people are bruised or wounded after getting electronic acupuncture because of inappropriate acupuncture time and strength.[1]. The pulse is considered an important factor in oriental medicine because observation of a person's pulse rate may reflect their health and illness. For example, if patient's heart stopped, it is very serious situation and this situation can be judged by pulse. Oriental doctors have considered pulse rates as important data in diagnosis. But the existing blood pressure pulse analyzer has some problem. It is uncertain whether the blood pressure pulse analyzing sensor is located precisely on the radial artery and it is difficult to

---

\* Corresponding author.

diagnose pulse exactly in different case of thick and thin forearm. Furthermore, the analogue type blood pressure pulse analyzer has problems with quantification of the blood pressure pulse. Although some people may have the same forearm length but the thickness of their blood vessel may differ. Therefore there is no set of data that is considered reliable enough to judge the accuracy of blood pressure pulse rates. Oriental doctors should not only judge the basic biological signals such as checking the pulse's size, strength, and speed, but should also consider the basic and quantitative analysis of the pulse in order to gain an accurate diagnosis. Also, the doctor should consider physical characteristics, such as the thickness of the skin and blood vessels, in order to reach an accurate conclusion. Therefore, measurement of the blood flow rate is a vital indicator in understanding the blood pressure rate and how the substances in the blood are transported.[2][3].

The method of exiting diagnosis has problem which cannot diagnose the old and the infirm exactly because it does not take into consideration the condition of patient's gender, age, skin. To solve this problem, we analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We proposed the algorithm that diagnoses patient optimally considering patient's condition using intelligent fuzzy technique.[4][5]. In this paper we developed adaptive wireless acupuncture system by using pulse diagnosis system to adjust strength and time of acupuncture and several acupuncture points of patients for whom intellectual fuzzy technology is applied. And also we proposed diagnosis system in which we could use simultaneously pulse and tongue diagnosis data, and measurement data of oxygen saturation using technology of bluetooth. The composition of this paper is as follows. Section 2 is about Intelligent pulse diagnosis algorithm, section 3 is about wireless electronic acupuncture system, section 4 is about the simulation of wireless electronic acupuncture, and finally section 5 concludes.

## 2 Intelligent Pulse Diagnosis Algorithm

The intelligent pulse diagnosis system composed of three parts. The first part is composed of the sensor to detect the conductance which corresponds with injured part of human body, and reference signal generator to moderate the signal generated from patients is included. The second part is composed of DSP (Digital Signal Processor) board in which the signals are measured and to do a sort using fuzzy algorithm.[5] The last part is composed of computer system that displays the signal from DSP board to the monitor, and analysis software to diagnose the patients. Fig.1 shows the whole diagnosis algorithm for electronic acupuncture.

Pulse is beat-wave pattern of chest wall and great arteries according to heartbeat. The main purpose of pulse is observation of cardiomotility and blood movement. Recently study using physical characteristics shows that pulse wave pattern can change according to condition of blood vessel and blood circulation. The pulse wave pattern can be obtained by second differentiation of digital plethysmogram using physical specific status such as uncertain inflection point. In this paper, we classified a patient's physical condition into three categories, dangerous, ordinary, normal condition adapting pulse diagnosis algorithm using acceleration pulse wave pattern.

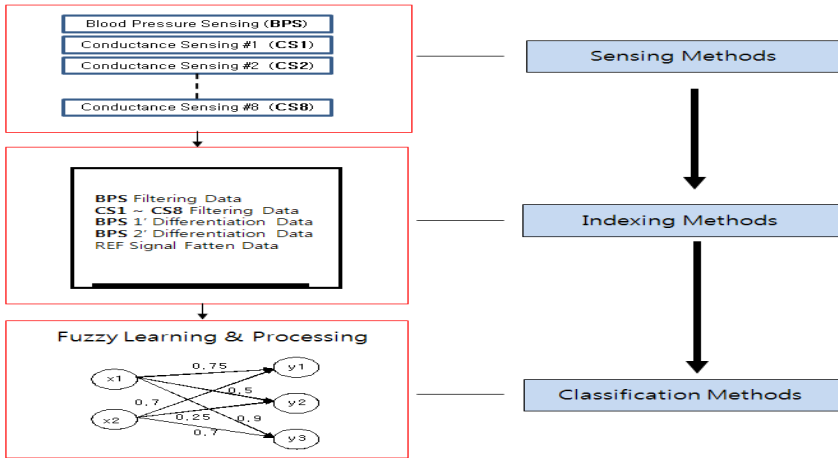


Fig. 1. Diagnosis algorithm for electronic acupuncture

Combination function of trust value: 1 and 2 type of fuzzy creation rule reduced type of 5 and 6 can come to the same node and conclusion through different inference path to infer fuzzy. In this node the same conclusion reach two of more different trust value. In this case combination function of trust value is used to recalculate trust value of conclusion.

$$\beta_c = \beta_{comb}(\beta_c, \beta_c^{old}) = \max(\beta_c, \beta_c^{old})$$

Here  $\beta_c^{old}$  is trust value of conclusion reached through inference path already,  $\beta_c$  is trust value of other conclusion reached through another inference path. If the 4 patients' (a, b, c, d) illness condition is end stages the value is displayed as 0.8-1.0 shown in the left, in case of middle stage the value is 0.4-0.7 and in case of first stage the value is displayed as 0.1-0.3. The value in the middle is displayed patient's physical condition. For example, if the patient's height is 150cm and weight is lower than 45kg the value is displayed as 0.1-0.3, in case of the height 151cm-170cm and the weight 46kg-70kg the value is displayed as 0.4-0.7, and in case of the height 171cm-200cm and weight 71kg-130kg the value is displayed as 0.8-1.0. In fig.1 the process to calculate fuzzy correction factor according to patient's physical condition is shown.

### 3 Wireless Electronic Acupuncture System

Wireless electronic acupuncture system with built in multi pad which can find out the condition of the patients automatically and treat the patients simultaneously. The system includes the function that it can treat the patients with acupuncture adjusted voltage, current, frequency oscillation automatically according to their physical conditions. To perform the function, the system has function to sense and treat acupuncture simultaneously, and required logical and statistical data processing technique using fuzzy and exact analysis.[6]

Fig.2 shows whole diagram of the system. Shown in the left of the fig.2 we can see the 5-pad installed underneath the palm. Installing the 5 circle pad underneath the palm we can exchange the signal, and then adaptive acupuncture treatment can be done. At this point, measurement of the signal use the wireless type instead of cable type. Because the wireless type has advantage of convenience to get acupuncture, reduction of noise by use cable connected computer system and prevention of electric shock according to abrupt high-tension electricity.[7][8][9]

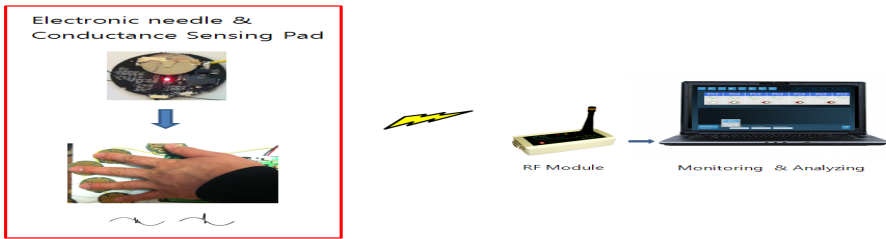


Fig. 2. Full diagram of the system

Information extraction of the human body to treat acupuncture is important not only data from body but also ages, sex, height and weight of patients. To do this, we make control variables using fuzzy algorithm before treatment of acupuncture.

Fig.3 shows Circuit of the Input signal AMP and Acupuncture signal. The part of sensing pad and contact point of the fingertip made of stripe array type to distribute contact point area evenly after plating with gold to reduce electric resistance.

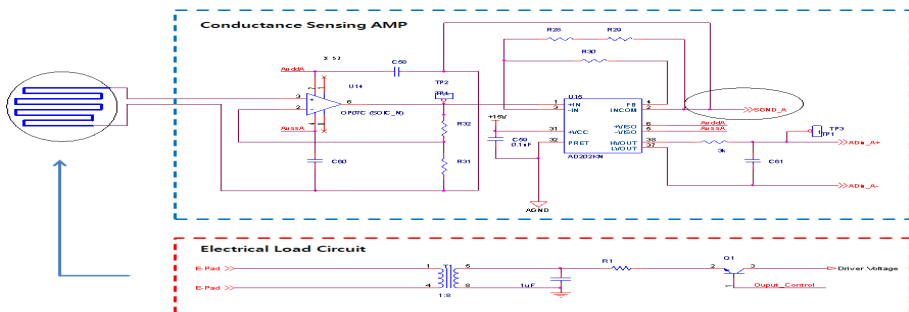


Fig. 3. Circuit of the Input signal AMP & Acupuncture signal

Fig.4 shows circuit of main processor and RF communication part. We made the compact circuit using 3V button battery, and provided expandability to measure another body part later.

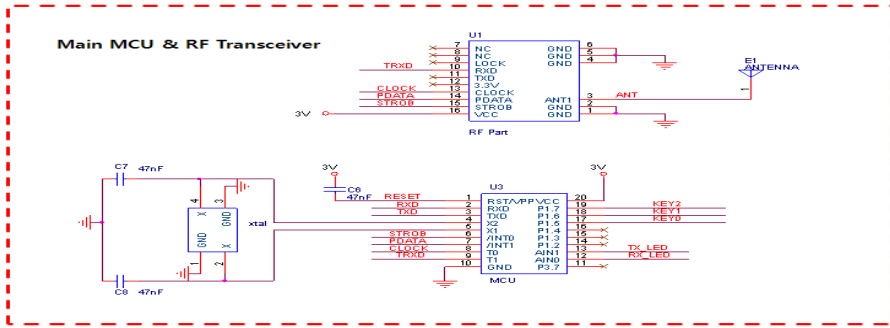


Fig. 4. Circuit of main processor and RF communication part

### 4 Simulation

In this paper, we proposed the optimal algorithm which could judge the remote medical diagnosis using fuzzy logic and fuzzy inference rules, and we simulated the process to calculate the optimal acupuncture time of body condition of patients. We produced the wireless communication part to transmit condition of patients’ pulse, skin conductance and oxygen saturation data to user’s terminal or remote medical terminal, and to receive the control signal from user’s terminal or remote medical terminal.

To do this, we made the sensing pad, the circuit of AMP and acupuncture signal, wireless communication module and charging circuit for storage battery. And also we proposed the software including algorithm of analysis and control using fuzzy technique. Existing acupuncture system using DSP has complex structure, uses up a lot of electricity and it’s big and expensive. But the adaptive wireless acupuncture system proposed in this paper is simple, inexpensive and safe. Fig.5 shows simulation of the glove type electronic acupuncture.

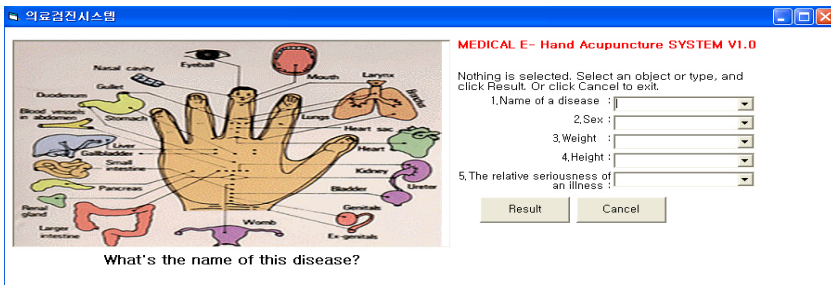


Fig. 5. Simulation of the glove type electronic acupuncture

To implement wireless system we used the way of RF data modem for wireless communication using Narrowband FSK. The feature of this way is robust to noise and it can transmit data easily by simple communication protocol. And this system is

adapt to design multi type data communication system and can be designed by low power, one 3V battery, in case of short distance. We considered not only resistance measurement but capacitive component to reduce error according to several condition of human body. To do this, we applied the pulse wave DC50V~200V, 500uA~1,500uA, intermittent stimulation of 5Hz~5KHz to main pad and fingertip and measured the voltage peak and phase frequency.

We used 470MHz band frequency and designed the system to change 21 physical frequency. And logical address of a channel corresponding to each adaptive acupuncture was assigned using polling technique and then called. The system supports half duplex communication. This way is suitable for the system because the system require low data and uses low speed communication relatively. The output power of wireless signal using button type battery is 1mW, and it is adequate to transmit data without noise. The speed of transmission is 1200~9600bps, wireless encoding uses way of Bi-phase Manchester code, communication between notebook computer and wireless modem uses RS232C.

Fig.6 shows wireless acupuncture system Android-based. In this system remote information is transmitted and received by process as follows.

- ① Measure bio information using sensor equipment for health care(pulse sensor, blood pressure/sugar sensor, ECG sensor, infrared thermometry sensor).
- ② Process measured bio information in the Hmote2420.
- ③ Transmit the information to Android platform through bluetooth using H-Andro210.

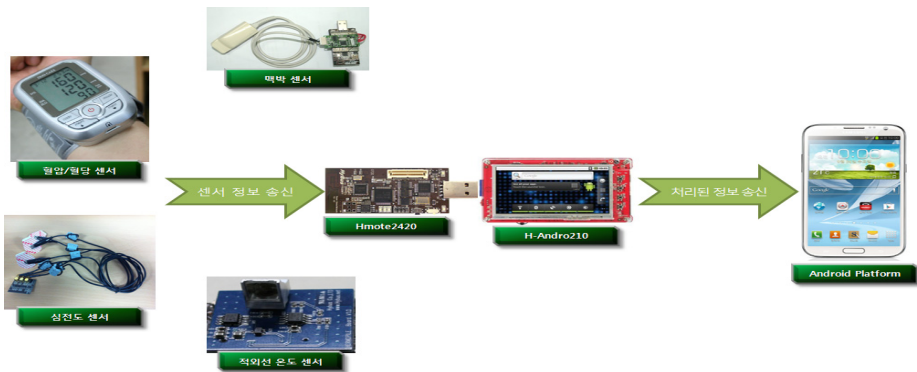
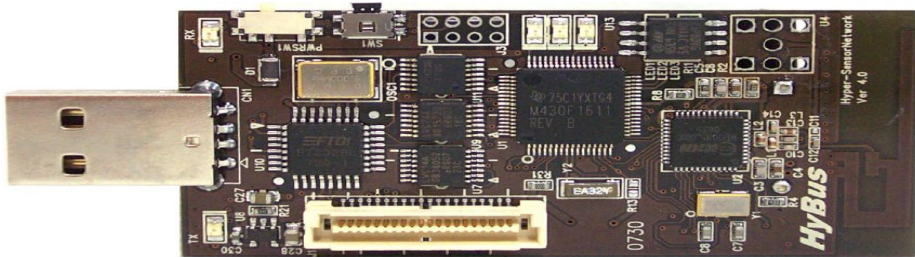


Fig. 6. Wireless acupuncture system

Fig.7 shows data the transmitter and the receiver using RF communication. For remote medical treatment, the transmitter acquire data from 4 sensors, and then transmit the data to receiver using RF communication.



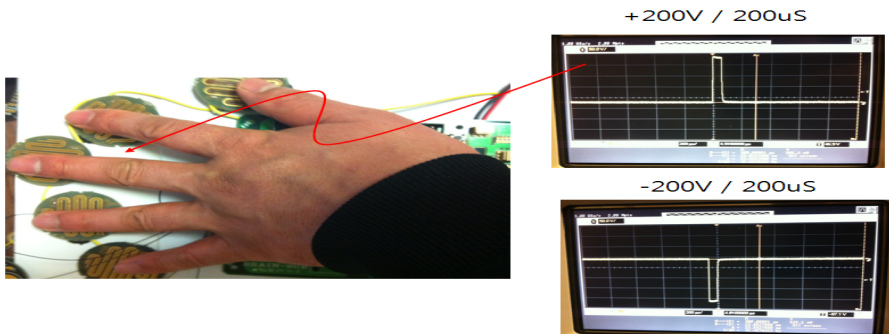
**Fig. 7.** Data transmitter and receiver using RF communication



**Fig. 8.** Transmit/receive system for ubiquitous network

Fig.8 shows transmit/receive system for ubiquitous network. It is made of MSP240CPU and CC2420 RF chip.

Fig.9 shows scene of electro stimulation to fingertips using pads. To obtain signal, we send a reference signal to palm, and then decide body condition of patients on the basis of data obtained from pre-investigation using sensing pads and MCU attached to fingertips. At the same time signal processing is completed, electric stimulation signal generated by fuzzy algorithm is transmitted to sensing pads.



**Fig. 9.** Scene of electro stimulation to fingertips

## 5 Conclusion

Existing acupuncture system using DSP has complex structure, uses up a lot of electricity and it's big and expensive. To solve this problems, we presented intelligent pulse diagnosis algorithm and wireless electronic acupuncture system.

Using proposed algorithm which judge the remote medical diagnosis based on fuzzy logic and fuzzy inference rules, we can calculate the optimal acupuncture time of body condition of patients. We made the sensing pad, the circuit of AMP and acupuncture signal, wireless communication module and charging circuit for storage battery. The intelligent wireless acupuncture system proposed in this paper is simple, inexpensive and safe compared with conventional acupuncture systems.

## References

- [1] Hong, Y.S., Kim, H.K., Kim, B.K.: Implementation of Adaptive Electronic Acupuncture System using Intelligent Diagnosis System. *International Journal of Control and Automation* 5(3), 141–152 (2012)
- [2] Lee, Y.J., Lee, J., Lee, H.J., Yoo, H.H., Choi, E.J., Kim, J.Y.: Study on the characteristics of blood vessel pulse area using ultrasonic. *Korea Institute of Oriental Medicine Researches* 13(3), 111–119 (2007)
- [3] Shaltis, P.A., Reisner, A.T., Asada, H.H.: Cuffless Blood Pressure Monitoring Using Hydrostatic Pressure Changes. *IEEE Trans. Biomed. Eng.* 55, 1775–1777 (2008)
- [4] Sik, H.Y., Kug, P.C.: Proc. of the sixth international fuzzy system association, IFSA, pp. 461–464 (1995)
- [5] Baruah, H.K.: The Theory of Fuzzy Sets: Beliefs and Realities. *IJEIC* 2(2), 1–22 (2011)
- [6] Kim, S.W., Choi, Y.G., Lee, H.S., Park, D.H., Hwang, D.G., Lee, S.S., Kim, G.W., Lee, S.G., Lee, S.J.: Improvement of Pulse Diagnostic Apparatus with Array Sensor of Magnetic Tunneling Junctions. *J. Appl. Phys.* 99, R908–R910 (2006)
- [7] Lee, S.S., Ahn, M.C., Ahn, S.H.: A New Measurement Method of a Radial Pulse Wave Using Multiple Hall Array Devices. *J. Magnet.* 14, 132–136 (2009)
- [8] Haykin, S.: *Modem Wireless communication*. Prentice-Hall (2003)
- [9] Swami, A., Hong, Y.: *Wireless Sensor Networks: Signal processing and communications*. Wiley (2007)



# Network Based Intelligent Agent for Ubiquitous Environments

Dong W. Kim<sup>1</sup>, Ho-Dong Lee<sup>2</sup>, Sung-Wook Park<sup>3</sup>, and Jong-Wook Park<sup>3</sup>

<sup>1</sup> Dept. of Digital Electronics, Inha Technical College, Incheon, South Korea  
dwnkim@inhatec.ac.kr

<sup>2</sup> Mitos Company

<sup>3</sup> Dept. of Electronics, University of Incheon, Incheon, South Korea

**Abstract.** The network based intelligent agent (NBIA) system consists of intelligent software components. The framework of an NBIA system warrants reliable and stable network connection with real-time response between software components. Intelligent software components work on the resources distributed on the network, and the NBIA uses the results from the resources. In this way, the NBIA can provide high-quality intelligent service to users. Thus, a TCP/IP-based interface is designed to integrate software components into an NBIA system. This interface uses the server-client architecture to solve the bottleneck problem with the transfer rate of a wireless LAN.

**Keywords:** Network based intelligent agent, Ubiquitous environment.

## 1 Introduction

High-quality intelligent services need more system resources and a variety of environments. Also, these intelligent services work on a variety of operating systems or OS's (Windows, Linux, etc.), as well as on diverse hardware. Therefore, integrating these services on the network is a very difficult task.

TCP/IP is a traditional method of integrating components via a network. It is reliable and stable, and it is very easy to develop its components. Various architectures and integration schemes have been proposed [1-5]. With the development of components-based software engineering, a middleware such as CORBA is used to integrate components by virtue of its abstract network layer. The developer can use unlimited resources via the network without considering different OS's and diverse hardware. Furthermore, the use of middleware increases software reusability. Numerous architectures that use middleware have been reported [6-10].

The network based intelligent agent (NBIA) system consists of intelligent software components. Thus, the framework of an NBIA system warrants reliable and stable network connection with real-time response between software components. Intelligent software components work on the resources distributed on the network, and the NBIA uses the results from the resources. In this way, the NBIA can provide high-quality

intelligent service to users. Thus, a TCP/IP-based interface is designed to integrate software components into an NBIA system.

In the following, communications interface design based on TCP/IP, will be introduced. This scheme uses the server-client architecture to solve the bottleneck problem with the transfer rate of a wireless LAN.

## 2 Communications Interface Design Based on TCP/IP

Communications TCP/IP(Transmission Control Protocol/Internet Protocol) is a well known and the most popular communications protocol for Internet connection. In this section, communications interfaces based on TCP/IP are designed for an NBIA.

### 2.1 TCP/IP Protocol

TCP/IP consists of an IP(Internet protocol), which is an Internet protocol that uses the packet communication method, and a TCP(transmission control protocol).

Layer	Protocol
Application Layer	TFTP, TLS/SSL, FTP, HTTP, IMAP, IRC, POP3, SMTP, SNMP, TELNET, PNRP, ...
Transport Layer	TCP, UDP, DCCP, SCTP, IL, RUDP, ...
Network Layer	IP (IPv4, IPv6)
Physical Layer	Ethernet, Wi-Fi, Token-ring, PPP, SLIP, FDDI, ATM, Frame Relay, SMDS, ...

**Fig. 1.** Internet protocol stack

An IP does not guarantee packet forwarding, sending, and receiving. Also, the order of the packets sent and received may differ (unreliable datagram service). The TCP protocol operates over the IP, and the TCP warrants the transfer of the data packet and the order of the data packet. HTTP, FTP, and SMTP, including a large number of IP-based application protocols, run on TCP, which is why we call the protocol TCP/IP.

Figure 1 shows the TCP/IP protocol’s stack and related protocols. The hierarchical structure is also shown.

### 2.2 Development of a Network Core Using TCP/IP Protocol

Environment data such as video streams and audio streams from an NBIA are sent to each resource via the main server. Thus, if number of service components increases, then the data throughput of the main server is increased. This situation is shown in Figure 2. Of course there is the multicast protocol, but it is not suitable for sending large amounts of data reliably. In addition, it is difficult to control each resource based on their role, and it is difficult to adjust each resource based on its usage because it has no structural design. Also, the programming complexity of a main server is increased when the main server manages the resources.

To avoid these situations, the main server needs some methods that can classify each service component according to its role.

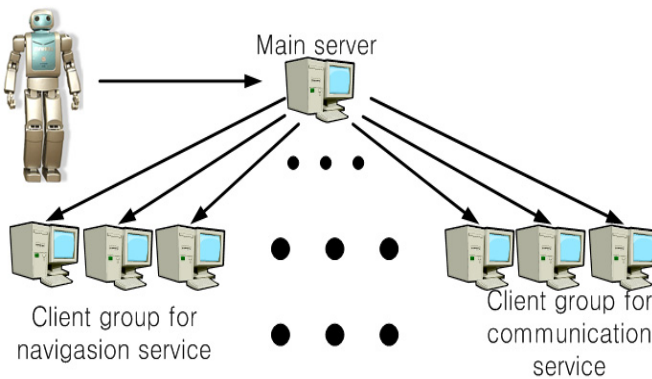


Fig. 2. A case of clients receiving information from the main server

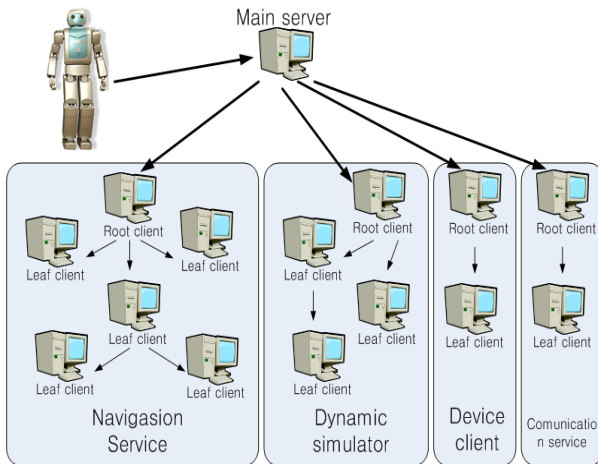


Fig. 3. An example of an NBIA system architecture that uses the tree structure

In this section, a structural design such as the tree structure is used to solve the aforementioned problems. Each resource is formed as a branch of a tree structure. Only one client, the root of the tree structure, is connected to the main server. Thus, the root client receives the data transmitted from the main server, and then transmits the received data to other clients. In this way, the burden of the main server can be reduced.

Figure 3 shows an example of an NBIA system with the proposed tree structure. The system in Figure 3 consists of four service groups, an NBIA and a main server. For each service group that consists of multiple clients, only the root client is connected to the main server. Other clients are connected to the root client of the service group. Therefore, the main server manages only four root clients. Furthermore, each client in the service group is formed within the tree structure. The client's location and mission are shown clearly and conceptually in Figure 3. In addition, the main server can control the service group by controlling the root client.

### 2.2.1 Network Core

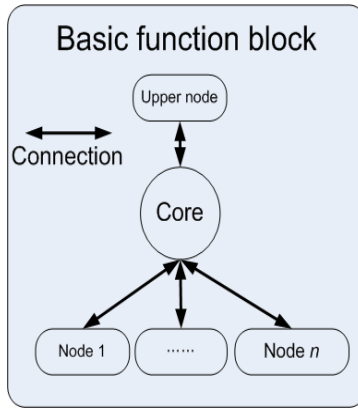
To develop the network framework described in Sec.2.2, a network core that can be configured into the network with a tree structure was designed. Also, the network core serves as a server and a client simultaneously. In Figure 4, the network core that was designed and implemented is shown. By default, the network core performs the roles of a server and a client. It can also, relay data from an input to multiple outputs. At this point, the kinds of data being entered are unlimited. With these structures, clients were configured with a tree structure on the network. Figure 5 shows an extension of a network core. An output of a network core can be an input of other network cores, and the direction of the data movement is bi-directional. Figure 6 shows an entire system configuration that uses this network core.

As shown in Figure 6, four network cores are connected to the main server. They are an input from an NBIA, a PDA, a client, and a service group that consists of four clients. At this point, the main server transmits an input from the NBIA to three clients, and the three clients that comprise the service group receive data from a root client that is connected to the main server. Thus, the root client that receives data from the main server is responsible for realizing the role of the main server with respect to other clients in the service group. Therefore, using the network core that serves as a server and a client simultaneously, each client will be able to organize itself into a tree structure. This reduces the load on the main server.

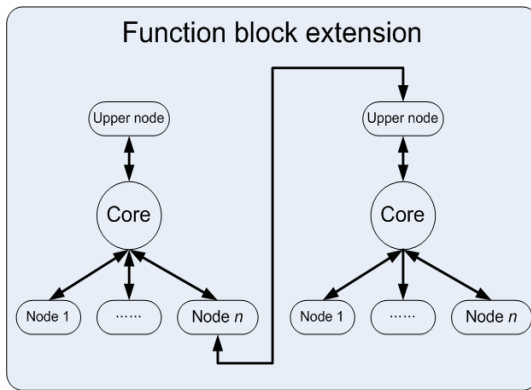
### 2.2.2 Basic Function of a Network Core

Table 1 describes the default functions of the proposed network core. A network core has four parts: the sender, receiver, parser, and sweeper.

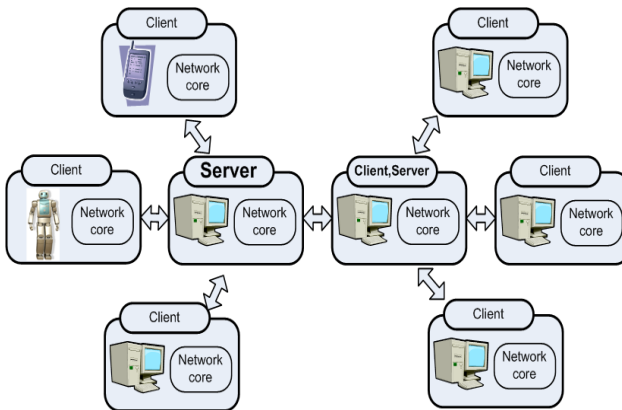
Figure 7 shows a block diagram of the network core. As shown in Figure 7, the sender transmits data packets to other network cores. The receiver receives the data packet from other network cores. The sweeper assembles the packets from the receiver to the data. The parser is responsible for processing the data. A network core has these four basic functions and management roles.



**Fig. 4.** Basic concept of a network core



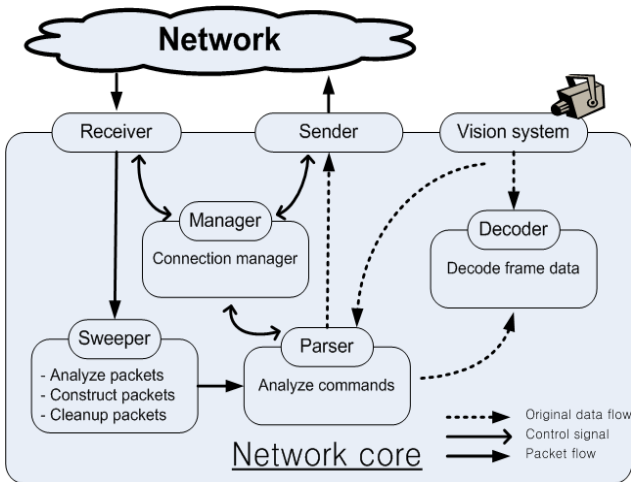
**Fig. 5.** Extension of a network core



**Fig. 6.** A system structure that uses a network core

**Table 1.** Default functions of a network core

Functions of a network core
One-input node
$n$ -output node
Server functions
Client functions
Transfer of any data
Bulk data transfer
Core-to-Core communication
Decompression of MPEG4 streams
Compression and decompression of JPEG images
There are no limitations in the tree depth
DLL implemented as easily as possible for use with other applications



**Fig. 7.** Block diagram of a network core

### 3 Performance Analysis and Discussion

Table 2 shows the network transmission delay of a network core. The transfer rate is measured between cores through a wired LAN, a wireless LAN, and a local machine, respectively. It shows also quick network core data transfer. Note that during the data transfer, the network core partitions and assembles the data.

There is little difference in the delay depending on the size of the transmitted data, but the delay in a wireless LAN increases remarkably. Therefore, data compression is necessary to reduce the size of the data when the data is being sent over a wireless LAN.

**Table 2.** Transfer rate using a network core

Size of transferred data	Send/ Receive ACK	Local machine	Internal machine	Wireless LAN
<b>Command only (28 bytes)</b>	<b>Send</b>	0.194 ms	0.053ms	0.038ms
	<b>Receive ACK</b>	2.152ms	2.161ms	2.838ms
<b>1000 bytes data with command</b>	<b>Send</b>	0.132ms	0.058ms	0.041ms
	<b>Receive ACK</b>	2.243ms	2.750ms	3.376ms
<b>4000 bytes data with command</b>	<b>Send</b>	0.232ms	0.101ms	0.071ms
	<b>Receive ACK</b>	2.218ms	3.209ms	11.166ms
<b>8000 bytes data with command</b>	<b>Send</b>	0.639ms	0.134ms	18.764ms
	<b>Receive ACK</b>	1.863ms	3.399ms	76.023ms

Using the proposed network core, the NBIA system can be configured quickly and easily, and the addition of new service components becomes very easy and fast.

The network complexity increases rapidly, however, with the addition of new service components and increased connectivity between service components. This phenomenon will occur when the service components and the network link increase. For example, if all the components make links to all the components, the network complexity will increase by  $N^2$ . This increased network complexity makes it difficult to understand the connections between the components. In addition, if a component caused the problem, it becomes hard to find the components and to recover the system. Moreover, the naming scheme that uses a fixed IP and a fixed port number based on the number enables a service component to execute on a specific resource and a specific environment. It will not be able to take full advantage of the distributed resources. To solve these problems, another trial will be needed in future.

## References

1. Schuehler, D.V., Moscola, J., Lockwood, J.: Architecture for a hardware based, TCP/IP content scanning system. *High Performance Interconnects*, 89–94 (2003)
2. Tang, W., Cherkasova, L., Russell, L., Mutka, M.W.: Customized library of modules for STREAMS-based TCP/IP implementation to support content-aware request processing for Web applications. In: *Advanced Issues of E-Commerce and Web-Based Information Systems*, WECWIS 2001, pp. 202–211 (2001)

3. Hansen, J.S., Riech, T., Andersen, B., Jul, E.: Dynamic adaptation of network connections in mobile environments. *IEEE Internet Computing* 2(1), 39–48 (1998)
4. Lee, K.B., Schneeman, R.D.: Internet-based distributed measurement and control applications. *IEEE Instrumentation & Measurement Magazine* 2(2), 23–27 (1999)
5. Liao, R.-K., Ji, Y.-F., Li, H.: Optimized Design and Implementation of TCP/IP Software Architecture Based on Embedded System. *Machine Learning and Cybernetics*, 590–594 (2006)
6. Wu, X.: A CORBA-based architecture for integrating distributed and heterogeneous databases. In: *Engineering of Complex Computer Systems, ICECCS 1999*, pp. 143–152 (1999)
7. Dhoutaut, D., Laiymani, D.: A CORBA-based architecture for parallel applications: experimentations with the WZ matrix factorization. In: *2001 Proceedings of the Cluster Computing and the Grid*, pp. 646–651 (2001)
8. Li, B., Lu, Z., Xiao, W., Li, R., Zhang, W., Sarem, M.: An architecture for multidatabase systems based on CORBA and XML. In: *Database and Expert Systems Applications*, pp. 32–37 (2001)
9. Curto, B., Garcia, F.J., Moreno, V., Gonzalez, J., Moreno, A.: An experience of a CORBA based architecture for computer integrated manufacturing. *Emerging Technologies and Factory Automation* 2, 765–769 (2001)
10. Nishiki, K., Yoshida, K., Oota, M., Ooba, M.: Integrated management architecture based on CORBA. In: *Network Operations and Management Symposium*, pp. 3–15 (2000)



# A Study on the Clustering Scheme for Node Mobility in Mobile Ad-hoc Network

Hyun-Jong Cha<sup>1,\*</sup>, Jin-Mook Kim<sup>2,\*\*</sup>, and Hwnag-Bin Ryou<sup>3</sup>

<sup>1</sup> Defense Acquisition Program, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701

<sup>2</sup> Division of Information Technology Education, Sunmoon University,  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708

<sup>3</sup> Department of Computer Science, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701  
{chj826,ryou}@kw.ac.kr, calf0425@sunmoon.ac.kr

**Abstract.** A mobile ad-hoc network is an autonomous collection of wireless mobile nodes that organizes a temporary network without any network infrastructure. Due to node mobility, it is a challenging task to maintain the network topology. In this paper, we propose a stable clustering algorithm that uses node mobility for cluster formation. In the proposed algorithm, the node mobility is measured by counting the time of nodes entering into leaving from its transmission range. The node having the lowest mobility is selected as a cluster head. For topology maintenance with reduced control overhead, the cluster head adaptively controls the broadcasting period of hello message to the measured node mobility. Through computer simulations, it is verified that the proposed algorithm outperforms previous clustering algorithms in terms of control overhead, the rate of node mobility changes and the number of cluster head changes.

**Keywords:** Mobile ad-hoc network, Cluster, Node Mobility.

## 1 Introduction

In an ad-hoc network (MANET), a temporary communication network is set up with mobile nodes alone without the need for an existing network infrastructure. As such, it is particularly useful for places where a network infrastructure can't be installed. MANETs are applicable in diverse areas such as battlefields, emergency situations, education, and conferences.[1]

However, in a MANET, frequent delivery of control messages and reliable delivery of information are difficult due to changes in the network topology caused by frequent node movement, limited available bandwidth and electric energy.

---

\* First author.

\*\* Corresponding author.

To overcome these limitations, a clustering scheme can be used, which manages mobile nodes by partitioning them into groups, resulting in the following benefits: efficient use of network resources, ease of mobility management, and reduction in control message overhead. Currently research into applying this clustering scheme is actively underway[2, 3, 4].

This paper addresses the problems of frequent changes to the topology caused by node movement in a 1-hop wireless ad-hoc network. Specifically, a cluster formation scheme is proposed, one that determines node mobility and elects the most stable node as the cluster head, as well as a clustering scheme which actively adjusts the HELLO message period (HP) according to the mobility of each cluster.

## **2 Related Work**

### **2.1 Clustering Scheme**

The most representative clustering schemes for a MANET are LID (Lowest ID Clustering) in which the node with the smallest ID is elected as the cluster head and CDS (Connected Domain Set) which is based on the domain set[5, 6].

However, in these clustering schemes 1-hop sized clusters are formed, and as such there is overhead from frequent cluster reformation. Although there is the 3hBAC (3-hop Between Adjacent Cluster heads) scheme in which 3-hop sized clusters are formed, it has the limitation of having to always form fixed 3-hop sized clusters.

Existing 1-hop cluster head election schemes for MANETs include LID (Lowest ID) and HD (Highest Degree)[5, 7].

LID is referred to as an address-based clustering scheme as unique node IDs are used to elect the cluster head. Under this scheme, each of the nodes periodically broadcasts its ID in order to elect the cluster head, and the node with the lowest ID is elected as the cluster head[5].

In HD, the cluster head is elected by taking into account node connectivity, and so it is referred to as a connectivity-based clustering scheme. Each node broadcasts the information of its neighboring nodes at the same interval, and the node with the most number of neighboring nodes (density) becomes the cluster head. If thereafter it is the case that the density of a different node that joins the cluster is greater than the density of the current cluster head, it is chosen as the new cluster head[7].

### **2.2 Measure the Mobility of the Nodes**

The MP-AOMDV protocol is based on the AOMDV protocol. It is a routing technique that selects a more stable route by selecting, among several nodes, one with less movement as an intermediate node using GPS information. As the route can be selected based on the type of data to transmit based on multiple routes, it has the advantage of increasing the transmission rate in the network. However, as movement is predicted using GPS signals, they must be processed, and it doesn't work indoors[8, 9].

During route construction, each of the nodes calculates its forward coordinates using location information obtained via GPS, and infers the next location information. The time it takes to break away from the transmission range of two nodes is calculated and the smallest value is found among the routes to the destination and set as the route effective time, and many routes are searched.

Afterward, in the route selection stage, the way the route is selected is different according to the type of data. For streaming data that has to be transmitted on a continual basis, the route with the greatest MET (transmission effective time) is selected. For the type of data that are small in size and must be transmitted quickly, the shortest route is selected based on the hopcount value. In the route maintenance stage, the route is reconstructed before getting broken (based on the route effective time), thereby allowing for continuous transmission.

### 3 Stable Clustering Scheme that Takes Mobility into Account

#### 3.1 Cluster Formation

**Broadcasting HELLO Messages.** In MP-AOMDV, mobility is measured at the time of data transmission request through a RREP. The constituent factors of a RREP are brought as they are. That is, separately from a RREP that occurs at the time of the transmission request, a HELLO message is periodically broadcast to neighboring nodes. The HELLO message contains the current location, the next location, and the transmission range, as with the contents added to a RREP packet in MP-AOMDV.

**Determining Node Mobility.** Each node calculates the time before the end of the transmission of the two nodes by calculating the RET (Route Effective Time) in MP-ADMDV through the HELLO message received from the area. The calculated RET is stored in the routing table; specifically the ID of the node and the RET value are stored.

**Electing the Cluster Heads.** In the cluster head election stage, cluster heads are elected based on the following procedure.

- ① All nodes broadcast a HELLO message at the initially set interval.
- ② Each node's own RET value is compared to the RET values of all other neighboring nodes to determine whether it is the cluster head or a member node.
  - If the node's own RET value < max. RET value of neighboring nodes then the node = cluster member
  - If the node's own RET value > max. RET value of neighboring nodes then the node = cluster head
  - If the node's own RET value = max. RET value of neighboring nodes then the node with the lowest ID = cluster head

## 4 Experiment

The following tests was done in order to check the performance of the clustering scheme proposed in this paper, which tested the number of times HELLO messages are broadcast as nodes move. In addition, the proposed scheme was compared with existing schemes LID and LIDAR.

Figure 1 shows the total number of HELLO messages sent by all nodes according to the node movement speed. As shown, as in LID cluster mobility is not reflected but a fixed HELLO message interval is used, the number of times a HELLO message is sent is constant. In contrast for LIDAR and the proposed scheme, as the HELLO message interval is adjusted according to mobility within the cluster, it changes as the movement speed of nodes changes.

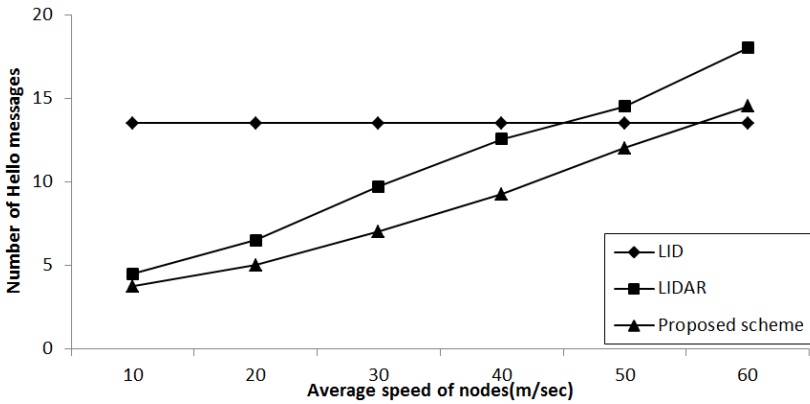


Fig. 1. Total number of HELLO messages as to maximum velocity of node

## 5 Conclusions

This paper proposed a scheme to form stable clusters in a mobile ad-hoc network by using node mobility. With existing clustering schemes, overhead occurs when clusters are reformed because node mobility is not taken into account, and because the size of clusters is small as well as fixed.

In this paper, node mobility is measured and clusters are formed based on cluster heads with low mobility, and as a result overhead from reforming clusters can be reduced, as well as routing overhead internal to the clusters.

For future work, a way of calculating mobility when the movement speed or direction of nodes arbitrarily changes in a mobile ad-hoc network needs to be studied, as well as optimal cluster formation based on this.

## References

1. Hong, X., Xu, K., Gerla, M.: Scalable routing protocols for mobile ad hoc networks. *IEEE Network* 16(4), 11–21 (2002)
2. Chatterjee, M., Sas, S.K., Turqut, D.: An on-demand weighted clustering algorithm (WCA) for Ad Hoc Networks. In: *Proc. of IEEE GLOBECOM*, pp. 1697–1701 (2000)
3. Jiang, M., Li, J., Tay, Y.C.: Cluster Based Routing Protocol (CBRP). In: *Internet Draft draft-ietf-manet-cbrp-spec-01.txt* (1999)
4. Wei, D., Chan, H.A.: A Survey on Cluster Schemes in Ad Hoc Wireless Networks. In: *Proc. of IEEE Mobility Conference*, pp. 1–8 (2005)
5. Ephermides, A., Wieselthier, J.E., Baker, D.J.: A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling. *Proc. IEEE* 75, 56–73 (1987)
6. Wu, J., Li, H.L.: On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks. In: *Proc. of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pp. 7–14 (1999)
7. Yu, Y.J., Chong, P.H.J.: A survey of clustering schemes for mobile Ad Hoc Networks. *IEEE Communications Surveys* 7(1), 32–48 (2005)
8. Cha, H.-J., Han, I.-S., Ryou, H.-B.: QoS routing mechanism using mobility prediction of node in ad-hoc network. In: *ACM MOBIWAC 2008*, New York, pp. 53–60 (2008)
9. Perkins, C., Royer, E.: Ad-hoc on-demand distance vector routing. In: *IEEE WMCSA 1999*, pp. 90–100 (1999)

# Design of User Access Authentication and Authorization System for VoIP Service

Ho-Kyung Yang<sup>1\*</sup>, Jin-Mook Kim<sup>2\*\*</sup>, and Hwang-Bin Ryou<sup>3</sup>

<sup>1</sup> Defense Acquisition Program, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701

<sup>2</sup> Division of Information Technology Education, Sunmoon University,  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708

<sup>3</sup> Department of Computer Science, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701  
{porori2000,ryou}@kw.ac.kr, calf0425@sunmoon.ac.kr

**Abstract.** VoIP, which is used to deliver voice data on the Internet, is being welcomed as a means of replacing the PSTN. In VoIP, voice data are converted to Internet protocol data packets so that they can be delivered in an ordinary IP network. Thus, compared to ordinary telephone networks, it is low cost and highly extensible.

As VoIP services gradually gain traction, issues are coming to the fore, specifically security vulnerabilities and lowered service quality. To mitigate this, in this paper an authentication system is designed which an AA (Attribute Authority) server has added to VoIP in order to increase security and discriminate user access.

**Keywords:** VoIP, Access Control, Authentication System, Authorization System.

## 1 Introduction

When it comes to multimedia techniques, as networking techniques advance, the link with the Internet - which connects the entire world - is accelerating. Demands on services such as video conference and VoIP (Voice over Internet Protocol), which use the same IP (Internet Protocol) network to deliver multimedia data, including audio and video data, are quickly increasing.

Although VoIP efficiently provides voice communication between terminals, for it gain greater use, a variety of services is needed. Examples of additional services include various types, including call transfer, call forwarding when busy or when there is no response, call reservation, call waiting and call filtering. As a signaling protocol of VoIP for users to register services that they want at any time using a simple way, SIP and H.323 in particular are getting the attention.[1]

---

\* First author.

\*\* Corresponding author.

Although increase in the number of users is expected for VoIP, there can be various problems with the packet network from a security standpoint in that anyone can access it as it is an open network. While a PSTN can be attacked only by physically accessing it, when it comes to a VoIP even remote attackers can easily alter signaling messages or wiretap voice packets. Standardization of the SIP began at the IETF by considering extensibility, component reuse and interoperability as key criteria.

In addition, RFC 3261 recommends the use of a stable security model as a security technique for SIP. SIP provides secure messaging services using digest user authentication, TLS, and S/MIME. Media security is implemented by using SRTP (Secure RTP), which is currently being drafted.

Although using a stable security model can secure security, there is a disadvantage that the quality is drastically reduced for users, making it inconvenient for use.

In this paper a system is designed that addresses security problems caused by the increase in the use of VoIP services and for providing discriminate services according to user access privileges. This paper is organized as follows: Chapter 1 gives the introduction; Chapter 2 is on related research; Chapter 3 describes the proposed technique and system; Chapter 4 implements the system and analyzes its the performance; and Chapter 5 gives the conclusions.

## **2 Related Research**

### **2.1 VoIP**

VoIP is a service that uses the packet network originally designed for data communications for Internet telephony. It is a communication service that converts voice data to Internet protocol data packets so that calls can be made over the ordinary telephone network. Compared to the traditional telephone network service, it is low cost, supports multiple users simultaneously over the cable, and is highly extensible. Some of the protocols used are SIP and H.323 [2][3][4].

### **2.2 Attribute Certificate**

The attribute certificate refers to a type of certificate that plays a special role according to the particular environment rather than the certificate for personal identification as information protection services of various purposes increase in e-commerce. This type of certificate is used only for a specific goal and has a shorter lifespan than certificates used for personal identification. It can be used along with personal identification certificates. It has diverse applications in many fields such as network access control, billing according to access to contents, and web page access control.[4][5][6]

### 3 Proposed Technique

The following are the prerequisite for the proposed technique. The AA server and KMS server goes through authentication beforehand and know each other's public key values. The user generates a public key and a private key based on the PKI authentication technique, registers the public key with the KMS server and requests for a certificate to be issued. The KMS server includes the public key value of the ADD server when issuing the certificate.

#### 3.1 User Registration Process

This is the process of registering the user before using the service. The register server issues a user certificate and the location server stores this. The register server and the location server are physically at the same location. The user registration algorithm, in which the user is registered with the register server and the certificate issued, is as follows.

#### 3.2 Service Operation Structure

Communication using SIP involves going through a call connection process, during which various pieces of information may be leaked such as sender/receiver information, encryption technique and method of communication. Therefore a secure call setup is needed. An authentication server and a KMS server are added based on the SIP call setup in the existing VoIP environment for the authentication process.

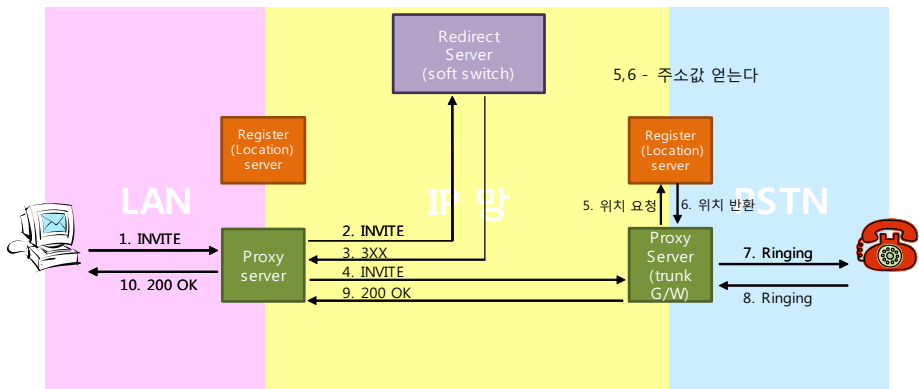


Fig. 1. SIP protocol session setup process

The servers authenticate each other beforehand and share their public key values. In the call setup stage the sender first sends a hello message and its certificate to the proxy server, which checks the certificate and sends a response message that messages have to be encrypted.



The user sends an INVITE message by obtaining the public key of the proxy server from the response message for encryption. Before sending the message, its public key is generated based on that public key.

The proxy server sends to the AA server the INVITE message and a public key certificate that includes a random number R and a hash value (R).

The user is identified using that certificate and the AA server sends the attribute certificate and the contents received from the proxy server. The SMS server receives that information, reviews the contents of the user certificate and the attribute certificate and sends the other party's address value and certificate.

The proxy server encrypts using the public key obtained from the other party's certificate and sends it. The proxy server on the receive side does authentication of the sender at SMS. Also, the sender's attribute certificate is verified at the AA server.

When this process is complete the proxy server sends a message to PSTN, and the telephone network sends the message using bell sounds.

If the process is successfully complete, a response message of "200 OK" is sent to indicate the call has been connected. The sender sends "ACK" to indicate that the message has been received successfully. This completes the call connection.

When secure call setup is complete, data transmission begins with the RTP protocol.

## 4 Implementation and Performance Analysis

### 4.1 Test Environment

For implementation of the proposed system, the RedHat 9.0, gcc V3.2.2 compiler of Linux Kernel V2.4.20.8smp was used.

The library was implemented based on text using a header file called "sip.h", which is freely provided on the Internet. TLS functionality was implemented using the openssl library.

### 4.2 Performance Analysis

This section describes the results of testing each of the following systems based on a test using a VoIP system based on the implemented SIP protocol: an ordinary VoIP system, a TLS-applied system, and the proposed system.

The tests were conducted based on the INVITE command which makes the CALL. It was assumed that account registration has been done.

**Comparison of Security Aspects of Each System.** Figure 2 compares the response times for different number of INVITEs in the VoIP system. For a typical system, the average response time is 5-6ms for 20 INVITE requests.

As shown in the figure, although the proposed system had a longer response time than a typical unsecure system, it was shorter than a TLS system.

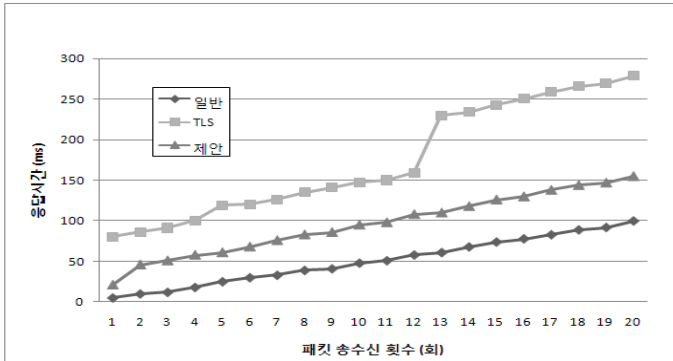


Fig. 2. The Response time with the number of INVITE

**Comparison of Advantages and Disadvantages of Each.** While the ordinary VoIP system has fast response speed and low load on the system, its level of security is poor, and as a result systems with TLS added have become almost a de facto standard.

But while these systems have excellent level of security, as a TLS session has to be set up every time a session is set up for each server, response time is slow and there is a lot of load on the system compared to ordinary systems.

For the proposed system, however, an adequate level of security is provided while having fewer loads on the system than TLS-based systems.

As shown in Figure 4.5, the proposed system allows for user-specific access control, billing can be easily set, and various additional services can be provided for each user. A disadvantage is that a stage is needed for setting attributes for each user.

## 5 Conclusions

VoIP services which deliver voice data on the Internet are being welcomed as a means for replacing the PSTN.

As VoIP services gain more traction, problems started to appear in terms of QoS and security. In this paper an authentication system is designed which is made secure and provides differentiated services according to user access. It does this by adding an AA server to the VoIP sessions setup stage. For future work, ways to increase QoS would need to be studied.

## References

1. RFC 2617, HTTP Authentication: Basic and Digest Access Authentication. IETF (1999)
2. RFC 2402, IP Authentication Header, IETF IPSec WG (1998)
3. RFC 2246, The TLS Protocol Version 1.0, IETF TLS WG (1999)
4. 임채훈, "VoIP 시스템에서의 보안기술", (주)퓨처시스템자료
5. RFC 3261, SIP: Session Initiation Protocol (June 2002)
6. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195-197 (1981),  
<http://www.ietf.org/html.charters/sip-charter.html>,  
 Session Initiation Protocol (sip) Working Group

# Approach of Secure Authentication System for Hybrid Cloud Service

Jin-Mook Kim\* and Jeong-Kyung Moon\*\*

Division of Information Technology Education, Sunmoon University  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708  
{calf0425, moonjk}@sunmoon.ac.kr

**Abstract.** Desire for cloud service is very increasing recently. But users are using cloud service actually very few. Because a private cloud service is very expensive and have a many restrict condition. On the other hand, public cloud service, if you can use it than you have to pay a usage fee when you needed external resources that have been published. However, public cloud service is a fear that the privacy of personal information or leaked, put to save the information that requires security. Therefore, the demand for hybrid cloud service has occurred. But, in order to provide a hybrid cloud service, is often difficult to be applied as an authentication system that was used in the cloud service existing. Therefore, I would like to propose a secure authentication system for the hybrid cloud service in this paper. We will secure authentication system for the hybrid cloud service that is provided security, availability, applicability. Our proposed system has a very good capability in suggested condition.

**Keywords:** Hybrid cloud service, Authentication system, Secure authentication, MITM.

## 1 Introduction

Public cloud service is a consideration economical to individuals and businesses that is intended for use with virtualization and (Virtualization) Technology Company that provides a cloud service professional is to build the system resources of software or hardware. It is a service that allows you to provide a service to pay. However, in providing the service, it has a problem that the user associated with the storage of information privacy and security issues. Therefore, public cloud user exists only about 22% currently.

The Private cloud service, if the operation is to allow only inside the company, was also set to live with the problem and cost security, ease of use, companies of about 20 percent is used in the country.

---

\* First author.

\*\* Corresponding author.

Thus, services that allow you to selectively use only advantage of Private cloud service and Public cloud service is Hybrid cloud service.

However, there is no service provider that provides (inter-cloud services) hybrid cloud services in the country.

Inter - To provide the (Inter-cloud) service cloud, between network and cloud service providers, database, security service providers, issues of cost and service models that can agree on is each other, ease of use, security real problem is because too many to solve the problem.

In order to provide a Hybrid cloud service, and accurate understanding of the security threat elements that occur in the public cloud service and private cloud service always, based on this, so that it can provide security services suitable for hybrid cloud services. There is a need to most important security service. an authentication service for the individual objects that make up the cloud services and user authentication.

2-Factor authentication services are provided in the previous studies in this context is proposed. However, this alone, there is a difficulty to provide authentication services that are suitable for cloud computing a rapidly changing environment. Therefore, we would like to propose an authentication system of the new cloud services with improved 2-Factor authentication system proposed in the existing studies.

We can then provide authentication differentiated services for Public cloud service using the 2-Factor existing authentication systems, the Private cloud service, and the authentication system that applies RADIUS (Remote Authentication Dial In User Service) I proposed. New authentication system which we proposed is able to improve the safety, availability and adaptability than conventional systems.

The paper is organized as follows. A related study, we describe cloud services, 2-Factor Authentication service, for RADIUS Section 2. In the third chapter, I'm writing about the new authentication system we provide to see the security threats that can occur in the cloud service existing. In the fourth chapter, we investigated on the basis of new cloud security authentication system that has been proposed, availability, and applicability. It is the conclusion in Section 5 at the end.

## **2 Related Works**

In this paper, we describe the related work of three. We hope we have demonstrated for the cloud services that you have proposed to exist. Describes the 2-Factor Authentication service of the second, at the end, it was written about RADIUS for mutual authentication in a wireless network environment of exist.

### **2.1 Cloud Services**

A cloud service is a service so that you can easily use resources of software and hardware of existing, so that you can share with the virtualization technology. Conventionally, by using a cloud service, it is possible to save the purchase costs of

resources, use resources effectively to manage and what was used to buy resources directly or software required hardware it has the advantage of being able to. Cloud services, can be divided into Hybrid cloud service Private cloud service, and Public cloud service, depending on the range to provide resources hardware or software. First Private cloud service is a system that is designed for a company to provide cloud services individually. There is a drawback, which can provide the strongest security, but the implementation cost is very high. Public cloud service is a system that can provide a cloud service published in the second. However, Public cloud service outflow of information of people who use the service, to provide security services have the drawback is difficult. Finally, Hybrid cloud service is a system that can be selectively applied to the advantages of Public cloud service and Private cloud service described above.

### 2.2 Two-Factor Authentication Service

The Authentication service, it is what service requestor have created the identity of their own to the system in advance, to examine the validity of the credentials of their own each time you request the service. The authentication service representative some ID-PW method, Token method, SSO, and PKI. It should be stored in the authentication server before the special user identification information each time the user requests a service, to check service utilization whether the request is valid, these, ID-PW, Token, Certificate it is a service to check and is correct. However, the authentication existing services, there is a point that is insufficient for use in the cloud service.

I think our authentication service in a variety of authentication system described above, suitable for the cloud with 2-Factor authentication service. 2-Factor authentication service is a method of mixing two or more ID-PW existing method, Token scheme, USIM method or MTM scheme can provide authentication services. Figure 1 shows an example of a method that can provide a 2-Factor authentication service.

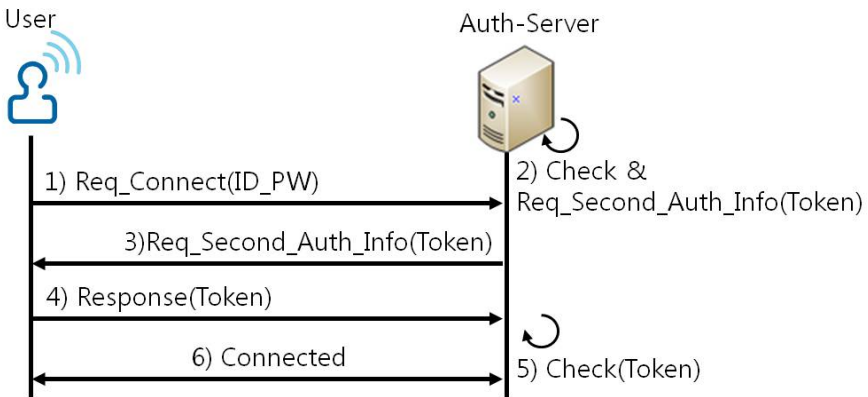
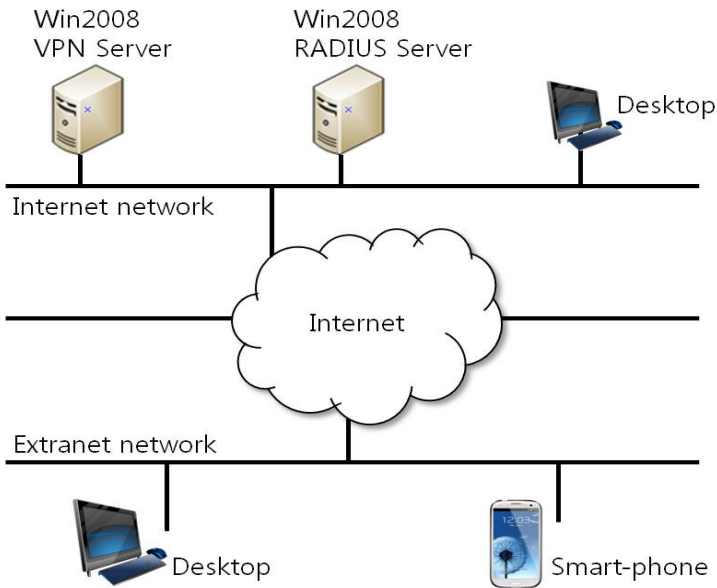


Fig. 1. Example of 2-Factor authentication service (used ID-PW and Token)

As shown in Figure 1, after requesting authentication information sequentially two. It was confirmed by the authentication server, and designed to provide a service requested by the user. First, let's look at the example checks the PW and ID, to check additional Token that had been decided in advance with the authentication server.

### 2.3 RADIUS (Remote Authentication Dial In User Service)

In a wireless network environment, RADIUS is the authentication service to be able to provide authentication services to and from the service provider and the user with a token. The authentication server can generate a token, not only the authentication for the user identification, which was requested by issuing the Token to verify the identity information between the service provider and the user in advance it is intended to be able to provide authentication services for the service. Figure 2 shows for the RADIUS.



**Fig. 2.** RADIUS

As shown in Figure 2, smart-phone existing outside network or in order to request a service to connect to a desktop that is present in the internal network via the wireless devices and network environment phone and via the RADIUS server first save pre-authentication information currently is confirmed the authentication. And It connect wireless devices on the external network and desktop internal network through a VPN service.

### 3 Proposal System

#### 3.1 Threats of Hybrid Cloud Service

In order to provide the appropriate authentication system for hybrid cloud service, the correct understanding of the potential security threats that occur in existing cloud services is essential. Threat was expected to occur in hybrid cloud services are as the following list.

- (1) External network, middle browser attacks and man-in-the-middle attack is capable of generating between authentication systems that exist in the internal network. To destroy the availability of authentication system that exists in
- (2) Internal network, the threat of distributed denial of service attack and denial of service attack is large. The wireless device that is connected to
- (3) External network, the movement of the position occurs so frequently, device authentication of the device is difficult. I have a vulnerability to attack by script
- (4) Internal attacker Just the user authentication system
- (5) Existing provide authentication services to the internal network, but the protection of information and authentication of the public cloud services. You have the vulnerability to external network.

As list shown above, also cloud service, vulnerabilities can occur authentication service has on an existing network not only exist, special vulnerabilities characteristic of cloud services is caused by its existence to.

#### 3.2 Proposal System

Authentication system suitable for hybrid cloud services environment, our proposal system has a structure in which five independent components. Two most important components of the five authentication system provided are the RADIUS server and a Hash Machine. Component of the two processes the authentication of service request and status of the user. I have shown in Figure 3.

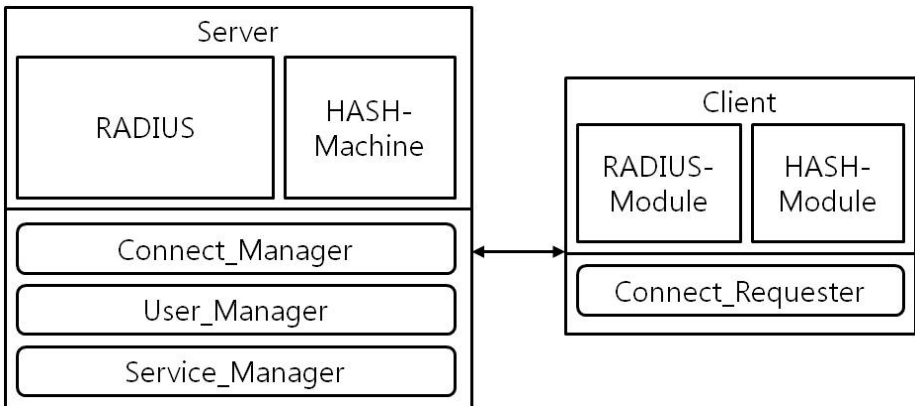


Fig. 3. Architecture of proposal system

Third component is the connection manager. When a user requests a cloud service, which manages the session - connection status - of the authentication service to complete the works. PW is the ID of the connection request user administrator; the user's fourth component is a component for creating users for the required information additional authentication, storage, and or management.

Finally, it is the creation of licensing authority information and services required for the authentication of the user service requirement, storage, service administrator to manage.

### 3.3 Procedures of Our Proposal System

Authentication system for hybrid cloud services which we proposed in the paper performs the processing procedure is divided into two stages increases. I showed in figure 4.

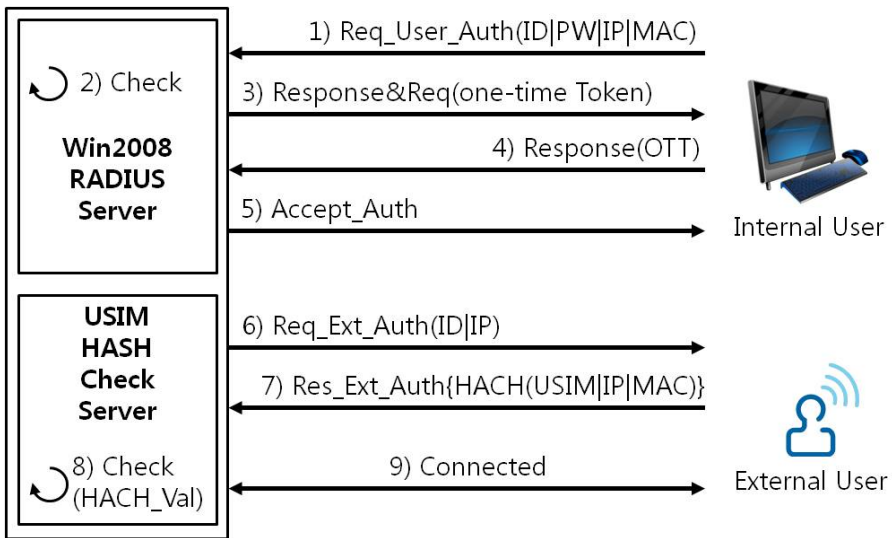


Fig. 4. Procedure of proposal system

The first process, even in the internal or external users, from step 1 to step 1, we describes the process of processing the service request in a wired environment. This was able to perform user authentication using the Token for temporary use with the ID-PW information and provide services.

- (1) Users to send the {ID-PW|IP|MAC} to the authentication server for request to user authentication.
- (2) RADIUS server checks the ID-PW value sent by the user.
- (3) User authentication request is valid, the authentication server, requesting to send a temporary token that responds that the user authentication request is valid, promised in advance.



- (4) The user generates and transmits a token to be used temporarily in the authentication server.  
Make sure the temporary token sent by the user is correct.
- (5) If it is correct, authentication server performs user authentication.

The second process of treatment is from step 6 to step 9. In this procedures, we shows the process when it requests a service using a wireless device. In order to check the user service requirement whether valid.

- (6) The server is required to verify the IP address and the user ID. USIM stored in the wireless device to the service request, IP.
- (7) The user responds to the value that is treated by the hash function that is promised in advance MAC.
- (8) Authentication server to verify the hash value that the user has transmitted. If the service request, user is legitimate.
- (9) System provides the appropriate services to the user.

## **4 Appraisalment of Proposed System**

### **4.1 Appraisalment of Security**

Authentication system of hybrid cloud proposed is a safe middle browser attacks and man-in-the-middle attacks. we in this paper. Between the authentication server and the user, by a malicious user receives the token temporarily or ID-PW, the user authentication system that we provide we, temporary token also perform man-in-the-middle attack since it is not possible to know the value of the secret used to encrypt the value of the token and method as promised in advance to generate, it is not possible to perform intermediate browser or attack middle attacks.

### **4.2 Appraisalment of Availability**

Our proposed hybrid cloud services to provide authentication system, that allocated before the firewall have advance against of DoS(denial of service attack) or DDoS(distributed denial of service attacks). And it cannot harm to our system placed in the DMZ area by the availability.

### **4.3 Appraisalment of Applicability**

In this paper, we propose a hybrid authentication system for cloud service that consist to private cloud service and public cloud service by complementary relationship. And we configure the network environment for hybrid cloud service in the DMZ area. By doing so, we offer cloud services to existing authentication systems, while maintaining the same applicability can be improved.

## 5 Conclusion

Virtualization Technology requires a computer system with an existing one cloud service to connect a lot of resources to improve the ability to provide services to users, is a technique that can be expected. But a private cloud service to connect with many dedicated resources to provide users high cost because it has the disadvantage that occurs. In contrast, public cloud service is installed on the external network to the existing hardware resources by connecting technology to provide services to public safety has a problem because it is.

In this paper, Hybrid cloud service to receive the most attention in the future was expected. Hybrid cloud service, but for the certification system and related services are until now not sleep.

We the above environmental constraints Hybrid cloud service was proposed for the authentication system. Our proposed Hybrid cloud service authentication system for 2-Factor authentication service with existing RADIUS service who presented in ways designed to take advantage of.

Our proposed Hybrid cloud service authentication systems for security, availability, in terms of applicability as compared to existing methods are expected to be excellent. In this paper, a hybrid cloud services, but restrictive for all matters were not considered. In future studies, in order to provide hybrid cloud services, which may occur for a variety of constraints further review of the existing authentication system to accept the lack of information about the in-depth research studies continue to perform better plans.

## References

1. Kim, T.H., Kim, I.H., Min, C.W., Yeom, Y.I.: Security Technical Trend of Cloud Computing. *Computer Science Magazine* 30(1), 30–38 (2012)
2. Jansen, W., Grance, T.: *Tuidelines on Security and Privacy in Public Cloud Computing* (2011)
3. Mannan, M., Kim, B.H., Ganjali, A., Lie, D.: Unicorn: Two-factor Attestation for data Security. In: *Proc. Of the 18th ACM conference on Computer and Communications Security* (2011)
4. Bang, Y.H., Jeong, S.J., Hwang, S.M.: Security Requirement Development Tools of Mobile Cloud System. *Information and Communications Magazine* 28(10), 19–29 (2011)
5. Hubbard, D., Sutton, M.: *Top Threats to Cloud Computing V1.0*. Cloud Security Alliance (March 2010)
6. Liao, I.-E., Lee, C.C., Hwang, M.S.: A password authentication scheme over insecure networks. *J. Comput. System Sci.* 72(4), 727–740 (2006)

# Intelligent Inference System for Smart Electronic Acupuncture

You-Sik Hong<sup>1</sup>, Chang-Hoon Choi<sup>2\*</sup>, and Baek-Ki Kim<sup>3</sup>

<sup>1</sup>Dept. of Computer Science, Sangji University, Korea

<sup>2</sup>School of Computer Information, Kyungpook National University, Korea

<sup>3</sup>Dept. of Information & Telecommunication Eng.,

Gangneung-Wonju National University, Korea

h5674korea@gmail.com, hoon@knu.ac.kr, bkkim@gwnu.ac.kr

**Abstract.** In this paper, we proposed the system that diagnoses a patient optimally considering the patient's condition using intelligent fuzzy technique. We designed the system to respond to the various patterns and to sense the situation which potential difference is changed according to the patient's painful part simultaneously. It contains the function that a patient can search the exact point of electronic acupuncture and check on optimal strength and time of electronic acupuncture considering the patient's body conditions. The system includes the hardware to provide protection function for safety and to support the multimode function of electronic acupuncture through change of control mode.

**Keywords:** inference, fuzzy rules, acupuncture, diagnose.

## 1 Introduction

Oriental doctors have considered pulse rates as important data in diagnosis. But the existing blood pressure pulse analyzer has some problem. It is difficult to standardize the pulse exactly because thickness of their blood vessels is different even if the thickness of two person's forearm is equal. And it is uncertain whether the blood pressure pulse analyzing sensor is located precisely on the radial artery. The analogue type blood pressure pulse analyzer has problems with quantification of the blood pressure pulse.[1] Therefore there is no set of data that is considered reliable enough to judge the accuracy of blood pressure pulse rates. In order to gain an accurate diagnosis, oriental doctors consider the patient's pulse by the basic biological signals such as checking the pulse's size, strength, and speed, and also the basic and quantitative analysis of the pulse. And the doctor should consider physical characteristics, such as the thickness of the skin and blood vessels, in order to reach an accurate conclusion.[2],[3] But the method of exiting diagnosis has problem which cannot diagnose the old and the infirm exactly because it does not take into consideration the condition of patient's gender, age, skin. Most of the conventional

---

\* Corresponding author.

electronic acupuncture systems are made by using low frequency and the rest are made by using momentary electro-stimulation. They can't treat the patients effectively because it uses uncertain and vague frequency. Furthermore, it can't find acupuncture points because it has no consideration for patient's sex, age, weight, illness, etc. And it causes problem that children and elderly people are bruised or wounded after getting electronic acupuncture because of inappropriate acupuncture time and strength. In this paper, to solve these problems, we proposed the algorithm that diagnoses patient optimally considering patient's condition using intelligent fuzzy technique. We analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We classified the patients by their body, illness and age, and calculated the exact time of electronic acupuncture suitable for patient's physical condition using fuzzy logic and inference. The composition of this paper is as follows. Section 2 is about disease inference algorithm, section 3 is about the simulation of the system, and finally section 4 concludes.

## 2 Disease Inference Algorithm

If human is taken with a disease, the electric resistance of the diseased part is higher than around part. The inherent current of human body does not flow well in the diseased part due to high electric resistance. Small current flows in the diseased part, as a result, absolute current of cell is decreased. In other words, if inflammatory reaction, various disease and cancer occur in the human body, then pain, part fever, edema and seizure are appeared. If human body is injured and got an infection in skin and bodily tissue, muscles are contracted to protect him. For instance, the reason of occurrence of pimple, atopic dermatitis and lentigo is that the electric resistance of these parts is high, and the parts become to status of nonconductor because of oxygen deprivation. So skin disorder appears, and gets an infection in the skin or skin tissue is dead. In these cases, if the patient gets electronic acupuncture in the diseased part for 1 minute, his blood circulation is promoted by penetrating blocked aeremia and supplying bioelectricity. The fine current from electronic acupuncture(13~500 $\mu$ A) strengthen the ATP five times, activate the tissue cell and go on smoothly metabolism. In the majority of cases, the intensity of blood pressure which shows the dynamics of blood flow from heart is measured by sensor pad attached to the heart. In oriental medicine case, it is measured by sensors attached to the arteries in the wrist.

DSP board presented in this paper for implementation of intelligent pulse diagnosis system is designed to respond to the various patterns and to sense the situation which potential difference is changed according to patient's painful part simultaneously.[4]

In this paper, if result H is not 100 percent in spite of evidence E in the rule; IF E THEN H, we can express this rule with conditional probability  $P(H|E)$  using methodology of probability. And this conditional probability can be found using Bayesian theory.[5],[6]

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E|H_j)P(H_j)}$$

Although this Bayesian theory is very clear theoretically, many problems can be occurred in case of application of the real issues. First, to know the conditional probability  $P(H_i|E)$ , we have to know  $P(H_i)$  and conditional probability  $P(E|H_j)$ . For example, let us suppose that  $E$  is symptom of patient's body and  $H_i$  is disease, to know the probability  $P(H_i|E)$  of a certain disease  $H_i$ , a prior probability of each disease  $P(H_j)$  and probability  $P(E|H_j)$  have to be given. But there are frequent occasions that the data of these cases is not enough. Second, in the equation at above, each disease  $H_i$  is has to mutually exclusive:  $P(H_i \cap H_j) = 0$ , but this assumption cannot be satisfied because three types of disease can be occurred in any case of patients simultaneously. Suppose the probability of disease  $H$  is 0.7 when three symptoms  $E_1, E_2, E_3$  are all true. This conditional probability is summarized as follows.  $P(H|E_1 \cap E_2 \cap E_3) = 0.7$ . Hear 0.7 is subjective value of probability. Be that as it may, it does not mean that the probability of not occurrence of disease  $H$  is 0.3 when the three symptoms are all true as below.

$P(H|E_1 \cap E_2 \cap E_3) = 0.3$ . The probability 0.3 is calculated based on the axiom of probability,  $P(H|E) + P(\bar{H}|E) = 1$ . Therefore, this means that 0.7 is not the value of probability. The reason why the above axiom is not true is that 0.7 is sure of disease  $H$ , but 0.3 does not mean the disease  $H$ . This means that trust and distrust are treated separately. 0.3 is just a unknown part. The meaning of 'Do not know' or 'ignore' is different from 'refute'. According to the axiom of probability, probability of disease  $H$  is assured as 0.7 and not disease as 0.3. If trust and distrust are coexisted together, it would be rather to lower the strength of trust. Namely the trust must be reduced as 0.4. If the probability is remained as 0.7, the rest probability 0.3 is not considered as distrust but unimaginable of unknown area. Fig.1 shows correction factor using fuzzy rules.[7]

Degree of belief represents degree of confirmation and it is expressed as difference of belief and disbelief as below.

$$CF(H,E) = MB(H,E) - MD(H,E)$$

Here  $CF$  means degree of belief about hypothesis  $H$  when the evidence  $E$  is given,  $MB$  means measure of increased belief about  $H$  caused by  $E$  and  $MD$  means measure of increased disbelief about  $H$  caused by  $E$ .  $CF$  means net increase of trust caused by given evidence.  $CF > 0$  means  $MB - MD > 0$  and a given evidence increases trust of hypothesis,  $CF = 1$  means that the hypothesis is proved clearly by evidence. There is two cases when  $CF = 0$ , in case of  $MB = MD = 0$  nothing can be trusted, in case of  $MB = MD > 0$  trust is offset by distrust.  $CF < 0$  means increase of trust that hypothesis is negation.  $CF = 0.7$  means trust 70 percent greater than distrust. The difference between  $MB$  and  $MD$  is important, not the each value.

	Weight Condition Y1	Age Condition Y2	Height Condition Y3
Patient Condition X1(Begin)	0.7	0.5	0.86
Patient Condition X1(End)	0.5	0.35	0.7

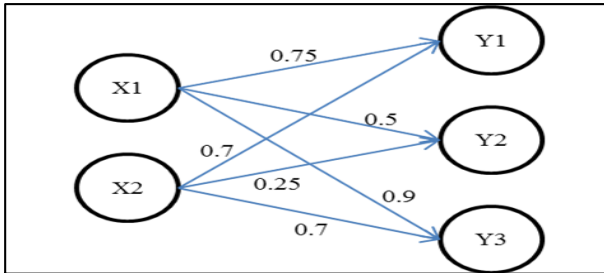


Fig. 1. Correction factor using fuzzy rules

### 3 Acupuncture System Using Intelligence

In this paper, we use the intelligent algorithm for pulse diagnosis as follows. In this paper, it tried to solve these problems using intelligent fuzzy rules.

$$e=R-Y$$

$$Ce=e2-e1$$

Where, Y: optimum pulse feeling judgment

R: Criteria Input

e: Error

Ce: Error Displacement

e2: Current Error

In this paper, in order to solve this kind of problem, it uses compositional inference while using the fuzzy rule. Fuzzy compositional rule of inference is applied to come up with a calibrating constant in order to derive an accurate result (considering the patient’s physical condition) in analyzing the blood pressure pulse. In existing method, an oriental doctor infers one pulse wave out of 28 pulse wave and diagnoses the patient. It is difficult to know whether pulse detection sensor is located in the radial artery exactly or not by using existing pulse checker. And in the case of different body type and the thickness of a forearm, it is difficult to take pulse exactly. And also It is difficult to standardize the pulse with analog pulse checker. For example, even if the thickness of two person’s forearm is equal, it is difficult to standardize pulse exactly because thickness of their blood vessels is different.

In this paper, we used TMS320VC33(TI) as main DSP of electronic acupuncture system, HY29F040 as flash ROM to store OP code, and 512K×32 capacity as Main operating RAM. The signal measured in sensor passes isolation amplifier(AD202) through primary filtering after being amplified at AMP stage. This is necessary and to prevent the fatal electrical accident occurred in body. The system was designed to control the changes of power supply such as voltage range of 10uV~1V and current range of 100uA~10mA using D/A and FET to experiment various patterns continuously base on the point of view that an electrical conductivity is varied with the body characteristics and patient’s affected part. The system includes the hardware to provide protection function for safety and to support the multimode function of electronic acupuncture through change of control mode.

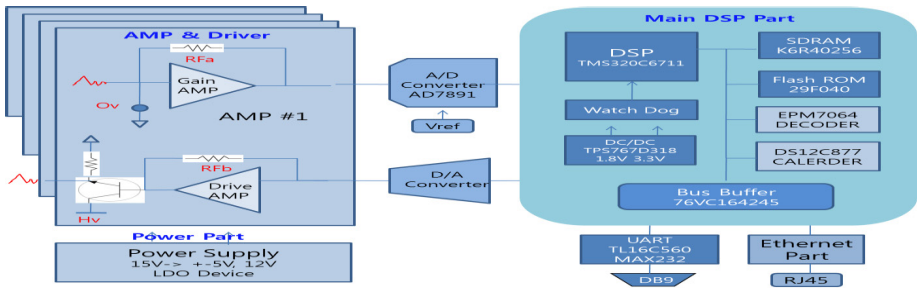


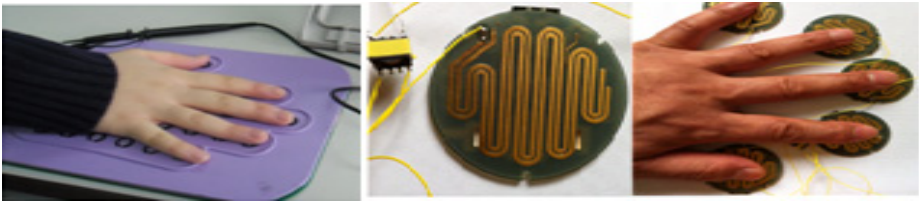
Fig. 2. Circuit diagram of acupuncture system

Fig.2 shows circuits of the electronic acupuncture system. If random order signal of selected function is entered control part, the signal is transformed and performs the output order corresponding to the signal data. Although pulse wave occurred in the human body is not fast in comparison with brainwave, the system requires fast A/D converter and DSP board with floating point arithmetic operation for real-time analysis, secondary differentiation and fuzzy relearning according to shape change of new pattern. But D/A converter does not need to be fast or accurate in comparison with A/D converter. For personal security this part also requires isolation of power supply, for this requirement, serial type D/A with 4 signals (CS, SCL, SDI, SDO) and photo coupler to isolate signal are added. The system was designed to control all data created in DSP board through RS232C or USB port and to accept result data in real-time. Dedicated serial controller was built in the system to support RS232C, communication speed is in the range of 1,200BPS to 115,000BPS. Also the system support USB2.0.

#### 4 E-Acupuncture Pad with Built-in Multi-active

What is a multi-pad with a built-active JEUNJACHIM (Electronic-Acupuncture) depending on a patient's current body status? Based on this information, the patient meets the voltage and current self-oscillation. The frequency with the ability to automatically advanced procedure is called JEUNJACHIM. In order to perform these

functions simultaneously with the sensing of JEUNJACHIM, one is required to possess the ability to perform surgery, derive accurate analysis from fuzzy logic and process statistical data. Electrical resistance of the body including long-term resistance, internal resistance and the surface can be divided into exposed skin. Resistance of the body when the DC voltage is based on the pure resistive component can be considered only based on the basis, when the impedance of the AC voltage should be considered. That body electrical conductors if you think skin, blood, muscles and other body each part of the voltage and current for the resistive component and capacity components are separated by impedance and its size, the electrical conduction path, the contact voltage, the contact area, and energizing time, is applied differently depending on the frequency may occur.



**Fig. 3.** Multipad with a built in electronic acupuncture

Figure 3 illustrates the basic theory of Electronic-Acupuncture. In addition, these changes in a person's age, gender, humidity, temperature, weight and fat accumulation is based on the changes. The requirements when considering the electrical resistance of human skin in general is based on the amount of approximately  $2500\Omega$ . However, the same voltage and current is applied even if the amount of contact area and pain change in resistance over time are different. In the electrical resistance of human body tissues, regardless of the DC and AC power is almost constantly appear if time longer JUAL heat due to temperature rise of tissue resistance is slightly reduced. When the electricity in the human body typically conduct a minimum of power to feel the flow of the AC voltage is  $1\text{mA} \sim 2\text{mA}$  for men. In contrast, direct the flow of power is smaller than the stimulus at least five double-road sensing current flow caused by the voltage applied, even though I do not feel the flow of electricity. Thus, in the treatment of **JEUNJACHIM**, electricity is AC rather than DC voltage with the voltage of the aneurysm and the frequency and voltage, over current change as a real hand acupuncture procedures, a small battery that has the same effect as a treatment is likely to be seen. In the experiment, according to AC current that can safely come off as self a man  $16\text{mA}$  ( $60\text{Hz}$ ) women  $10.5\text{mA}$  ( $60\text{Hz}$ ) is about the human body can withstand DC current is approximately  $74\text{mA}$  men for women is approximately  $50\text{mA}$ . But it also including a person's body size and weight may appear slightly different depending on the requirements. In this paper, the voltage between  $15\text{V} \sim 50\text{V}$  AC voltage to the change of  $5\text{Hz} \sim 1.2\text{KHz}$  and current  $500\text{uA} \sim 1500\text{uA}$  given in the current experiments were carried out. Figure 4 illustrates the electronic acupuncture circuit.



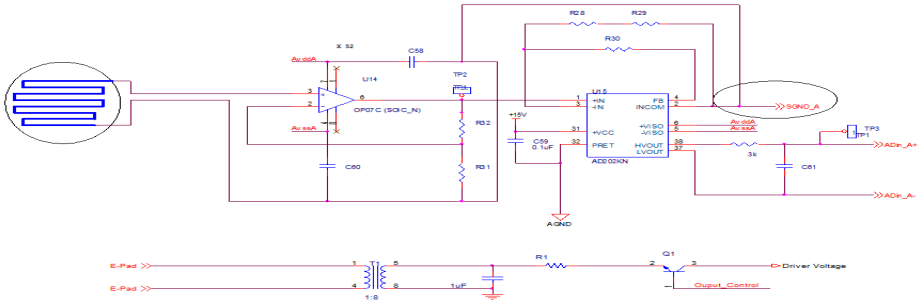


Fig. 4. Electronic acupuncture circuit

## 5 Conclusion

In this paper, we proposed the system that diagnoses a patient optimally considering the patient’s condition using intelligent fuzzy technique. We analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We classified the patients by their body, illness and age, and calculated the exact time of electronic acupuncture suitable for the patient’s physical condition using fuzzy logic and inference.

We designed the system to respond to the various patterns and to sense the situation which potential difference is changed according to the patient’s painful part simultaneously. It contains the function that a patient can search the exact point of electronic acupuncture and check on optimal strength and time of electronic acupuncture considering the patient’s body conditions. The patients can be treated with optimally, and acupuncture time can be shortened and strength of acupuncture can be moderated considering their condition. This system is useful for remote medical examination and treatment.

**Acknowledgement.** This research was supported by Kyungpook National University Research Fund, 2012.

## References

- [1] Haider, A.W., Larson, M.G., Franklin, S.S., Levy, D.: Systolic Pressure, Diastolic Blood Pressure, and Pulse Pressure as Predictors of Risk for Congestive Heart Failure in the Framingham Heart Study. *Ann. Inter. Med.* 138, 10–17 (2006)
- [2] Shaltis, P.A., Reisner, A.T., Asada, H.H.: Cuffless Blood Pressure Monitoring Using Hydrostatic Pressure Changes. *IEEE Trans. Biomed. Eng.* 55, 1775–1777 (2008)
- [3] Lee, Y.J., Lee, J., Lee, H.J., Yoo, H.H., Choi, E.J., Kim, J.Y.: Study on the characteristics of blood vessel pulse area using ultrasonic. *Korea Institute of Oriental Medicine Researches* 13(3), 111–119 (2007)

- [4] Hong, Y.S., Kim, H.K., Kim, B.K.: Electronic Acupuncture system with built-in Multi-pad using intelligence. In: Proc. of the 2012 Advanced Information Technology and Sensor Application, p. 162 (2012)
- [5] Sik, H.Y., Kug, P.C.: Proc. of the sixth international fuzzy system association, IFSA, pp. 461-464 (1995)
- [6] Garg, M.L., Ahson, S.I., Gupta, D.V.: A Fuzzy Petri-nets for Knowledge Representation and Reasoning. Information Processing Letters 39, 165-171 (1992)
- [7] Hong, Y.S., Kim, H.K., Kim, B.K.: Implementation of Adaptive Electronic Acupuncture System using Intelligent Diagnosis System. Internation Journal of Control and Automation 5(3), 141-152 (2012)
- [8] Leung, K.S., Lam, W.: Fuzzy Concepts in Expert Systems. IEEE Computer, 43-56 (September 1988); [10] Looney, G.C., Alfize, A.A.: Logical Controls via Boolean Rule Matrix Transformation. IEEE Trans. on SMC 17(6), 1077-1082 (1987)
- [9] Looney, G.C.: Fuzzy Petri Nets for Rule-based Decision Making. IEEE Trans. on SMC 18(1) (January/February 1988); [10] O'Rourke, M.F., Kelly, R.P., Avolio, A.P.: The Arterial Pulse, 1st edn. Lea & Febiger, Philadelphia (1992)
- [10] Gong, Y., Chen, H., Pu, J., Lian, Y., Chen, S.: Quantitative investigation on normal pathological tongue shape and correlation analysis between hypertension and syndrome. China Journal of Traditional Chinese Medicine and Pharmacy (2005)

# Electric Braking Control System to Secure Braking Force in the Wide Speed Range of Traction Motor

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2,\*</sup>, and Ok-Hwan Kim<sup>1</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Budaе-dong, Cheonan-si, Chungnam, 330-717, Korea  
{yckim59, owkim}@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College, 316, Daehak Road,  
Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

**Abstract.** In this paper, a vehicle stopping method using an electric brake until a traction motor is stopped is studied. At the moment of vehicle stop, electric brake is changed to control mode wherein torque is reduced at a low speed. Gradient is controlled by estimating the load torque of motor thereby traction motor is not rotated after stop. In addition, coasting operation and brake test were performed from normal-opposite operation and start using a small-scale model comprising the inertial load equipment and the power converter. Further, traction motor was made to be equipped with a suspension torque. Pure electric braking that makes traction motor stopped by an air brake at the time of stop was also implemented. Constant torque range and constant power range were expanded during braking so that braking force was secured with the electric brakes even in high speed region. Therefore, vehicle reduction effect could be expected by reducing parts related with an air brake which is not used frequently by using a pure electric brake in the M car in wide speed region. Further, maintenance of brake system could be reduced, Besides ride comfort of passenger in the electric rail car, energy efficiency improvement, and noise reduction effect could be additionally expected.

**Keywords:** Electric brake, Traction motor, Constant torque.

## 1 Introduction

Brake technology in high-speed region is executed in parallel with air brake that supplements electric braking force. Still, air brake use frequency has to be minimized to improve performance of vehicles. Since air brake in the electric rail car is basically the equipment which applies mechanical friction, the noise and dust generated in the process of brake are caused to reduce the performance of vehicle.[1],[2]

In the braking system of electric rail car, braking is extended in the high-speed region, thus operating range is expanded provided that traction motor secures

---

\* Corresponding author.

insulation that withstands overvoltage to secure electric braking power in a high-speed region.

Therefore, in this study, the control method that secures brake force of electric rail car in a wide speed range where traction motor is driven with an assumption that a device which absorb regenerative power is installed in electric rail cars. By using electric brake until securing braking force and stop the vehicles at high-speed region, air brake which was essential due to shortage of existing electric brake needs not be used.[2],[3]

The possibility of securing brake force having a constant power region was tested in the braking test for the small-scaled test system. Also, pure electric brake for a wide speed range of traction motor was realized by using an auxiliary power converter which sends regenerative power equivalent to regenerative motion of main power converter for drive and terminal voltage increases due to expansion of constant power drive to the DC bus.

## 2 Method of Electric Brake

The electric brake method suggested in this study is presented in fig. 1. In high-speed region, drive range was expanded and drive at constant torque control region was expanded to secure electric brake force to minimize deceleration changes as much as possible. At stop mode, suspension torque was given to prevent driving.

After stopping electric rail cars, it is made to be stopped by air brake.

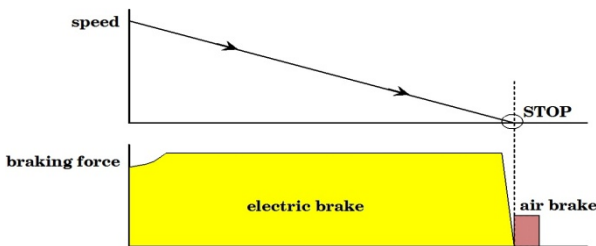


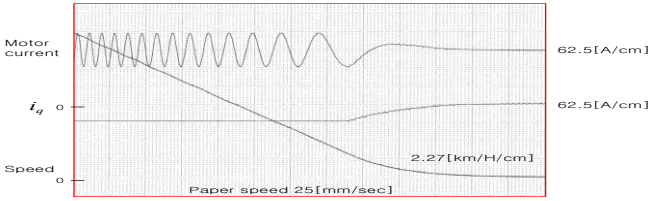
Fig. 1. Electric brake method

In this study, a drive motor is comprised of permanent magnet type synchronous motor (PMSM) in the device comprised of motor for inertial mass and load. Drive motor was stopped by electrical brake till stop by vector control was used. Electric braking was executed until stopping the drive motor. Stop control was executed by position detection of rotor by resolver and the mode of assuming the speed and load torque was used.

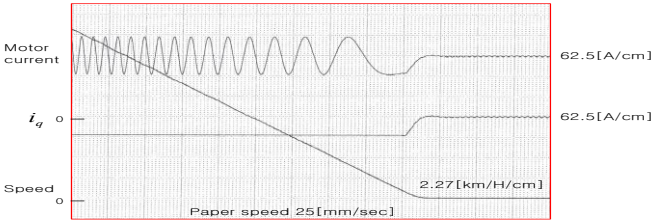
A brake test was conducted by the proposed method after accelerating drive motor and then driving it at constant speed. Test equipment was a small-scale model of direct drive system. When rotational speed is 824[rpm], the speed of the electric rail car is equivalent to 120[km/h].

The gain of velocity feedback and time constant of filter are related with the time which reduces braking torque at the moment of the vehicle stop. During this time,

since the deceleration ratio is changed, it considerably affects on the ride comfort. Therefore, it is desirable to increase feedback of speed. Fig. 2 and Fig. 3 show the measurement results of stopping motion according to velocity feedback and time constant of filter. With increase in the time constant of filter, the oscillation of current was reduced and stable drive was observed. Also, the larger the velocity feedback, the shorter was the duration of torque reduction, i.e., deceleration change duration became shortened.



**Fig. 2.** Velocity feedback: 50, filter time constant: 0.0384[s]



**Fig. 3.** Velocity feedback: 250, filter time constant: 0.0384[s]

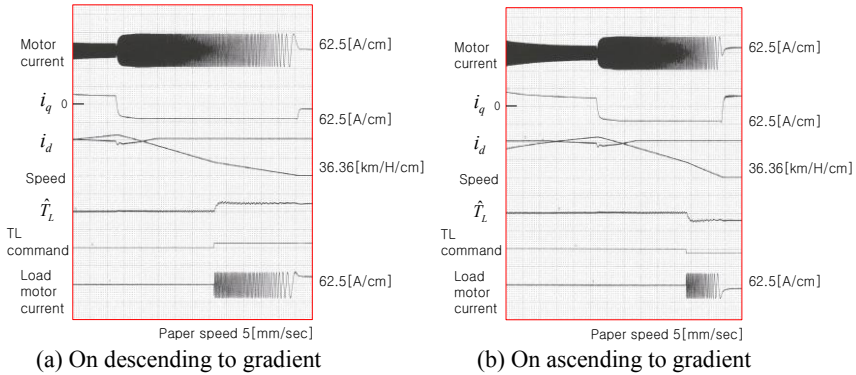
In case a load torque is existed, gradient of transfer at stop control acts as load torque. Fig. 4 shows the measurement result of brake after acceleration till 80[km/h]. It shows the brake condition under positive(+) torque and negative (-) torque of a loaded motor. These are corresponding to up motion and down-motion of electric rail cars on the gradient.

In Fig. 4, load torque is assumed even after a vehicle stops. The suspension torque is generated under a gradient condition. Therefore, a pure electric brake is executed to remove the electric brake after stopping by the air brake. This kind of vehicle can be environment-friendly by reducing noise and dust and can improve performance of electric rail car.

### 3 Securing the Braking Force and Test Result

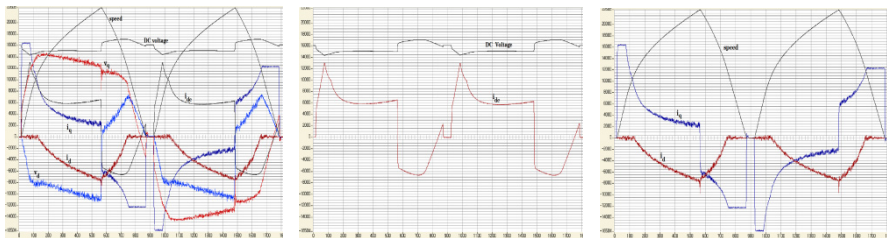
Because voltage of the inverter can be secured with the brake of motor and therefore brake of constant torque in total interval can be achieved, motor should be driven by limiting voltage to be saturated under any voltage conditions. During brake, increases in the brake force for the trolley voltage, limit in constant power driving by flux driving, and magnifying method of brake pattern by inserting series resistance were proposed.

Fig. 5 shows the case of characteristics driving by constant power driving. Driving speed was 65[km/h] with the braking characteristics. It shows driving from the maximum speed 120[km/h]. Voltage variation characteristics of transfer also affects on the braking characteristics. The higher the voltage variation, the higher is the brake force.



**Fig. 4.** Response of load torque estimator (velocity feedback) : 150, filter time constant: 0.084[s] (Gain of estimator  $k_i = 58.6, k_p = 62$  of estimator)

Fig. 5 shows the case of characteristics driving by constant power driving. Driving speed was 65[km/h] with the braking characteristics. It shows driving from the maximum speed 120[km/h]. Voltage variation characteristics of transfer also affects on the braking characteristics. The higher the voltage variation, the higher is the brake force.



Condition: ratio: 0x3500, Characteristics drive conversion point: 0x23c0 65[km/H], transfer conductance: 0.1[Ω]

**Fig. 5.** Transfer drop and constant power brake

Since resistance is made short according to speed, brake force is secured in high-speed region. On the contrary, when speed becomes reduced, resistance becomes short which improves efficiency. The series resistance expands constant drive region, thus it is possible to secure brake force for the wide speed range.

Constant torque brake is regarded as ideal for wide speed range. Fig. 6 shows the vector diagram wherein an armature resistance of motor is ignored. Here, point A

indicates the starting point of constant torque brake using a resistance drop, while point B indicates the case where all the resistances were short. Resistance drop can be estimated using Eq. (1).

$$Ri_q = k\omega\phi - \sqrt{v_{\max}^2 - (\omega Li_q)^2} \tag{1}$$

At point B where all the series resistance becomes short, the resistance drop should be 0. Therefore, it is when Eq. (1) becomes  $\omega$  that is 0 and this is the maximum point of constant torque driving. Meanwhile, the regenerative power of motor becomes Eq. (2).

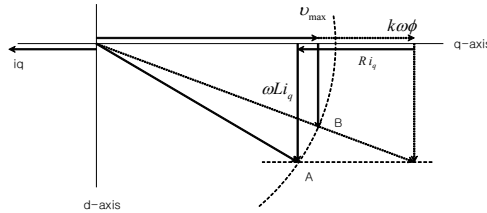


Fig. 6. Vector diagram during brake

$$k\omega\phi i_q = (Ri_q + \sqrt{v_{\max}^2 - (\omega Li_q)^2})i_q \tag{2}$$

Within the parentheses bracket in Eq. (2), the first term is the resistance drop and the second term is the component which is in-phase with induced electromotive force in inverter voltage. Since resistance drop has a constant current, it brings voltage drop in proportion with resistance, thus it has a constant size regardless of speed.

### 4 Conclusion

The electric brake is a method used to stop the vehicle by changing torque at low speed to control the mode at the moment of vehicle stop. The load torque for the gradient was assumed so as vehicle location is not rotated after stopping.

The braking test was carried out with a small-scale model comprised of an inertial load system and a power converter. The model was tested from starting till coasting operation and braking test including normal-opposite operation.

Suspension torque after stop was implemented in the model and a pure electric brake that stops the model with air brake. Therefore, if there is no problem like emergency shutdown when there is a fault in the electrical system, the air brake can be omitted in the M car. Besides, it is possible to reduce weight of the vehicles.

Pure electric brake was also realized in the system. During braking the car, the brake force could be secured only with electric brakes even in high speed region by expanding constant torque region and constant power region.

## References

1. Kim, Y.-C., Song, H.-B., Cho, M.-T., Lee, C.-S., Kim, O.-H., Park, S.-Y.: A Study on the Improved Stability of Inverter through History Management of Semiconductor Elements for Power Supply. In: Kim, T.-h., Ramos, C., Kim, H.-k., Kiumi, A., Mohammed, S., Ślęzak, D. (eds.) ASEA/DRBC 2012. CCIS, vol. 340, pp. 155–162. Springer, Heidelberg (2012)
2. Kim, Y.-C., Song, H.-B., Cho, M.-T., Moon, S.-H.: A Study on Vector Control System for Induction Motor Speed Control. In (Jong Hyuk) Park, J.J., Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. LNEE, vol. 181, pp. 599–612. Springer, Heidelberg (2012)
3. Kovudhikulrungsri, L., Koseki, T.: Speed Estimation in Low-Speed Range for an Induction Motor to Realize Pure Electric Brake. I.E.E. Japan Joint Technical Meeting on Transportation & Electric Railways and Linear Drive, TER-00-38 LD-00-65, 19–24 (July 2000)



# Optimized Design of Charger for Electric Vehicles with Enabled Efficient CCCV Mode Movement

Ji-Yong Chun<sup>1</sup>, Young-Choon Kim<sup>2\*</sup>, and Moon-Taek Cho<sup>3</sup>

<sup>1</sup> Div. of Automotive Engineering, Ajou Motor College  
jychun@motor.ac.kr

<sup>2</sup> Div. of Mechanical and Automotive Engineering College of Engineering, Kongju National Univ. 275 Budae-dong, Cheonan-si, Chungnam, 330-717, Korea  
yckim59@kongju.ac.kr

<sup>3</sup> Dept. of Electrical & Electric Engineering, Daewon University College, 316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

**Abstract.** This paper presents a charger technology for the electric vehicles that enables charge and discharge not only for low voltage and high voltage but also for any battery type by using a high performance DSP. The proposed charger was made to function as generalized fast and low battery by using PWM buck converter that runs with CCCV(Constant Current Constant Voltage) mode. Besides, by designing the controller as fixed-type and varied suiting to the load characteristics, constant output was ensured even during power trip. Also, by controlling the battery type, charge, and discharge, a/s becomes easy. This battery would be possibly implemented not only in the Off Board Charger of the electric vehicles but also in the On Board Charger of EREV(Electric Range Extender Vehicle) in future.

**Keywords:** Electric vehicle, DSP, PWM, CCCV, EREV.

## 1 Introduction

A battery charge/discharge system proposed by this study has a computation part to automatically control itself by perceiving mode change during charging and discharging the battery by using DSP Controller MC56F8345. Also, a module was constructed so that charging and discharging are quickly executed.[1],[2]

For that, a system was made to adjust charging current of various steps according to the load by supplying voltage and current suiting to the imposed load by adjusting the time ratio of PWM(Pulse Width Modulation) during control even if battery voltage is different from Li-ion and lead storage type. In addition, battery for the electric vehicle was made as plug-in system in case the battery voltage is a slow charger-type. The capacity of battery was set up as 2[kVA] level considering that DC output voltage is 310[V] when AC voltage is input. At the same time, in case of fast

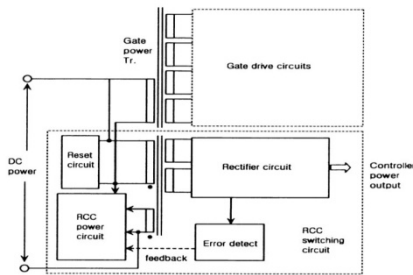
---

\* Corresponding author.

chargers, capacity was designed as 10[kVA]. Therefore, input current into the battery 1[C-rate] was set as 3[A] based on the current input into the battery around 35[A] so that charge and discharge mode of minimum ten steps could be switched.

## 2 Design of EV Charger

Since there is one MOSFET switching device inside power converter, and microprocessor controller source is required, a SMPS(Switching Mode Power Supply) that supplies multi-source required in each circuit like gate circuit of controller and power converter by RCC(Ring Choke Converter) as in Fig. 1 by dc source from battery was constructed.



**Fig. 1.** MOSFET Gate and Power Circuit

For the driving source, ac source 380/200[V] was commutated for fast charger. Maximum output voltage de 310[V] was also set during diode commutation of 220[V] at second side.

Insulation element photocoupler was used so that signal from output port of microprocessor transmits signal to high power circuit at drive circuit terminal. Circuits were insulated by OP-AMP to reduce the effect of noise at controller terminal to control voltage and current. DC offset was made to be adjusted to convert power to digital value. Also, the dc voltage input is an important factor in PWM modulation when output voltage is controlled in DC/DC converter. In addition, since dc voltage is high voltage-type, control circuit is necessarily required. Therefore, a hall element was used to electrically insulate the dc signal.

To filter peak voltage by carrying out single phase full wave of ac voltage  $100\sqrt{2}$  at source side, 20 numbers of electrolytic condensers having capacity 2200[μF]/63[V] were used. Since load is battery and it is working as voltage source at output side, capacitor was not needed. It was arranged by programming so that output side can function as CC(Constant Current) mode.

Meanwhile, iron core was used as an inductor. Also, a current constant-type reactor capacity 1[mH] was used to stably maintain current 15[A] needed by load.

### 3 Controller Design

Buck chopper was constructed as 10[kW] level, and power part was constructed to drive as continuous mode under load higher than 10[%] at switching frequency 15[kHz]. Output voltage was set as ripple less than 1[%].

The switching frequency was set as 15[kHz] considering temperature restriction by switching loss at IGBT. Requirement for transient response in power converter was not included since the expected change of load condition was huge.

The voltage controller that includes feed forward compensator is same as Eq. (1). The control diagram is presented in Fig. 2.  $k_f$  is the feed forward compensation gain and optimum value was chosen by frequency response characteristics of output voltage for the load current.

$$d(t) = k_p(v_{ref}(t) - v_o(t)) + k_i \int (v_{ref}(t) - v_o(t)) dt - k_f \frac{di_o(t)}{dt} \tag{1}$$

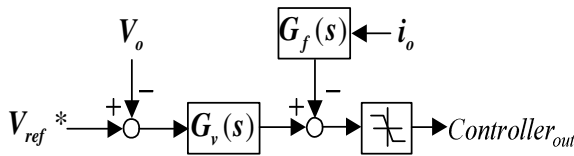


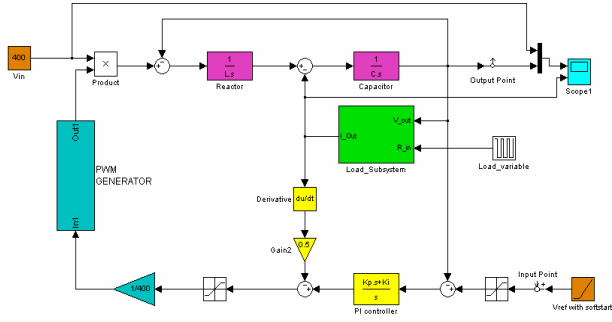
Fig. 2. System Control Block Diagram Using Feedforward Compensator

### 4 Simulation

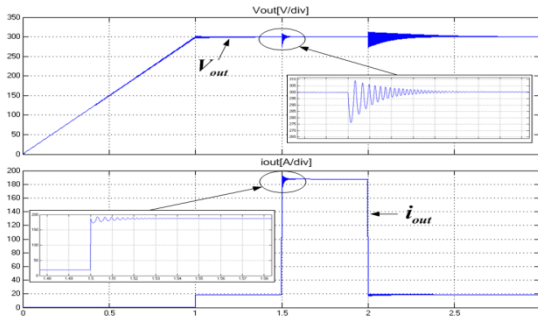
The algorithm using feed forward compensator proposed in this study was implemented in the PWM buck chopper. To examine the overall control characteristics of proposed algorithm, a control block was constructed by using MATLAB/Simulink as shown in Fig. 3.

The overall control system which was used in the simulation was constructed as shown in Fig. 3. System parameter for the output voltage was set as 300[V] when input voltage was 400[V] as in Table 1. As a test condition, loads were input in the sequence of 10[%](16.36[Ω])→100[%](1.636[Ω])→10[%](16.36[Ω]).

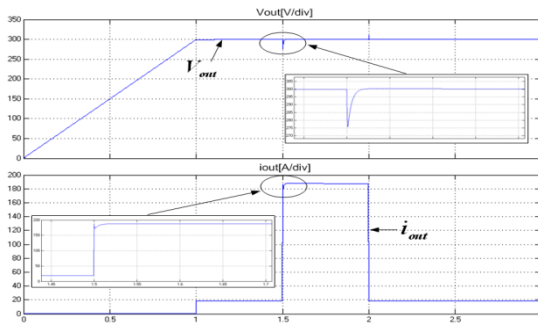
When the feed forward compensator of load current was implemented from the simulation results as in Fig. 4 and Fig. 5, the transient response and steady-state characteristics were significantly improved. Therefore, the characteristics equivalent to the dynamic characteristics required by the system proposed in this paper could be obtained.



**Fig. 3.** SIMULINK Block Diagram of PWM Buck Chopper



**Fig. 4.** With PI Controller



**Fig. 5.** With Load Current Feedforward Compensator

## 5 Test Results

Fig. 6(a) shows a current waveform between IGBT emitter and base during PWM modulation(upper figure) and the voltage wave form which was input during battery charging(lower figure). Fig. 6(b) shows the voltage(upper figure) and current wave form(lower figure) which were input during battery charging.

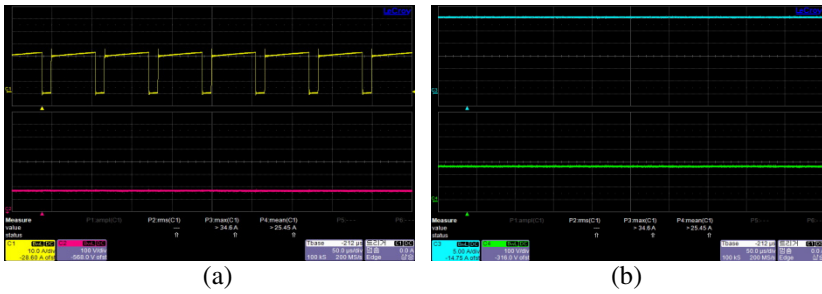


Fig. 6. (a) PWM Modulation Waveform and (b) Output Voltage and Current Waveform

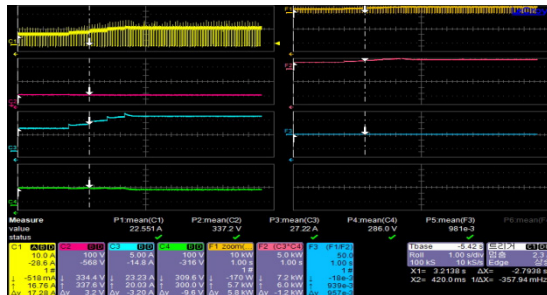


Fig. 7. Input Voltage, Output Voltage, and Current Waveform in Load Test / Efficiency Measurement

The load test was performed to check the stability when load was changed in the sequence of no load→light load→ heavy load→ full load under the same condition as shown in Fig. 7. The efficiency measurement shows that the efficiency of more than 95[%] was confirmed. As can be seen from Fig. 7, the efficiency characteristics of 98.1[%] was obtained.

## 6 Conclusion

The present paper presents a charger technology for the electric vehicles that enables charge and discharge not only at lower voltage and high voltage range but also irrespective of battery type by using a high performance DSP.

The charger technology proposed in this study could be equipped with generalized fast and low function by comprehensively implementing high voltage semiconductor device MOSFET for large power as a low voltage type and IGBT for high voltage type. The generalization of battery was realized by using PWM buck converter that runs with CCCV(Constant Current Constant Voltage). In addition, by designing the controller fixed type and variable type suiting to the load characteristics, continuous output was ensured even during power trip. A/S was also convenient by controlling the battery type, charge, and discharge. Further, it was applied in the power supply unit linking with battery that enabled the synchronization of power converter and drive power design technology for the interface.

## References

1. Maharjan, L., Yamagishi, T., Akagi, H.: Active-Power Control of Individual Converter Cells for a Battery Energy Storage System Based on a Multilevel Cascade PWM Converter. *IEEE Transactions on Power Electronics* 27, 1099–1107 (2012)
2. Zhang, Z., Xu, H., Shi, L., Li, D., Han, Y.: A unit power factor DC fast charger for electric vehicle charging station. In: 2012 7th International Power Electronics and Motion Control Conference (IPEMC), vol. 1, pp. 411–415 (2012)

# Ubiquitous Mobile Computing on Cloud Infrastructure

DongBum Seo<sup>1</sup>, In-Yong Jung<sup>2</sup>, Jong Hyuk Park<sup>3</sup>, and Chang-Sung Jeong<sup>2,\*</sup>

<sup>1</sup> Visual Information Processing

<sup>2</sup> Department of Electrical Engineering,

Korea University, Seoul, 136-713, Korea

{treeline,dekarno,csjeong}@korea.ac.kr

<sup>3</sup> Department of Computer Science and Engineering,

Seoul National University of Science and Technology,

Seoul, 139-743, Korea

jhpark1@seoultech.ac.kr

**Abstract.** The demand for ubiquitous mobile application on cloud computing platform is increasing due to the rapid advancement of cloud technology. In this paper, we shall present a new architecture for ubiquitous computing environment on cloud computing platform for mobile application. Our system provides a combined architecture of ubiquitous and cloud computing which supports a powerful framework for mobile applications which requires high performance. Also, we describe two models for software mobility and cloud computing for mobile applications, and based upon them show the experimental results for mobile applications which requires high performance. We have presented the ubiquitous computing environment on cloud computing platform for mobile application. Also, the experimental results have shown that the computation intensive mobile applications can be executed very efficiently.

**Keywords:** Ubiquitous Computing, Cloud Computing, PaaS, Software Mobility, Ubiquitous Mobility.

## 1 Introduction

Ubiquitous computing needs the dynamic execution of applications. Especially, in mobile applications where a user moves to a new environment, they need to discover characteristics of their execution environment dynamically, and then either configure aspects of system and application behavior for the efficient and robust execution or adapt behavior during program execution. Recently, cloud computing has dramatically changed how business applications are built, and eliminates the costs and complexity of evaluating, buying, configuring, and managing all the hardware and software, and supports high performance computing platform by sharing and coordinating the shared resources. In this paper, we shall present a new architecture for ubiquitous computing environment on cloud computing platform for mobile application. Our system provides a combined architecture of ubiquitous and cloud computing

---

\* Corresponding author.

which supports a powerful framework for mobile applications which requires high performance.

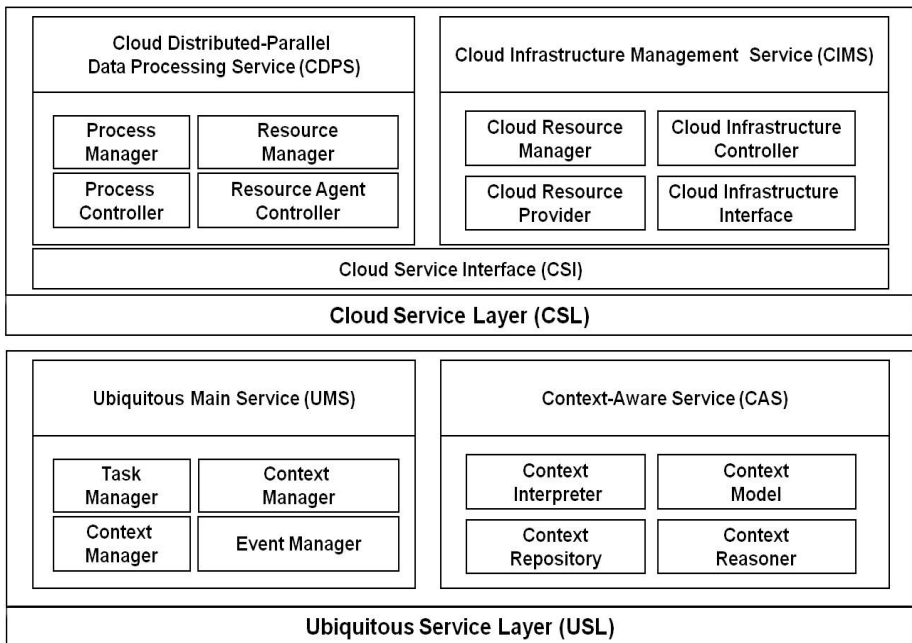
## 2 System Architecture

### 2.1 Introduction

In this section, we describe architecture of our system so called UMOF-C(Ubiquitous Mobile Computing Framework on Cloud Computing Infrastructure), which is composed of two layers: Cloud Service Layer (CSL) and Ubiquitous Service Layer (USL) as shown in Fig.1. We shall explain in detail about them in the subsequent subsections respectively [1-4].

### 2.2 Cloud Service Layer

The Cloud Layer [11] consists of two components:



**Fig. 1.** Architecture of UMOF-C

Cloud Distributed Parallel data processing Service (CDPS) and Cloud Infrastructure Management Service (CIMS). CDPS provides an agent based processing platform for executing various distributed models through four components: process manager, process controller, resource manager and resource agent controller. This layer provides a software execution environment, and manages execution of user mobile



application. CIMS provides on-demand VM allocation for dynamic execution of mobile processes through Cloud Resource Manager which receives resource requirements for mobile application from USL, and launches VM instances via Resource Provider.

### 2.3 Cloud Service Layer

The Ubiquitous Service Layer [9] consists of two components: Context-Aware Service (CAS) and Ubiquitous Man Service (UMS). It is responsible for shielding the user from the underlying complexity and variability through intelligent context-aware infrastructure and automatic ubiquitous service in the self-tuning environment for mobility and adaptation for ubiquitous application. In mobile computing, whenever the user moves from one place to another, the tasks such as cloud resource allocation, environment variable configuration, and remote service execution can be automatically set up and performed by UMS and CAS. CAS provides a context-aware infrastructure which supports the gathering of context information from different sensors and the delivery of appropriate context information to mobile applications.

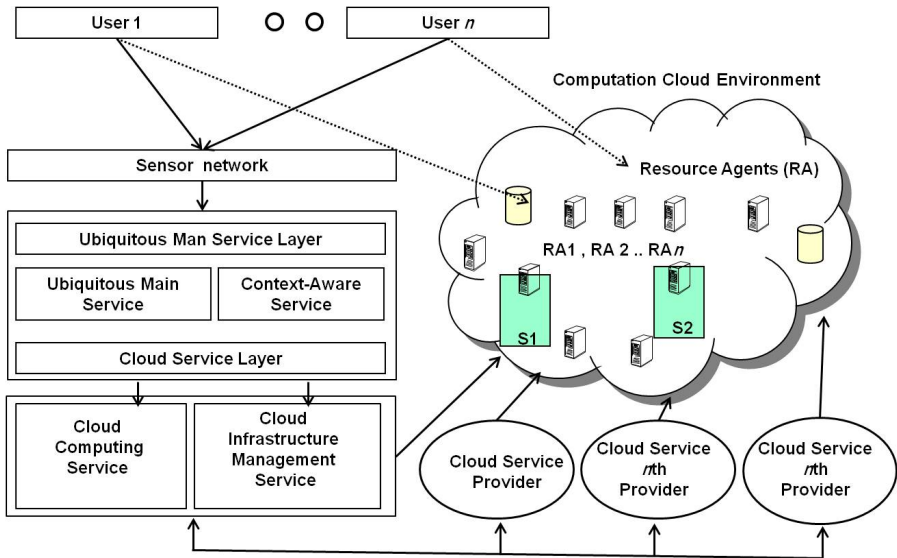


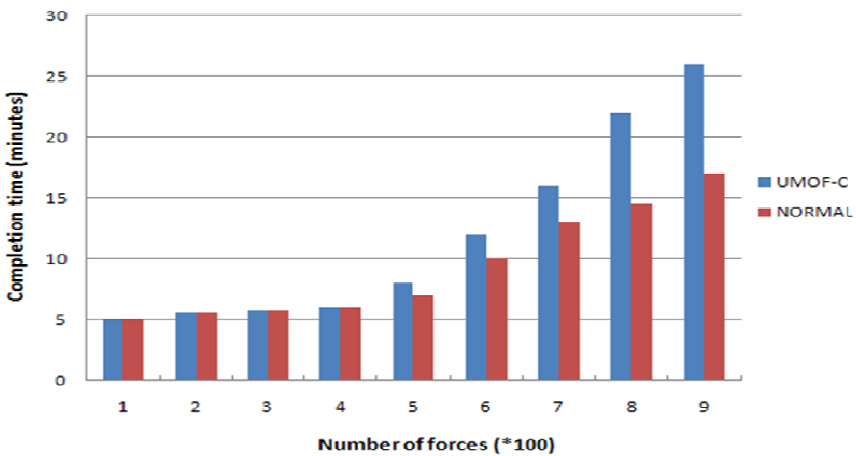
Fig. 2. Ubiquitous Mobile Application on Cloud Computing

## 3 Ubiquitous Mobile Computing

For ubiquitous mobile computing, when a user moves from one place to other, we need a dynamic configuration of software environment on his desktop or on cloud resources or both, i.e. support software mobility. When sensor detects any signal such as leave and entrance of person to other place, CAS derives the proper context by reasoning using context interpreter, context reasoned based on context model, and asks UMS to carry out its corresponding service appropriate to the context [10-16].

## 4 Experiment

In the first experiment, we evaluate the completion time of DOWS on UMOF-C measured as the number of tasks corresponding to that of actors (or forces) increases. The completion time means the loading and execution of task. This experiment is carried out to confirm key services of UMOF-C. In Fig.3, UMOF-C is better than the normal system without cloud service as the scale of simulation is increasing, although the time consumption of initialization has an effect on the state of small forces. UMOF-C can utilize abundant computing power and adapt for various environments, as well as provide convenient user interface. This brings a fast response time, and the demand becomes larger.



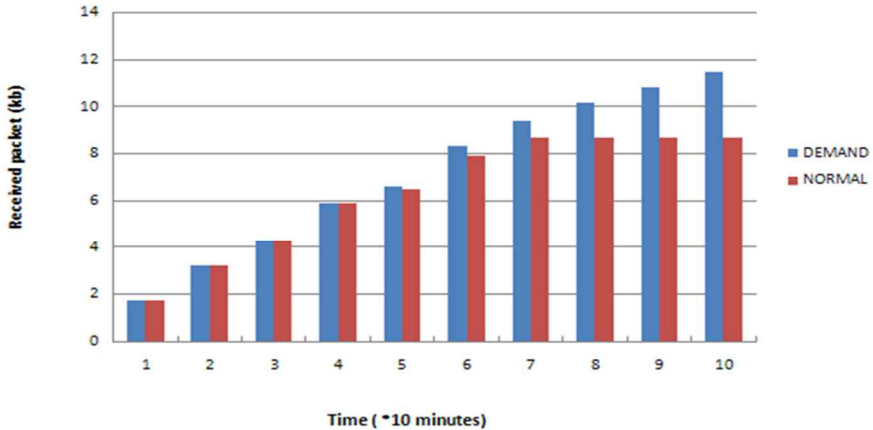
(a) Software mobility model

Fig. 3. Nomal system of UMOF-C

In Fig.4, to verify the dynamic migration service, we execute a second experiment. The information related to execution like object information is stored periodically, and then the application resigns from the execution.

The application sends the stop message to applications before resignation. RM gathers the information, and DM sends the application to the migrated host. The restart messages are sent to all applications, which in turn receive the data stored before failure. After failure, UMOF-C migrate DOWS to the new hosts, and completes its execution, while the other normal system without the migration service is halted.

In general, for a system as its response time and loss ratio are decreased, and its mobility time is decreased, then the demand for its service is increasing.



(b) Dynamic migration Service

Fig. 4. Migration Service of UMOF-C

## 5 Conclusion and Future Work

In this paper, we have presented the ubiquitous computing environment on cloud computing platform for mobile application. It supports a combined architecture of ubiquitous and cloud computing which provides a powerful framework for mobile applications which requires high performance.

This paper has presented architecture of distributed-parallel data processing platform and infrastructure platform on cloud, and shown how it can be used to process massive data and provide scalable adaptive software ecosystem [6-7]. Our proposed system provides an integration layer between context aware layer, ubiquitous main layer and PaaS, and helps the users to develop and execute their massive distributed-parallel applications for processing various social media data. The proposed architecture recovers various aspects including users request, resource negotiation, adaptive resource creation, software deployment and execution.

**Acknowledgements.** This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2010-0020723).

## References

1. Wendorf, R.G., Bodlaender, M.P.: Proceedings of the International Conference on Consumer Electronics, pp. 232–233 (2001)
2. Seo, D.B., Lee, T.D., Jeong, C.S.: Proceeding of the International Multi Conference of Engineers of Engineers and Computer Scientists, IMECS, Hong Kong, vol. I, pp. 19–21 (March 2008)

3. Seo, D.B., Lee, T.D., Jeong, C.S.: IAENG International Journal of Computer Science 35(3), IJCS\_35\_3\_09, Advance Online Publication (August 21, 2008)
4. Seo, D.B., Lee, T.D., Jeong, C.S.: Trends in Communication Technologies and Engineering Science. LNEE, vol. 33, pp. 221–234. Springer, Heidelberg (2009)
5. Jung, I.Y., Lee, D.K., Han, B.J., Kim, K.H., Jeong, C.S.: Proceedings of the 3rd International Conference on Internet, Sepang, Malaysia, December 15-18 (2011)
6. Borthakur, D., Muthukkaruppan, K., Ranganathan, K., Rash, S., Sarma, J.S., Spiegelberg, N., Molkov, D., Schmidt, R., Gray, J., Kuang, H., Menon, A., Aiyer, A.: Proceedings of the 2011 International Conference on Management of Data, Athens, Greece, June 12-16 (2011)
7. Baun, C., Kunze, M.: Proceedings of the IEEE Cloud, 4th International Conference on Cloud Computing, Washington, DC USA, July 04-July 09 (2011)
8. Ghosh, P., Roy, N., Das, S.K., Basu, K.: Proceedings of 18th International Parallel and Distributed Processing Symposium, Snata Fe, New Mexico (April 2004)
9. Abowed, G.D.: Proceedings of the International Conference on, New York, USA, May 16-22, pp. 75–84 (1999)
10. Glesner, M., Hollstein, T., Murgan, T.: ICM Proceedings, and the 16th International Conference on Microelectronics, December 6-8, pp. 11–14 (2004)
11. Kindberg, T., Fox, A.: System Software for Ubiquitous Computing. IEEE Pervasive Computing 1(1), 70–81 (2002)
12. Griswold, W.G., Boyer, R., Brown, S.W., Truong, T.M.: Proceedings of the 25th International Conference, May 3-10, pp. 363–372 (2003)
13. Foster, I., Kesselman, C.: Globus: A Metacomputing Infrastructure Toolkit. Intl J.Supercomputer Applications 11(2), 115–128 (1997)
14. Roy, N.: Providing Better QoS Assurance to Next Generation ubiquitous Grid Users. MS Thesis, University of Teaxs at Arlington, USA (April 2004)
15. Roy, N., Das, S.K., Basu, K., Kumar, M.: Proceedings of the 9th IEEE International Parallel and Distributed Processing Symposium, April 04-08, p. 92a (2005)
16. Chohan, N., Bunch, C., Pang, S., Krintz, C., Mostafa, N., Soman, S., Wolski, R.: First International Conference on Cloud Computing, Munich, Germany, October 19-21 (2009)

# Identifying Invalid Data within Operating System for Higher Flash Utilization

June Young Kim and Seung-Ho Lim

Department of Digital Information Engineering,  
Hankuk University of Foreign Studies  
slim@hufs.ac.kr

**Abstract.** The physical limitation of NAND flash memory is covered by management of Flash Translation Layer (FTL) between logical space of Operating System and physical space of NAND flash memory. However, the FTL management makes unintentional illusion that logically obsolete data is physically valid, in many cases, which causes unnecessary flash-internal operations. The logical invalidation can be immediately applied to physical invalidation using TRIM command. In Operating System, there are lots of logical invalidation region, however, these are not deleted physically if flash memory is used as storage devices. Physical invalidation for logically invalidated data gives higher flash utilization. In this paper, we identify data invalidation within Operating System as two parts; explicit data invalidation and implicit data invalidation. These identified data invalidation can be invalidated immediately by calling TRIM command, which can increase utilization of NAND flash memory-based storage systems, and thus increase bandwidth of storage systems.

**Keywords:** NAND Flash Memory, FTL, TRIM, data invalidation.

## 1 Introduction

NAND flash memory is becoming main storage media not only for mobile embedded system, but also for general computing systems, with the help of low power consumption, low density, high capacity, and high IO bandwidth in comparison with hard-disk drives. However, there are two main physical limits for NAND flash memory. The one is out-of-place update nature, which means that write operation should be preceded by erase operations. The other is mismatch of operation unit, which means that read/write operations is based on *page* unit, while erase operation is based on *block* unit, which is much larger than page. Due to the physical characteristics, the storage sub-systems should be more complex than traditional systems.

To compensate the limitations of flash memory, Flash Translation Layer (FTL) [1][2] is employed between host systems and flash device, and translates logical address and physical address. The FTL remaps incoming logical address of write request to free physical address, just like logging mechanism. If there is no free region

to write incoming request, FTL tries to make available free block by selecting one block, moves valid page to other region, and erasing that block. This is called garbage collection (GC). IO performance of flash storage system is dependent on the FTL's mapping algorithm and GC algorithm.

However, since FTL provides logical address to operating system and hides physical address, operating system has no idea about the real location of data. Even the operating system thought that a data is considered as delete, FTL might recognize the data is being valid. For example, when a file is deleted by delete operation from user, file system deletes it by just deleting the metadata of the file, leaving data area of the file alive. In this case, the data is considered as valid within flash memory until the region is rewritten by operating system. This case is referred as logical invalidation. Besides the example, there are a lot of logical invalidations in operating system. Since the logically invalidated data is considered as valid within flash memory, the data decreases flash utilization not only space utilization, but also bandwidth utilization. Since the data is valid, the available free space region within flash memory is lessen, which makes more GC operations. When GC occurs, the data should be moved from source block to target block in GC operation since the data is considered as valid in flash memory. Thus these logically invalidated data reduces flash utilization. In other words, physical invalidation for logically invalidated data can increase flash utilization in both of space and bandwidth.

The physical area of logically invalidated area can be invalidated by newly introduced command, called TRIM command [4]. Thus, it is important to identify the logical invalidation within operating system and apply TRIM command to the region, which can enhance performance of flash-based storage systems. In this paper, we investigate the nature of logical data invalidation within Operating System. In Operating System, there are lots of logical invalidation region, however, these are not deleted physically if flash memory is used as storage devices. From the view of invalidation, there occurs in both of implicit invalidation and explicit invalidation within IO operation of operating system. The identified logical invalidation can be really invalidated by calling 'trim' command, as a result, flash has invalid region as many as possible. It leads overall flash utilization.

## 2 Background

In NAND flash memory, the read and write operations are done with page level, while erase operation is done as block level, so the granularity of mapping table maintained by FTL can be page or block. Regardless of mapping granularity of FTL, GC is required to make available free region. During GC operation, valid data should be copied from victim block to available other region. Thus it is said that the smaller the number of data to be copied, the higher of GC efficiency. FTL identify data validation by checking whether the logical address of incoming write requests are overwritten or not. When a logical address of incoming write is already mapped to physical address, FTL recognize that incoming write is logically overwritten, allocate new physical address, remaps logical address to the new physical address, and invalid old physical address.

Generally, OS generate logical invalidation, that is, some data are not used anymore, whereas FTL has no idea about the logical invalidation since it identifies data invalidation with only overwriting of logical address. The logically invalid data is considered as valid within NAND flash memory without any other information, until the logical addresses of these are overwritten. To inform these from host to NAND flash memory, TRIM command is introduced by ATA standard Technical Committee T13[4]. The TRIM command is composed of logical address and length of address to be invalidated. Data invalidation at outside flash device is achieved by interface command, such as trim, that allows an operating system to inform a flash device which blocks of data are no longer considered in use and can be wiped internally. The invalidation of duplicated data leads reducing of write and erase operations, and garbage collection efficiency for flash device.

### 3 Data Invalidation and Trim

The IO bandwidth for NAND flash storage can be increased as much as the amount of logical invalid region is identified within operating system, if the logically invalidated data is also invalidated physically since it increases flash utilization. There are many types of IO requests generated by Operating System, and almost of these are from file system operations, memory caching, and virtual memory IO operations. File systems generates normal write operations such as file write, truncate, delete, move, copy, directory create, delete, move, attribute updates, as well as metadata write operations. Memory management system in Operating System is also one of the major resources for IO request which includes cache management, demand paging mechanism, and swap operations for virtual memory. Some of these IO requests generated by

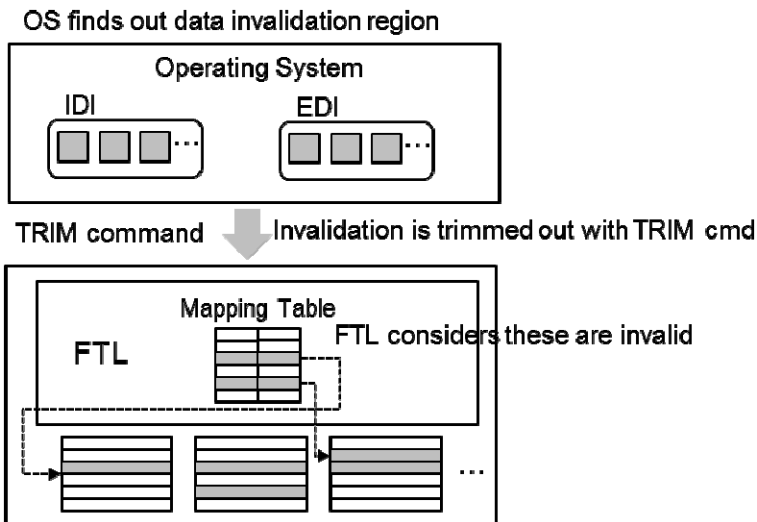


Fig. 1. Physical data invalidation of logically invalidated area with TRIM command

Operating System generate logical invalidation, which can be physically invalidated via TRIM commands.

The logical invalidate IO requests can be explicitly identified, or implicitly identified by Operating System. The explicit logical invalidation is the invalidation for IO request that is explicitly generated invalid data by Operating System, or some modules within Operating System. There are several well-known explicit logical invalidations within Operating System. When file is deleted, the file system just reset the file's metadata, and leaves data of the file unchanged. In this case, file's data are considered as valid in flash storage although these are not used anymore by host. The temporary files, temporary buffer region for file downloads also generate explicit logical invalidation when there are not used anymore. Applications that use data region used for transient data location generate explicit logical invalidation. These explicit logical invalidation data can be physically invalidated by calling explicit TRIM command. When calling TRIM, the proper valid logical address and size of the invalidation is identified explicitly from corresponding part or Operating System, and these are transferred to flash controller. The flash controller easily eliminate the region by altering FTL mapping table and related meta-information from valid to invalid without large overhead.

Implicit logical invalidation means the unintentionally generated invalidation by internal operations of operating system. Virtual memory would generate this type of invalidation when victim pages are selected. The victim block is not used anymore in VM, so this can be eliminated from swap region, however virtual memory does not notify to swap region that is deleted. Flash memory recognizes the physical region of victim block still has valid data although it is not used by operating system. Journal operation of file system also generates implicit logical invalidation. The successive commit operations generate duplication of same journal blocks that don't need to be synced to home location. Checkpoint operation also generates massive logical invalidation, since all data within journal region are synced home location, although file system does not recognize it generates data invalidation. These can also be physically invalidated by calling TRIM command.

The physical invalidation for logically invalidated data is described in Fig.1. In our design, the implicit and explicit logical invalidated data is managed with IDI list and EDI list, respectively. For each identifying, the to-be-invalidate data is inserted into the list along with its type. And At some point, the elements within the lists are transferred to flash memory via TRIM command to be invalidated physically. The invalidation sequence is circulated. By doing this, utilization of flash memory storage would be increased.

## 4 Conclusion and Further Work

NAND flash memory is becoming main storage media in computing systems. To compensate the limitations of flash memory, Flash Translation Layer (FTL) is employed between host systems and flash device, and translates logical address and physical address. Since FTL provides logical address to operating system and hides



physical address, operating system has no idea about the real location of data. Even the operating system thought that a data is considered as delete, FTL might recognize the data is being valid. The physical area of logically invalidated area can be invalidated by newly introduced command, called TRIM command. Thus, it is important to identify the logical invalidation within operating system. In Operating System, there are lots of logical invalidation region, however, these are not deleted physically. In this paper, we identify data invalidation within Operating System as two parts; explicit data invalidation and implicit data invalidation. These identified data invalidation can be invalidated immediately by calling TRIM command, which can increase utilization of NAND flash memory-based storage systems, and thus increase bandwidth of storage systems. In this paper, we just investigate the logical invalidation with Operating System and classify these with two types. Also we show the pilot design of the invalidation management mechanism within Operating System. For the further work, we implement this algorithm in Linux Operating System with more specific design and detailed implementation issues.

**Acknowledgement.** The Research was supported by the NRF grant funded by the MEST (No. 2010-0021094).

## References

1. Ban, A.: Flash file system optimized for page-mode flash technologies. U.S. Patent 5,937,425 (Filed October 16, 1997)
2. Intel Corporation, Understanding the flash translation layer (FTL) specification, <http://developer.intel.com/>
3. Choi, H.J., Lim, S.-H., Park, K.H.: JFTL: A flash translation layer based on a journal remapping for flash memory. *ACM Transactions on Storage* 4(4), 1–22 (2009)
4. Shu, F.: Data set management commands proposal for ata8-acs2. In: T13 Technical Committee, United States: At Attachment: e07154r1 (2007)
5. Samsung Electronics. Nand flash-memory datasheet (2011), <http://www.datasheetcatalog.com/samsungelectronic/41/>
6. Avantika Mathur, M.C., Bhattacharya, S.: The new ext4 filesystem: current status and future plans. In: *Proceedings of the Linux Symposium*, pp. 21–33 (2007)
7. Intel Corporation, Intel® High Performance Solid State Drive - Advantages of TRIM. Intel.com (2010)
8. Microsoft, Windows 7 Enhancements for Solid-State Drives. Microsoft downloads. Microsoft Corporation (2009)
9. Doytchev, N.: HOWTO: Configure Ext4 to Enable TRIM Support for SSDs on Ubuntu and Other Distributions (2012), <https://sites.google.com/site/lightrush/random-1/howtoconfigureext4toenabletrimforssdonubuntu>

# Moving Human Detection Using Motion Depth in Depth Image Sequences

Bong-Hwa Hong<sup>1</sup> and Kiseo Park<sup>2</sup>

<sup>1</sup> Dept. of Digital Media Engineering, Kyung Hee Cyber University,  
26, Kyungheedaero, Dongdaemun-gu, Seoul, Korea

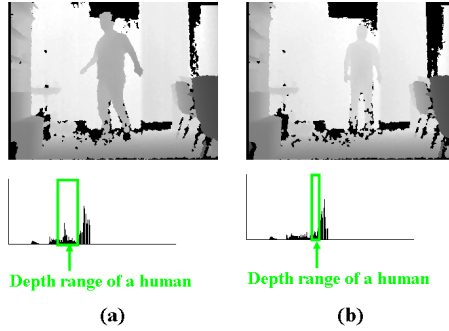
<sup>2</sup> Dept. of Convergence Biomedical Engineering, Daelim University College,  
29, Imgok-ro, Dongan-gu, Anyang-si, Gyeonggi-do, Korea  
bhong@khu.ac.kr, arsenide@daelim.ac.kr

**Abstract.** We propose a novel method that detects moving objects in depth image sequences using background images and motion depth (MD). The background image represents the camera view with no moving objects and the MDs are the depth values of moving objects. Foreground regions can be easily detected by background subtraction; however, the foreground regions have some noise and do not contain a regional diffusion of depth values. Therefore, we apply both background subtraction and motion depth detection to detect moving objects. Experimental results show the proposed method robustly detects moving objects, even if the moving region exists in close proximity to the background surfaces. Therefore, the method can be widely used to detect moving objects in depth image sequences.

**Keywords:** Depth segmentation, Object detection, Motion depth.

## 1 Introduction

Depth cameras such as Microsoft Kinect [1] can easily measure a distance to the nearest surface at each pixel by the time-of-flight (TOF) principle [2] or structured light imaging [3]. Moving objects may be more easily detected by depth cameras than visual cameras because moving objects are closer than background surfaces. However, moving objects cannot be easily detected in cluttered or cramped rooms. Background subtraction methods [4] or range segmentation methods based on depth histograms [5] have been used to detect moving objects; however, moving objects existing in close proximity to background surfaces are not easily detected by these methods as shown in Fig. 1(b) because the distance is short between the background surfaces and the moving object. In order to overcome this drawback, we propose a novel detection method for the depth image sequences using both background estimation and motion depth (MD) distribution, which is an updated distribution of depth values for moving regions.



**Fig. 1.** Depth images for moving bodies; (a) Body depth with isolated range, and (b) Body depth included in background range

## 2 Motion Depth Detection

The MDs are variable depth values of the moving regions, so we use an MD distribution as a Gaussian distribution given by

$$\eta_t(m) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{(m - \mu_t)^2}{2\sigma_t^2}\right), \tag{1}$$

where  $m$  denotes the MD value,  $\mu_t$  and  $\sigma_t$  denote the mean and standard deviation at time  $t$ , respectively. In order to update  $\eta_t(m)$ , a set of valid MDs,  $M^v = \{m_i \mid i = 1, 2, \dots, n\}$ , are detected by the following condition:

$$d_{t-1}(x, y) - d_t(x, y) > \tau^m, \tag{2}$$

where  $\tau^m$  denotes a positive threshold value, and MDs having  $d_t(x, y) > d_{t-1}(x, y)$  are not selected because the moving objects are closer than the backgrounds with respect to the camera. If any depth value of pixels satisfies condition (2), then the depth value is appended to an element of  $M^v$  and the frequency of the depth value,  $F(m_i)$ , is also computed for updating  $\eta_t(m)$ .

Using  $M^v$ ,  $\eta_t(m)$  is updated as follows:

$$\mu_{t+1} = \mu_t + \alpha^m \rho_t, \tag{3a}$$

$$\sigma_{t+1}^2 = (1 - \alpha^m)\sigma_t^2 + \alpha^m \phi_t^2, \tag{3b}$$

where  $\alpha^m$  denotes an updating gain ranging  $0 < \alpha^m < 1$ ,  $\rho_t$  is a weighted summation of  $m_i - \mu_t$  and  $\varphi_t$  is a weighted summation of  $(m_i - \bar{m})^2$ , where  $\bar{m}$  is the weighted mean of valid MDs calculated by

$$\sum_{i=1}^n m_i w_i . \tag{4}$$

Each weight of  $m_i$  computing for  $\rho_t$ ,  $\varphi_t$  and  $\bar{m}$  is the frequency proportion of each  $m_i$  to all elements of  $M^v$ , that is

$$w_i = F(m_i) / \sum_{k=1}^n F(m_k) . \tag{5}$$

Since  $\eta_t(m)$  presents the depth of moving objects, we detect moving objects using

$$r_t(x, y) = \begin{cases} 1 & \text{if } |d_t(x, y) - \lambda \mu_t| < \sigma_t , \\ 0 & \text{otherwise} \end{cases} , \tag{6}$$

where  $\lambda$  is a factor for ranging the width of motion depth.

### 3 Moving Human Detection

Backgrounds of depth sequences represent the surfaces farthest from a camera, so foregrounds can be detected by

$$f_t(x, y) = \begin{cases} 1 & \text{if } b_t(x, y) - d_t(x, y) > \tau^b , \\ 0 & \text{otherwise} \end{cases} , \tag{7}$$

where  $f_t(x, y)$ ,  $d_t(x, y)$  and  $b_t(x, y)$  are foreground, depth and background pixels at  $(x, y)$  and time  $t$ , respectively, and  $\tau^b$  is a positive threshold value.

The initial background  $b_0(x, y)$  is  $d_0(x, y)$ , but  $b_t(x, y)$  must be updated over time because false background values of  $b_0(x, y)$  must be corrected and depth values are not always stable. Thus, we update the background as follows:

$$b_{t+1}(x, y) = (1 - \alpha^b) b_t(x, y) + \alpha^b d_t(x, y) , \tag{8}$$

where  $\alpha^b$  denotes an updating gain ranging  $0 < \alpha^b < 1$ , and the larger the gain the faster the background is updated by the depth values of recent frames.

In order to detect moving objects that exist in close proximity to the background surfaces, a small value of  $\tau_b$  must be utilized; however,  $f_t(x, y)$  involves not only

moving objects, but also noisy regions. However,  $r_t(x, y)$  is able to determine the depth of moving objects and we detect moving regions using  $f_t(x, y) \wedge r_t(x, y)$  as shown in Fig. 2.

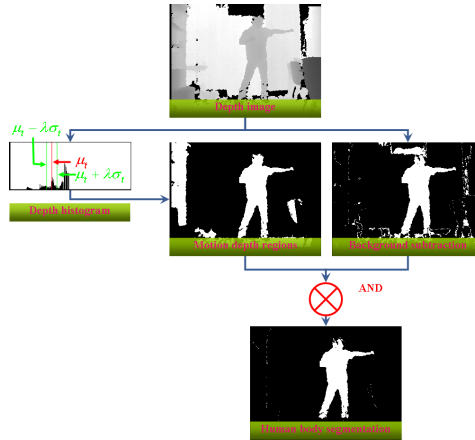


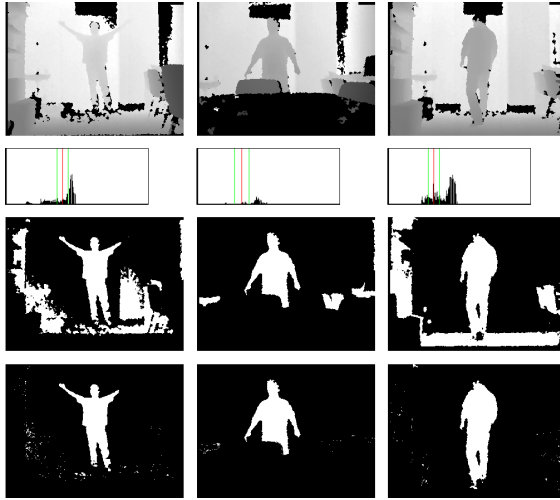
Fig. 2. An example of moving object detection

## 4 Experimental Results

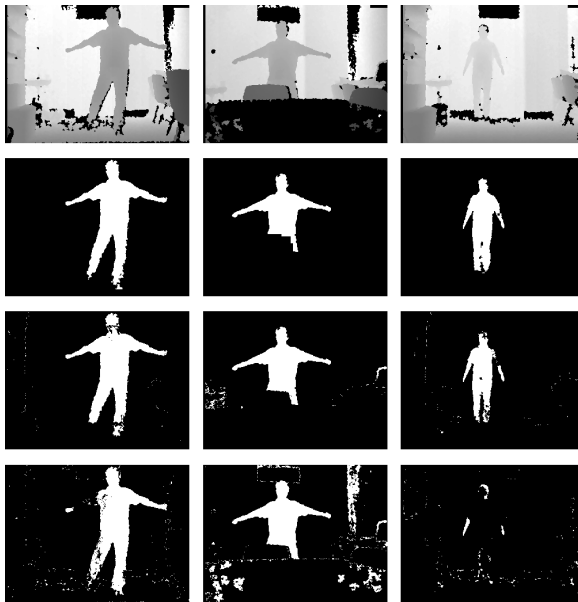
The proposed method was tested on a Pentium PC (Core™ 2 Duo, 2.40 GHz). Three depth image sequences were acquired from a Microsoft Kinect camera using Microsoft Kinect SDK [6]. The resolution of the test images was  $320 \times 240$  (13bits/pixel). All test sequences were acquired in a room where the distance from the camera to the front view surface was about 4m; a person taken the whole body moves in the first (DB #1) and third (DB #3) sequences, and a person taken the upper body moves in the second (DB #2) sequence. Examples of the results of the proposed method are shown in Fig. 3, where the estimated motion depth regions are shown in the third row.

To evaluate the performance of the proposed method, we compared the method with the detection of the background subtraction in [7] (MoG, or mixture of Gaussian, method) and ground truth images were manually segmented. We used  $\alpha^b = 0.001$  and  $\alpha^m = 0.2$ , where  $\alpha^m$  should be a large value because the MD distribution must contain recent MD information. In the experiments, our method successfully detected human regions as shown in Fig. 4. Specifically, our method was able to segment human regions close to the background surfaces, as shown in the last image of Fig. 4 and Fig. 5.

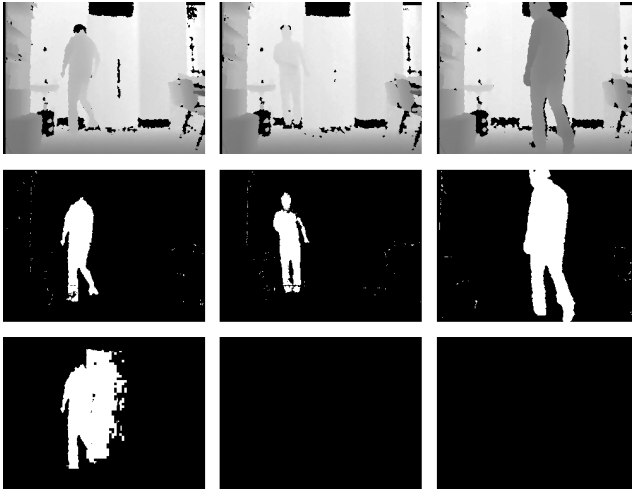
The MSE (mean squared error) with the ground truth image are shown in Table 1. The detection accuracy of the proposed method is higher than that of the MoG method. Consequently, the false detection of background subtraction can be recovered by detecting moving regions using MD distribution.



**Fig. 3.** Human body segmentation results (depth, depth histogram, motion depth region and segmentation result from the top to the bottom rows, respectively and sample frames in DB #1, DB #2 and DB #3 from the left column)



**Fig. 4.** Comparison with the background subtraction method (depth, ground truth, segmentation of the proposed method and segmentation of the MoG method from the top to the bottom rows, respectively and sample frames in DB #1, DB #2 and DB #3 from the left column)



**Fig. 5.** Comparison with the background subtraction method (depth, segmentation of the proposed method and segmentation of the MoG method from the top to the bottom rows, respectively)

**Table 1.** Segmentation comparison of the proposed method and the MoG method

DB	Method	MSE					
		1	2	3	4	5	Ave.
#1	Proposed	0.0228	0.0201	0.0021	0.0021	0.0011	0.0096
	MoG	0.0514	0.0747	0.0637	0.0509	0.0439	0.0569
#2	Proposed	0.0093	0.0068	-	-	-	0.0081
	MoG	0.0554	0.0389	-	-	-	0.0472
#3	Proposed	0.0048	0.0028	0.0270	0.0099	0.0129	0.0115
	MoG	0.0218	0.0531	0.0655	0.0660	0.0703	0.0553

## 5 Conclusion

We proposed a novel method that detects moving objects using background subtraction and motion depth detection using the MD distribution in depth image sequences. Since a distribution of MDs is updated as the Gaussian distribution, the moving regions that are in close proximity to the background surfaces can be successfully detected. Therefore, our method can significantly contribute to human action recognition systems based on depth cameras. In future work, we will develop a human action tracking system using the proposed method, object modeling and region tracking.

## References

1. Microsoft Kinect, <http://www.xbox.com/Kinect>
2. Lange, R., Setiz, P.: IEEE Journal of Quantum Electron. 37, 3 (2001)
3. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), pp. 195–202 (2003)
4. Crabb, R., Tracey, C., Puranik, A., Davis, J.: Real-time foreground segmentation via range and color imaging. In: IEEE CS Conference of Computer Vision and Pattern Recognition Works, pp. 1–5 (2008)
5. Parvizi, E., Wu, Q.M.J.: Multiple object tracking based on adaptive depth segmentation. In: Canadian Conference of Computer and Robot Vision, pp. 273–277 (2008)
6. Microsoft Kinect SDK beta,  
<http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk/download.aspx>.
7. Zivkovic, Z., Heijden, F.: Pattern Recognition Letter 27, 7 (2006)



# Microbial Fuel Cell for STEAM Teaching Tools and Method

Yeonghae Ko and Namje Park\*

Major in Elementary Computer Education, Department of Primary Education,  
Graduate School of Education, Jeju National University,  
61 Iljudong-ro, Jeju-si, Jeju Special Self-Governing Province, 690-781, Korea  
{smakor, namjepark}@jejunu.ac.kr

**Abstract.** This paper is about microbial fuel cell for teaching tool and the method for fabricating the cell, particularly the cell and the method designed to help students to understand the principle of generating electricity in an easier and more fun way using the principle of transferring the electrons obtained from the microorganisms in the course of metabolizing organic matters while recognizing the importance of developing alternative energy.

**Keywords:** STEAM, Microbial Fuel, Cell, Elementary School.

## 1 Introduction

Bioenergy is the energy obtained from biomass such as tree, food waste, and animal excretion. Since the petroleum, the major fossil fuel currently used, is expected to exhaust within few decades, bioenergy is received as the realistic alternative to oil.

Bioenergy uses the plant which grows permanently as the source of energy that it is free from the issue of depletion and the carbon dioxide produced by the use of bioenergy is reabsorbed into the plant in the course of growth that it emits little carbon dioxide.

In major advanced nations, new and renewable energy accounts for 3~7% of primary energy and also bioenergy accounts for 30~60% of the new and renewable energy which is quite high. Korea conducted extensive studies on the development of biomass energy technologies. Some technologies were commercialized and others are in the basic or the applied research stage. The typical technological developments include the energy generation using organic waste and the production and utilization of biodiesel.

This paper is about microbial fuel cell for teaching tool and the method for fabricating the cell, particularly the cell and the method designed to help students to understand the principle of generating electricity in an easier and more fun way using the principle of transferring the electrons obtained from the microorganisms in the course of metabolizing organic matters while recognizing the importance of developing alternative energy.

---

\* Corresponding author.

## **2 Relationship between Bioenergy and STEAM Education**

Bioenergy is the alternative energy which will take up a large part of our living in the future. Currently the environmental issue is dealt with quite seriously in the elementary school, and the bioenergy as the new science and technology to resolve energy and environmental issues is the IT subject that can attract both interest and attention of students.

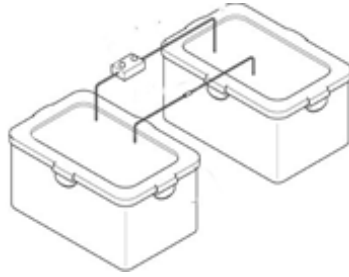
STEM Education, the predecessor of STEAM Education, was stimulated by low interest in science and math and low achievement rate in such subjects. Considering the objective of STEAM Education, increasing student's interest in science, technology, and engineering through new approaches, bioenergy is the new technology that students can approach with interest. For it is related with the technologies that may be commercialized in a few years and with the problems found in living, students can change their attitude toward science and technology in real life and also apply what they learned at home, which are the advantages of this education program.

## **3 Teaching Tool Prepared with Microbial Fuel Cell**

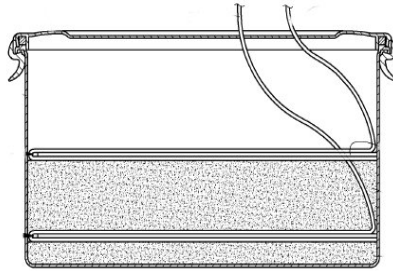
A microbial fuel cell was developed as in (Figure 3) for STEAM Education on the theme of bioenergy for elementary school students who are the subject of study. Microorganisms were cultured using mud and mechanical pencil lead was used as the electrodes. The electricity generated was collected with condenser and used to light LED to check the result visually. Each cell produced 250mV-320mV voltage, and 2 cells were connected in series.

### **3.1 Microbial Fuel Cell for Teaching Tools**

The exhaustion of energy resources such as petroleum and coal is getting intensified, and the development of alternative energy resources that can replace oil and coal is urgently needed. Currently developed alternative energy sources are generally divided into natural energy and artificial energy. The natural energy is again divided into the solar energy, wind energy, and tidal energy that can be obtained from the nature, and body energy, and the artificial energy is divided into the nuclear fusion energy using hydrogen and helium, the hydrogen energy that can be obtained through the electrolysis of water and the pyrolysis of natural gas, and the fuel cell that can convert the chemical energy of fuel into electric energy. This study developed the teaching tool using a fuel cell that can help students understand the principle of fuel cell more easily and learn about the subject with fun to enhance their learning capacities. This study suggested a teaching tool prepared with microbial fuel cell and the method for preparing the tool which is designed to resolve the said issue with the intention of help students learn about the principle of fuel cells in an easier and more fun manner and also the teacher to easily prepare a teaching tool to enhance the learning capacities of students.



**Fig. 1.** Perspective drawing representing the microbial fuel cell for teaching tool



**Fig. 2.** Sectional drawing representing the section of the cell body

### 3.2 Process of Fabricating a Teaching Tool Using Microbial Fuel Cell

The following explains the process of fabricating a teaching tool using microbial fuel cell dealt with in this paper. According to Figures 1 and 2, the microbial fuel cell for teaching tool in this study has the main body of the cell which is an air-tight container to accommodate combined mud (20) which is composed of 10~15 wt.% of water, 1~2 wt.% of salt, and mud for remaining portion. The main body of the cell is composed of the lid and the container. The combined mud above must be composed of 10 wt.% of water, 1 wt.% of salt, and mud for remaining portion to form the most ideal microbial fuel cell for teaching tool.

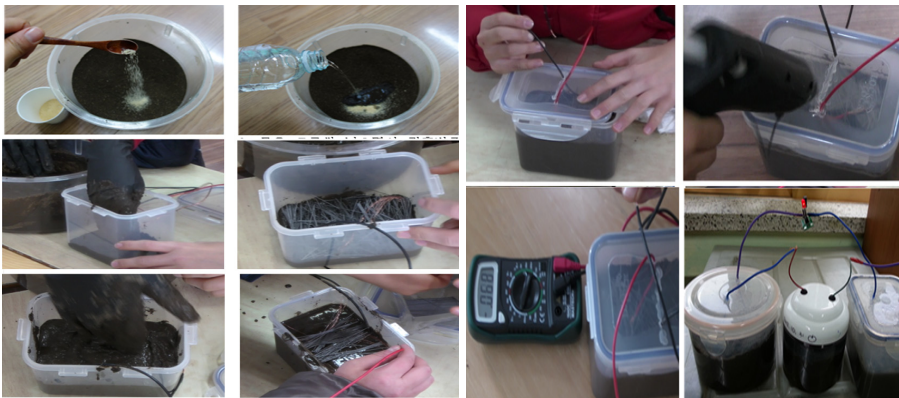
The combined mud shall be divided into Layer 1, Layer 2, and the Bottom Layer, and Layer 1 is placed on top of Bottom Layer and Layer 2, on top of Layer 1 while keeping a distance to each other and the cathode is placed between Bottom Layer and Layer 1 and anode between Layer 1 and Layer 2. In detail, the cathode is located between Bottom Layer and Layer 1 while the other end is composed of Wire 1 which is drawn to the outside of the body and Conductor 1 which is electrically connected to an end of Wire 1.

Also, the anode is located between Layer 1 and Layer 2 while the other end is composed of Wire 2 which is drawn to the outside of the body and Conductor 2 which is electrically connected to an end of Wire 2. Wires 1 and 2 are composed of coated wires containing multiple strands of small wires, and the parts that come in contact with each conductor are stripped. Wires are placed around the conductor to ease the

flow of current. In other words, wires are stripped up to approximately twice the length of the conductor to cover the top and bottom side of the conductor. Mechanical pencil lead or ordinary pencil lead is recommended for Conductors 1 and 2. This is because the structure suggested in this study is the teaching tool that can be built with the materials that can be easily found.

On the other hand, more than 1 main body of the cell can be prepared and connected in series. Connecting cathodes with anodes can produce electricity with higher voltage using multiple microbial fuel cells.

The main body of the cell can use condenser to store certain level of energy produced from the cell. This enables the cell to gather up the electric energy obtained from the metabolism of organic matters in the combined mud for later use. In this case, the electrons produced through the metabolism of organic matters by microorganisms included in the combined mud are received by cathode and then transferred to the anode along the wire which is to be described later to produce voltage. When the number of microorganisms increase the voltage of the microbial fuel cell can rise up to 0.4~0.7V and 1.8V by using condenser which is sufficient for lighting LED. However, there is limit in connecting cells that the boosting transformers can be added to Wires 1 and 2 in case higher voltage is required.



**Fig. 3.** Process of preparing a teaching tool using microbial fuel cell

Students in low to high grades can understand and learning about the principle of the microbial fuel cell which produces energy through microorganisms in an easier and fun manner while understanding the importance in developing alternative energy by using the teaching tool prepared with microbial fuel cell suggested in this paper and the method for preparing the cell.

## 4 Conclusion

The STEAM Education Program on the theme of bioenergy suggested in this study can increase students' interest in STEAM Education and bioenergy to resolve

problems in real life and develop knowledge of and positive attitude toward science and technology. Furthermore, the program deals with the contents students have already learned during in the school curriculum, and therefore students can incorporate the contents they learned into the process of solving problems. Students can also develop self-directed learning capabilities and problem solving capabilities while planning, implementing, and evaluating the process of producing microbial fuel cell and designing biohouse by themselves.

## References

1. Park, N., Kwak, J., Kim, S., Won, D., Kim, H.: WIPI Mobile Platform with Secure Service for Mobile RFID Network Environment. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 741–748. Springer, Heidelberg (2006)
2. Park, N.: Security scheme for managing a large quantity of individual information in RFID environment. In: Zhu, R., Zhang, Y., Liu, B., Liu, C. (eds.) ICICA 2010. CCIS, vol. 106, pp. 72–79. Springer, Heidelberg (2010)
3. Park, N.: Secure UHF/HF Dual-Band RFID: Strategic Framework Approaches and Application Solutions. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS, vol. 6922, pp. 488–496. Springer, Heidelberg (2011)
4. Park, N.: Implementation of Terminal Middleware Platform for Mobile RFID computing. *International Journal of Ad Hoc and Ubiquitous Computing* 8(4), 205–219 (2011)
5. Park, N., Kim, Y.: Harmful Adult Multimedia Contents Filtering Method in Mobile RFID Service Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS, vol. 6422, pp. 193–202. Springer, Heidelberg (2010)
6. Park, N., Song, Y.: AONT Encryption Based Application Data Management in Mobile RFID Environment. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS, vol. 6422, pp. 142–152. Springer, Heidelberg (2010)
7. Park, N.: Customized Healthcare Infrastructure Using Privacy Weight Level Based on Smart Device. In: Lee, G., Howard, D., Ślęzak, D. (eds.) ICHIT 2011. CCIS, vol. 206, pp. 467–474. Springer, Heidelberg (2011)
8. Park, N.: Secure Data Access Control Scheme Using Type-Based Re-encryption in Cloud Environment. In: Katarzyniak, R., Chiu, T.-F., Hong, C.-F., Nguyen, N.T. (eds.) Semantic Methods for Knowledge Management and Communication. SCI, vol. 381, pp. 319–327. Springer, Heidelberg (2011)
9. Park, N., Song, Y.: Secure RFID Application Data Management Using All-Or-Nothing Transform Encryption. In: Pandurangan, G., Anil Kumar, V.S., Ming, G., Liu, Y., Li, Y. (eds.) WASA 2010. LNCS, vol. 6221, pp. 245–252. Springer, Heidelberg (2010)
10. Park, N.: The Implementation of Open Embedded S/W Platform for Secure Mobile RFID Reader. *The Journal of Korea Information and Communications Society* 35(5), 785–793 (2010)
11. Park, N., Ko, Y.: Computer Education's Teaching-Learning Methods Using Educational Programming Language Based on STEAM Education. In: Park, J.J., Zomaya, A., Yeo, S.-S., Sahni, S., et al. (eds.) NPC 2012. LNCS, vol. 7513, pp. 320–327. Springer, Heidelberg (2012)

12. Ko, Y., An, J., Park, N.: Development of Computer, Math, Art Convergence Education Lesson Plans Based on Smart Grid Technology. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H., et al. (eds.) SecTech, CA, CES3 2012. CCIS, vol. 339, pp. 109–114. Springer, Heidelberg (2012)
13. Ko, Y., Park, N.: Experiment and Verification of Teaching Fractal Geometry Concepts Using a Logo-Based Framework for Elementary School Children. In: Kim, T.-h., Adeli, H., Slezak, D., Sandnes, F.E., Song, X., Chung, K.-i., Arnett, K.P., et al. (eds.) FGIT 2011. LNCS, vol. 7105, pp. 257–267. Springer, Heidelberg (2011)
14. An, J., Park, N.: The Effect of EPL Programming Based on CPS Model for Enhancing Elementary School Students' Creativity. In: Park, J.J(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. Lecture Notes in Electrical Engineering, vol. 181, pp. 237–244. Springer, Heidelberg (2012)
15. An, J., Park, N.: Computer Application in Elementary Education Bases on Fractal Geometry Theory Using LOGO Programming. In: Park, J.J., Arabia, H., Chang, H., et al. (eds.) IT Convergence and Services. LNEE, vol. 107, pp. 241–249. Springer, Heidelberg (2011)
16. Kim, Y., Park, N.: Development and Application of STEAM Teaching Model Based on the Rube Goldberg's Invention. In: Yeo, S., Pan, Y., Lee, Y.S., et al. (eds.) Computer Science and its Applications. LNEE, vol. 203, pp. 693–698. Springer, Heidelberg (2012)
17. Kim, Y., Park, N.: The Effect of STEAM Education on Elementary School Student's Creativity Improvement. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H., et al. (eds.) SecTech, CA, CES3 2012. CCIS, vol. 339, pp. 115–121. Springer, Heidelberg (2012)
18. Kim, Y., Park, N.: Elementary Education of Creativity Improvement Using Rube Goldberg's Invention. In: Park, J.H(J.), Kim, J., Zou, D., Lee, Y.S. (eds.) ITCS & STA 2012. LNEE, vol. 180, pp. 257–263. Springer, Heidelberg (2012)
19. Park, N., Cho, S., Kim, B., et al.: Security Enhancement of User Authentication Scheme Using IVEF in Vessel Traffic Service System. In: Yeo, S., Pan, Y., Lee, Y.S., et al. (eds.) Computer Science and its Applications. LNEE, vol. 203, pp. 699–705. Springer, Heidelberg (2012)
20. Kim, K., Kim, B.-D., Lee, B., Park, N.: Design and Implementation of IVEF Protocol Using Wireless Communication on Android Mobile Platform. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H., et al. (eds.) SecTech, CA, CES3 2012. CCIS, vol. 339, pp. 94–100. Springer, Heidelberg (2012)
21. Kim, G., Park, N.: Program Development of Science and Culture Education Tapping into Jeju's Special Characteristics for Adults. In: Kim, T.-h., Stoica, A., Fang, W.-c., Vasilakos, T., Villalba, J.G., Arnett, K.P., Khan, M.K., Kang, B.-H., et al. (eds.) SecTech, CA, CES3 2012. CCIS, vol. 339, pp. 133–138. Springer, Heidelberg (2012)
22. Hong, J., Park, N.: Teaching-learning Methodology of STS Based on Computer and CAI in Information Science Education. In: Yeo, S., Pan, Y., Lee, Y.S., et al. (eds.) Computer Science and its Applications. LNEE, vol. 203, pp. 733–738. Springer, Heidelberg (2012)

# Research on Educational Use of Smart-Phone Applications with Smart Clicker Technique

Ji-Hye Bae and Sung-Ki Kim

Division of Information Technology Education, Sun Moon University  
100 Kalsan-ri, Tangjeong-myeon, Asan-si, Republic of Korea 336-708  
{angdoo98, skkim}@sunmoon.ac.kr

**Abstract.** Advancement of smart-phone and its market expansion according to the era of rapidly changing ubiquitous, led to the development of numerous educational applications. The biggest advantage of smart-phone is that learning can be achieved anywhere and at any time without limit, and it can create an effective learning environment for instructors and students. In this research, we developed instructional strategies for learning and interactive communications by utilizing smart clicker applied apps on smart-phone for effective communications between instructors and students, and with satisfied educational results, as well as the attempt to measure the students' studying results by using such strategies in actual education fields. As a result, educational effects can be obtained for the usefulness of learning, motivation, and interactivity.

**Keywords:** Smart-phone, Smart Clicker App, Learning, Instructional Strategy, Educational Effects.

## 1 Introduction

In modern era, knowledge expansion system is quickly changing due to the rapid development of mobile devices based next generation IT convergence technology, and information acquiring is progressing actively and swiftly as well. In particular, the advancement of mobile devices such as smart-phones, smart-pads and others increased the number of users, and accordingly, numerous apps in various fields were released and are in use [1]. Among these apps, apps for educational use are in the development stage and can be regarded as one field of the infrastructure constructions for the enhancement of education quality [2]. However, the expected effects cannot be met even though diversified apps' contents for educational use are provided, unless properly used by instructors and students. In addition, it is extremely difficult to improve quality of communication between instructors and students, and learning immersion, academic achievements while meeting diverse demands of students with completely different learning styles. After all, there is a need for an instructional strategy to become the center of communications with students by designing the learning procedures and methods for diversified students in the actual education fields. In this paper, we suggested instructional strategies for learning and interactive

communication by utilizing Smart Clicker applied apps on smart-phone for effective communication between instructors and students, and for satisfying educational results, as well as the attempt to measure students' learning effects by applying such strategies to actual education fields.

## **2 Theoretical Background**

### **2.1 Examples of Educational Use of Smart-Phone Apps**

Self-directed and positive learning were promoted through smart-phones and users favoring mobile devices as a learning device tend to increase. This is because such system provides superior quality of learning during traveling, and customized learning is supported for each individual. Moreover, necessary information can be obtained through social networking services. Smart-phone equipped with such features can be regarded as having the best condition as an educational medium. According to the research of [2], among the results researched by 148Apps.biz in 2011, the number of educational app programs developed for iPhone accounts for approximately 8.5% of all developed apps. Due to the boost of such tendency, it is anticipated that the educational app development fields will be further promoted in the future.

However, as most of educational apps used on smart-phones comprise of repetitive exercise types, it does not appear that students use the interactive functions on smart-phones [3]. Research examples that achieved effective interaction results in the education fields by using smart-phones are as follows.

Research [4] found that in the research results regarding usefulness, improved aspects, and others by implementing smart-phone app that provides creative idea generation technique, the developed smart-phone app was effective in the value aspect and for bidirectional interactions with students' creating and sharing ideas easily. As the interactions between instructors and students are very important in large scaled lectures, research [5] aimed at raising educational effects in the large scaled group education by developing Android operating system based apps to overcome the limits of interactions. In research [6], they designed a smart-phone app that applies the concept of differentiated teaching aid tools, proposed an instructional strategy to diagnose and satisfy every student's demands, and presented the educational importance of effective and practical teaching aid tools for instructors. In research [7], it suggests that academic achievements and learning interest can be raised because of real time interactions and feedbacks, as a result of testing after implementing a smart-phone app enabling a collaborative learning system possible for interactions at all times for students and for real time feedback functions.

However, these researches are about the apps mainly for specialized learning, and are, in some degree, different from apps' generally used functions and for supporting easy and simplified operations. Unlike various educational use examples of



smart-phone apps as suggested, this research aims to focus on verifying the educational effects by properly utilizing previously developed apps with smart clicker technique for instructional strategy mainly targeting at students' interactions.

## 2.2 Overview of Smart Clicker App

Generally speaking, the clicker system in the educational environment means a device that can respond to questions of instructors in real time using a proper button style device during class hours. Instead of using this button style device, smart clicker incorporated student respond functions using app programs on smart-phones. Among those apps using smart clicker technology, Socrative app [8] is a typical one.

This app is a bidirectional communication lecture solution, can be downloaded free of charge, guarantees anonymity, and above all, it offers the advantage of obtaining students' responses without delay in the forms of visual materials. In addition, because of the ease to use and operate, even instructors not familiar with use of devices or software can give customized quiz and carry out communication strategies sufficiently. It is implemented using convenient User Interface (UI) for simplified response to instructors' lecture pattern after students download apps easily and connects to the system. Using this Socrative app, interaction problems in the large scaled lecture room, which is difficult for physical communication between students and instructors, can be supplemented, it appeared appropriate for education targeting at self-controlled college students over 20's who are used to smart-phone devices.

## 3 Educational Use Designs for Smart-Phone Apps with Smart Clicker Technique

### 3.1 Learner Focused Study Design System

Smart-phone is a tool not restricted by time and space and can use multimedia freely, but it also has disadvantages of relatively small size of screen and limited input devices compared with regular desktop PCs. There are many methodologies suggested to provide effective user inputs to overcome such limits, and Human Interface Guidelines [9] of Apple Company is regarded as a very useful instruction manual for smart-phone apps development. Among the contents, this research determined "Focus on the Primary Task" and "Give People a Logical Path to Follow" as a study design strategy like the design principle suggested in research [4], and implemented smart clicker based learning system using Socrative app.

Entire system design using such learning design strategy is shown in Fig.1. Based on the suggested UI design strategy, instructors develop learning contents for students' learning activities through apps, implement them on apps in addition to implementing occasional questions and quiz as a strategy for interactive communication. In this case, types such as '*true or false*', '*multiple choice*', '*short answer*' can be set up for feedbacks based on situations.

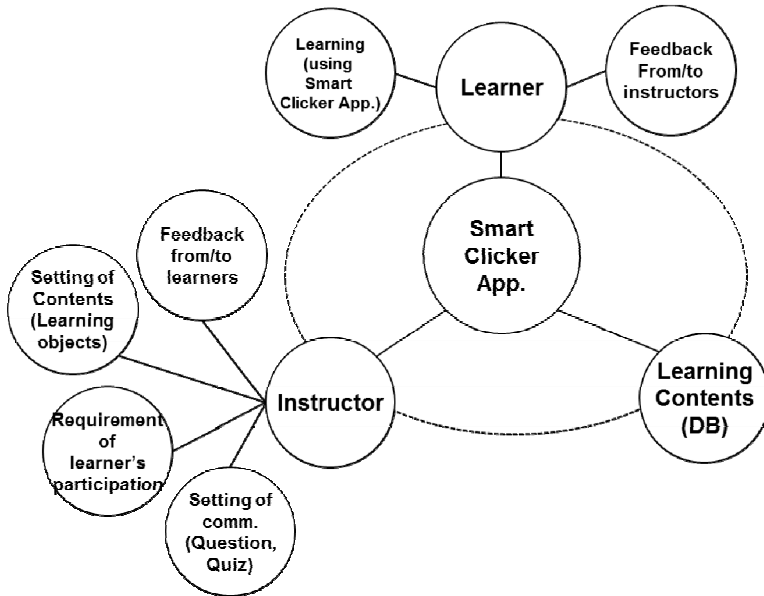


Fig. 1. Educational use designs for Smart-phone apps with Smart clicker technique

### 3.2 Research Method

The subjects of this research are 60 students taking computer classes as a liberal arts course at Korea's four year universities, and are comprised of freshmen in their early 20's and familiar with smart-phone device operations. For experimental comparison, we conducted the experiments in two groups with 30 students each. While typical memorization based lecture and paper based quiz/feedback were used for the controlled group, part of learning was proceeded with an app using smart clicker technique for the experimental group, and likewise, quiz/feedbacks were conducted through the app. For both groups, experiments were conducted for one week, and actual learning time was 3 hours each.

### 3.3 Evaluation Method

For evaluation method of research result analysis, a questionnaire among others was used as a typical method in this research, and contents were gathered by referring to suggested evaluation items in [4] for academic achievement and concentration degree analysis. Table 1 is the questionnaire used in the research and consists of 10 questions. It uses Likert style 5 points criteria, and is composed of "very true" 5 points, "true" 4 points, "some degree" 3 points, "not true" 2 points, and "not at all" 1 point. Credibility of questionnaire was shown as .81 in the calculation of Cronbach  $\alpha$  coefficient, and contents validity examination was performed by an instructional design expert.

**Table 1.** Questionnaire details

Evaluation area	Specific evaluation details
Usefulness of learning (3)	<ul style="list-style-type: none"> <li>- Feel that I learned new things.</li> <li>- This learning method will help my study even in the future.</li> <li>- It will be helpful for learning contents of a different class.</li> </ul>
Interest/Motivation(3)	<ul style="list-style-type: none"> <li>- Like to use this learning method for next class.</li> <li>- Liked the continuous engagement for effective learning.</li> <li>- Became more interested in learning with this method.</li> </ul>
Interactivity (3)	<ul style="list-style-type: none"> <li>- Can obtain feedbacks immediately with this learning method.</li> <li>- Responses from instructors to my answer sheet were useful and helpful.</li> <li>- Can communicate with other students smoothly.</li> </ul>
Convenience (1)	<ul style="list-style-type: none"> <li>- Was convenient to use tools.</li> </ul>

## 4 Educational Effects Investigation

In the result of questionnaire analysis, students in experimental group showed positive responses in overall items in comparison with students in controlled group, the difference was noticeable particularly in the aspect of interactivity. In the *t-test* results conducted for each item of learning usefulness, interest/motivation, interactivity, and convenience, as shown in Table 2, all three items, except the convenience item, indicated significant difference ( $p < .05$ ).

**Table 2.** Questionnaire results analysis among groups

Item	Controlled group	Experimental group	Difference in average score
Learning usefulness	3.50	4.35	0.85
Interest/Motivation	3.12	4.19	1.07
Interactivity	2.65	4.67	2.02
Convenience	4.20	4.31	0.11

Item	Group	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
Learning usefulness	Experimental group	30	4.35	.403	7.183	.01
	Controlled group	30	3.50	.508		
Interest/ motivation	Experimental group	30	4.19	.453	8.636	.01
	Controlled group	30	3.12	.499		
Interactivity	Experimental group	30	4.67	.383	18.935	.00
	Controlled group	30	2.65	.442		
Convenience	Experimental group	30	4.31	.455	1.018	.31
	Controlled group	30	4.20	.350		

## 5 Conclusion

This research suggests the effects on academic achievement by using apps with Smart Clicker technique in actual education fields. Smart Clicker technique based learning system using Socrative app was implemented by determining "Focus on the Primary Task" and "Give People a Logical Path to Follow" as instructors' major study strategy, and after testing the system in two groups, it was noted to have effects in the aspects of learning usefulness, motivation, and interactivity. However, this research pinpoints one problem, that UI design expansion proper to students' environment cannot be achieved further as it focuses on the utilization of the apps only. Therefore, for research in the years to come, instead of applying existing apps, it is necessary to implement new type of apps that emphasizes more on UI design and communication functions.

## References

1. Kim, E.-S., Park, J.-S.: A Study on Educational App Development using the App Authoring Tool. *The Journal of Digital Policy & Management* 10(5), 1–6 (2012)
2. Lee, J., Choi, J.: Implementation of Application for Vocabulary Learning through Analysis of Users Needs Using Smart Phone. *The Journal of Korean Association of Computer Education* 15(1), 43–53 (2012)
3. Si, J., Park, D., Chae, A., Kim, D.: Discussion-based Interface Design Research on the Smart phone at Cyber Universities. *The Journal of Korean Association of Computer Education* 14(5), 81–96 (2011)
4. Lee, H., Jung, E.: A Design and Implementation of a SmartPhone App Providing the SCAMPER Method. *The Journal of Korean Association of Computer Education* 14(5), 29–37 (2011)
5. Kim, I.-M.: Android phone app. development for large scale classes. *The Journal of Digital Policy & Management* 9(6), 343–354 (2011)
6. Cha, H., Ahn, M.-L.: New Design of a Smart-phone Version of PAL Tool. *The Journal of Educational Information and Media* 17(1), 91–108 (2011)
7. Lee, M.-S., Son, Y.-E.: Design and Implementation of Android-based Cooperative Learning System using Social Network Service. *The Journal of Korean Association of Computer Education* 14(5), 71–79 (2011)
8. Socrative App., <http://www.socrative.com>
9. Apple Inc., iOS Human Interface Guidelines, Apple Inc. (2011)

# Hierarchical Customization Method for Ubiquitous Web Applications

Wonjae Lee<sup>1,\*</sup>, Hyun-Woo Lee<sup>1</sup>, Min Choi<sup>2</sup>, Jong Hyuk Park<sup>4</sup>, and Young-Sik Jeong<sup>3</sup>

<sup>1</sup> Broadcasting & Telecommunications Media Research Laboratory,  
Electronics and Telecommunications Research Institute,  
Daejeon, South Korea

{russell,hwlee}@etri.re.kr

<sup>2</sup> School of Information and Communication Engineering, Chungbuk National University,  
Cheongju, South Korea

mchoi@cbnu.ac.kr

<sup>3</sup> Department of Multimedia Engineering, Dongguk University,  
Seoul, South Korea

ysjeong@dongguk.edu

<sup>4</sup> Department of Computer Engineering, Seoul National University of Science and Technology,  
Seoul, South Korea

jhpark1@seoultech.ac.kr

**Abstract.** Web-based application delivery has many advantages in ubiquitous computing. Customizations of Web applications are required to support anytime/anywhere/anymedia paradigm of ubiquitous computing. In this paper, we present a hierarchical customization method for ubiquitous Web applications. Contexts and namespaces for those contexts are hierarchically modeled in the method. Inheritance and polymorphism are applied to namespaces. Customizations for complex context hierarchies are supported by multilevel inheritance. Our implementation of the hierarchical customization method supports runtime customizations of ubiquitous Web applications for contexts.

**Keywords:** ubiquitous Web applications, context-aware Web applications, adaptive hypertext.

## 1 Introduction

Web-based application delivery has many advantages in ubiquitous computing [1]. Developers do not have to build applications for different platforms because Web applications are platform agnostic. Web applications generally have shorter development cycle than native applications do. Web applications are cheaper to develop and more manageable. As more applications migrate to the Web, more devices begin to support Web standards.

---

\* Corresponding author.

To support ubiquity, an application needs to adapt to its context which characterizes the environment the application is running in. Customizations are required to those adaptations. A ubiquitous Web application is adapted to a specific context by customization. Presentation, navigation structure, and contents can be customized to provide appropriate adaptation.

There are three orthogonal dimensions in the design space of customization: the kind of context, the granularity of adaptation, and the degree of customizability [2]. There are a user context, device context, network context, time context, and location context in terms of the kind of context. The granularity of adaptation ranges from micro adaptation to macro adaptation. The degree of customizability expresses that context and adaptation can be either static or dynamic. A customization method for ubiquitous Web applications needs to cover all three dimensions.

The Web Unified Modeling Language (WUML) [3] has been introduced to support ubiquity through customization. A generic customization model, a context model, and other models has been proposed in the work. In other work [4], a model of context called activity-centric context has been introduced to focus on developing context-aware applications that support cognitive activities.

In this paper, we present a hierarchical customization method for ubiquitous Web applications. The method supports customizations of ubiquitous Web applications with hierarchically modelled contexts.

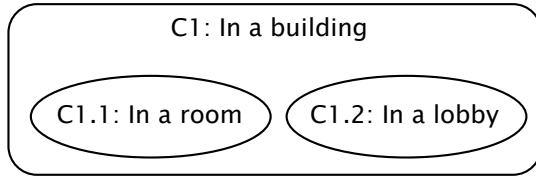
In the proposed customization method, contexts and namespaces for those contexts are hierarchically modeled. A namespace inherits contents from its ancestor namespaces like inheritance in object-oriented programming. Customizations for a context are achieved by adding contents to the corresponding namespace. Customizations for complex context hierarchies are supported by multilevel inheritance.

The hierarchical customization method can be used to implement macro adaptation and micro adaptation. For example, the method can be applied to file system for macro adaptation. For micro adaptation, it can be applied to retrieval of values.

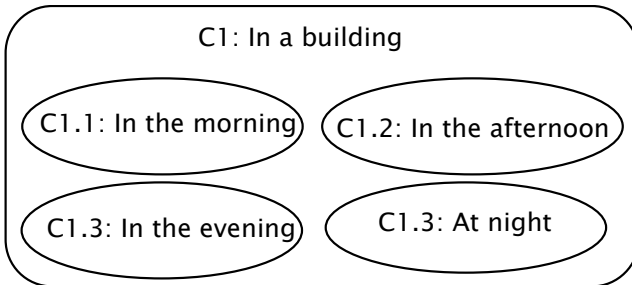
In our implementation, a Web application server dynamically executes the user interface codes and business logic codes that are transferred from a file server. This dynamic execution allows modification and customization of ubiquitous Web applications in runtime.

## 2 Context Model

Contexts can be hierarchically modeled as in Fig. 1 [4]. When a user is in a building, the context C1 can be applied. When the user is in a room of a building, the more specific context C1.1 can be applied. Different kinds of context can be used in the hierarchical context model as in Fig. 2.



**Fig. 1.** An example of the hierarchical context model. In this example, a location context is modeled. The generic context C1 is applied when a user is in a building with no more specific location information is available. When a more specific location of the user is available, the more specific con-text like C1.1 or C 1.2 can be applied.



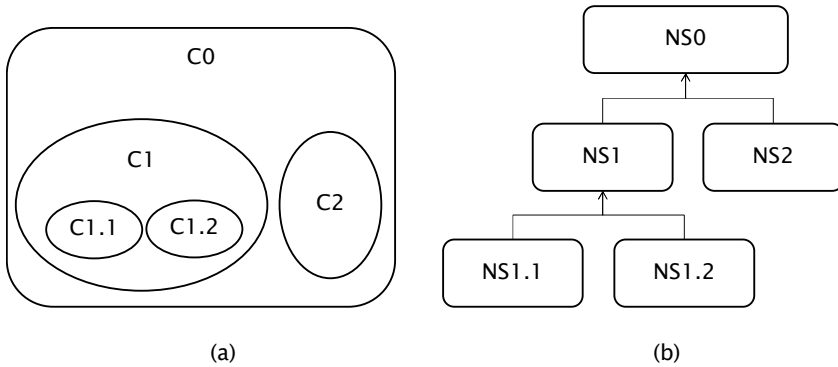
**Fig. 2.** An example of the hierarchical context model with mixed kinds of context. In this example, a location context and time context is modeled.

### 3 Hierarchical Customization Method

We propose a hierarchical customization method to support customization for the hierarchical context model. In the method, a namespace is assigned to each context. A namespace inherits contents from its ancestor namespaces as in object-oriented programming [7]. When customization is required for a context, you add contents to the corresponding namespace. Contents in a namespace override contents with same names of its ancestor namespace. When contents are requested for a user in a context, the contents of the corresponding namespace are provided.

The proposed method can be applied to different aspects of ubiquitous Web application development. When the method is applied to retrieval of values, it is suitable for micro adaptation. When the method is applied to retrieval of files, customization of images and style sheets can be easily done.

For an example of applying the method to retrieval of files, let’s assume that home.html and logo.gif files exist in the NS0 namespace of the Fig. 3. The logo.gif file is used in the home.html. Let’s assume that a different logo.gif file is uploaded into the NS1 namespace. When a user in the C1 context requests home.html, the home.html file of the NS0 namespace will be provided because the NS1 namespace inherits the home.html from the NS0 namespace. However, when the user in the context C1 requests logo.gif, the logo.gif in the NS1 namespace, which override the inherited logo.gif of the NS0 namespace, will be provided. In this way, a user is provided a customized image file for the context C1.



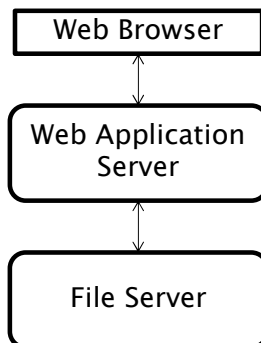
**Fig. 3.** Contexts and namespaces. A namespace  $NS_n$  is assigned to a context  $C_n$ . Contents in namespaces are inherited and overridden as in object-oriented programming.

Code reuse is promoted by the proposed hierarchical customization method. When a ubiquitous Web application is customized for a context, codes for higher level namespaces can be reused by inheritance. Contents that require more customization can be added to the namespace for the context.

The proposed customization method supports staged delivery [9] of a ubiquitous Web application. Customizations for more general contexts are developed and delivered in early stages. Then customizations for more specific contexts can be developed and delivered in later stages. In this way, time-to-market and risk can be minimized, and the value to the users can be maximized.

#### 4 Implementation

We have applied the hierarchical customization method to the file system. The architecture of the system is illustrated in Fig. 4.



**Fig. 4.** The architecture of the implemented system



The Web application server handles requests from users. When a user's request is sent to the Web application server, the Web application server needs to retrieve files from the file server to process the user's request. So, file requests are sent from the Web application server to the file server. When the Web application server receives the required files, the received files are sent to the user or dynamically executed.

Python program files [5] are used to implement business logics, and Chameleon template files [6] are used as executable user interface codes. Server-side dynamic pages are provided by the Chameleon template engine. When user interface codes and business logic codes are executed in the Web application server, the Web application server may access the database that contains user data.

In the proposed system, image files, HTML (Hypertext Markup Language) files, JavaScript files, Cascading Style Sheets (CSS) files, Python program files, and Chameleon template files are stored in the file server.

When the file server receives a request, the file server searches namespaces for a required file. The search procedure is similar to method resolution procedure in object-oriented programming. The search starts with the namespace for the most specific user context, and continues to ancestor namespaces until the file is found. Then the found file is sent to the Web application server.

The dynamic execution of application codes enables developers to modify and customize ubiquitous Web applications in runtime. Multiple Web application servers can be used to serve large number of users. Multiple Web application servers also improve the availability of the system.

We have used the Pyramid Web framework [8] to implement the Web application server and the file server. However, the proposed methods are not tied to the Pyramid Web framework. Other server-side Web frameworks can be used to implement the system.

In general, the file-based hierarchical customization method is suitable to implement macro adaptation. Micro adaptation can be implemented using client-side/server-side dynamic pages and JSON (JavaScript Object Notation) services.

## 5 Conclusion

We presented a hierarchical customization method for ubiquitous Web applications. Namespaces, inheritance, and polymorphism are used to implement customizations for contexts. Contexts and namespaces for those contexts are hierarchically modeled. Contents of a namespace are inherited to its descendant namespaces, and customizations are applied by overriding contents of ancestor namespaces or by adding new contents. Multilevel inheritance allows customization for complex context hierarchies. Both macro adaptation and micro adaptation are supported by the method. Code reuse is promoted by inheritance in the proposed customization method. The hierarchical nature of the customization method supports staged delivery of a ubiquitous Web application. In our implementation, user interface codes and business logic codes are dynamically executed so that runtime modification and customization of ubiquitous Web applications are possible.

**Acknowledgment.** This research was supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the R&D program supervised by the KCA (Korea Communications Agency) (KCA-2013-12-912-03-001)

## References

1. Pendyala, V.S., Shim, S.S.Y.: The Web as the Ubiquitous Computer. *Computer* 42(9) (2009)
2. Kappel, G., Pröll, B., Retschitzegger, W., Hofer, T.: Modeling ubiquitous web applications - a comparison of approaches. In: *The International Conference on Information Integration and Web-based Applications and Services*, Austria (2001)
3. Kappel, G., Pröll, B., Retschitzegger, W., Schwinger, W.: Modelling ubiquitous web applications - the wuml approach. In: Arisawa, H., Kambayashi, Y., Kumar, V., Mayr, H.C., Hunt, I. (eds.) *ER Workshops 2001*. LNCS, vol. 2465, pp. 183–197. Springer, Heidelberg (2002)
4. Prekop, P., Burnett, M.: Activities, Context and Ubiquitous Computing. *Computer Communications* 26(11), 1168–1176 (2003)
5. Python Programming Language, <http://python.org/>
6. Chameleon templates, <http://chameleon.repoze.org/>
7. Stroustrup, B.: *The C++ Programming Language*. Addison-Wesley, Reading (2000)
8. Pyramid Web framework, <http://www.pylonsproject.org/>
9. McConnell, S.: *Software Project Survival Guide*. Microsoft Press (1998)

# PV System for Medical Devices in the Hospital

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2,\*</sup>, and Ok-Hwan Kim<sup>1</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Buda-dong, Cheonan-si, Chungnam, 330-717, Korea  
{yckim59, owkim}@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College,  
316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

**Abstract.** In this thesis, a boost chopper using pv system and PWM(Pulse Width Modulation) voltage type power converter were constructed to provide a pleasant environment to the patients in the hospital wards by controlling temperature, humidity and air-conditioning and heating. For the stable modulation of solar cell, synchronizing signal and control signal were processed using one chip microprocessor. Power converter system was constructed with booster chopper and voltage source inverter and test was carried out for both devices. Constant voltage control method was used to track a maximum power point at boost converter control. For the inverter control, synchronizing signal and control signal were processed by microprocessor according to the switching theory of SPWM(Sinusoidal pulse Width Modulation), directions to each sector and stable modulation. Test was carried out for inverter control using SPWM control method. In addition, grid voltage was detected and this grid voltage and inverter output were operated at the same phase for the phase locking with PWM voltage source inverter so that surplus power could be linked to grid. This characteristics were applied on the temperature and humidity sensors in the general buildings and buildings of specific purpose such as hospitals. The good dynamic characteristic of inverter could be obtained by these applications.

**Keywords:** PV, SPWM, inverter, solar cell.

## 1 Introduction

Output of solar cell with the photovoltaics system is a dc voltage. Therefore, if solar cell has to be linked to grid, it is needed to convert this output into ac. and the sinusoidal current having unity power factor and voltage have to be supplied into solar cell. Also, PWM modulator should perform a stable modulation even if disturbance such as distortion or noise in grid source voltage waveform which is synchronizing signal is included. Besides, when synchronizing signal and control signal are processed by microprocessor, time difference is existed between sampling timing and carrier wave, thus compensation method is required for this time difference.[1][2]

---

\* Corresponding author.

Photovoltaics system is categorized into two types according to linking method with utility line. A parallel connection system refers to a system wherein photon is always electrically connected. Whereas, a grid change-over system refers to a system which enables reverse power transmission of surplus power which is generated by photovoltaics. It is always electrically separated and is connected only when generated output is in shortage. In this system, reverse power transmission is not possible and it supplies power only on the load.

In this thesis, we intended to control boost chopper so that maximum output point can be always tracked regardless of insolation and temperature changes by changing time ratio based on the power comparison after constructing a grid connected photovoltaics system as voltage type inverter. Also, inverter was controlled as phase driving the grid voltage and inverter output by detecting grid voltage in order to synchronize phase so that power of high power factor and low-frequency harmonic were supplied to the load and system.

## 2 Grid Connect PV System

### 2.1 Equivalent Circuit Solar Cell

Fig. 1 shows an equivalent circuit of solar cell by using a photovoltaic effect. In an ideal case, the relationship between voltage and current during light projection becomes as in Equation (1).

$$I = I_{ph} - I_0 \left[ \exp\left(\frac{qV}{nKT}\right) - 1 \right] \tag{1}$$

Practically, series resistance  $R_s$  and parallel resistance  $R_{sh}$  are connected as shown in Fig. 1. While current becomes as in Equation (2).

$$I = I_{ph} - I_0 \left[ \exp\left(\frac{q(v + IR_s)}{nKT}\right) - 1 \right] - \frac{V + IR}{2} \tag{2}$$

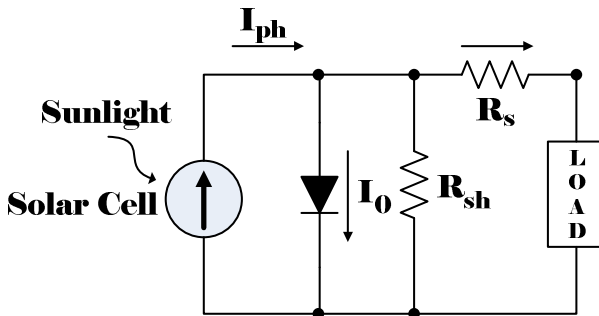


Fig. 1. Equivalent Circuit of Solar Cell

Where,  $I$  is the output current,  $I_{ph}$  is the photoelectric current,  $I_0$  is the diode saturation current,  $n$  is the diode constant,  $k$  is the Boltzmann constant, and  $q$  is the charge of one electron.

### 2.2 Grid Connected Boost Chopper Circuit

Grid connected inverter uses an utility line to maintain power balance between dc source and ac load. When inverter output is in short as compared with power required by load, the shortage is automatically supplied from utility system. While, if inverter output is larger than power required by load, the surplus power is supplied to grid so that it can be supplied to other loads on the grid. Therefore, without using storage battery which is costly and inefficient to store dc power from alternative energy source like solar cell or fuel cell, power can be always supplied even during night or rainy days. Besides, since insulating transformer of 60[Hz] is installed between inverter output and grid, insulation is provided so that dc component leakage towards grid can be prevented during internal accident at dc side. Further, the transformer itself plays a role of series impedance to lower harmonic current which is input into grid at less than allowable level, thus it makes interface between inverter and grid voltage easy.

A current controller is implemented to control input voltage of inverter to match reference current at sampling point  $n + 1$  and actual current based on the voltage equation of inverter.

### 2.3 Equivalent Circuit of System and Its Analysis

When grid is connected, the equivalent circuit of inverter system is same as in Fig. 2 and it satisfies Equation (3).

$$E_i = E_L + E_u \tag{3}$$

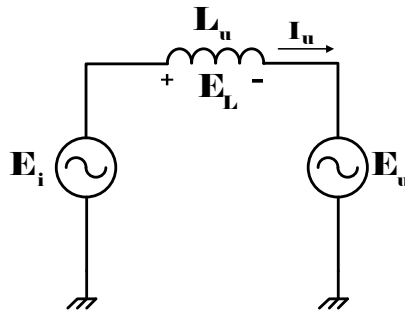


Fig. 2. Equivalent circuit of inverter system

### 2.4 Control of Grid Connected Power Converter

PI controller makes current of desired size flowing by controlling voltage at both terminals of reactor as in Equation (4).

$$V_L = V - E = XI^* + \left(\frac{ki}{s} + kp\right)(I^* - I) \tag{4}$$

The voltage at ac side of power converter can be freely changed by dc voltage and PWM modulation factor, thus it can be expressed as in Equation (5).

$$E = (V - XI^*) - \left(\frac{ki}{s} + kp\right)(I^* - I) \tag{5}$$

The first item in right side in Equation (5) is the voltage after subtracting reactance drop from source voltage, thus it means voltage vector of power converter under steady state. The second item of right side controls the reactance voltage by PI controller so that set current is maintained.

### 2.5 PV System

Fig. 10 show a construction of controller in the stand-alone type source system using solar cell. To control tracking the maximum output point, after detecting voltage and current of solar cell, solar cell array was made rotated to follow location of sun by controlling stepping motor to generate maximum power all the time.

Also, boost chopper was controlled to boost voltage of solar cell. Controller was constructed to maintain a constant voltage regardless of input voltage changes and load changes so that output voltage of voltage source inverter is under constant voltage controlled state.

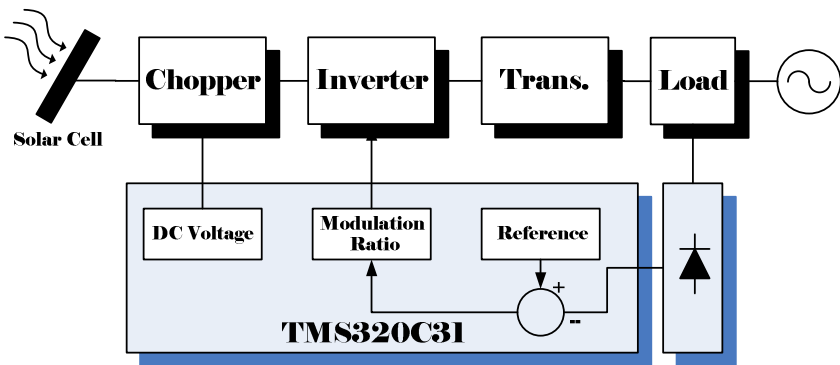


Fig. 3. Block diagram of solar energy power generating system

### 3 Experiment Results

Fig. 4 shows a sinusoidal wave which is carrier of SPWM and triangular wave which is a reference. In this system, modulation index was controlled near to one and frequencies were adjusted to 60[Hz] and 600[Hz].

Fig. 5 shows modulation waveform which is a gate input signal for the carrier sinusoidal wave. As in sinusoidal pulse width modulation theory, it was clear that modulation wave forms are correctly matching with sinusoidal period.

Fig. 6 shows output voltage and current measurement results when only L load (temperature and humidity sensors in the general buildings and hospitals, 15[W] and 53[W] electric fans, lamp load of 180[W], air-conditioning & heating devices (applied in 1[KW] radiator)) is used. From the Figure, it was clear that a relatively low frequency was included and good waveform was displayed.

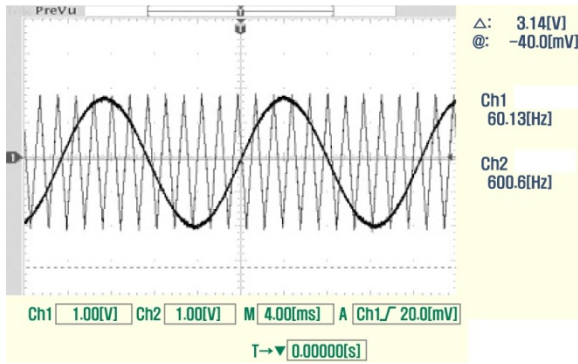


Fig. 4. Carrier and reference waveform of SPWM

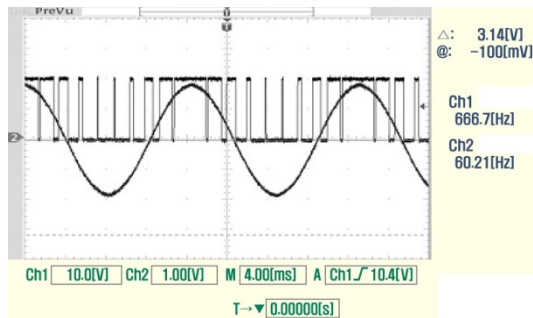


Fig. 5. Carrier and modulation waveform of SPWM

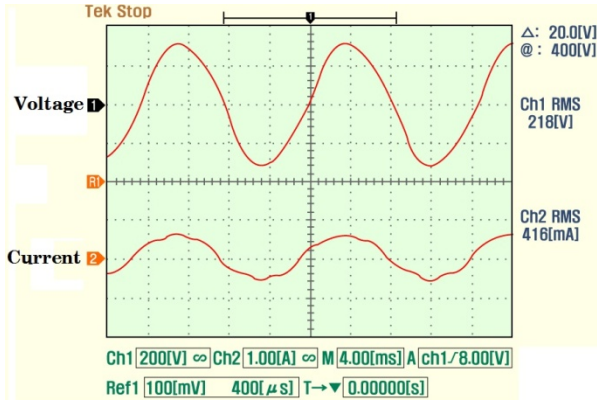


Fig. 6. Inverter output voltage and current waveform(with L load)

## 4 Conclusion

Experiment was performed using a microprocessor in the proposed system and conclusion was drawn as follows:

(1) With operation of PWM voltage inverter linking to boost chopper, voltage generated by conversion of dc-ac was applied on the temperature and humidity sensors and air-conditioning and heating load such as hospital. The applied devices showed a stable operation(applied on the temperature and humidity sensors and 15[W] and 53[W] electric fans, lamp load of 180[W], and air-conditioning & heating device(radiator) of 1[KW]).

(2) Voltage and current of solar cell were measured and the calculated optimum movement voltage was set as a reference of dc voltage. After that, the maximum voltage by boost was made to be operated at near maximum output point for solar cell.

## References

1. Kim, Y.-C., Cho, M.-T., Song, H.-B., Kim, O.-H.: Regeneration Break Control in the High-Speed Area using the Expanding of the Constant Torque Region and Power Region. *International Journal of Control and Automation* 6(4), 347–356 (2013)
2. Kim, Y.-C., Song, H.-B., Cho, M.-T., Lee, C.-S., Kim, O.-H., Park, S.-Y.: A Study on the Improved Stability of Inverter through History Management of Semiconductor Elements for Power Supply. In: Kim, T.-h., Ramos, C., Kim, H.-k., Kiumi, A., Mohammed, S., Ślęzak, D. (eds.) *ASEA/DRBC 2012. CCIS*, vol. 340, pp. 155–162. Springer, Heidelberg (2012)
3. Kim, Y.-C., Song, H.-B., Cho, M.-T., Moon, S.-H.: A Study on Direct Vector Control System for Induction Motor Speed Control. In: Park, J.J.(J.H.), Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) *EMC Technology and Service. LNEE*, vol. 181, pp. 599–612. Springer, Heidelberg (2012)



# Color Image Retrieval Using Fuzzy Measure Hamming and S-Tree

Thanh The Van\* and Thanh Manh Le

HCMC University of Food Industry, Hue University  
140 Le Trong Tan, Tay Thanh ward,  
Tan Phu district, HCM City, Vietnam  
thanhv@cntp.edu.vn, lmthanh@hueuni.edu.vn

**Abstract.** This chapter approaches the image retrieval system on the base of the colors of image. It creates fuzzy signature to describe the color of image on color space HSV and builds fuzzy Hamming distance (FHD) to evaluate the similarity between the images. In order to reduce the storage space and speed up the search of similar images, it aims to create S-tree to store fuzzy signature relies on FHD and builds image retrieval algorithm on S-tree. Then, it provides the content-based image retrieval (CBIR) and an image retrieval method on FHD and S-tree. Last but not least, based on this theory, it also presents an application and experimental assessment of the process of querying similar image on the database system over 10,000 images.

**Keywords:** CBIR, Image Retrieval, FHD, Signature, S-Tree.

## 1 Introduction

It is difficult to find similar images in a large database of digital images. There is a solution to solve this problem, such as: text-based image retrieval (TBIR) based on the keywords [4, 5] but it is time-consuming and unfeasible for many different applications. Moreover, the process of labeling depends on the semantic description of the image. So the image retrieval system on the base of the content is developed to extract visual feature to describe the content of image. A number of image retrieval system was built as: QBIC, ADL, Virage, Alta Vista, SIMPLYcity,...

In recent years, the works of query image regarding CBIR, such as: the image retrieval system on the base of the color histogram [4, 5], the similarity measure of image on the base of combining color and texture image [7, 8], image retrieval technique Variable-Bin Allocation (VBA) using signature and the S-tree [6],...

In the approach of the paper will create the fuzzy signature of an image. The content of the paper will aim to efficient query "similar images" in a large database system of digital image. There are two major targets are used to reduce the amount of storage space and speed up the query image on large database systems. It also builds an evaluation method of similarity measure of image on fuzzy measure Hamming and builds an image retrieval method on the S-tree.

---

\* Corresponding author.

## 2 The Related Theory

### 2.1 Fuzzy Signature

The fuzzy signature  $F$  with a length  $m$  is a vector  $(f_1, f_2, \dots, f_m)$ , with  $f_i \in [0, 1]$ ,  $i = 1, \dots, m$  ([1], [2]). The conjunction of fuzzy signatures  $F^i$  and  $F^j$  is a fuzzy signature:  $F^i \wedge F^j = (f_1^i \wedge f_1^j, \dots, f_m^i \wedge f_m^j)$ , with  $f_r^i \wedge f_r^j = \min\{f_r^i, f_r^j\}$ ,  $r = 1, \dots, m$ . The disjunction of fuzzy signatures  $F^i$  and  $F^j$  is a fuzzy signature:  $F^i \vee F^j = (f_1^i \vee f_1^j, \dots, f_m^i \vee f_m^j)$ , with  $f_r^i \vee f_r^j = \max\{f_r^i, f_r^j\}$ ,  $r = 1, \dots, m$

### 2.2 S-Tree

S-tree [1, 3] is a tree with many branches that are balanced; each node of the S-tree contains a number of pairs  $\langle sig, next \rangle$ , where  $sig$  is a binary signature and  $next$  is a pointer to a child node. Each node root of the S-tree contains at least two pairs and at most  $M$  pairs  $\langle sig, next \rangle$ , all internal nodes in the S-tree at least  $m$  and at most  $M$  pairs  $\langle sig, next \rangle$ ,  $1 \leq m \leq M/2$ ; the leaves of the S-tree contain the image's binary signatures  $sig$ , along with a unique identifier  $oid$  for those images. The S-tree height for  $n$  signatures is at most  $h = \lceil \log_m n - 1 \rceil$ . The S-tree was built on the basis of *inserting* and *splitting*. When the node  $v$  is full, it will be split into two.

### 2.3 FHD Distance

For two vector  $n$ -dimensional real value  $x$  and  $y$ , difference fuzzy set as  $D_\alpha(x, y)$ , with membership function as  $\mu_{D_\alpha(x, y)} = 1 - e^{-\alpha(x-y)^2}$ . The *fuzzy Hamming distance* [1, 2] between  $x$  and  $y$  with sign  $FHD_\alpha(x, y)$  is the fuzzy cardinality of the difference fuzzy set  $D_\alpha(x, y)$  and has membership function matching parameter  $\alpha$  as  $\mu_{FHD(x, y)}(\alpha) : \{0, 1, \dots, n\} \rightarrow [0, 1]$ . Moreover  $\mu_{FHD(x, y)}(k; \alpha) = \mu_{Card(D_\alpha(x, y))}(k)$ , with  $k \in \{0, 1, \dots, n\}$ ,  $n = |Support(D_\alpha(x, y))|$ .

## 3 Building Data Structures and Image Retrieval Algorithms

### 3.1 Creating a Fuzzy Signature of the Color Image

*Step 1:* Choose a color set  $C = \{c_1, \dots, c_n\}$  to calculate the color histogram of the images. To quantify the image  $I$  in order to retain the dominant colors  $C_i = \{c_1^i, \dots, c_{n_i}^i\}$ , the color histogram vector of image  $I$  is  $H_i = \{h_1^i, \dots, h_{n_i}^i\}$ . *Step 2:* Calculate the color histogram vector standardizes  $H = \{h_1, \dots, h_n\}$ , where  $h_i = h_j^i / \sum_j h_j^i$  if  $c_i \in C \cap C_i$ , otherwise  $h_i = 0$ . *Step 3:* Each color  $c_j^i$  will be described into a fuzzy signature as  $f_1^j, \dots, f_m^j$ . Therefore, fuzzy signature of the color image  $I$  will be  $f_1^1, \dots, f_m^1, f_1^2, \dots, f_m^2, \dots, f_1^n, \dots, f_m^n$ , in which:  $f_j^i = h_j$  if  $j = \lceil h \times m \rceil$ , otherwise  $f_j^i = 0$ . Setting  $F^j = f_1^j \dots f_m^j$ , the fuzzy signature of the color image will be  $FSig = F^1 \dots F^n$

### 3.2 The Similar Measure FHD

Each fuzzy signature  $FSig = F^1 \dots F^n$  as vector  $V_i = (v_1, \dots, v_n)$ , in which  $v_i = weight(F^i) = \sum_{k=1}^m w_k^i$ , with  $F^i = f_1^i \dots f_m^i$  and  $w_k^i = f_k^i + (k/m) \times 100$  if  $f_k^i \neq 0$ , otherwise  $w_k^i = 0$ . Setting  $J$  as an image to calculate the similarity in comparison with the image  $I$ , the fuzzy Hamming distance FHD between two vector  $V_i = (v_1^i, \dots, v_n^i)$  and  $V_j = (v_1^j, \dots, v_n^j)$  is as follows:

$FHD_\alpha(V_i, V_j) = Card(D_\alpha(V_i, V_j)) = \sum_{i=0}^n i / (\mu_{Card(D_\alpha(V_i, V_j))}(i))$ . In there,  $\mu_{Card(D_\alpha(V_i, V_j))}(i) = \mu(i) \wedge (1 - \mu(i+1)) = \min\{\mu(i), (1 - \mu(i+1))\}$ ,  $\mu(i)$  as an  $i$ -th largest value of function  $\mu_i$  corresponds to fuzzy set  $D_\alpha(V_i, V_j) = \sum_{i=1}^n i / \mu_i$  and  $\mu(0) = 1, \mu(n+1) = 0$ . At that time, the different levels of  $V_i$  and  $V_j$  on  $k$  the component is as follows:  $\mu_{FHD_\alpha(V_i, V_j)}(k, \alpha) = \mu_{Card(D_\alpha(V_i, V_j))}(k)$ , with  $k \in \{0, 1, \dots, n\}, n = |Support(D_\alpha(x, y))|$

### 3.3 Creating S-Tree Based on FHD Distance

The process of creating the S-tree is based on *inserting* and *splitting* the node in the tree [3, 6]. The algorithm to create the S-tree to store the fuzzy signature is as follows:

**Input:** FS = {<FSig<sub>i</sub>, Oid<sub>i</sub>> | i = 1, ..., n}

**Output:** The S-tree

**Algorithm1.** Gen-FStree(FS, Root)

Step 1. v = Root;

**If** FS =  $\emptyset$  **then** STOP;

**Else** Choosing <FSig, Oid>  $\in$  FS; FS = FS \ <FSig, Oid>;

    To go step 2;

Step 2. **If** v **is** leaf **then**

**Begin**

        v = v  $\cup$  <FSig, Oid>; UnionSignature(v);

**If** v.count > M **then** SplitNode(v); To go back Step 1;

**End**

**Else**

**Begin**

    FHD(SIG<sub>0</sub>  $\rightarrow$  FSig, FSig) = min{FDH(SIG<sub>i</sub>  $\rightarrow$  FSig, FSig) | SIG<sub>i</sub>  $\in$  v};

    v = SIG<sub>0</sub>  $\rightarrow$  next; To go back Step 2;

**End**

The splitting node relies on  $\alpha$ -seed,  $\beta$ -seed is done as follows:

**Input:** Node v

**Output:** The S-tree after splitting node

**Algorithm2.** SplitNode(v)

Create the node  $v_\alpha$  and  $v_\beta$  contains  $\alpha$ -seed and  $\beta$ -seed

v = v \ { $\alpha$ -seed,  $\beta$ -seed}

**For** (SIG<sub>i</sub>  $\in$  v)

**If** FHD(SIG<sub>i</sub>  $\rightarrow$  FSig,  $\alpha$ -seed) < FHD(SIG<sub>i</sub>  $\rightarrow$  FSig,  $\beta$ -seed) **then**  $v_\alpha = v_\alpha \cup$  SIG<sub>i</sub>;

**Else**  $v_\beta = v_\beta \cup$  SIG<sub>i</sub>;

```

 $s_\alpha = \sqrt{SIG_i^\alpha}$ , with  $SIG_i^\alpha \in v_\alpha$ ;  $s_\beta = \sqrt{SIG_i^\beta}$ , with  $SIG_i^\beta \in v_\beta$ ;
If ( $v_{parent} \neq \text{null}$ ) then  $v_{parent} = v_{parent} \cup s_\alpha$ ;  $v_{parent} = v_{parent} \cup s_\beta$ ;
If ( $v_{parent}.count > M$ ) then SplitNode( $v_{parent}$ );
If ( $v_{parent} = \text{null}$ ) then Root = {  $s_\alpha$ ,  $s_\beta$  };
End.
Procedure UnionSignature( $v$ )
Begin
 $s = \sqrt{SIG_i}$ , with  $SIG_i \in v$ ;
If ( $v_{parent} \neq \text{null}$ ) then
Begin
 $SIG_v = \{SIG_i \mid SIG_i \rightarrow next = v, SIG_i \in v_{parent}\}$ ;
 $v_{parent} \rightarrow (SIG_v \rightarrow FSig) = s$ ; UnionSignature( $v_{parent}$ );
End
End.

```

### 3.4 Image Retrieval Algorithm Based on the S-Tree

The process of querying provides the signature of the image on the base of browsing the S-tree with the similarity measure FHD is as follows:

```

Input: Query signature  $QSig$  and the S-tree
Output:  $SIGOUT = \{ \langle SIG_i, Oid_i \rangle \}$  reference to images
Algorithm3. Search-Image-Sig( $QSig$ , S-tree)
 $v = \text{root}$ ;  $SIGOUT = \emptyset$ ;  $Stack = \emptyset$ ; Push( $Stack$ ,  $v$ );
while(not Empty( $Stack$ )) do
Begin
 $v = \text{Pop}(Stack)$ ;
If( $v$  is not Leaf) then
begin
For( $SIG_i \in v$  and  $SIG_i \rightarrow FSig \wedge QSig = QSig$ ) do
FHD( $SIG_0 \rightarrow FSig$ ,  $QSig$ ) =  $\min\{FHD(SIG_i \rightarrow FSig, QSig) \mid SIG_i \in v\}$ ;
Push( $Stack$ ,  $SIG_0 \rightarrow next$ );
end
Else  $SIGOUT = SIGOUT \cup \{ \langle SIG_i \rightarrow FSig, Oid_i \rangle \mid SIG_i \in v \}$ ;
end
return  $SIGOUT$ ;

```



Fig. 1. Some results of the process query image in COREL database over 10,000 images

## 4 Experiment

### 4.1 Model Application

#### Phase 1: Perform Preprocessing

Step 1: Quantize images in the database and convert to a color histogram. Step 2: Convert the color histogram of the image in the form of fuzzy signatures. Step 3: Respectively calculate the similarity measure FHD distance of the fuzzy signatures and insert into the S-tree.

#### Phase 2: Implementation Query

Step 1: For each query image, calculate the color histogram and convert into fuzzy signatures. Step 2: Perform fuzzy signature query on the S-tree consisting of the similar image signature at the leaves of the S-tree through the FHD measure. Step 3: After finding similar images, conduct arrangement of similar levels from high to low and make the title match with the images arranged on the basis of FHD distance.

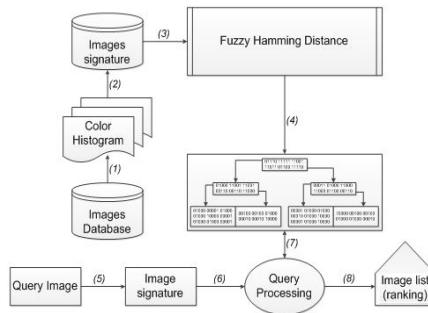


Fig. 2. Model image retrieval system using FHD and S-Tree

### 4.2 The Experimental Results

Each image will calculate the color histogram based on 16 colors: BLACK, SILVER, WHITE, GRAY, RED, ORANGE, YELLOW, LIME, GREEN, TURQUOISE, CYAN, OCEAN, BLUE, VIOLET, MAGENTA, RASPBERRY.

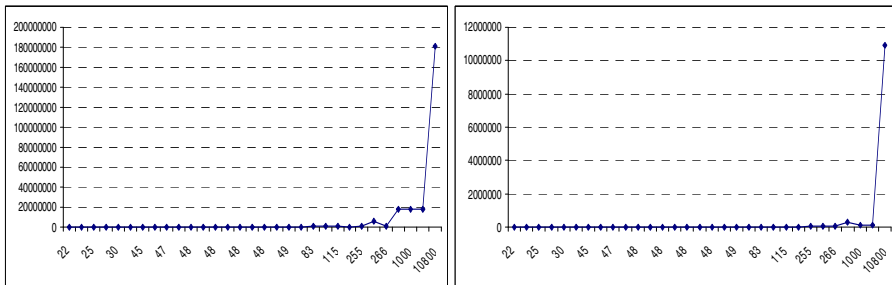
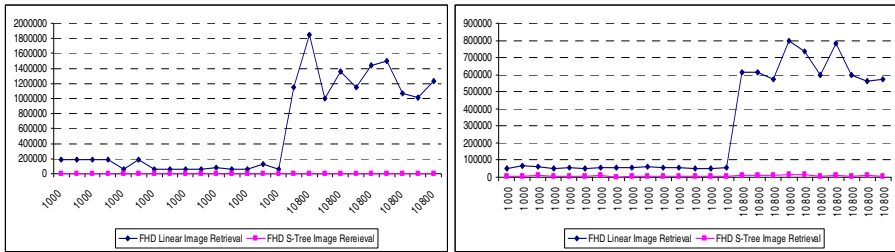


Fig. 3. Number of comparisons to create S-tree; the time to create S-tree (in milliseconds)



**Fig. 4.** Number of comparisons to query; the time to query (in milliseconds)

According to the experiment shows the process of creating the S-tree from the fuzzy signature of the image takes time-consuming, but the query image relies on the S-tree do much faster than the linear search method on the base of FHD distance.

## 5 Conclusion

In the experiment, the paper created the fuzzy signature to describe the color's image and showed the similar image retrieval algorithm, at the same time, experimented to query image on the base of the content. However, using the distribution of the image's color will result in inaccuracy in the case of images with the same percentage of color pixels, but the color distribution location does not correspond to each other. The next development will assess the similarity of the image with location distribution of the percentage of color pixels and compare objects in the contents of image to increase accuracy when querying the similar images.

## References

1. Snasel, V., et al.: Fuzzy Signatures Organized Using S-Tree. In: Proceedings of the IEEE International Conference on Digital Object Identifier, Alaska, October 9-12, pp. 633–637 (2011)
2. Ionescu, M.M., et al.: Image Clustering for a Fuzzy Hamming Distance Based CBIR System. In: Proc. of the 16th Midwest AI and Cog. Sc. Conf., Dayton, pp. 102–108 (April 2005)
3. Chen, Y., Chen, Y.: On the Signature Tree Construction and Analysis. *IEEE Trans. Knowl. Data Eng.* 18(9), 1207–1224 (2006)
4. Neetu Sharma, S., Paresh Rawat, S., Jaikaran Singh, S.: Efficient CBIR Using Color Histogram Processing. *Signal & Image Processing: An Inter. Jour.* 2(1), 94–112 (2011)
5. Malik, F., et al.: Feature Analysis of Quantized Histogram Color Features for Content-Based Image Retrieval Based on Laplacian Filter. In: *Int. Conf. on Sys. Eng. and Modeling*, vol. 34, pp. 44–49 (2012)
6. Nascimento, M.A., Tousidou, E., Chitkara, V., Manolopoulos, Y.: Image indexing and retrieval using signature trees. *Data & Knowledge Eng.* 43(1), 57–77 (2002)
7. Mehta, R., Mishra, N., Sharma, S.: Color - Texture based Image Retrieval System. *International Journal of Computer Applications* 24(5), 24–29 (2011)
8. Varshney, G., Soni, U.: Color-Based Image Retrieval in Image Database System. *International Journal of Soft Computing and Engineering* 1(5), 31–35 (2011)

# Pipeline Verification via Closed-Loop Feedback

Danghui Wang, Yongda Li, and Xiaoping Huang

School of Computer Science, Northwestern Polytechnical University, Xi'an China  
{wangdh, huangxp}@nwpu.edu.cn, liyongda@mail.nwpu.edu.cn

**Abstract.** With the growth of the depth and complexity of pipeline, it needs more and more effort to verify a pipeline. Closed-loop Feedback (CLF), a method which can verify complex pipeline, is proposed in this paper. In CLF, activities of pipeline are fed back to verification pattern generator, where it can be used to generate the following verification patterns dynamically and automatically. Compared with non-feedback by simulation, CLF can achieve the goal of verification much more quickly.

**Keywords:** Pipeline, Closed-Loop Feedback, Verification.

## 1 Introduction

With the development of technology and the increasing requirement of human life, it needs more and more diversities of embedded system. There are many application specific processors in these embedded systems, these processors are powerful and with complex pipelines. With the increasing of complexity in pipeline design, a lot of troubles have been met in the verification of pipeline.

There are two widely used methods for pipeline verification, one of which is formal verification, and the other is simulation-based verification [1]. Formal verification can achieve 100% coverage, but it is not suitable for complex system as it uses exhaustive method, whose complexity grows in exponential order as the circuit scale. Simulation-based verification can detect some unpredictable errors like an error in handling window overflow exception. But in common used simulation-based approach, the initial test pattern can creating many effective instructions, in later even with a new seed, it is less likely to generate instructions that reaches uncovered areas of the pipeline. So the major drawback of common used simulation-based verification is the difficulty to produce effective stimuli for pipeline [4].

According to these problems, we proposed closed-loop feedback (CLF), a method of simulation-based verification which employs a random instruction generator. For guidance in the instruction generation, CLF monitors activities of pipeline and feeds the activities back to instruction generator, this is a closed-loop feedback which is used to guide the generation of instructions. CLF generate instruction in heterogeneous distribution, it adjusts the generation of instructions dynamically during verification to produce effective patterns. Compared with verification without feedback, CLF can find out errors more efficiently because it can persistently generate

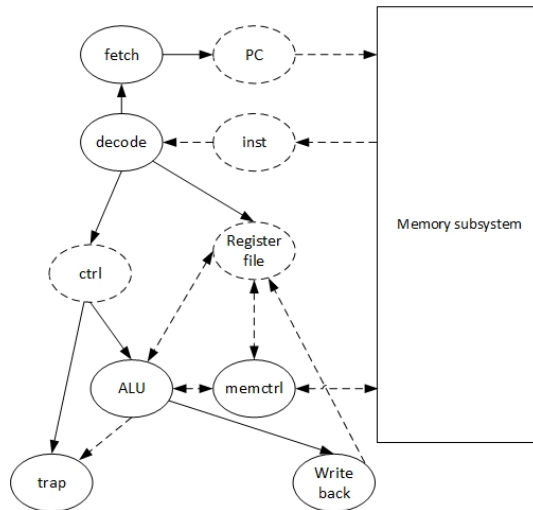
instructions which effectively cover missing function points of pipeline. Compared with dedicated feedback tools, CLF is much simpler and suitable for common simulation tools. As CLF is implemented in System Verilog, it can be adapted to other digital designs easily.

The verification system introduced in this paper is designed for the Space oriented Embedded Pipeline (SEP), which is applied to aerospace exploration. The pipeline should be verified completely and stable in aerospace. In order to verify the pipeline completely and effectively, we developed the verification frame which utilizes CLF, and the result provides that our method is very efficient.

## 2 Modeling of Space Embedded Pipeline

SEP is a typical RISC pipeline [5], a configurable and synthesizable 32-bit pipeline which is compatible with the IEEE-1754 (SPARC V8) [5]. The most importance feature of SEP is its scalability. Because there are many combinations of configuration, we must design a combinatorial method which can significant reduce the number of tests, this method is showed in section three.

A structured pipeline can be modeled by a block diagram, which is a DAG (Directed Acyclic Graph) graph  $G S = (VS, E S)$  [7]. In the graph, nodes are used for functional blocks and edges are used for signal transfer. In this kind of graph, combinational logic and sequential logic can be distinguished easily. For illustration, fig. 1 shows a simple graph-based model for this pipeline, in which, the oval boxes denote combinational logic, the dotted ovals are storage unit, the bold edges are pipeline activities, and the dotted edges are data-transfer.



**Fig. 1.** Block diagram of pipeline



### 3 Structure of Closed-Loop Feedback

The structure of CLF consists of four units, which are monitor, generator, driver and comparator, as shown in fig. 2. Generator and monitor are the most important parts of CLF. The generator supplies stimulus to DUT and reference model in parallel, the monitor collects activities of DUT and feed the activities back to the generator, this is a closed-loop, and this loop will guide the generation of instructions automatically. The monitor sets watch points in some important signals or units of DUT, and records the switching activities of these watch points, the records collected by the monitor are used as feedback to the generator. Generator adjusts every probability of pipeline according to the feedback and the probability is used to control the generation of effective instructions. The driver receives instructions from the generator and feeds them to DUT and reference model in parallel. The comparator compares the results of DUT with the reference model. Environment corresponds to the configuration of verification. Because the pipeline has up to ten configuration options, it is infeasible to do exhaustive verification, from the research, Most errors of configuration can be discovered in the interaction between the combination of two configurations (65-97%)[7]. To balance cost and risk, CLF select a subset of combinations at the level of two-way interaction, this interaction can be modeled by Cartesian product, and CLF chooses 18 combinations that cover all the pairs of configurations.

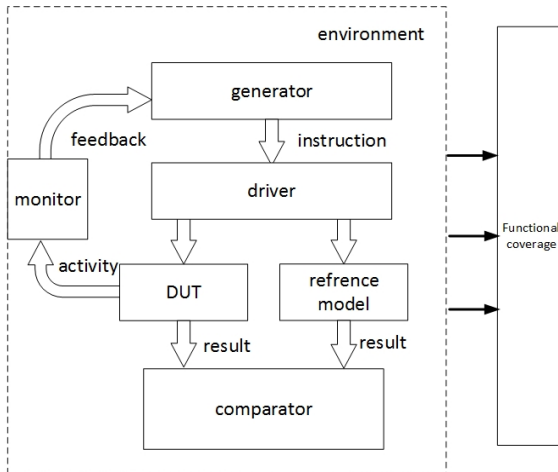


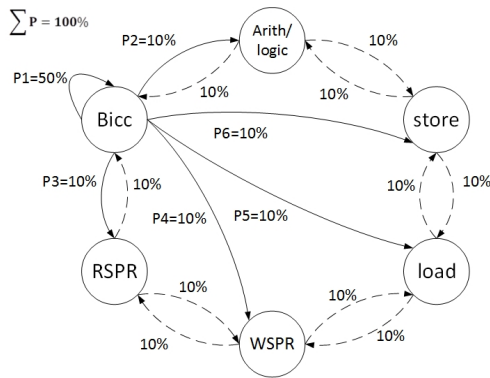
Fig. 2. Structure of verification

### 4 Design of CLF

There are two most important functional blocks in verification system, generator and monitor. Both of which are critical to the performance of CLF.

## 4.1 Generator

The generator adopts the method which implements heterogeneous possibility. The heterogeneous possibility can be adjusted dynamically to increase the possibility of undiscovered cases and lower the possibility of cases that has been verified. Instructions are generated randomly, so it is unknown what instruction will be executed later, In order to control the possibility of the next instruction, the generator use dist operator of System Verilog to control the weighted distribution of instruction, The dist operator takes a list of values and weights, separated by the “:=” operator, CLF can adjust the possibility of instructions dynamically with dist by control the weight of instructions.



**Fig. 3.** Possibility of instructions

The possibility distribution of instruction is illustrated in fig. 3, this figure only describes possibility of bicc instruction and some possibilities of other instructions are omitted.

A starting vertex is selected randomly. The system produces the corresponding inputs and randomly selects a transition to the next vertex. The probability adjustment algorithm considers a weighted sum of the switching activity reported by the monitor and converts it to a value ( $w_{\text{monitor}}$ ). The value relates to an edge (indicated by  $\langle \text{instruction}, \text{transition} \rangle$  pair). The feedback value is subtracted from the previous weight ( $w_{\text{old}}$ ), then the weight is normalized. After this step the sum of possibility will not be equal to 1.0, so the probabilities should be normalized.

## 4.2 Monitor

Monitor is located in the feedback path. Ilya and Valeria [4] have implemented a monitor based on feedback depth, but this monitor is a software tool, which cannot be used in System Verilog verification. The structure of the proposed monitor is much more universal. The monitor tracks the activities of watch points and composes them into simple data structures. The feedback of CLF is only related to the instruction which triggers the activity.

CLF will set many watch points in pipeline, so it is likely to detect activities which are tracked from more than one watch point. From the observation of different levels (2 or 3 watch points at the same time, etc.), if CLF deals with multiple watch point at the same time, the complexity of monitor would increase exponentially. If CLF deals with multiple watch point concurrently, the performance of the Monitor will be reduced to 70%, because it needs to add the logic to distinguish the activities of them. According to the analysis, the monitor only deals with single watch point at one time.

Instructions are generated many clock cycles before the corresponding activity be observed. CLF use an instruction queue to cache the instruction execute in pipeline. When sampling an observed activity at a watch point, the opcode of the instruction of the relevant pipeline stage also be tracked. The opcode is used to match an <instruction, transition> pair from the queue.

### 5 Experimental Results

To evaluate the performance of CLF, we experiment three verification methods by simulation SEP on VCS (a simulation tool of Synopsys), which are feedback verification, no feedback verification and simple random verification. No feedback verification is the verification which don't totally random instructions, it use the System Verilog language to describe the format of the stimulus (like "address is 32-bits; opcode is ADD, SUB or STORE; length < 32 bytes"), and generator picks up instructions that meet the constraints. Simple verification is the verification without constraint, it has only a fully definition of ISA (instruction set architecture). The three verification systems are developed in System Verilog language. Fig. 4 shows the statistics of first one hundred hours of the three random verification approaches applied to SEP. For each of them, the figure illustrates the time versus the number of error., in first forty hours, No feedback and Simple random verification would be hanged by fatal errors, it needs manual intervention to fix these errors, but CLF can walk around these errors by monitor which can insert correct value to watch point, this phenomenon makes the statistic line of CLF more curve than the other two verifications. As shown in Fig. 4, CLF is far more efficient than no feedback.

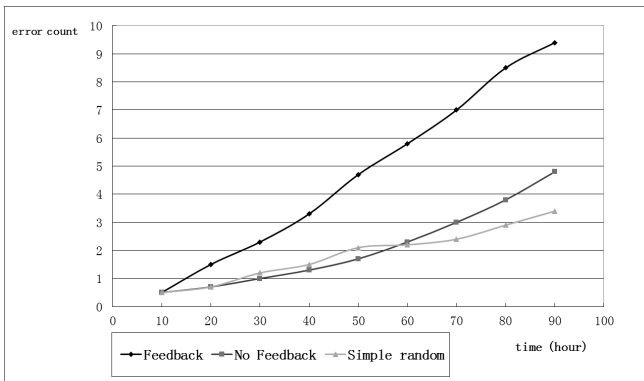


Fig. 4. Verification result

## 6 Conclusion

Increasing complexity of pipeline in embedded computing makes verification a primary concern for designer. This paper proposes a closed-loop feedback method to decrease the overhead of pipeline verification. It can feed the activities of key signals back to direct the generation of stimulus. This method adjusts the generated instruction dynamically, and has reduced the overhead of verification dramatically. The verification system is a closed-loop feedback and fully automated, so it only requires very little human intervention. Because we use system Verilog to implement the verification system, it can be used in many simulation tools. The experimental results show that our technique can significantly improve the verification.

**Acknowledgement.** This work is supported in part by National Natural Science Foundation of China (NSFC, Grant No. 61272122) and Basic Research Project of Northwestern Polytechnical University (JC201256).

## References

1. Lam, W.: *Hardware design verification: simulation and formal method-based approaches*. Prentice Hall (2005)
2. Qadeer, S., Tasiran, S.: Promising directions in hardware design verification. In: *Proceedings of ISQED* (2002)
3. Adir, A., Almog, E.: *Genesys-Pro: Innovations in Test Program Generation for Functional Processor Verification*. IEEE Design & Test of Computers (2004)
4. Wagner, I., Bertacco, V.: *Microprocessor Verification via Feedback-Adjusted Markov Models*. IEEE TCAD (2007)
5. SPARC International Inc. *The SPARC Architecture Manual Version 8 (R)*. 535 Middlefield Road Suite 210. SPARC International, Menlo Park (1992)
6. Patterson, D.A., Hennessy, J.L.: *Computer Architecture: A Quantitative Approach*, 5th edn. Elsevier Inc. (2012)
7. Mishra, P., Dutt, N.: *Functional Coverage Driven Test Generation for Validation of Pipelined Processors*. IEEE Date (2005)

# Fuzzy Rule-Based to Predict the Minimum Surface Roughness in the Laser Assisted Machining (LAM)

M.R.H. Mohd Adnan<sup>\*</sup>, Azlan Mohd Zain, and Habibollah Haron

Soft Computing Research Group, Faculty of Computing,  
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia  
aban.ridhwan.adnan@gmail.com,  
{azlanmz, habib}@utm.my

**Abstract.** Laser Assisted Machining (LAM) has been used to predict the surface roughness of Aluminium Oxide ( $Al_2O_3$ ) workpiece using solid neodymium-doped yttrium aluminum garnet (Nd:YAG) laser cutting. The rule-based reasoning and fuzzy logic are used to develop a model to predict the surface roughness values. The process parameters considered in this study are depth of cut, rotational speed, feed, and pulsed frequency, each has three linguistic values. The fuzzy rule-based model is developed using MATLAB fuzzy logic toolbox. Nine IF-THEN rules are created for model development. The relationship between experimental results, predicted results of the proposed model and statistical results gave a good agreement with the correlation 0.994. The differences between experimental results and predicted results have been proven with estimation error value 0.072. The findings indicate that the best predicted value is located at combination 0.2mm (depth of cut), 1500rpm (rotational speed), 0.02mm/rev (feed) and 40 (pulsed frequency).

**Keywords:** Laser assisted machining, fuzzy rule-based, surface roughness, linguistic values.

## 1 Introduction

Light Amplification Stimulation Emission of Radiation or the common name is Laser is one of non conventional machining which uses an intense laser beam as the heat source and changing the ceramic deformation behavior by focusing the laser beam on the workpiece materials. There are different types of laser cutting such as solid laser, liquid laser and gaseous laser. Solid states Nd:YAG and gaseous  $CO_2$  are mostly used for laser cutting; Although  $CO_2$  lasers have wide application in industries but it is not common in used due to its poor absorptivity compared to Nd:YAG laser with shorter wavelength and highly absorbed when falling even on a reflective surface [1]. There are different workpiece materials have been investigated by laser cutting such as such as metals sheet [2], ceramics [3-4] and composites [5]. Prediction is very important in application of the laser process in order to produce the desired product. The desired

---

<sup>\*</sup> Corresponding author.

product quality can be obtained and improved to predict the machining performance before start the actual process. In machining process, the machining performance is an indicator to describe the quality of the manufactured product. The quality of the laser cutting is affected determine by the laser cutting parameters [6]. The important factor of laser process in evaluating the quality of products is surface finish; and surface roughness ( $R_a$ ) is indicator to determine the surface finish [7-8].

In modeling the prediction technique of  $R_a$ , it involve three models such as experimental models which may involve trial and error experimental with higher cost operation, analytical model such as Taguchi technique and response surface methodology (RSM), and Artificial Intelligence (AI) models such as Artificial Neural Network (ANN) and Fuzzy Logic (FL) [9-10]. AI has produced a number of powerful tools, many of which are of practical use in engineering to solve difficult problems normally requiring human intelligence. The predictions resulted from the AI approaches are more accurate than the non AI ones and it is clear that each of the state-of-the-art modeling, inference and decision making methods are able to predict the  $R_a$  in a non intrusive manner [11]. AI has been successfully applied in predicting  $R_a$  values for laser cutting and give accurately representing  $R_a$  values with respect to experimental results [12-13]

FL is a convenient way to map an input space to an output space; it also is a precise logic of imprecision and approximate reasoning [14]. FL is very useful in modeling complex and imprecise systems. This technique is capable performing simple and fast solution to predict the  $R_a$  by modeling human behavior even in case of derived a mathematical model is impossible. The combination of incomplete, imprecise information and the imprecise nature of the decision-making process proved the FL ability in modeling complex model. FL may be viewed as an attempt at formalization of two incredible human capabilities: First, the capability to converse, reason and make rational decisions in an environment of imperfect information and Second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations [14]. Therefore, in order to predict  $R_a$ , FL is capable to handle an environment of imperfect information of process parameters and give desired machining performance. The FL is a formation of fuzzy IF-THEN rules which uses the concept of pure FL system. It consist set of rules of the IF-THEN form. This experiments uses fuzzy rule-based by manually construct the rules based on the experimental results.

This paper focuses on prediction of  $R_a$  value for  $Al_2O_3$  ceramics workpieces machined by laser assisted machining using a Nd:YAG laser by constraining the input using four process parameters that are depth of cut, rotational speed, feed, and pulsed frequency.

This paper outlines an understanding of how FL operates to develop the model for prediction of  $R_a$  in the laser process. Since there are no considerations of studies taken on the laser process of  $Al_2O_3$  ceramics workpieces using FL, it can be considered as the contribution for this paper. The objective of this paper is to study the result for the prediction of  $R_a$  by applying FL as modeling technique. Real machining experimental results by Chang and Kuo [3] is considered as the case study for FL modeling purposes.

## 2 Development of FL Model

The development of FL model for predicting  $R_a$  involves six phases which are data preparation, parameter setup, fuzzification, rule creation, development of inference engine, and defuzzification.

A set of machining experimental data by Chang and Kuo [3] is referred. First phase is data preparation which needs the real experimental data to be analyzed. Second phase is parameter setup that will be used in the development of fuzzy logic. There consists of the default parameter setting included the determination of the number of parameter that is based on the default of MATLAB fuzzy logic toolbox. The determination of the value of parameters is based on the analysis of the real machining experiment results collected in the first phase.

Fuzzification which defined in phase 3 is formed by analyzing and classifying numerical measurement into fuzzy set by defining triangular membership function (MF) in order to assign the linguistic value and fuzzy interval. All these four inputs have been assigned with three linguistic values namely Low (L), Medium (M), and High (H).  $R_a$  is assigned with nine linguistic values namely NegativeLow (NL), Low (L), PositiveLow (PL), NegativeMedium (NM), Medium (M), PositiveMedium (PM), NegativeHigh (NH), High (H), and PositiveHigh (PH). For linguistic values and fuzzy interval, the determination of parameter list and its values is based on analysis for both of the data set collected in Step 1 and the default parameters in MATLAB fuzzy logic toolbox.

Rule creation defined in phase 4 involved analysis of the data preparation from phase 1 and phase 3. Based on the analysis, a list of rules consists of collection IF and THEN is created. Nine rules are created based on the analysis of real experimental result.

Phase 5 involves the development of inference engine that employed the list of rules created in phase 4. There are two components involved namely implication and aggregation. The implication uses the IF part meanwhile the aggregation uses THEN part of the list of rules. Finally, phase 6, defuzzification took place and extract fuzzy set into representative fuzzy value by applying centroid. These fuzzy values will lead more precise predicted  $R_a$  value.

## 3 Result and Discussion

The development of proposed FL model may reduce time and operational cost through ‘trial and error’ experiments in order to get a good combination value of process parameters to obtain minimum value of  $R_a$ . Table 1 shows the result of real machining experiment and proposed FL model including the difference values between experimental and predicted results. The differences between predicted results showed a better  $R_a$  values according to the lower values than the experimental results. It has been proven by the value of the estimation error which is 0.072.

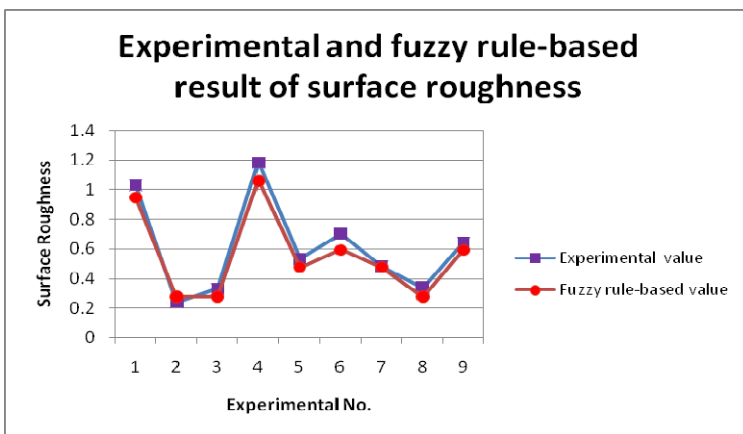
The relationships between the experimental values and predicted values are plotted as shown in Figure 1. The figure indicates a much closed result between experimental and predicted values. It shows the ability of FL in predicting the  $R_a$ . It is evident from

the Figure 1, that eight measured results are less than experimental results. The remaining one predicted  $R_a$  value which is at experimental number 2 is higher than measured values. Even the remaining one predicted  $R_a$  value is higher, it shows that the combination at experimental number 2 is lower than remaining experimental numbers. These experiments are still giving a good agreement where the average of the experimental values and predicted values are 0.61 and 0.552, respectively.

The proposed FL model with the combination of three linguistic values for each input and nine linguistic values for an output has given a better compared to real machining experiments. It has been proven by using the combination of these linguistic values, the correlation coefficient value is  $0.994 \approx 1$  between experimental and predicted result. The statistics and paired sample test between experimental and proposed FL model are summarized in Table 2.

**Table 1.** The predicted and differences result

Test case	Value of Process Parameters				Surface Roughness		Differences between Experimental Result and Predicted Result
	Depth of cut (mm)	Rotational speed (rpm)	Feed (mm/rev)	Pulsed frequency	Real Experimental $R_a$ values ( $\mu\text{m}$ )	Predicted fuzzy $R_a$ values ( $\mu\text{m}$ )	Experimental vs. fuzzy
1	0.2	1000	0.01	30	1.03	0.945	0.085
2	<b>0.2</b>	<b>1500</b>	<b>0.02</b>	<b>40</b>	0.24	0.276	0.036
3	0.2	2000	0.03	50	0.33	0.276	0.054
4	0.5	1000	0.02	50	1.18	1.060	0.120
5	0.5	1500	0.03	30	0.53	0.475	0.055
6	0.5	2000	0.01	40	0.70	0.593	0.107
7	1	1000	0.03	40	0.48	0.475	0.005
8	1	1500	0.02	50	0.34	0.276	0.064
9	1	2000	0.01	30	0.64	0.593	0.047
$R_a$ minimum					0.24	0.276	
$R_a$ maximum					1.18	1.060	
$R_a$ average					0.61	0.552	
Estimation error							0.072



**Fig. 1.** Experimental vs. Predicted results of surface roughness



**Table 2.** Statistics and Paired sample test

Comparison of real experimental and predicted values for fuzzy rule-based

Variable	Mean	N	Std. Deviation	Std. Error Mean	Correlation
Real Experimental	0.6078	9	0.32050	0.10683	0.994
Fuzzy rule-based	0.5521	9	0.28641	0.09547	

Experimental data vs. fuzzy rule-based

Variable	Paired differences							
	Mean	Std.deviation	Std.error mean	95% Confidence interval of the difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Real Exp_Laser and Fuzzy_Laser	0.05567	0.04842	0.01614	0.01845	0.09288	3.449	8	0.009

## 4 Conclusion

The ability of FL model has been proven by using the combination of three linguistic values for each input and nine linguistic values for an output. The predicted  $R_a$  results are more accurate by selecting optimal combination of the fuzzy sets.

In this experiment, the best predicted  $R_a$  value using FL is located at combination 0.2mm (depth of cut), 1500rpm (rotational speed), 0.02mm/rev (feed) and 40 (pulsed frequency). It is proved that rotational speed had the most dominant effect on LAM performance. The proposed FL model is suitable in order to give best predicted  $R_a$  values since the estimation error has been proven with the close to zero value. The accuracy of proposed FL model as shown in Table 2 proved that the ability of proposed FL model in predicting  $R_a$  values.

It should be noted that the selection of combination numbers for input and output linguistic values in MF is an important part that lead to the best predicted  $R_a$  value. This prediction model proved that combination at experimental number 2 gave very optimal combination machining parameters values to obtain minimum  $R_a$  values.

**Acknowledgments.** The authors greatly acknowledge the Research Management Centre, UTM and Ministry of Higher Education (MoHE) for financial support through the Exploratory Research Grant Scheme (ERGS) Vot. No. R.J130000.7828.4L087.

## References

1. Dubey, A.K., Yadava, V.: Experimental study of Nd: YAG laser beam machining—an overview. *Journal of Materials Processing Technology* 195(1), 15–26 (2008)
2. Pandey, A.K., Dubey, A.K.: Taguchi based fuzzy logic optimization of multiple quality characteristics in laser cutting of Duralumin sheet. *Optics and Lasers in Engineering* 50(3), 328–335 (2012)
3. Chang, C.W., Kuo, C.P.: Evaluation of surface roughness in laser-assisted machining of aluminum oxide ceramics with Taguchi method. *International Journal of Machine Tools and Manufacture* 47(1), 141–147 (2007)
4. Samant, A.N., Dahotre, N.B.: Laser machining of structural ceramics—a review. *Journal of the European Ceramic Society* 29(6), 969–993 (2009)
5. Wang, Y., Yang, L.J., Wang, N.J.: An investigation of laser-assisted machining of Al<sub>2</sub>O<sub>3</sub> particle reinforced aluminum matrix composite. *Journal of Materials Processing Technology* 129(1), 268–272 (2002)
6. Syn, C.Z., Mokhtar, M., Feng, C.J., Manurung, Y.H.: Approach to prediction of laser cutting quality by employing fuzzy expert system. *Expert Systems with Applications* 38(6), 7558–7568 (2011)
7. Zain, A.M., Haron, H., Sharif, S.: Prediction of surface roughness in the end milling machining using Artificial Neural Network. *Expert Systems with Applications* 37, 1755–1768 (2010)
8. Zain, A.M., Haron, H., Sharif, S.: Application of GA to optimize cutting conditions for minimizing surface roughness in end milling machining process. *Expert Systems with Applications* 37, 4650–4659 (2010)
9. Benardos, P.G., Vosnaikos, G.C.: Predicting surface roughness in machining: a review. *International Journal of Machine Tools and Manufacture* 43, 833–844 (2003)
10. Mohd Adnan, M.R.H., Sarkheyli, A., Zain, A.M., Haron, H.: Fuzzy logic for modeling machining process: a review. *Artificial Intelligence Review*, 1–35 (2013)
11. Torabi, A.J., Joo, E.M., Xiang, L., Siong, L.B., Lianyin, Z., Jie, P.S., Junhong, Z., Sheng, S.L.H., Tijo, J.T.T.: A Survey on Artificial Intelligence Technologies in Modeling of High Speed End-Milling Processes. In: *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 320–325 (2009)
12. Madić, M., Radovanović, M.: An Artificial Intelligence Approach For The Prediction Of Surface Roughness In Co<sub>2</sub> Laser Cutting. *Journal of Engineering Science and Technology* 7(6), 679–689 (2012)
13. Muhamad, M.N., Kadirgama, K., Ruzaimi, M., Rejab, M.: Artificial Intelligent Model to Predict Surface Roughness in Laser Machining. In: *International Conference on Recent Advances in Materials, Minerals & Environment* (2009) ISBN 978-983-3986-57-6
14. Zadeh, L.A.: Is there a need for fuzzy logic? *Information Sciences* 178, 2751–2779 (2008)

# GRT-Multigraphs for Communication Networks: A Fuzzy Theoretical Model

Siddhartha Sankar Biswas<sup>\*</sup>, Bashir Alam, and M.N. Doja

Department of Computer Engineering  
Jamia Millia Islamia University, New Delhi – 110025, India  
{ssbiswas1984,babashiralam}@gmail.com,  
ndoja@yahoo.com

**Abstract.** There are many real life problems of network which can not be modeled into graphs but into multigraphs. In this paper we introduce a theoretical notion of ‘Generalized Real Time Multigraph’ (GRT-Multigraph) whose construction is based on fuzzy set theory. The GRT multigraph can be regarded as a data structure which is an improvement of the notion of RT multigraph (real time multigraph). Any RT-multigraph can be viewed as a special case of a GRT-multigraph. The present day networks of communication systems, be it a giant or not, contain a lot of uncertainties, in particular regarding attack or damage from internal or external sides. Consequently, the existing links(arcs) of a given network are not always in their excellent condition, rather in a weaker condition at any real instant of time. Our proposed model of GRT-multigraph is expected to play a vital role in any network of communication system because of its high potential in considering such real time data and information, and may open a new direction for rigorous research in communication systems.

**Keywords:** FN, Multigraphs, RT-multigraphs, GRT-multigraphs, Neighbor node, TBL, LSV, LSC, TBN, RN, Communicable node, CF, EC.

**MSC Code (2000):** 05C85.

## 1 Introduction

Many of the communication network of present days are of small size, many are of medium or large size and many are like giants, in terms of their number of nodes and/or links, and also in terms of cost of the links. There are many real life problems of networks, in particular of computer science, communication systems, transportation systems, electrical networks, metagraph nets, biological neuron systems, optimization techniques, library and information systems, etc. which can not be modeled into graphs, but into multigraphs only, and then can be easily solved. In the present days of ‘big data’ flow, the networks are expanding very fast in huge volumes in terms of their nodes and links/arcs. In many cases, there are migration of data

---

<sup>\*</sup> Corresponding author.

from one giant network to another giant network too. But, in reality, for a given alive network its complete topology may not be always available to the communication systems at a given point of time because of the reason that few or many of its links/arcs may be temporarily disable or under-performer owing to some amount of damage or external attack or blockage upon them, and of course they are under repair at that point of time. One link may be good for communication process at this moment of time, but may get damaged after few hours because of several possible reasons which are usually unpredictable and hence ill defined in general. Besides that, in most of the cases the cost parameters corresponding to its links are not crisp numbers, rather fuzzy numbers (FNs). Thus at any real time instant, the complete multigraph is not available but a submultigraph or a weaker submultigraph of it is available to the system for executing its communication or packets transfer.

There was no mathematical model available in the existing literature to represent such type of real time sub-networks (which are variable networks) of the original network. As a solution, in our previous work in [8] we introduced a mathematical model for such types of multigraphs called by 'Real Time Multigraphs' (RT-multigraphs) in which real time information (being updated every  $q$  quantum of time) are incorporated so that the communication/transportation system can be made very efficiently with optimal results. In this paper we make further consideration of real time data and information to introduce a theoretical model 'Generalized Real Time Multigraphs' (or GRT-Multigraphs) as a generalization of RT-multigraphs equipped with fuzzy theory. The notion of GRT-multigraphs is the most generalized form of the crisp multigraphs, and will surely play a major role in networks, in particular in computer science, communication systems, transportation systems, electrical engineering, metagraph nets, biological neuron systems, optimization techniques, library and information systems, etc. Clearly a GRT-multigraph is a variable representation of a network with respect to the parameter time.

## 2 Preliminaries

In this section we present basic preliminaries of the existing notion of multigraphs. A multigraph  $G$  is an ordered pair  $(V, E)$  which consists of two sets  $V$  and  $E$ , where  $V$  or  $V(G)$  is the set of vertices (or, nodes), and  $E$  or  $E(G)$  is the set of edges (links or arcs). Here, although multiple edges or arcs might exist between pair of vertices but in our discussion in this paper we consider that no loop exists.

Multigraphs may be of two types: undirected multigraphs and directed multigraphs. In an undirected multigraph the edge  $(i, j)$  and the edge  $(j, i)$ , if exist, are obviously identical unlike in the case of directed multigraph. For a latest algebraic study on the theory of multigraphs, the work [7] may be seen along with the books ([1],[2],[3],[5]). Fig.1.(a) shows below a directed multigraph  $G = (V, E)$ , where  $V = \{A, B, C, D\}$  and  $E = \{AB_1, AB_2, BA, AD, AC, CB, BD, DB\}$ . Fig.1.(b) shows a submultigraph of it.

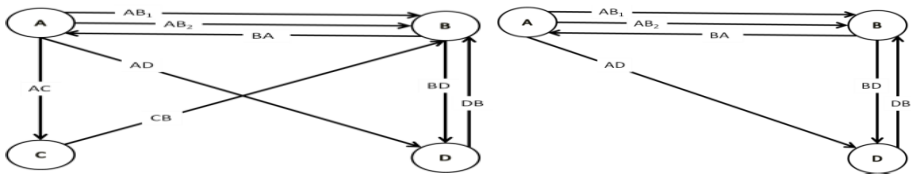


Fig. 1. (a) A multigraph G; (b) A submultigraph H of G

### 3 Generalized Real Time Multigraph (GRT-Multigraph) : A Fuzzy Model

In most of the real life problems of networks, be it in a communication model or a transportation model, the weights of the arcs are not always crisp but fuzzy numbers (FNs). For example, the Fig.2 below shows a public road transportation model for a traveler where the cost parameters for travelling each arc have been considered as FN involving a pre-estimated degree: degree of acceptance.

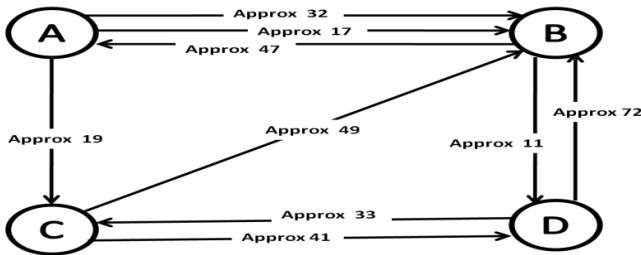


Fig. 2. A multigraph G with fuzzy weights of arcs

#### 3.1 ‘Neighbour’ Node

For a given node  $u$  the node  $v$  will be designated as a ‘neighbour’ node of  $u$  if  $u$  has at least one link from  $u$  to  $v$ . In our work here we consider more real situations which are actually and frequently faced by the present communication systems. For example, consider an Adhoc Network or a MANET in which there may exist multiple paths between two neighbour nodes, but because of some reasons one or more number of paths may not be in the ideal condition (may be partially damaged, or temporarily damaged). Thus, although they are available for transmission of packets by a node  $u$  to its neighbor node  $v$ , but will cause the communication delay (for a damaged condition, there will be no scope for communication). This is a very useful information to the communication system if available to the sender node in advance. For this we introduce a new parameter corresponding to each link (edge) called by ‘Condition Factor’ (or ‘link status’).

### 3.2 ‘Condition Factor’(CF) of a Link

Consider a node  $u$  and its neighbor node  $v$ . Suppose that there are  $n (\geq 1)$  number of links from  $u$  to  $v$  outward which are  $uv_1, uv_2, \dots, uv_n$ . Let us designate them as  $1^{st}, 2^{nd}, 3^{rd}, \dots, n^{th}$ . For each link  $uv_i$ , we define the **condition factor (CF)** or **link status** at this instant of time in the following way :-

(i) The ideal (i.e. best) ‘condition factor’ for each link  $uv_r$  is 1, if it is available at its original condition without any damage or attack internally/externally at this moment of time and the node  $v$  is functional at the real time under consideration.

We write  $CF(uv_r) = 1$ .

(ii) The worst ‘condition factor’ for each link  $uv_r$  is 0, if either it is not available, i.e. it is at a condition of fully damaged/blocked at this moment of time and thus having no feasibility for communication or the node  $v$  itself is non-functional at the real time under consideration. We write  $CF(uv_r) = 0$ .

(iii) Otherwise, the ‘condition factor’ for each link  $uv_r$  is in between 0 and 1, which means that the link is available but not in its original/ideal condition and the node  $v$  is functional at the real time under consideration. The link  $uv_r$  could be partially damaged or in a traffic-jam or in a similar one out of many other real time circumstances (could be a temporary problem, and expected to become OK soon after repair) which may cause the communication to be at slower pace.

Thus, for the pair of nodes  $u$  and  $v$  here,  $0 \leq CF(uv_r) \leq 1 \quad \forall r$ . If at some real period of time  $CF(uv_r)$  is close to 1, then it signifies that the available link  $uv_r$  is in its good condition for communication almost like its original condition. If  $CF(uv_r)$  is close to 0, the available link  $uv_r$  is in a very bad condition for communication. If  $CF(uv_r) = 0$ , then either the link  $uv_r$  is ‘not available’ or the node  $v$  itself is non-functional at the period of time under consideration. Clearly, if the node  $v$  is non-functional then  $CF(uv_r) = 0 \quad \forall r$ , but the converse is not necessarily true.

Consider the following directed RT-Multigraph [8]  $G$  where the fuzzy weights (FNs) are shown against each link (Fig.3).

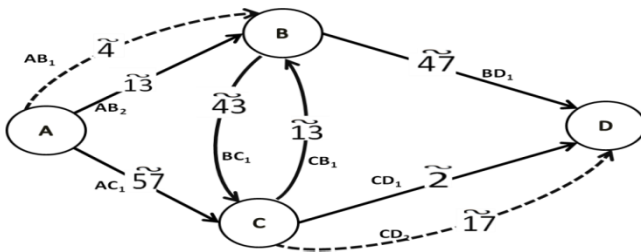


Fig. 3. A RT-multigraph  $G$  having few links damaged partially/temporarily

In our proposed mathematical model of GRT-multigraphs (see Fig.4), we incorporate further the real time data from the network regarding the condition of each and every link(arc) to make the RT-multigraphs more dynamic, more useful, and hence more efficient to the users for making an optimal strategy for communication.

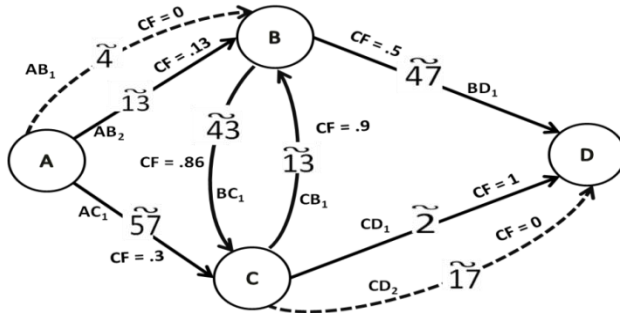


Fig. 4. A GRT-multigraph G with various CF at some instant of time

Suppose that there are  $n (\geq 1)$  number of links from  $u$  to  $v$  outward which are  $uv_1, uv_2, \dots, uv_n$ . Let us designate them as  $1^{st}, 2^{nd}, 3^{rd}, \dots, n^{th}$ . In real life situation, because of natural phenomenon (flood, earthquake, thunderstorm, solar storm, etc. etc.) or because of some kind of external attack or technical failure or because of an predictable/unpredictable fully or partial damage of the link, it may happen in reality that during a period of time the  $r^{th}$  link  $uv_r$  of the node  $u$  to its neighbour  $v$  is not available at its best condition ( $r = 1, 2, 3, \dots, n$ ). In our proposed model of GRT-multigraphs, this is a precious information and is available with the node  $u$  here in advance. If the node  $u$  has the node  $v$  as a neighbour node then  $u$  carries the following information handy with it :-

- (i) Every node of the multigraph carries an information vector corresponding to each of its neighbour nodes. Corresponding to every neighbour node  $v$  of  $u$  in a GRT-multigraph, there exist a **Link Status Vector (LSV)**  $I_{uv} = (i_1, i_2, i_3, \dots, i_n)$  of  $u$ , where at any given point of time  $i_r$  happens to be some value from the closed interval  $[0,1]$  for  $r = 1, 2, 3, \dots, n$  where  $i_r = CF(uv_r)$ .
- (ii) If at a given time  $i_r$  happens to be 0, i.e. if the link  $uv_r$  is completely non-functional then we say that the link  $uv_r$  is a **temporarily blocked link (tbl)** from  $u$ . The CF value  $i_r$  is also called the **'link status'** of the link  $uv_r$  as mentioned earlier.

In real situation the complete multigraph thus may not be available due to existence of non-functional or under-performance status for few links, i.e. due to existence of few tbls and few under-performer links. Few or many of the available links may not be available with  $CF = 1$ , but available with less amount of CF. Consequently not the complete topology but a sub-multigraph or a weaker multigraph of it be available for communication (Example: for communication of packets in an Adhoc Network/ MANET, or for a salesman to travel many cities, or for a buss/truck carrying goods/passengers in a transportation network, etc.).

If a node  $u$  has  $k (\geq 0)$  number of neighbour nodes  $v_1, v_2, v_3, \dots, v_k$ , then  $u$  carries  $k$  number of LSV:  $I_{uv_1}, I_{uv_2}, I_{uv_3}, \dots, I_{uv_k}$ . In our mathematical model of GRT-multigraph, we propose that there is a system  $S$  for the multigraph which

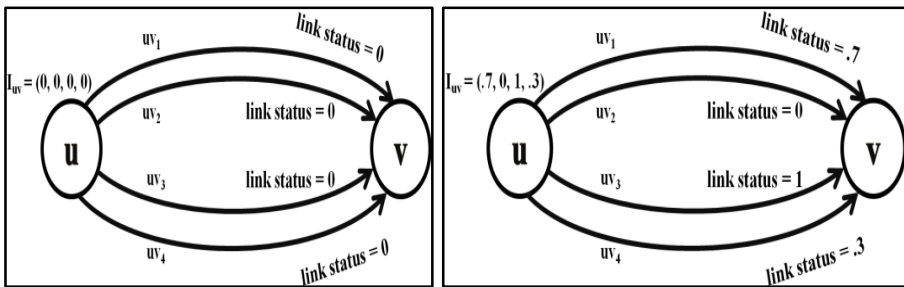
updates all the information vectors of all the nodes after every quantum time  $\tau$ . This quantum  $\tau$  is fixed (can be reset) for the system  $S$  in a multigraph but different for different multigraphs, in general depending upon the various properties of the physical problem for which a multigraph is modelled.

**(iii) Link Status Class (LSC)**

For a given node  $u$ , the collection of all LSV are called ‘Link Status Class’ (LSC) of  $u$  denoted by  $I_u$ . If a node  $u$  has  $k (\geq 0)$  number of neighbour nodes  $x_1, x_2, x_3, \dots, x_k$ , then  $I_u = \{I_{ux1}, I_{ux2}, I_{ux3}, \dots, I_{uxk}\}$ .

**(iv) Temporarily Blocked Neighbour (tbn) & Reachable Neighbour (rn)**

If  $v$  is a neighbour node of a given node  $u$ , and if  $I_{uv}$  is a null vector at a given instant of time then  $v$  is called a **temporarily blocked neighbour (tbn)** of  $u$  for that instant.



**Fig. 5.** (a) A tbn  $v$  of the node  $u$ ; (b) A rn  $v$  from the node  $u$

However, since it is a temporary phenomenon, and if any of the links be repaired in due time, then obviously a ‘blocked neighbour’ may regain its ‘neighbour’ status at some later stage. If a neighbour  $v$  is not a tbn, then it is called a **reachable neighbour (rn)** of  $u$ . Thus  $v$  is a reachable node from  $u$  if there is at least one link having non-zero CF (see Fig.5(a),(b) above).

**(v) Communicable Node**

For a given node  $u$ , if  $I_u \neq \phi$  and at least one member of  $I_u$  is non-null at a given time, then the node  $u$  is called a communicable node for that instant of time. If  $u$  does not have any neighbor node then  $I_u = \phi$ , and in that case it is trivial that further communication is never possible. However, if  $I_u \neq \phi$  and all the members of  $I_u$  are null vectors at a point of time, then it signifies that further communication is not possible temporarily.

**(vi) Information Update Periodically**

There is a system  $S$  of the multigraph such that all the real time information mentioned/defined above will get automatically updated at every node of the multigraph



at every  $q$  quantum of time. This quantum  $\tau$  is fixed (can be reset) for the system  $S$  in the multigraph but different for different multigraphs, in general depending upon the various properties of the physical problem under exercise for which the multigraph is modelled.

**(vii) Effective Cost (EC) of a Link**

Consider a node  $u$  and its neighbor node  $v$ . Suppose that there are  $n (\geq 1)$  number of links from  $u$  to  $v$  outward which are  $uv_1, uv_2, \dots, uv_n$ . Let us designate them as  $1^{st}, 2^{nd}, 3^{rd}, \dots, n^{th}$ . Corresponding to each link  $uv_r$ , there is a cost (weight) of the link which is a FN  $n_r$ . If this link is not available at its best condition, then a fraction of its ideal condition is available to the system for communication. Consequently, if this link is chosen for communication of a packet from node  $u$  to node  $v$ , the effective cost of this link will not be in reality equal to  $n_r$  but a little higher side, depending upon the condition of the link  $uv_r$  at that real instant of time. Then the **effective cost (EC)** of the link  $uv_r$  will be defined by

$$EC(uv_r) = n_r / CF(uv_r), \quad \text{where } CF(uv_r) \neq 0.$$

If  $CF(uv_r) = 0$ , then we say that  $EC(uv_r) = \infty$ . Since  $n_r$  is a fuzzy number,  $EC(uv_r)$  is also a fuzzy number for each  $r$ .

**3.3 GRT-Multigraphs**

Such type of multigraphs is called ‘**Generalized Real Time Multigraphs**’ or ‘**GRT-multigraphs**’ as they contain all real time information of the networks with time. Consequently, for a given network the GRT-multigraph is not a static multigraph but changes with time, i.e. becomes weaker sometime, regain ideal condition back, again becomes weaker, so on. As a special case, if a network can be modelled into a graph (need not be a multigraph) then we call our proposed model as ‘**Generalized Real Time Graph**’ or ‘**GRT-graph**’ as a special case of ‘**Generalized Real Time Multigraphs**’ or ‘**GRT-multigraphs**’.

**4 Conversion of a GRT-Multigraph into RT-Multigraph**

A GRT-multigraph cannot be converted physically into an equivalent RT-multigraph. But for the purpose of solving various problems of soft-computing for implementing very effective communication system via a GRT-multigraph, one can mathematically convert it into an equivalent RT-multigraph and apply all the algorithms/theories of RT-multigraph [8] to find the final solutions for the GRT-multigraph. Consider the GRT-multigraph  $G_{grt}$  of Figure-4 at some instant of time. Now, for each link of this  $G_{grt}$  if we replace the existing cost by a new value equal to the corresponding EC value of it, we get a new multigraph  $G_{rt}$  with common  $V$  and  $E$ . This new multigraph  $G_{rt}$  (Fig.6) can be viewed as an equivalent RT-multigraph of the GRT-multigraph  $G_{grt}$ , because in  $G_{rt}$  we shall view the condition factor to be either 0 or 1 only for each link.

Thus in  $G_{rt}$ , corresponding to every neighbour node  $v$  of a given node  $u$ , the **Link Status Vector (LSV)** of  $u$  is  $I_{uv} = (i_1, i_2, i_3, \dots, i_n)$ , where at any given point of time  $i_r$  takes any of the two values only from  $\{0,1\}$  for  $r = 1, 2, 3, \dots, n$  with the following significance :-

$$i_r = 0, \text{ if either the link } uv_r \text{ or the node } v \text{ is non-functional.}$$

$$= 1, \text{ if the link } uv_r \text{ and } v \text{ both are functional.}$$

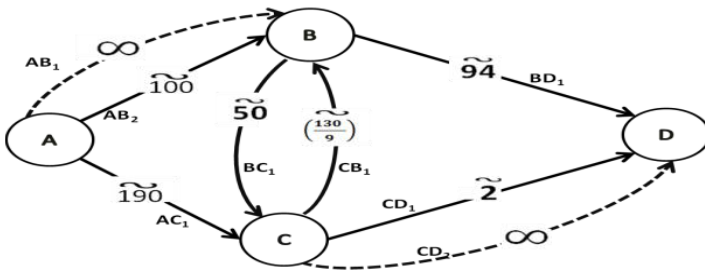


Fig. 6. GRT-multigraph (of figure-4) converted into an equivalent RT- multigraph

## 5 Conclusion

In this paper we have introduced a theoretical notion of Graph Theory called by “GRT-multigraphs” which is a generalization of the notion of RT-multigraphs [8]. In a RT-multigraph, at a given instance of time, a link is either available (status = 1) or not available (status = 0), i.e. the status of a link or CF of a link is always one score from the set  $\{0,1\}$ . In a GRT-multigraph, the status or CF of a link can be from a closed interval  $[0,1]$ , depending upon its condition or status for communication at that real instant of time. A GRT-multigraph can be mathematically converted into a RT-multigraph for computing purposes while searching for solutions for the various problems on the GRT-multigraphs, in particular in the areas of computer science, communication systems, transportation systems, electrical networks, metagraph nets, biological neuron systems, optimization techniques, library and information systems, etc. Our future research work will be to implement fuzzy CF in GRT-multigraphs instead of crisp CF values because of the reason that condition factor of a link may not be always a precise quantity but an ill-defined quantity.

## References

1. Balakrishnan, V.K.: Graph Theory. McGraw-Hill (1997)
2. Bollobas, B.: Modern Graph Theory. Springer (2002)
3. Diestel, R.: Graph Theory. Springer (2000)
4. Dubois, D., Prade, H.: Fuzzy Sets and Systems: Theory and Applications. Academic Press, New York (1980)
5. Harary, F.: Graph Theory. Addison Wesley Publishing Company (1995)

6. Jenson, P., Barnes, J.: Network Flow Programming. John Wiley and Sons, NY (1980)
7. Biswas, S.S., Alam, B., Doja, M.N.: A Theoretical Characterization of the Data Structure 'Multigraphs'. *Journal of Contemporary Applied Mathematics* 2(2), 88–106 (2012)
8. Biswas, S.S., Alam, B., Doja, M.N.: Intuitionistic Fuzzy Real Time Multigraphs for Communication Networks: A Theoretical Model. In: *Proceedings of 2013 AASRI Conference on Parallel and Distributed Computing Systems (DCS 2013)*, Singapore (2013); © 2013 Published by Elsevier B.V. Selection and/or peer review under responsibility of American Applied Science Research Institute
9. Zadeh, L.A.: Fuzzy Sets. *Inform. And Control* 8, 338–353 (1965)

# Exploring the Key Determinants of Successful ICT Innovation Adoption: A Case Study of a Fishing Community in Thailand

Heng Wei Lee<sup>1,\*</sup>, Maziani Sabudin<sup>1</sup>, Rosnah Idrus<sup>1</sup>, Mohd Azam Osman<sup>1</sup>,  
Thongchai Kruahong<sup>2</sup>, and Nattayanee Leanjay<sup>2</sup>

<sup>1</sup> School of Computer Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia

<sup>2</sup> Faculty of Science and Technology, Suratthani Rajabhat University, 84100

Suratthani Province, Thailand

lhwips10\_man0@student.usm.my,

{maziani,rosnah,azam}@cs.usm.my,

thongchai@sru.ac.th, mod\_mis@hotmail.com

**Abstract.** This study aims at shedding some light on identifying the key determinants of successful innovated ICT adoption in a fishing community in Thailand. The innovated ICT delivered a medium for fishing community and the authorities to communicate over many kinds of information like tides and appropriate fishing strategy, new fishing techniques and practices, government announcement, news sharing, and more. In this study, interviews were carried out with the fishermen to recognize the factors involved in their adoption process. The result of the analysis showed that the adoption process is first initiated by the task-technology fit which examine how the technology characteristics fit with the task requirement. As the fit is met, livelihood assets like human capital, social capital, and financial capital will be gained as the result of the usage of the technology. Fishermen then utilized the livelihood assets to develop living strategies and subsequently improve their living quality.

**Keywords:** Technology adoption, Mobile technology, Livelihoods, Qualitative method.

## 1 Introduction

The International Fund for Agricultural Development (IFAD) showed that one of the causes that lead to poverty is the limited access to information and poor communication technology [1]. Many researches shown that mobile technology is able to lessen the information asymmetries, increase market efficiency by reducing risk and uncertainty, promote market integration, and enhance user's networking capability. In order to grasp the benefits of mobile technology, a collaboration between School of Computer Sciences, University Sains Malaysia and the Faculty of Science and Technology, Suratthani Rajabhat University had initiated a project called M-Community

---

\* Corresponding author.

GreenEve2Peace for Fishing Community, to incorporate the mobile and Internet technologies into the management and the preservation of aquaculture industry in Khun Talay fishing community, Thailand. M-Community GreenEve2Peace for Fishing Community serves as a platform to allow the authority to disseminate information to the fishing community through a more convenience and cost-effective way. The range of information includes the schedule of fishing activities like fish feeding and harvesting, tides levels, weather forecast, and also the appropriate strategy to be used by the fishermen to fish. All these information will be communicated and broadcasted via Short Message Service (SMS).

## **2 Study Objectives**

The purpose of this study is to explore the key determinants of the adoption of the innovated ICT system deployed within the M-Community GreenEve2Peace for Fishing Community projects. Past literature has focused on examining the factors that impact users' adoption mainly from the hedonic intensive technology (usage is mainly for individual pleasure) for example, e-commerce, e-learning, social networking, and more. However, very few of the studies were done specifically to address the acceptance of utilitarianism intensive technology (usage is mainly to maximize happiness as a whole). Unlike most of the technology adoption literature, this study focused on a utilitarianism intensive technology as the usage of the system is mainly aimed to facilitate and improve the current management of the aquaculture practices in Khun Talay fishing community as a whole. As a result, the findings of this study are per-haps to fill up the gaps in the current literature and identifying the impediments which affect the adoption of utilitarianism intensive technology among the fishermen.

## **3 Research Methodology**

A convenience sampling strategy was used to recruit members from the fishing community to participate in the face-to-face interviews. We managed to recruit a total of 37 fishermen to participate in this interview. All of the fishermen have been using GreenEve2Peace for at least six months. The interviews were carried out over a period of two months. In sum, eight rounds of interview were conducted with approximately four to five interviewees per session. Each interview took place in a single session and took about 50 minutes each. Most of the interviews were conducted in Thai language and a few in English. Every interviewee consented to the interview and to being audio-recorded.

The interview was conducted using mostly an unstructured format consisting of two parts to elicit fishermen's belief regarding their adoption behavior in using GreenEve2Peace. In the first part of the interview, the objective of this study was explained to the interviewees. Then, each respondent was asked to describe their working experiences, usage of ICT in work and during leisure time, and how they became involved in this project. During the process, interviewees were encouraged to talk freely about their attitude and behaviors regarding the GreenEve2Peace usage.

The interviewees' initial replies and probing for further elaboration would guide the flow of the interview process. In the second part of the interview, more direct questions were asked about the difficulties and troubles interviewees faced when using GreenEve2Peace. The way they solve their problems were also discussed subsequently.

## 4 Data Analysis

The audio-recordings of the interviews were translated into English and transcribed into an interview report after each session. After that, the transcribed interviews were subjected to a coding process using computer software packages designed for qualitative analysis - ATLAS.ti version 5.0. During the coding process, open-coding was applied wherein common themes related to issues of GreenEve2Peace adoption were identified and passages from the interviews were marked with the corresponding thematic codes. Next, theme analysis was performed wherein the codes were compiled so that similarities and differences among codes could be justified and categorized into different themes. After all thematic codes were identified, verification and validation of the results was done in two ways: peer debriefing and inter-coder reliability testing. The debriefing meetings were carried out repeatedly, step by step, until the final report was ready. The data was then submitted to inter-coder reliability testing. In this study, inter-coder reliability testing was divided into two rounds. In the first round, two independent reviewers were asked to evaluate the meaning of each thematic code and determine if it was best fitted with the quotations attached to it. Inter-coder agreement is present when both of the coders agree that the quotations listed exactly reflect the thematic codes attached to them. In the second round, the codes on which the testers did not agree were modified by attaching them to different quotations and/or by expanding the meaning of the codes to reduce confusion and ambiguity. The process of comparison of result continued until agreement exceeded 0.6. The percentage of agreement in this study was recorded as 0.7. A percentage between 0.6 and 0.8 indicated a good level of agreement. Thus, the findings of this study are considered sufficient to provide solid grounds for accepting the identified themes as real issues pertaining to GreenEve2Peace adoption.

## 5 Findings

Themes emerged with regards to the issues experienced by the fishermen in adopting GreenEve2Peace in practice. Identified themes fell into the following categories: (a) Task- Technology fit, (b) Financial Capital, (c) Human Capital, (d) Social Capital, and (e) Satisfaction.

### 5.1 Task-Technology Fit

Task-Technology Fit (TTF) is defined as to the capability of a technology to assist users in performing his portfolio of tasks [2]. The fishermen's tasks basically falls

into this five categories, which are task complexity (categorized tasks into simple and complex tasks), task effort (the amount of effort without requiring much cognitive workload to perform a task), task frequency (the number of recurring), task importance (the view of certain task to be more salient than others) and task time criticality (time sensitive task that required immediate execution). For instance, the task to decide which fishing strategy to use depending on the tides conditions would indicate high task complexity for fishermen. They need to listen to news in order to get the information about tides. Although, information on tides can be easily accessed through Internet but unfortunately most of them are computer-illiterate. Therefore, task effort would be high for them to collect information on tides from other sources. Despite the task of collecting information on tides seems to be complex, required lot of efforts, high reoccurring task, important, and time critical, this study showed that fishermen believed the attributes of GreenEve2Peace is capable to updated tides information quickly so that they can decide on which appropriate strategy to fish, this include using the correct fishing nets and fish at the right timing. Some of the fishermen said that:

*“GreenEve2Peace is quick; it can provide me with updated information.” [P3]*

*“GreenEve2Peace make my work easy because it provides detailed information and I can get the information when I need it” [P20]*

*“The information provided in GreenEve2Peace is relevant, and it is clear and understandable.” [P15]*

## **5.2 Financial Capital**

Financial capital is related to the financial resources available which provide livelihood enhancement [3]. Increased profit margin can be easily traced when the cost of operating reduced or the numbers of catches increased, from the fishermen’s perspective. This study revealed that GreenEve2Peace is able to improve fishermen’s financial capital where it helps them to reduce cost in three extents; reducing the reliance on external vendors for information, saving fuel money by reducing numbers of fishing trips, and save time for other part-time jobs. When the fishermen use GreenEve2Peace, they can communicate among community to get more information regarding the market price of the fishes, tides level, new technology in fishing, and modern fishing techniques. Meanwhile, with accurate and instant information provided by GreenEve2Peace, fishermen can have better catch because of correct timing of fishing, and appropriate fishing nets can be used according to the tides level. Consequently, fishermen can reduce the number trips to fish in order to catch the minimum amount of fishes to support their living. As was said by some of the fishermen:

*“I get more money when is use GreenEve2Peace because I get more information” [P1]*

*“Using GreenEve2Peace help me to save my petrol cost.” [P17]*

*“I can do part-time because I have more time with GreenEve2Peace.” [P36]*

### 5.3 Human Capital

Human capital development is constituted by the improved quality of the fisher-men. It can be determined by the education, skills, and also health conditions [3]. Human capital provided an outcome-based perspective on the adoption behavior of the fishermen when using GreenEve2Peace. Fishermen will appreciate GreenEve2Peace when the functionality of the technology created a net benefit to the fishermen after deducting the cost of using it. This includes the added knowledge, and skills the fishermen acquired during the process of adopting GreenEve2Peace:

“ My fishing practices improved and my cost also lower after I use GreenEve2Peace.” [P13]

“I learn more about fishing.” [P16]

“Using GreenEve2Peace allows me to learn new innovative practices in fishing” [P22]

Fishermen learned new fishing techniques, innovated fishing practices to improve catches, and increased awareness of the importance of sustainable fishing via GreenEve2Peace. All this values are the outcomes of using GreenEve2Peace which further reinforce the usability of the technology itself.

### 5.4 Social Capital

Social capital is the benefit derived from the membership in the community which includes the support and assistance from their counterparts and the authorities during their daily activities. One of the members in the community said that:

“I can communicate better with other fishermen and the authority when I use GreenEve2Peace.” [P30]

Besides that, other fishermen also claimed that:

“Using GreenEve2Peace could strengthen my connectivity and contact with other fishermen.” [P37]

“My access to government institution become easier if I use GreenEve2Peace” [P34]

“ I can easily get assistance in fishing when I use GreenEve2Peace” [P31]

Hence, we can deduce that GreenEve2Peace provides social resources (networks, social relations, affiliations, and association) to the fishermen to draw upon in order to develop different livelihood strategies which require coordinated actions. Established on the existing skills which they already learned (how to SMS using their phone), GreenEve2Peace is considered to be more user friendly and subsequently encourage more usage from the members of the fishing community.

### 5.5 Satisfaction

DeLone and McLean Model on system success showed that information, system, and service qualities will result in user satisfaction and then affect their intention to use [4]. The success of GreenEve2Peace adoption largely depends on how the system effectively improves fishermen livelihood. At the end of the interviews, each of the



interviewees was asked about their satisfaction level towards GreenEve2Peace on a 5-point scale, with 1 being very dissatisfied and 5 being very satisfy. The result showed that on average the fishermen satisfaction level towards GreenEve2Peace was recorded at 3.97 with standard deviation of 0.31. Consequently, the data indicated that fishermen are satisfied with the performance of GreenEve2Peace in improving their living quality. The result of the case study is summarized in figure 1.0.

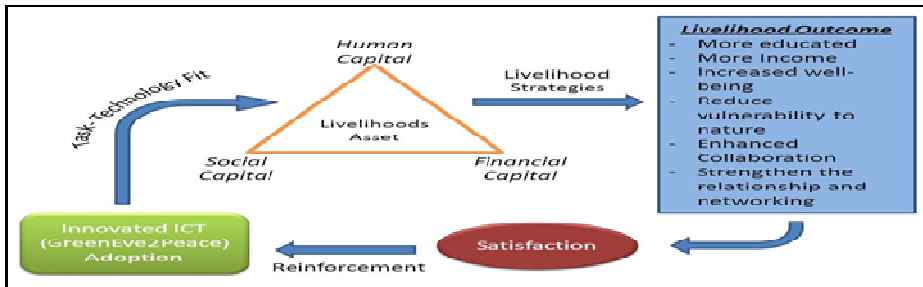


Fig. 1. Show GreenEve2Peace adoption process

## 6 Conclusion

Exposure to climate variability, limited information, and lacking in technology-enabled medium for communication are posing substantial risk to the fishermen in the Khun Talay fishing community. The introduction of GreenEve2Peace to the fishing community is expected to improve their living quality by reducing the risk they will encounter when they want to earn a living. Through, closely communication and collaboration among the fishermen and the authorities, fishermen now are more informed and better prepared to the unpredicted nature disaster.

## References

1. Grimshaw, D.J., Kala, S.: Strengthening Rural Livelihoods: The Impacts of Information and Communication Technologies in Asia. Practical Action Publishing, Warwickshire (2011)
2. Goodhue, D.L., Thompson, R.L.: Task Technology Fit and Individual Performance. *Management of Information System* 31(6), 213–236 (1995)
3. Chambers, R., Conway, G.: Sustainable Rural Livelihoods: Practical Concepts for the 21st Century. IDS, Brighton (1992)
4. DeLone, W., McLean, E.: Information System Success: The Quest for The Dependent Variable. *Information System Research* 3(1), 60–95 (1992)

# Dynamic Bandwidth Allocation Algorithm for Ethernet Passive Optical Networks Based on Traffic Prediction

Chengwei Xiao\*, Hong Fan, and Dongjing Liu

School of Optoelectronic Engineering,  
Nanjing University of Posts and Telecommunications, Nanjing, 210023, China  
xcw@njupt.edu.cn,  
{fanh.njupt, liudongjing55}@163.com

**Abstract.** In this paper, an algorithm based on traffic prediction is proposed, in bid to predict instantaneous traffic in ONU, guide higher-priority bandwidth allocation, and avoid "T+2" queuing delay. That is the Dynamic Bandwidth Allocation (Pre\_DBA) for EPONs, which also includes a prediction corrected mechanism, to maximize bandwidth utilization. As is verified in the designed simulation, this algorithm, apart from ensuring effectiveness and fairness of the premise, is quite able to alleviate delay and jitter in higher-priority services as well.

**Keywords:** Ethernet Passive Optical Network (EPON), Dynamic Bandwidth Allocation (DBA), Priority, Traffic Prediction, Delay.

## 1 Introduction

Ethernet Passive Optical Network (EPON) has been widely considered as a promising technology for implementing the FTTx solutions to the "last mile" bandwidth bottleneck problem. Different from other standardized organizations, EPON is merely specified by IEEE802.3 in physical layer and data link layer of its system while the others are all excluded, giving rise to the concentration of academic researches regarding EPON mainly on the standard-unspoiled areas. And the Dynamic Bandwidth Allocation (DBA) technology research, however, constitutes a both crucial and overheated direction of EPON system.

## 2 Major DBA Algorithms for EPON

Currently, a dazzling array of state-of-the-art DBA algorithms have been put forward, for instance, IPACT(Interleaved Polling with Adaptive Period Time) and BGP (Bandwidth Guarantee Polling)[1], which, although have significantly enhanced the bandwidth utilization compared with the static approach, give little thoughts to QoS still.

---

\* Corresponding author.

Another algorithm proposed for EPON is CBR(Constant Bit Rate), which, featuring differentiated support, is capable of providing reliable QoS support for delay-sensitive services. In this algorithm, intra-ONU bandwidth scheduling is controlled at OLT, by referring to the grant-report-based polling program in IPACT, while inter-ONU scheduling is completed at ONU. CBR includes three different priority queues, for services that shall be coped with respective strategies, namely, the voice, the video and the data services. A strict priority-focused concept should be applied in inter-ONU and intra-ONU scheduling so as to produce smaller delay and packet loss in real-time services such as voice services than those in data, and to further perfect QoS in higher-priority services.

Nevertheless, those DBA algorithms in various forms can still be improved in the following two areas:

(1) Performance improvement necessitates all-round perspectives. All existing DBA algorithms are fundamentally shaped for a certain problem or a defined context, making them unavoidably one-sided to some extent, for one field can be undermined when the other is undergoing improvement. One evidence is that in CBR the arrival rate of higher-priority services must be acknowledged in advance.

(2) "T+2" queuing delay needs to be addressed. According to GATE/REPORT mechanism, having requested bandwidth through REPORT in current cycle (T) and achieved grant of GATE, ONU can send data in the next cycle (T+1). This is called a "T+1" transmission approach. However, for ONU data acquired after the REPORT being sent, "T+2" approach is the only option.

### 3 Pre\_DBA Algorithm Based on Traffic Prediction for EPON

#### 3.1 Description

By way of drawing upon CBR, Pre\_DBA succeeds in having three different priority queues lined up for each ONU cache, respectively,  $P_0$ ,  $P_1$ ,  $P_2$ . And the traffic predicted by OLT from ONU is nothing but the amount of packets arriving at an ONU between two consecutive pollings based on the self-similarity characteristic of network traffic.

The total bandwidth  $ONU_i$  acquired in Cycle  $k$  can be interpreted as follows.

$$B_i(k) = B_{i,0}(k) + B_{i,1}(k) + B_{i,2}(k) \quad (1)$$

In Formula (1),  $B_i(k)$  refers to the total bandwidth in Cycle  $k$ ;  $B_{i,j}(k)$  to the bandwidth assigned for  $ONU_i$  Queue  $j$  in Cycle  $k$  by OLT;  $J=0, 1, 2$ , namely to three different priority queues in  $ONU_i$ ,  $P_0$ ,  $P_1$ ,  $P_2$ , respectively.

To avoid the upstream channel being dominated by a certain ONU in the long run, Pre\_DBA sets for  $B_{w,i}(k)$  a maximum, following that the total bandwidth acquired by  $N$  ONU in which  $ONU_i$  is included shall be no more than  $T_{\max}R_N$ .

$$T_{\max}R_N = NB_g + \sum_{l=i+1}^N B_l(k-1) + \sum_{l=1}^{i-1} B_l(k) + B_{w,i}(k) \quad (2)$$

In this Formula,  $R_N$  represents the line transmission rate;  $B_g$  represents the guaranteed bandwidth;  $T_{\max}$  represents the maximum polling cycle;  $N$  represents the number of ONU.

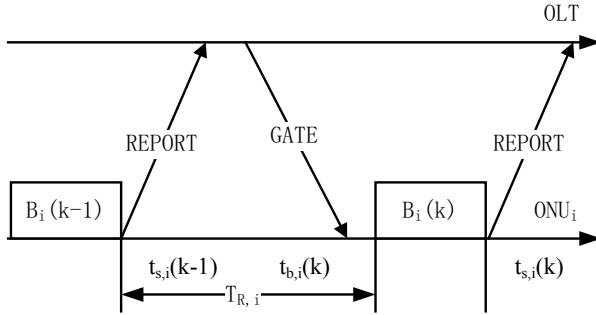
Constrained by the above-mentioned  $B_{w,i}(k)$ , in integrating bandwidth percentage allocated in each queue  $\phi_j$ , Pre\_DBA makes a further definition of the defined higher-, medium- and lower-priority bandwidth assigned to ONU<sub>*i*</sub>.

$$B_{i,0}(k) = \min\{B_{R,i0}(k) + B_{p,i}(k), B_{w,i}(k)\phi_0\} \quad (3)$$

$$B_{i,1}(k) = \min\{B_{R,i1}(k), B_{w,i}(k)\phi_1\} \quad (4)$$

$$B_{i,2}(k) = \min\{B_{R,i2}(k), B_{w,i}(k)\phi_2\} \quad (5)$$

In Formula (4) and (5),  $B_{R,ij}(k)$  is the bandwidth requested by ONU<sub>*i*</sub> for its Queue *j* in Cycle *k*;  $B_{p,i}(k)$  the pre-assigned bandwidth for voice and other real time services in higher-priority queue of ONU<sub>*i*</sub> in Cycle *k*.



**Fig. 1.** Diagram on REPORT Cycles

The pre-allocation of traffic bandwidth available only to higher-priority services by Pre\_DBA is called traffic prediction. The interval after ONU finishing sending one REPORT frame and before it starting the next is a REPORT Cycle, marked as  $T_R$ . To work out the value of  $B_{p,i}(k)$  as is explained above, the traffic produced in ONU within  $T_R$  must be predicted first by OLT. In Fig. 1,  $t_{s,i}(k-1)$  stands for the local time when bandwidth is requested in Cycle *k* after ONU<sub>*i*</sub> has sent REPORT;  $t_{b,i}(k)$  for ONU<sub>*i*</sub> timeslot start time in Cycle *k*, as is seen in the diagram.

$$T_{R,i} = t_{b,i}(k) - t_{s,i}(k-1) + B_i(k) / R_N \quad (6)$$

$$T_{R,i} = t_{b,i}(k) - t_{s,i}(k-1) + \frac{B_{R,i0}(k) + B_{p,i}(k) + B_{i,1}(k) + B_{i,2}(k)}{R_N} \quad (7)$$

As the size and the arriving time distribution of packets obey the exponential distribution, i.e., data packets generated by each source means a Poisson process. And superimposing several data sources can add up their stability, the amount of arriving packets  $B_{p,i}(k)$  in current  $T_R$  can be predicted by a reference to the (average) value in previous cycle(s).

### 3.2 Prediction Corrected Mechanism

It is worth to mention that the fluctuation of actual transmission rate in higher-priority services can lead to the deviation of the prediction value  $B_{p,i}(k)$  from the real traffic value. To enlarge the bandwidth utilization, a prediction corrected mechanism is introduced in Pre\_DBA algorithm, which is elaborated as follows.

$$\Delta Q_i(k-1) = B_{i,0}(k-1) - [B_{R,i0}(k-1) + A_i(k-1)] \quad (8)$$

In Formula(8),  $B_{i,0}(k-1)$  stands for the bandwidth allocated by OLT for  $ONU_i$  higher-priority queue in Cycle  $(k-1)$ ;  $A_i(k-1)$ , for the actual traffic arrived in  $T_R$  interval for  $ONU_i$  higher-priority queue;  $\Delta Q_i(k-1)$ , for the deviated value between the bandwidth allocated by OLT for  $ONU_i$  higher-priority queue in Cycle  $(k-1)$  OLT and that in actual need.

$B_{p,i}(k)$  can be corrected in the following formula.

$$B_{p,i}(k) = B_{p,i}(k-1) + \Delta_i(k) \quad (9)$$

In this equation,  $\Delta_i(k)$  refers to the corrected value for  $B_{p,i}(k)$ .

All in all, the Pre\_DBA algorithms can be concluded in these final formulas.

$$B_{i,0}(k) = \min\{B_{R,i0}(k) + B_{p,i}(k) + \Delta_i(k), B_{W,i}(k)\phi_0\} \quad (10)$$

$$B_{i,1}(k) = \min\{B_{R,i1}(k), B_{W,i}(k)\phi_1\} \quad (11)$$

$$B_{i,2}(k) = \min\{B_{R,i2}(k), B_{W,i}(k)\phi_2\} \quad (12)$$

## 4 Simulation

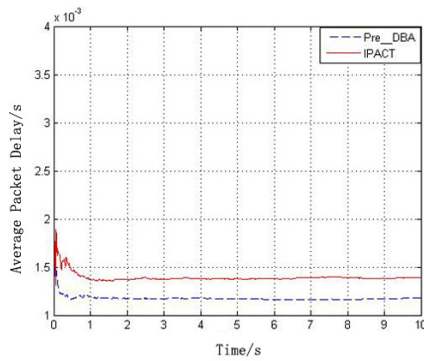
In EPON systems consisting of 1 OLT and 16 ONU, OLT-to-ONU data transmission rate ( $R_N$ ) is 1000Mbps, user-to-ONU data transmission rate ( $R_U$ ) is 100Mbps. The polling cycle is 2ms; protection slot between the upstream data 5 $\mu$ s; the maximum transmission window ( $W_{max}$ ) 212500 bytes; the cache in one ONU queue 1 megabyte.

### 4.1 Analysis of Delay in Higher-Priority Services

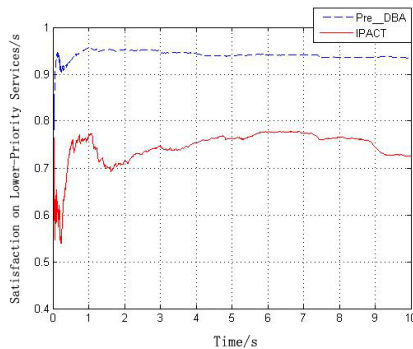
Delay variance in both Pre\_DBA and IPACT is calculated in this study. While the value in the former is  $5.11E-07$  that in the latter is  $5.11E-07$ , giving vivid descriptions of their delay jitter in higher-priority services. The fact that variance in Pre\_DBA is significantly smaller than that in IPACT, has very much indicated that Pre\_DBA is a better improvement compared with IPACT in terms of delay jitter.

### 4.2 Analysis of Average Packet Delay

As is shown in the fig. 2, Pre\_DBA and IPACT are compared in perspective of average package delay in higher-priority services. The figure for Pre\_DBA higher-priority services is roughly 1.2ms, while that for IPACT is 1.4ms larger, making the former very much better in this regard.



**Fig. 2.** Average Package Delay



**Fig. 3.** Satisfaction on Lower-Priority Services

### 4.3 Analysis of Satisfaction on Lower-Priority Services

Fig. 3 implies the satisfaction on lower-priority services while the upstream bandwidth utilization is approximately ranging from 60% to 70%. It is also evident that in IPACT, the result is around 75% while it keeps up to a much higher number in Pre\_DBA, 95%. Conspicuously, traffic prediction designed for improving performances in higher-priority services, can come to its own end, and simultaneously help upgrade medium- and lower-priority services to some degree.

## 5 Conclusion

This study, in improving CBR algorithm, designs a DBA algorithm based on traffic prediction for higher-priority services like voice and other real-time services, which is then followed by pre-allocation of bandwidth, and constant predict correction in order to alleviate delay and jitter in higher-priority services, and also improve the bandwidth utilization effectively. And EPON model which is built in OPNET Modeler software verifies that the research is not only feasible, but more effective to alleviate delay and jitter in higher-priority services, and to better satisfy network demands from diversified services and varied priorities.

## References

1. Xue, C., Sun, S.-H., Dong, L.: Ethernet Passive Optical Networks, pp. 35–51, 157–196. Beijing University of Posts and Telecommunications Press, Beijing (2007)
2. Yan, D.-S., Bian, E.-J., Xu, W., Liu M.: EPON-New Generation Optical Access Technology and Application, pp. 76–80, 114–116. Mechanical Industry Press, Beijing (2007)
3. Zou, J.-N., Xiong, H.-K., Wang, M.: Traffic Estimation and Pre-allocation Supported Dynamic Bandwidth Allocation Algorithm for EPON. *Journal of Shanghai Jiaotong University* 41(6), 940–943 (2007)
4. Ding, Z.-Z., Zhu, N., Fang, L.-D.: A EPON dynamic bandwidth assignment algorithm based on the multiple traffic prediction. *Optical Network* 12, 30–31 (2008)
5. Zheng, J., Mouftah, H.T.: A Survey of Dynamic Bandwidth Allocation Algorithms for Ethernet Passive Optical Networks. *Optical Switching and Networking* 6, 151–162 (2009)
6. Wang, W.-B., Zhang, J.-W.: OPNET Modeler and Network Simulation. People's Posts and Telecommunications Press, Beijing (2003)

# Game Based Learning for Teaching Electrical and Electronic Engineering

Michael Callaghan<sup>1</sup> and Mark Cullen<sup>2</sup>

<sup>1</sup> Intelligent System Research Center, University of Ulster,  
Derry, Northern Ireland

<sup>2</sup> School of Creative Arts, University of Ulster,  
Derry, Northern Ireland  
{mj.callaghan,m.cullen}@ulster.ac.uk

**Abstract.** In recent years there has been significant growth in the use of video games technologies and game mechanics for teaching and learning. These environments and techniques offer the ability to create complex, highly interactive simulations with solid theoretical underpinnings to present teaching material in new and highly interactive ways. This paper discusses and practically demonstrates how video games mechanics can be used to create highly immersive and engaging user experiences to teach engineering related material. The Circuit Warz project is introduced and demonstrates how a game-based approach, using a competitive format, can be used to create immersive, highly engaging student but pedagogically sound learning experiences.

**Keywords:** Virtual Worlds, Engineering Education, Virtual Learning Environments, Game Based Learning.

## 1 Introduction

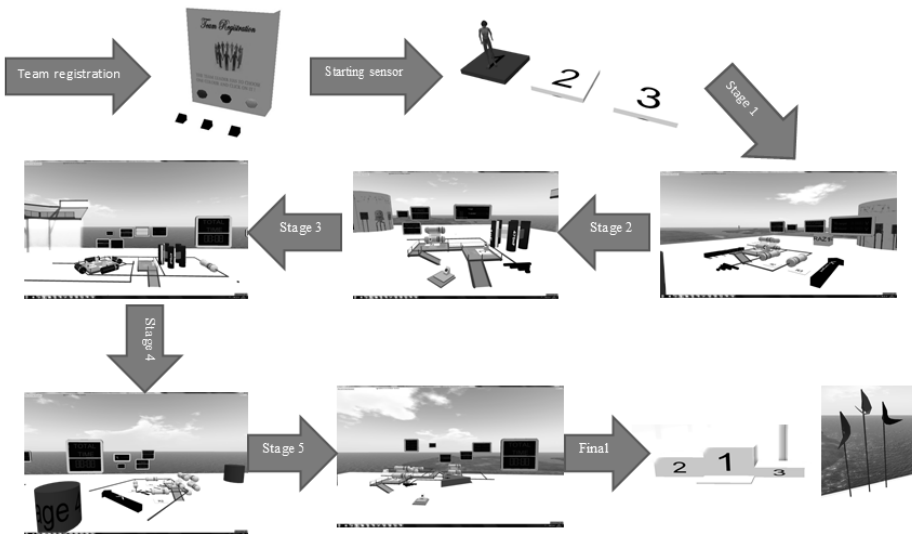
Gamification is a term used to describe the application of video game mechanics to non-game processes in order to improve user engagement. This type of game based learning is increasingly been used in educational settings and is widely predicted to become mainstream in the next 3-5 years [1-3]. This paper discusses a practical example of using game mechanics for educational and teaching purposes in the context of electrical and electronic engineering. It demonstrates how a commercial games engine e.g. Unity3D can be used to rapidly prototype simulations to teach advanced electronic/electrical circuit theory. A game based learning experience is used inside a 3D immersive game world where the students compete to bias electronic circuits.

Section 2 of the paper discusses recent University of Ulster research in virtual worlds and serious games, section 3 looks at an extension to this research to create a cross platform Unity3D based version of the Circuit Warz game. Section 4 concludes the paper.



## 2 Game Based Learning in Virtual Worlds

Internet-based 3D virtual worlds are immersive environments which facilitate an advanced level of social networking where residents can explore and socialize by participating in individual and group activities [4]. The Serious Games & Virtual Worlds research team at the Intelligent Systems Research Center (ISRC), University of Ulster focus on the potential of virtual worlds and video games technologies for undergraduate/postgraduate teaching of electrical and electronic engineering related subjects [5]. In this context the Circuit Warz project was conceived with the overall objective to investigate if creating a compelling, engaging, immersive and competitive environment to teach electrical and electronic theory and principles would increase student engagement. The project was created using the OpenSim virtual world simulator integrated with the Moodle virtual learning environment and SLOODLE [6]. The game was a team based exercise where groups of students worked together collaboratively to compete competitively against other teams to complete a virtual assault course. In practice this was a series of electronic and electrical circuits (puzzles) which need to be solved (i.e. biased correctly) in order to complete the game and progress to the next level (Fig1).



**Fig. 1.** Circuit Warz virtual assault course

The virtual assault course was made up of five stages of increasing more complex electronic and electrical circuits which needed to be correctly biased/solved to proceed. The stages implemented include a series/parallel resistor circuit (Fig.2), R/C filter circuit, Graetz bridge, Wheatstone bridge and a weighted summing amplifier circuit. Each stage had a series of learning outcomes e.g. in the series/parallel resistor circuit simulator the key learning objective of the experiment was to understand

parallel and series resistances. Learning was achieved by creating exercises where the student would run simulations, within a given time constraint, to calculate the correct value of  $R_1$  given fixed values of  $R_2$ ,  $R_3$  and  $V_o$  in varying configurations (Fig.3). The arrows and pie/bar charts in the simulation provided real-time feedback to the students via visualization of the relative voltage drops across the circuit as the value of  $R_1$  changed. Students are assessed on how quickly they could identify the correct value for  $R_1$  by applying their theoretical knowledge practically. The feedback enabled the students to understand the circuit operation on two levels i.e. using a rule of thumb to visualize the relationship between small and large  $R_1$  values and circuit output and how to calculate the exact values of  $R_1$  for specific  $V_o$  values.

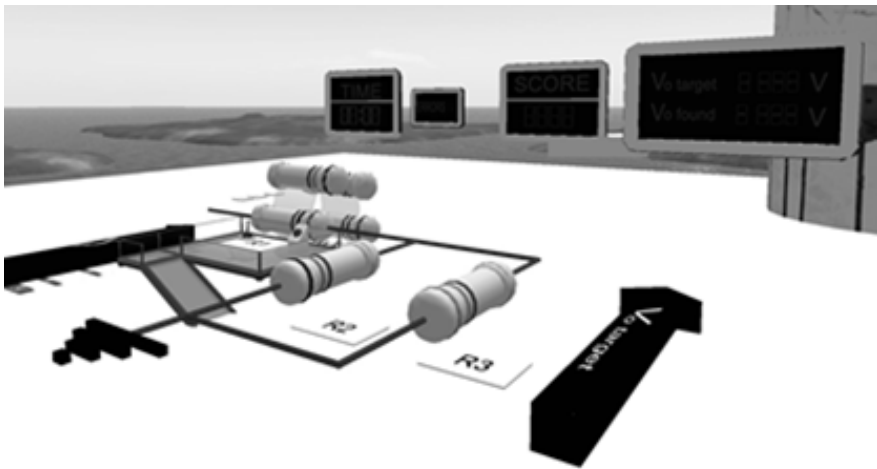


Fig. 2. Series/parallel resistor circuit

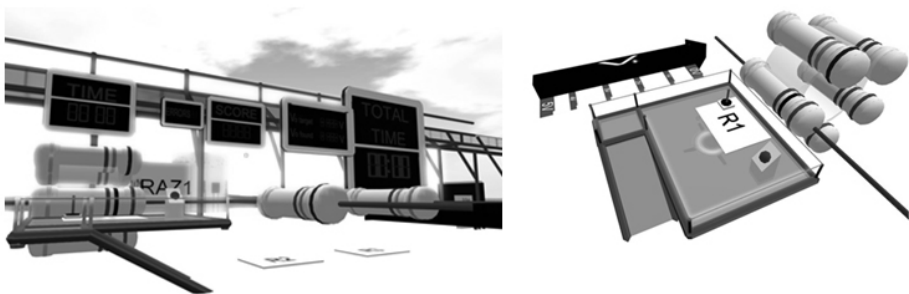


Fig. 3. Calculate value of  $R_1$  with real time visualization of relative voltage drops

This approach allowed the presentation of abstract circuit theory in new and highly interactive ways allowing students to experiment with different biasing setups and visualize the resulting circuit phenomenon. Students are assessed/scored on how close they match the cut off frequency to the target value under a given time constraint.

Visual feedback on performance is displayed on the score board as a percentage of accuracy related to calculating circuit output. The other simulations described earlier followed a similar approach.

## 2.1 Initial Evaluation and Limitations

The initial evaluation process focused on user acceptance of this type of environment as teaching tools from both an educator and student perspective. The overall feedback was positive. The cohort were familiar with social networking and technology in general and after a short learning curve readily accepted the game based virtual world as just another tool and complementary resource to add to their repertoire of learning resources with minor reservations e.g. granularity of navigation controls and interactions. In summary the students enjoyed the collaborative group aspect of the project and the ability to interact with the simulations and visualize circuit theory/operation in new ways. In addition they strongly felt that the competitive team based element of the project helped reinforce the theoretical material learnt as they had to practically apply this knowledge. The academic staff involved in this stage of the evaluation, were very positive about the potential this approach had once the initial learning curve was overcome. In particular they felt the collaborative working facilities offered by the 3D immersive environment was the most useful and warranted the extra effort required to create the content. However there were a number of shortcomings identified with the project particularly related to the use of the OpenSim platform e.g. installation and updating of the client software, networking/ports restrictions and lack of support for deployment in browsers and on mobile devices e.g. iPhone, iPad and Android platforms.

## 3 Cross Platform Application Development

To address the previous shortcomings the project was redesigned and repurposed for deployment using Unity3D, a cross platform game engine which allows the game to be easily ported to browsers and mobile devices. An additional two stages were added to the game creating seven increasing difficult levels for the student to complete. The game layer/client was integrated into the Moodle virtual learning environment and an

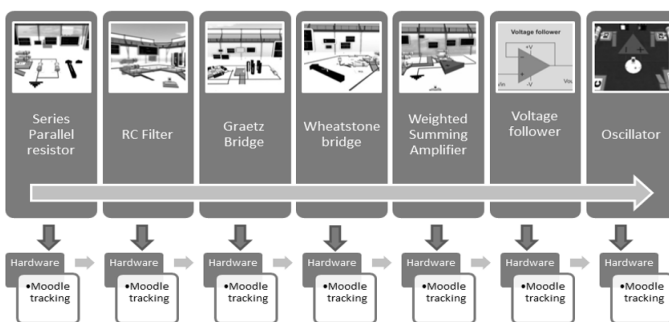


Fig. 4. Seven stage game with hardware/virtual learning environment integration

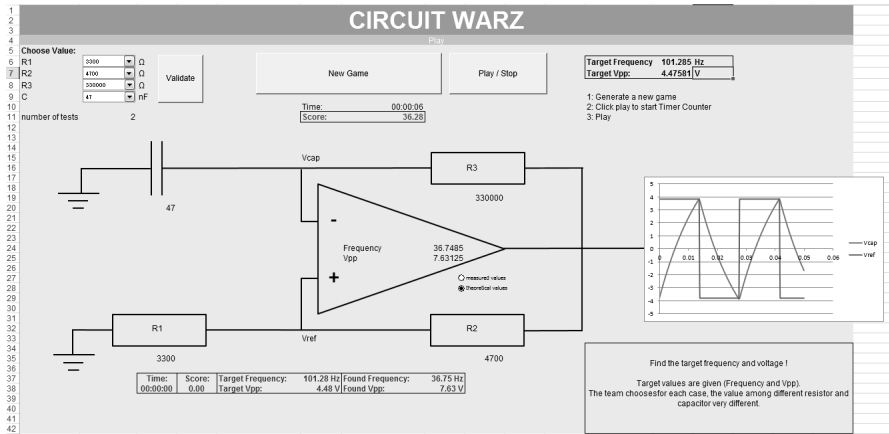


Fig. 5. Level 7, the oscillator circuit modelled in Excel

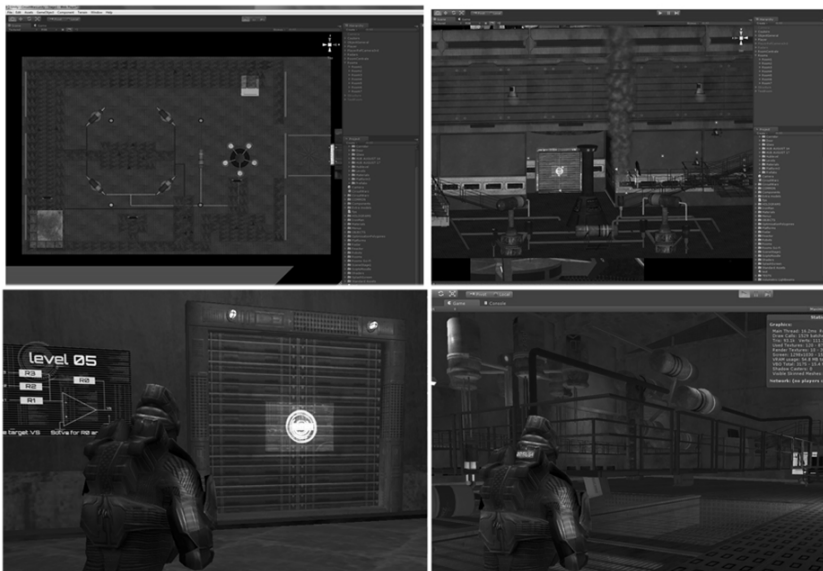


Fig. 6. Theoretically correct physical layout of in-game and real world circuits in level 5

underlying hardware infrastructure (Fig.4). Each stage of the game was modelled using Excel to determine/fine tune the core game play and determine the relative values of the components/formulas before the main development started (Fig 5).

The physical layout of the circuit's was accurately recreated inside the Unity level editor (Fig. 6). This element of the game design was important as the circuit layout and physical operation had to accurately reflect the constraints/requirements of the real world counterparts. Figure 6 shows screenshots from in-world game play (in this instance level 5, the weighted sum amplifier). Figure 7 shows the Unity3D client/game integrated into the Moodle virtual learning environment and linked to a range of test instrumentation and circuits/hardware.

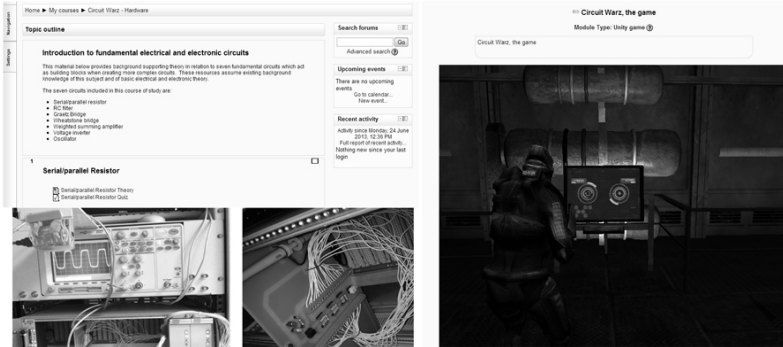


Fig. 7. Game/Unity3D integrated with the Moodle learning environment and hardware

## 4 Conclusion

This paper provided an overview on ongoing research at the Intelligent Systems Research Center, University of Ulster, Northern Ireland into the use of virtual worlds/games and virtual learning environments for teaching. The Circuit Warz project was introduced and a number of complex, highly interactive and engaging simulations described which make effective use of game play mechanics to engage students. This approach potentially offers a new engaging and highly interactive way to teach engineering related material.

Overall this technology is maturing rapidly and reaching the stage where it is robust and reliable for wide scale deployment as an enhancement and extension to virtual learning environments. Barriers to widespread adoption relate to educator awareness, the inherent learning curve, and acceptance of the possible benefits of using these environments for teaching and a willingness to explore innovative technologies in educational practice.

## References

1. Kirriemuir, J.: An update of the July 2007 'snapshot' of UK higher and further education developments in Second Life Eduserv Foundation, Bath (2007)
2. Chang, Y., Aziz, E.-S., Esche, S.K., Chassapis, C.: Game-based laboratory for gear design. *Computers in Education Journal* 22(1) (2012)
3. Aziz, E.-S., Esche, S.K., Chassapis, C.: Review of the state of the art in virtual learning environments based on multi-player computer games. *Computers in Education Journal* 20(1), 22–35 (2010)
4. Kemp, J., Livingstone, D.: Putting a Second Life 'Metaverse' skin on learning management systems. Second Life Education Workshop at SLCC San Francisco (2006)
5. Callaghan, M.J., McCusker, K., Losada, J.L., Harkin, J., Wilson, S.: Using Game-Based Learning in Virtual Worlds to Teach Electronic and Electrical Engineering. *IEEE Trans. on Industrial Informatics* 9(1), 575–584 (2013)
6. Callaghan, M.J., McCusker, K., Losada, J., Harkin, J.G., Wilson, S.: Engineering Education Island: Teaching Engineering in Virtual Worlds. *ITALICS (Innovation in Teaching and Learning in Information and Computer Sciences)* 8(3) (November 2009)

# Mobile WiMAX Resource Allocation Design Goals: Key Features/Factors/Issues

Zaid G. Ali\*, R.B. Ahmad, Abid Yahya, and L.A. Hassnawi

School of Computer and Communication Engineering,  
University Malaysia Perlis,  
02600 Arau, Perlis – Malaysia  
{zaid679, laythmail}@yahoo.com,  
{badli, abidyahya}@unimap.edu.my

**Abstract.** The technical design challenge of Mobile WiMAX is to exploit the limited available spectrum of the downlink subframe efficiently while achieving the desired goals behind the resource management algorithms. Various goals can be achieved when the resource management algorithms exploited in the right direction. This paper presents the coordination between MAC layer scheduler objectives and PHY layer burst allocation algorithm design parameters. Utilizing PHY layer operation mode with MAC layer coordination leads to increase the proportion of the desired objectives achievement. The presented methodologies of the burst allocation algorithm design can achieve higher satisfaction of MAC requirements and avoid the limitations. The study has demonstrated that resource allocation algorithm requires many factors and parameters to be employed in the correct direction to enable efficient use of the downlink resources to satisfy MAC layer requirements.

**Keywords:** Burst allocation, Downlink subframe, Resource allocation algorithm design, Mobile WiMAX.

## 1 Introduction

The Mobile WiMAX system is attracting huge interest as a promising solution for delivering mobile broadband wireless access services. The IEEE standards 802.16e-2005 [1] specifies the requirements for the MAC and PHY layers of the Mobile WiMAX BS systems. These two layers provide great potential for satisfying users and operators needs. From user perspective the wireless system should deliver data with the required QoS level, while operator aim is to maximize network capacity and revenue. These two goals are achieved by resource management algorithms [2]. Resource management in the OFDMA-based Mobile WiMAX network includes Call Admission Control (CAC), transmission algorithms, and handover algorithms. The CAC algorithm handles system overloading and satisfies users' QoS by limiting the number of users in the system. Transmission algorithms enable QoS guaranteed opportunistic

---

\* Corresponding author.

data transmission over a wireless medium. They include general scheduler, bandwidth management, power control, and time-frequency resource allocation. The horizontal handover guarantees continuous service by assigning new serving BSs to the users during mobility and system load variations.

In the transmission algorithms, the key problems are computational intensity due to number of degrees of freedom, complicated frame structure, and complex MAC and PHY layers processing procedures [2]. The transmission algorithms for an IEEE 802.16e system should guarantee efficient bandwidth utilization while providing fairness between users and responding correctly to the QoS constraints [3]. Simple and efficient designs with low computational complexity are recommended for the transmission algorithms [4].

The objective of this paper is to present the features and parameters that affect the design of the resource allocation algorithm (called burst allocation algorithm hereafter) that degraded the achievement of MAC layer requirements.

The main contribution of this paper is to categorize the design of burst allocation algorithm to comply with the MAC layer requirements. The categorization includes a full detail of the required objectives, PHY layer parameters, type of service, deployment, complexity, mechanisms and so on. To the best knowledge of the author this is the first study that explaining the dark issues of the resource allocation algorithms.

## 2 Related Work

The motivation of any resource allocation management can be classify into two main categories. Firstly is to minimize the total transmit power with a constraint on the user data rate. Secondly is to maximize the total data rate with a constraint on total transmit power [5], [6]. The first category is consistent with the demands of the operators, while the second category is consistent with the demands of the users [7].

A wide range of burst allocation algorithms was reviewed in [8]. The study concludes that the motivations of the proposed allocation algorithms are basically to achieve QoS guarantees for all service classes, to maintain the fairness, to maximize the system goodput, to minimize power consumption and finally to have low complexity algorithm. The mentioned motivations are lie within the two main demands categories, which are users and operators demands.

A survey of the downlink resource allocation algorithms for OFDMA Mobile WiMAX was presented in [9]. The study reviewed the motivation of the resource allocation algorithm to classify them into five classes; packing efficiency (with burst fragmentation and without), QoS and traffic priority, power consumption, overhead reduction and other mixed approaches. Author found that designing an efficient downlink resource allocation algorithm to improve a specific class is at the cost of scarifying other class, such as grouping multiple users' data within single burst to minimize overhead size at the cost of increase power consumption at the recipients, efficient allocation impact the QoS adversely, cross layer design at the cost of higher complexity and so on. The research concludes that there is lack of comprehensive comparative analysis of the published algorithms.

The main challenge of the resource allocation algorithm is the motivations contradiction as conclude in [7], [8], [9]. Utilizing the operation mode of the PHY layer in the correct direction with MAC layer coordination if necessary, leads to increase the proportion of the desired goal achievement.

### 3 Scheduling and Resource Allocation Mechanisms

The general scheduler is located at each MAC layer of the base station to enable rapid response to traffic requirements and channel conditions [10]. In order for the MAC general scheduler to make an efficient management and provide the desired QoS in the downlink, the MSs must feedback accurate and timely information as to the traffic conditions and QoS requirements [7]. Each connection is associated with a single data service with a set of QoS parameters that quantify the aspects of its behavior [5]. Consequently, the data packets are associated to service flows with well defined QoS parameters in the MAC layer so that the general scheduler can correctly determine the packet transmission ordering over the air interface. The Channel Quality Indicator Channel (CQICH) provides fast channel information feedback to enable the general scheduler to choose the appropriate coding and modulation for each allocation.

The general scheduler is supported by burst allocator algorithm that can operate on different PHY layer types of subchannelization scheme [7]. The general scheduler that supported by burst allocator based on distributed permutation such as Partial Usage of Subchannels (PUSC) which provide similar quality subchannels can support a QoS with fine granularity and flexible time frequency resource allocation [11]. While the general scheduler that supported by burst allocator based on adjacent permutation such as Adaptive Modulation and Coding (AMC) permutation that has different subchannels attenuation can interact with burst allocator. The interaction is in term of frequency selection to allocate mobile users to their corresponding strongest subchannels [7],[11].

### 4 Resource Allocation Goals and Mechanism

To meet the highest proportion of the objectives achievement, it is important to figure out the compatibility between the desired objectives of the general scheduler and the design parameters of the burst allocation algorithm. Based on the scheduler objectives, the burst allocation algorithm design can be divided into two main scenarios. Both scenarios seeks to enhance the system performance, but in different point of view.

The first scenario: appropriate for operator demands, which is mainly focuses on increase the frame utilization. In this scenario the burst allocation is a PHY layer decision taking into account preserves the general scheduler requirements pertaining to the users' priority and QoS. The burst allocation algorithm design in this scenario focuses on managing the available frequency (subchannels) and time (slots) resources



to pack the provided users data within Downlink (DL) subframe in the form of bursts with the objective of increasing the frame utilization. The general scheduler is aware about subchannels conditions through MSs feedback. Consequently, the scheduler calculates users' QoS and priority according to user subchannels state to include or exclude them within the scheduling procedure. The distributed permutation is preferred to make sure that all the subchannels have almost equal adequate condition for all the associated users. Equal condition of the subchannels means the calculated QoS requirements can be applied to any subchannel. Consequently the burst allocation algorithm is free to allocate any subchannel set to the users, which lead to constrain the user data rate. The general scheduler can assign less robustness modulation and code rate to increase data rate whenever user's feedback indicates higher level of subchannel condition.

It is clear the objective is to find a higher possibility of frame utilization, while user satisfaction comes in the second priority and binds to the MAC layer general scheduler only. The burst allocation algorithm could be NP-problem or P-problem algorithm depends on the designer. The NP-problem is a complexity class of decision problem that can be solved by non-deterministic algorithm in polynomial time, and for the P-problem solved by deterministic algorithm in polynomial time[12],[13].

The second scenario: appropriate for user demand, which is mainly focuses on increase user satisfaction parameters. In this scenario the burst allocation is a cross layer decision between general scheduler (in MAC layer) and burst allocation algorithm (in PHY layer). The general scheduler utilizes burst allocator information to conduct user scheduling and allocation. The calculated QoS requirements (such as latency, data rate, packet error rate etc.) for individual user require a special subchannels conditions to accommodate these QoS requirements (such as the ability to provide the required constellation level, code rate and power level). The burst allocation algorithm in this scenario based on adjacent permutation AMC, which provide multiple levels of subchannel conditions to support multi-user diversity. The general scheduler and the burst allocation algorithm find the compatibility between the available PHY layer resources and the QoS requirements then allocates the resources. Consequently, this scenario requires combined adaptive scheme between MAC layer general scheduler and PHY layer burst allocation algorithm to determine the best location within frame spectrum for each user in the DL subframe that can achieve the desired objectives.

This is the fact behind the cross layer design that interested in adjacent permutation which can provide multiple levels of subchannel conditions to support user diversity. It is clear the objective is to achieve higher proportion of user satisfaction parameters, while efficient utilization of the DL subframe resources comes in the second priority. Usually cross layer design produces a NP-problem algorithm.

Table 1 summarizes the burst allocation algorithm categories, the served objectives, mechanisms, features and the required efficiency.

**Table 1.** Mobile WiMAX resource allocation goals and mechanism

<b>Category</b> [5]	<b>Category1:</b> Minimize the total transmit power with a constraint on the user data rate. (appropriate for operator demands)	<b>Category 2:</b> Maximize the total data rate with a constraint on total transmit power. (appropriate for user demands)
<b>Resource allocation mechanism</b> [7]	<b>First scenario:</b> the algorithm obey the general scheduler instruction related to (QoS, priority, power control, burst profile) to allocate resources with objective of high frame utilization.	<b>Second scenario:</b> based on the adaptive general scheduler (QoS, priority, power control, burst profile) combined with a burst allocation algorithm to develop sophisticated algorithms for determining which users to schedule, how to allocate subcarriers to them, and how to determine the appropriate power levels for each user on each subcarrier.
<b>Scheduler objectives</b> [7]	1. Maximization of operator’s revenue. 2. Simple and fast fairness implementation. 3. Power consumption.	1. Satisfy higher QoS guarantees provided to end users. 2. Maximize user throughput. 3. More sophisticated fairness to end users.
<b>System deployment</b> [7,14]	Adopted for rural environments, in which a system is deployed specifically to serve roads or railways	Adopted for urban cells in which the traffic source is dominated by pedestrians or stationary users.
<b>Burst allocation features and criteria of the design</b>	Permutation type support frequency diversity (PUSC) [8].	Permutation type support user diversity (AMC) [8].
	More convenient for but not limited to fixed-rate applications [5].	More convenient for bursty applications [5].
	Isolation between MAC & PHY layer functions [8] .	Cross layer design: adaptive combined between MAC & PHY layers [7].
	Strictly commitment to the users' priorities and QoS provided by the MAC layer [14].	Users' priorities and QoS are negotiated between MAC & PHY layers [14].
	Control the available downlink resources frequency and time.	Control the resources to satisfy the required QoS, transmit power, transmit rate (constellation) and coding rate [7].
	Less complexity.	Higher complexity
	Less BLER [5].	Higher BLER [5].
	Immunity against channel fading and subcarrier interference [15].	Violable by channel fading and subcarrier interference [15].
	Usually the algorithm serves single user at a time [14].	Usually the algorithm serves set of users at a time [14].
	Constraints on user data rate.	Constraints on transmit power.
	The algorithm considers the frame is ready for transmission and switch to the next new frame when there is no more space available to pack additional use.	The algorithm considers the frame is ready for transmission and switch to the next new frame when users set is fulfill or when the remaining DL resource cannot support the requirements of the remaining users.
	Power consumption at MSs can be supported.	Power consumption at MSs usually cannot be supported.
<b>Desired efficiency</b>	Increase frame usability.	Increase user satisfaction parameters

The first scenario of the burst allocation algorithm follows the standard recommendation in which the PHY layer works to reliably deliver information bits from the transmitter to the receiver. Usually, the PHY layer is not informed of QoS requirements and is not aware of the nature of the application, such as VoIP, HTTP, or FTP. While the MAC layer is responsible for controlling and multiplexing various such links over the same physical medium [5].

Power consumption minimization at the MSs can be achieved when the allocation algorithm can reduce the working time of MSs within DL subframe duration. The first scenario is free to allocate data bursts in any orientation vertical and/or horizontal. Consequently, first scenario can support MS power consumption if this objective considered in the design. However, the second scenario looking for the best subchannels set that can meet the MAC layer requirements. Consequently, the allocation is based on these subchannels which they are physically horizontal, and thus the second scenario unable to support MSs power consumption [8].

## 5 Conclusion

This paper presented the features of the burst allocation algorithm and its compatibility with MAC layer general scheduler requirements considering the PHY layer parameters. The ambiguous area between MAC layer general scheduler and PHY layer resource allocation algorithm has been illustrated. The methodology of optimum utilization of MAC and PHY layers entities in the design of the resource allocation algorithms has been introduced.

It has been concluded that the mismatch between general scheduler desired objectives and the correct chose of the appropriate PHY layer parameters leads to increase the design complexity and reduce the proportion of the desired objectives achievement. Moreover, the burst allocation algorithm that can collect all of the objectives of the MAC layer general scheduler in a single solution can be called as the 100% optimal solution; otherwise it is a relatively optimal solution. On the other hand, the optimal solution is very difficult or possibly to say it is impossible to achieve, because some of the objectives are contradictory to each other. Thus the proposed algorithms in this field are relative solution that depends on the achieved objectives in the individual category of the general scheduler.

## References

1. Amendment to IEEE Standard for Local and Metropolitan Area Networks. In: IEEE Std P802.16e-Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems—Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands: IEEE Group (2005)
2. Xiao, Y.: WiMAX/MobileFi: advanced research and technology. Taylor & Francis Group, New York (2008)

3. QoS Scheduling in WiMAX Networks,  
<http://wimax-made-simple.blogspot.com/2010/09/qos-scheduling-in-wimax-networks.html>
4. Awad, M.: Resource Allocation for Broadband Wireless Access Networks with Imperfect CSI. University of Waterloo (2009)
5. Jeffrey, J.G.A., Andrews, G., Ghosh, A., Muhamed, R.: Fundamentals of WiMAX. Prentice Hall, United States (2007)
6. Sadr, S., et al.: Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems. IEEE Communications Surveys & Tutorials 11, 92–106 (2009)
7. Tang, S.-Y., Muller, P., Sharif, H.R.: WiMAX Security and quality of Service an End-to-End Perspective. John Wiley & Sons Ltd. (2010)
8. So-In, C., Jain, R., Tamimi, A.-K.: Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey. IEEE Journal on Selected Areas in Communications 27, 156–171 (2009)
9. Shabani, A.M.H., et al.: Survey of Down Link Data Allocation Algorithms in IEEE 802.16 WiMAX. International Journal of Distributed and Parallel System 3 (2012)
10. IEEE Std 802.16<sup>TM</sup> for Local and metropolitan area networks. In: Part 16: Air Interface for Broadband Wireless Access Systems. IEEE 3 Park Avenue New York, NY 10016-5997. IEEE Std., USA (2009)
11. WiMAXForum. WiMAX<sup>TM</sup> System Evaluation Methodology V2.1. WiMAX Forum®, USA (2008)
12. M.T.C.S.: JIS, Computers and intractability A Guide to the Theory of NP-Completeness (1979)
13. Israeli, A., et al.: On the complexity of sequential rectangle placement in IEEE 802.16/WiMAX systems. Information and Computation 206, 1334–1345 (2008)
14. Zubow, A., et al.: Greedy scheduling algorithm (GSA) – Design and evaluation of an efficient and flexible WiMAX OFDMA scheduling solution. Computer Networks 54, 1584–1606 (2010)
15. WiMAX-Forum. Service Recommendations to Support Technology Neutral Allocations. In: FDD/TDD Coexistence (2007)

# Hybrid Trust Framework for Loss of Control in Cloud Computing

Tang Jia Li and Manmeet Mahinderjit Singh

School of Computer Science,  
Universiti Sains Malaysia,  
Penang, Malaysia

tjl1104072@student.usm.my, manmeet@cs.usm.my

**Abstract.** Today cloud computing is prone to security and privacy threats. Decrement in security within clouds contributes to a loss of clients-confidence, lack of control and loss of trustworthiness in cloud service providers. Loss of control occurs when clients lose their control over their own resources in the hand of service provider. As lack of authentication and access control placed by providers, loss of control contributes to greater security concern. In this research, a novel trust framework is proposed which focus on tackling the loss of central concern by granting back control to the cloud clients in ensuring their service provider environments fully secure and trustworthy. Secondly, the trust model also functions in ensuring that both client and service providers transaction follows set of security mechanism in place. The results proved that both trust need to be enforced is any cloud. Thus the proposed trust model contributes as a mean to solve the security challenges known as loss of control.

**Keywords:** Cloud computing, Trust, Loss of Control, Authentication, Access control.

## 1 Introduction

Cloud computing is a model to a shared pool of configurable computing resources for network access enable convenient and minimize service provider interaction (Dillon et al. 2010). This is because the cloud computing has plenty advantages such as on-demand self-service, high commonality, broad network access, low cost and more (Dillon et al. 2010 and Wang, 2011). Issues and challenges of cloud computing are security, privacy, charging model, service level agreement, and more (Dillon et al. 2010 and Lar et al. 2011). In this research, the main concern is security issues because the system can easily attack by hackers, data loss, phishing and more (Dillon et al. 2010).

This research mainly focuses on loss of control. Many users are aware of the danger of letting data control out of their hands and storing data with other cloud provider. There are four challenges in dealing with loss of control within cloud computing. Among others are 1) user have multiple account associated with multiple service provider; 2) undesirable mapping of identity to a user; 3) accessing of services by the user once authenticated is complicated with multiple accounts and authenticate could

lead to deception and 4) sharing some of sensitive identity information between services can lead to the undesirable mapping of identity to the user. This is a very sensitive function because users are often stored and processing sensitive data using Cloud Computing services. Loss of control decreases the trustworthiness in cloud computing technology because users are not fully confident in using cloud especially in term of the security properties offered within the technology.

Trust in cloud computing (Ko et al. 2011) is the confidence of users in using the cloud and increase trust by mitigating technical and psychological to using cloud computing. The gap of trust management in cloud computing shows that there is lack of hybrid trust (both hard trust and soft trust) employed together. Thus the aim of this paper is to propose a hybrid trust framework for cloud computing environment. In addition, the second aim would be to tackle the issue of loss of control that occurs in cloud computing by employing the trust approach.

The objectives of this paper are 1) to propose a trusted authentication and access control approach and 2) to calculate the overall trust score and recommendation. The main contribution is to proof the concept of the importance of hybrid trust in cloud computing. This paper constructs as follows. Section 2 discusses the background. Section 3 introduces our proposed method. Section 4 illustrates extended trust cloud framework. Section 5 provides evaluation and discussion. Section 6 the conclusions.

## **2 Background – Authentication, Access Control and Trust Cloud in Cloud Computing**

### **2.1 Authentication**

Authentication (Winkler 2011) is used to synchronize identity information with the enterprise. Federated Identity Management (FIM) (RE-SEARCHER 2012 and Jensen 2012) is a secure way to make up an identity in cloud computing. However, federated identity uses claim-based token model that needs to be supported by a federated token model (Winkler 2011). Lum and Brandon (2010) had created a framework that lets companies easily use of Zero-Knowledge Authentication which allows for secure login without transmitting the password or hash over the network. The zero knowledge protocol can also be used for authentication. It is a popular concept in cryptography systems (Lum and Brandon 2010).

### **2.2 Access Control**

The most common access control models (Winkler 2011) in cloud computing are discretionary access control (DAC), role based access control (RBAC) and mandatory access control (MAC). Role Based Access Control (RBAC) is used in cloud computing. The three main components of RBAC are profiled, roles and authorizations. For example, apply RBAC to work out Cloud User Role Assignment (CURA) & the Role Permission Assignment (RPA), which is called Cloud based RBAC. The component of Cloud based RBAC are cloud user, access permission, role and session (Wang 2011).

### 2.3 Trust Cloud

Trust concept in cloud (Ko et al. 2011) is the confidence of users in utilising cloud computing by mitigating technical and psychological factors. There will be five types of trust metric (Zhang et al. 2011) which are binary state, scaled, probability, hybrid or multi-metric trust.

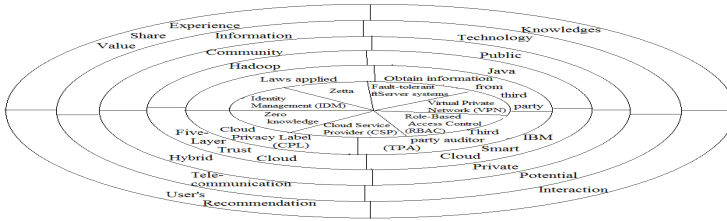
There are two different trusts which are soft trust and hard trust (Lin and Varadharajan 2007). Lin and Varadharajan (2007) had mentioned that hard trust is represents the trust relationships which derived from cryptography based on security mechanisms and soft trust is based on trust relationships through social control mechanisms to derive from the localized and external of system behavior. Among trust systems for cloud computing are peer trust (Firdhous et al. 2011), eigen trust (Firdhous et al. 2011), CuboidTrust (Firdhous et al. 2011). Mahinderjit-Singh and Li (2009) had proposed seven-layer trust framework to tackle security and privacy concern within Radio Frequency Identification (RFID) enabled supply chain which is employ both soft and hard trust. The framework consist of attributes such as authenticity (layer 1), privacy (layer 2), data level (layer 3), detection (layer 4), monitoring as auditing (layer 5), and experiences (layer 6 and 7).

### 3 Proposed Trust Model – User-Centric Trust Model

In this research, we propose a user-centric trust model by adding feedback score and score rating by the user. The trust model allows client to rate the cloud provider. This practice ensures that the cloud provider follows certain security measurements in hand. We have a client and server on each side. First client will be authenticated and authorized in order to utilize the resources in the cloud provider. Once the client had access the resource, the client will then rate the service based on both soft and hard trust parameters. In addition, cloud provider will also rate the system accordingly. Finally, both soft trust and hard trust will be evaluated against trust threshold and a recommendation value will be stored as well. The end result is either the system is trusted or not. The trust model will be calculated by combining both soft trust and hard trust components. The hard trust value will use absolute value or modulus operators, so all hard trust value will be positive score or non-negative score.

### 4 Proposed Trust Framework-Hybrid Trust Cloud Computing Framework

We have transformed the existing seven layer trust framework designed by Mahinderjit-Singh and Li (2009) for security and privacy threat such as loss of control issue in cloud computing. Six layer trust framework for cloud computing is introduced as Fig. 3.



**Fig. 1.** Hybrid Trust Cloud Computing Framework

There are also six relevant layers of this framework. We will explain each of the layers in-depth. First and second layer are known as physical level core function include the security layer and privacy layer. Security layer consists of security services and other security challenges within cloud such as availability, confidentiality, data integrity, access control, audit, loss of control and authentication. The solution for authentication is zero knowledge. The access control challenge solution is Role-Based Access Control (RBAC).

Layer 2 which is Privacy layer consist the solution to some privacy concern within a cloud application involving context such as legal issues, multi-location issues, user control and unauthorized secondary usages. Next two layers are service core functions which consist of middleware level and deployment layers. Layer 3 which stand for some existing open-source software framework for cloud computing implementation such as Hadoop which execute under Apache. Layer 4 which is deployment model layer consists of being private cloud, public cloud, hybrid cloud and community cloud.

Finally, the third core is the application-based core which includes layer 5 (community based and system policy) and layer 6 (User experience and Feedback layers). Social aspect such as community beliefs and attitude towards a technology such as cloud computing can contribute to positive shared experience in the next layer of business decision. Layer 7 which is the feedback layer aim to evaluate users and providers usage experience involves context such as user interaction and knowledge feedback.

## 5 Evaluation and Discussion

In order to evaluate our proposed trust model, case studies scenarios would be applied. Cases would be classified as trustworthy or not trustworthy based on the value of soft trust and hard trust. The system was implemented using Java. The system is used to support in the proving the concept of both hard and soft trust importance. For the cloud system, assumption is made that both authentication and access control is in place within the system. The main page of the system consists of a login page. The user logs into the system with options such as 1) with authentication and access control (full capability); 2) with authentication and no access control (partial capability or 3) without authentication and without access control (no security). The third option is not allowed in the system as no security mechanism is disallowed in any cloud system. The user also can choose to view the trust score result. After login into the system, a page showing the resources offered by the provider will be shown. The



resource page offers a pool of application files. A user is authorized to either view the files or view and write on the application. This is based on the role-based authorization designed by the client of the cloud computing. After viewing or view and copy the resource, the user need to click trust rating button to give a trust score rate for the server and also trust rating for the client side. In the trust rating server page there will be three rating criteria which are service, availability and timely. After the user had chosen the score for these three criteria, the user could then click the submit button to submit the result. Within the trust rating client page there are two rating criteria; 1) user rating and 2) successful transaction. When the user chooses the score for user rate and successful transaction the recommended score will be automatically calculated. Finally, the users can log out of the system. The sample size for the experiment is 80 users. Even though the size of the sample is small, it is sufficient enough to show the proof concept aim to demonstrate the importance of hybrid trust and the importance to tackle the loss of control.

### 5.1 Results and Evaluations

In this section we will demonstrate on how the evaluation is done by using the user-centric trust model. There are five different types of cases are applied in the trust analysis (Table 3). Three cases were recorded as trustworthy and one case is not trustworthy. Then the other case is depending on user rating so it can be trustworthy or not trustworthy. The result is further discussed in section 5.2.

**Table 1.** Case Studies for User-centric Trust Model

Scenario	Description	Result
Case 1	<ul style="list-style-type: none"> <li>Hard trust: User access system with full authentication + full access control (Read + Write). So hard trust is 1.</li> <li>Soft trust: Feedbacks from users are according to their own rating while using the services. Then soft trust is 1.</li> </ul>	<ul style="list-style-type: none"> <li>Trust score <math>\geq 0.75</math>.</li> <li>Trustworthy</li> </ul>
Case 2	<ul style="list-style-type: none"> <li>Hard trust: User access system with full authentication + partial access control (Read). So hard trust is 0.</li> <li>Soft trust: Feedbacks from users are according to their own rating while using the services. Then soft trust is 0.</li> </ul>	<ul style="list-style-type: none"> <li>Trust score <math>&lt; 0.75</math>.</li> <li>Not trustworthy</li> </ul>
Case 3	<ul style="list-style-type: none"> <li>Hard trust: User access system with full authentication + full access control (Read + Write). So hard trust is 1.</li> <li>Soft trust: Feedbacks from users are according to their own rating while using the services. Then soft trust is 0.</li> </ul>	<ul style="list-style-type: none"> <li>Trust score <math>\geq 0.75</math>.</li> <li>Trustworthy</li> </ul>
Case 4	<ul style="list-style-type: none"> <li>Hard trust: User access system with full authentication + partial access control (Read). So hard trust is 0.</li> <li>Soft trust: Feedbacks from users are according to their own rating while using the services. Then soft trust is 1.</li> </ul>	<ul style="list-style-type: none"> <li>Trust score <math>\geq 0.75</math>.</li> <li>Trustworthy</li> </ul>
Case 5	<ul style="list-style-type: none"> <li>Hard trust: Some user access system with full authentication + full access control (Read + Write). So hard trust is 1. Or some user access system with full authentication + partial access control (Read). So hard trust is 0.</li> <li>Soft trust: Feedbacks from users are according to their own rating while using the services. Then soft trust can be 0 or 1.</li> </ul>	<ul style="list-style-type: none"> <li>Trust score <math>\geq 0.75</math> or <math>&lt; 0.75</math>.</li> <li>Trustworthy or Not trustworthy</li> </ul>

## 5.2 Discussions

The system is trustworthy when both or either hard trust or soft trust scored as one. Hard trust will show a score one when user authenticate with authentication and access control. The system is not trustworthy when both hard trust and soft trust are scored zero. Hard trust scored zero when the user authenticates with authentication and partial access control. However, the soft trust it rated by the user themselves which is feedback score. If hard trust and soft trust score record 0/1 as score, then the result is either trustworthy or not trustworthy. There will only two possibilities shown the system which is not trustworthy. First possibility is the scores for both hard trust and soft trust are zero. The other possibility is hard trust and soft trust scores are 0/1 score.

The six layers trust cloud framework has both hard trust and soft trust as core function levels for those layers. Thus the combination of hard trust and soft trust core could determine the system trust or distrust. Hybrid trust cloud framework is a solution to increase trustworthiness for user core functions at three main levels. In addition this hybrid trust is a means to tackle the loss of control. By returning back the control to the users in evaluating the services provided by the cloud provider, a more controllable and manageable cloud computing transaction is achieved.

## 6 Conclusion and Future Work

As a conclusion, the main contribution of this research is a user-centric trust model to tackle the loss of control. We have also proposed a hybrid trust framework for cloud computing environment aim to increase user trustworthiness in using cloud services. Trust is essential to encourage more users to use cloud computing because it will increase the confidence and safety for the users. By implementing the trust model and proposing the hybrid trust framework, we have tackled the loss of control issue in cloud computing.

Among the future work, we will employ the trust framework for other security and privacy threat for cloud computing such as confidentiality, availability, data integrity and more. In addition, this framework could also be implemented in real-life cloud environments supporting any type of applications regardless its security requirement level.

## References

1. Dillon, T., et al.: Cloud Computing: Issues and Challenges. In: 24th IEEE International Conference on Advanced Information Networking and Applications, Perth, WA (2010)
2. Wang, Z.: Security and privacy issues within the Cloud Computing. In: International Conference on Computational and Information Sciences, Chengdu, China (2011)
3. Lar, S.-U., et al.: Cloud Computing Privacy & Security Global Issues, Challenges, & Mechanisms. In: 6th International ICST Conference on Communications and Networking in China (CHINACOM 2011), Harbin (2011)
4. Ko, R.K.L., et al.: TrustCloud: A Framework for Accountability and Trust in Cloud Computing. In: World Congress on Services 2011, Washington, DC, pp. 584–588 (2011)

5. Winkler, V.(J.R.): Data security in cloud computing - Part 3: Cloud data protection methods,  
<http://www.eetimes.com/design/embedded-internet-design/4218591/Data-security-in-cloud-computing-Part-3-Cloud-data-protection-methods?pageNumber=0> (accessed August 8, 2011)
6. Re-Searcher. Federated Identity Management in Cloud Computing. Researcher's Blog,  
<http://clean-clouds.com/2012/04/25/federated-identity-management-in-cloud-computing-2/>  
(accessed April 25, 2012)
7. Jensen, J.: Federated Identity Management Challenges. In: Seventh International Conference on Availability, Reliability and Security (2012)
8. Lum, J.J., Brandon: Implementing Zero-Knowledge Authentication with Zero Knowledge (ZKA\_wzk), p. 9 (2010)
9. Lin, C., Varadharajan, V.: A Hybrid Trust Model for Enhancing Security in Distributed Systems. In: The Second International Conference on Availability, Reliability and Security, ARES 2007, pp. 35–42 (2007)
10. Firdhous, M., et al.: Trust Management in Cloud Computing: A Critical Review. International Journal on Advances in ICT for Emerging Regions 04(02), 24–36 (2011)
11. Mahinderjit-Singh, M., Li, X.: Computational Model for Trust Management in RFID Supply Chains. In: IEEE 6th International Conference on Mobile Adhoc and Sensor Systems, MASS 2009, Macau, pp. 734–740 (2009)
12. Zhang, P., et al.: Survey of Trust Management on Various Networks. In: 2011 International Conference on Complex, Intelligent, and Software Intensive Systems, Seoul, pp. 219–226 (2011)

# Secure Authorized Proxy Signature Scheme for Value-Added Service in Vehicular Ad Hoc Networks

Iuon-Chang Lin<sup>1,2</sup>, Chen-Hsiang Chen<sup>1</sup>, and Pham Thuy Linh<sup>1</sup>

<sup>1</sup> Department of Management Information Systems,  
National Chung Hsing University

<sup>2</sup> Department of Photonics and Communication Engineering, Asia University  
iclin@nchu.edu.tw, jameskoyou@hotmail.com,  
phamlinhhc@gmail.com

**Abstract.** In recent years, value-added services in vehicular ad hoc networks (VANETs) have become more popular. As it brings more convenient and interesting uses for users, there still are security problems that need to be addressed and solved. In this paper, our proposed scheme solves the bottlenecking problem at the service provider (SP) which reduces the verification speed. In addition, our scheme not only provides basic security requirements such as “mutual authentication and message integrity”, “conditional anonymity”, and “traceability and revocation”, but also lets the scheme be more scalable in VANETs.

**Keywords:** Vehicular ad hoc networks (VANETs), Proxy signature, Conditional privacy, Mutual authentication, Security.

## 1 Introduction

The uses of vehicular ad hoc networks (VANETs) have developed in recent years. Traditionally, VANETs have been used for road-safety broadcasting. For example, if there is an emergency on the road where an ambulance is travelling, the ambulance can broadcast a message to a road-side unit (RSU) to control the traffic light and notify other vehicles that there has been an accident. Recently, business and entertainment owners have realized the practical importance of VANETs; therefore, the applications of VANETs have changed.

There are basically three components in the VANETs: trust authority (TA) or service provider (SP), on-board unit (OBU) and road-side unit (RSU). The TA (SP) is a unit which is in charge of deployment of RSUs and registration of legal vehicles. OBU is a device in the vehicle where applications are installed, after registration with the TA, the OBU will receive the messages from the RSUs or other vehicle's OBUs, and relays messages to the above units. RSU is an infrastructure which is deployed by TA. RSUs are in charge of data exchange between OBUs and external Internet. RSUs have storage and computational capability. When the OBUs are moving at high speed, due to the frequent data exchange from one RSU to another; therefore, the RSUs have to solve the rapid handoff requirement.

Since the VANETs' signal are transmitted by air, there are existing threats in wireless networks, such as eavesdropping, man-in-the-middle, and replay attacks which are harmful to VANETs. Based on the attacks which are malicious to VANETs, the design scheme of VANETs should be secure and to preserve the user's privacy. The followings are the security requirements for the aforesaid [2, 4, 8, 9, 10]:

- (1). Mutual authentication and message integrity: When receiving messages from OBUs, RSUs should check whether the sender is legal or not, and also check whether the data is completely void of modified content. Moreover, vice versa with OBUs when receiving messages from RSU or other vehicles as well.
- (2). Conditional anonymity: To protect user's privacy, including from attackers, nobody can inform who the real sender is. But if the sender is malicious, the TA can trace the malicious user and find out his/her identity.
- (3). Traceability and revocation: When a malicious attack happened, the TA can trace the malicious user and revoke his/her identity.

In addition, scalability is another issue for VANETs. During heavy traffic and rush hours, there can be several cars in the same area. If every vehicle simultaneously requests service from the service provider in the same time period, the service provider will encounter a verification speed reduction, a bottlenecking problem. In Li et al. scheme, all requests of value-added services first verified by the SP, it will cause a bottleneck at the SP. In Yeh et al.'s scheme [17], they proposed an enhanced communication protocol, which is avoiding the communication with the SP. In Choi et al.'s scheme [6], they used a delegation of authorities to solve the scalability problem, although it is applied for road-safety messages, still causes a bottlenecking problem at the TA.

In this paper, we propose a secure authorized proxy signature scheme for VANETs. Our scheme can be applied to value-added services, achieve above mentioned requirements, and reach scalability.

## 2 Related Works

Li et al. provided a secure and efficient communication scheme [13] with privacy preservation (SECSPP) for non-safety applications in VANETs. They use Chaum et al.'s blind signature scheme [7] to design SECSPP. SECSPP contains two phases: access authorization phase and access service phase. During the access authorization phase, the vehicle user receives authorized credential, which contains information about service list from the service provider. The access service phase is when the vehicle user requests service using the authorized credential received during the access authorization phase and receives service via an RSU. Yeh et al. [17] pointed out that only a single service provider may verify authorized credentials; therefore, there will has a bottlenecking or Distribute/Denial-of-service (D/DoS) attack may occur. Yeh et al. enhanced SECSPP and proposed a portable privacy-preserving authentication and access control protocol (PAACP), which eliminates the communications between RSUs and service providers. Choi et al. proposed secure and efficient protocol [6] with privacy preservation, which provided scalability for VANETs.

A delegator could enable another user to impersonate a delegator by leaking some credential information. Therefore, the TA or SP should allow RSUs to verify the vehicle as well. While there is still a bottlenecking of information, when users are using the service in the same time interval, they will also request a delegated key from TA or SP at the same time. Although Choi et al.'s scheme is deployed for the road-safety messages, we think it is useful for value-added service as well.

### 3 Proposed Scheme

#### 3.1 System Model

In our system, the SP may provide various services to vehicle users. In order to satisfy users' requirements, the SP provides a time interval service for the users. For example, one week or one month service for users. The SP sets a deadline for users who want to buy service and generates pseudo-id for the user. Until the deadline is due, the SP will continue to collect users' identities and generate a warrant containing all users who request to purchase the same time interval service. In order to avoid bottlenecking at the SP, SP will assign certificates to vehicle users and RSUs; therefore, vehicle users can request services directly with the RSUs.

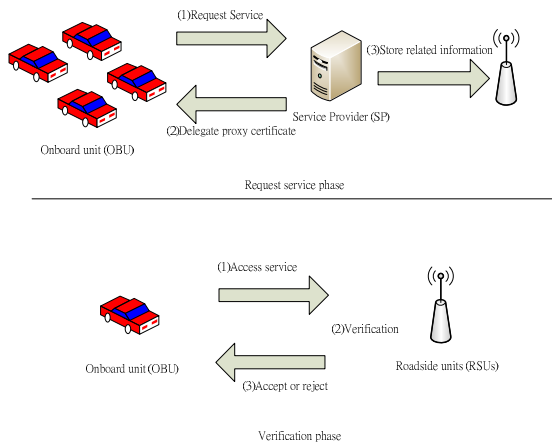


Fig. 1. System model of our scheme

Our proposed scheme for the system model has two phases: request service phase and verification phase. First, each vehicle user must request a certification in request phase. After receiving certification, vehicle users can use certificate to access service to neighboring RSUs. RSUs can verify whether the user is legal or not. Once the vehicle's user is authorized, the RSUs delivers service to the authorized user; otherwise, the RSU rejects the request and sends report to the SP. The SP then can trace the malicious user's identity and thereby revoking it. Fig. 1 shows our system model.

### 3.2 System Parameters

The system parameters are defined as follow:

- $p$  is a large prime,  $q$  is a prime factor of  $p-1$ ,  $g$  is an element of order in  $Z_p^*$ ;
- $priv_{V_i}$  is vehicle user  $V_i$ 's private key and public key is  $pub_{V_i}$ ,  $pub_{V_i} = g^{priv_{V_i}} \pmod{p}$ ;
- $(pri_R, pub_R)$  are the private/public key pair that the  $SP$  generates for each  $RSU$ ;
- $(pri_O, pub_O)$  are the private/public key pair of the  $SP$ ;
- $h$  is an one-way hash function;
- $m_w$  is a warrant, it records some information about the  $SP$ , and all of the vehicle users of the group;
- $PerID$  is period id, it presents the service period when the vehicle users want to use the service. For example, if users want to use the service during June 1<sup>st</sup> to June 30<sup>th</sup>, the  $PerID$  is 06010630;
- $PID$  is a pseudo id, it provides user anonymity to prevent user from revealing their true identity;
- $SK$  is a session key, vehicle users and RSUs can use it to communicate or exchange data.

### 3.3 Request Service Phase

In this phase, vehicle users want to buy a service during a period, the  $SP$  sets a deadline and collects user identities that they want to buy in the same period. After receiving identities from users, the  $SP$  generates pseudo ids, warrants, and certificates to vehicle users.

Step 1: Vehicle user  $V_i$  requests a service with the  $SP$ , sending service request message  $ServReq$ , his/her vehicle id  $VID_i$ , and wanted service id  $SID_l$  to the  $SP$ .

Step 2: There is a deadline set by the  $SP$ , users who request to use the same period service, they will send request before deadline. After receiving the request from the  $V_i$ , the  $SP$  generates a pseudo-id  $PID_i$  and sets  $PerID_i$  for every  $V_i$ . The  $SP$  will then collect  $VID_i$  from users until the deadline is due. Therefore, there are  $i$  users ( $i=1, \dots, n$ ) desiring to request service in the same period. The  $SP$  will records all users'  $PID_i$  in  $m_w$ , storing  $PID_i$  and  $VID_i$  in the  $SP$ 's database. At last, the  $SP$  sends  $PerID_i$ ,  $PID_i$ , and  $m_w$  back to the  $V_i$ .

Step 3: After receiving  $PerID_i$ ,  $PID_i$ , and  $m_w$  from the  $SP$ , each  $V_i$  uses  $m_w$  information and generates function  $f_i(X) = pri_{V_i} + a_{i0} + a_{i1} \cdot X \pmod{p}$ , where  $a_{i0}$  and  $a_{i1}$  are random numbers, and  $i=1, \dots, n$ . The  $V_i$  computes  $y_{i,1} = f_i(PID_1) = pri_{V_i} + a_{i0} + a_{i1} \cdot PID_1 \pmod{p}$ , ...,  $y_{i,n} = f_i(PID_n) = pri_{V_i} + a_{i0} + a_{i1} \cdot PID_n \pmod{p}$ . Then  $V_i$  computes  $A_{i0} = g^{a_{i0}} \pmod{p}$ ,  $A_{i1} = g^{a_{i1}} \pmod{p}$ , and  $Y_i = (y_{i,1} \parallel \dots \parallel y_{i,n})$ . Finally,  $V_i$  sends  $(Y_i, A_{i0}, A_{i1}, SID_l, PerID_i, VID_i, m_w)_{pub_O}$  to the  $SP$ .

Step 4: The  $SP$  collects messages from all  $V_i$  ( $i=1, \dots, n$ ), decrypts all messages  $(Y_i, A_{i0}, A_{i1}, SID_l, PerID_i, VID_i, m_w)_{pub_O}$  by  $SP$ 's private key  $pri_O$ . Then the  $SP$

computes  $y_A = \sum_{\alpha=1}^n y_{i,\alpha} \pmod{p}$ , verifies whether  $g^{y_A} = pub_{V_i}^n \cdot A_{i0}^n \cdot A_{i1}^{(PID_1 + \dots + PID_n)} \pmod{p}$  is legal or not, if equation is correct, the  $V_i$  is a legal user; otherwise, the SP cancels procedure. The SP computes  $s_i = \sum_{\alpha=1}^n y_{\alpha,i} = Z + a_0 + a_1 \cdot PID_i \pmod{p}$  for each period service user, where  $Z = \sum_{i=1}^n pri_{V_i} \pmod{p}$ ,  $a_0 = \sum_{i=1}^n a_{i0} \pmod{p}$ , and  $a_1 = \sum_{i=1}^n a_{i1} \pmod{p}$ . Later, the SP computes parameters  $y_G = g^Z \pmod{p}$ ,  $A_0 = g^{a_0} \pmod{p}$ , and  $A_1 = g^{a_1} \pmod{p}$ . After computing parameters, the SP chooses random number  $k$  and computes  $K = g^{a_1} \pmod{p}$ . The SP generates a proxy certificate  $\sigma = e \cdot pri_0 + k \pmod{q}$ , where  $e = h(m_w, K)$ . The SP uses this certificate  $\sigma$  to compute  $\sigma_i = \sigma + r \cdot PID_i$  for each period service user, where  $r$  is a random number. After computing parameter  $B = g^r \pmod{p}$ , the SP stores  $(y_G, A_0, A_1, e, B, SID_i, PerID_i, m_w)$  in RSU's database and sends  $(\sigma_i, s_i, K, B)_{pub_{V_i}}$  to  $V_i$ .

Step 5: After receiving  $(\sigma_i, s_i, K, B)_{pub_{V_i}}$  from the SP, the  $V_i$  uses his/her private key  $pri_{V_i}$  to decrypt it. The  $V_i$  verifies whether  $g^{\sigma_i} = pub_0^{h(m_w, K)} \cdot K \cdot B^{PID_i} \pmod{p}$ , if equation is not correct, the procedure will be cancelled; otherwise, the  $V_i$  computes share proxy certificate  $\sigma'_i = \sigma_i + s_i \cdot h(m_w, K) \pmod{q}$ .

### 3.4 Verification Phase

In this phase, the  $V_i$  requests to receive service with a neighboring RSU. The RSU only checks the validity of the certification, without communicating with the SP. The  $V_i$  and RSU will agree upon session key to exchange data.

Step 1: The  $V_i$  randomly selects a number  $d$ , computing  $D = g^d \pmod{p}$ . Then the  $V_i$  uses the RSU public key to encrypt  $(SID_i, PID_i, D)$  and sends  $(SID_i, PID_i, D)_{pub_R}$  to RSU.

Step 2: The RSU uses its private key to decrypt  $(SID_i, PID_i, D)_{pub_R}$ , and computes  $U = g^u \pmod{p}$ , where  $u$  is a random number selected by the RSU. Then the RSU computes session key  $SK = D^u \pmod{p}$ . In order to check whether if  $V_i$  is legal or not, the RSU generates function  $f'_R(X) = b_{R0} + b_{R1} \cdot X \pmod{p}$  and computes  $y'_{R,i} = f'_i(PID_i) = b_{R0} + b_{R1} \cdot h(PID_i \parallel U) \pmod{p}$  and parameters  $B_{R0} = g^{b_{R0}} \pmod{p}$ ,  $B_{R1} = g^{b_{R1}} \pmod{p}$ . RSU uses session key  $SK$  to encrypt  $(y'_{R,i}, B_{R0}, B_{R1})$  and sends  $U, (y'_{R,i}, B_{R0}, B_{R1})_{SK}$  to the  $V_i$ .

Step 3: The  $V_i$  uses  $U$  to compute session key  $SK = U^d \pmod{p}$ , decrypting  $(y'_{R,i}, B_{R0}, B_{R1})_{SK}$ . The  $V_i$  verifies  $g^{y'_{R,i}} = B_{R0} \cdot B_{R1}^{h(PID_i \parallel U)} \pmod{p}$  is correct or not, if correct, the  $V_i$  can authenticate the RSU is legal; otherwise, the procedure will be cancelled. Then the  $V_i$  generates function  $f'_i(X) = b_{i0} + b_{i1} \cdot X \pmod{p}$  and computes  $y'_i = f'_i(PID_i) = b_{i0} + b_{i1} \cdot h(PID_i \parallel U) \pmod{p}$ ,  $\beta = y'_i + y'_{R,i} \pmod{p}$ . Next, the  $V_i$  uses his/her share proxy certificate to compute  $\gamma_i = \beta \cdot \psi + \sigma'_i \cdot h(SID_i, PID_i, PerID_i, t) \pmod{q}$ , where  $t$  is a timestamp,  $\psi = g^{b_{i0}} \cdot B_{R0} = B_{i0} \cdot B_{R0} \pmod{p}$ ,  $\pi = g^{b_{i1}} \cdot B_{R1} = B_{i1} \cdot B_{R1} \pmod{p}$ . At last, the  $V_i$  sends  $(\gamma_i, SID_i, PerID_i, \psi, \pi, t)_{SK}$  to the RSU.



Step 4: After decrypting  $(\gamma_i, SID_i, PerID_i, \psi, \pi, t)_{SK}$  by session key  $SK$ , the RSU uses  $SID_i$  and  $PerID_i$  to find  $m_w$  and check timestamp  $t$  is in a legal delay time. Then the RSU verifies  $g^{y_i} = (\psi \cdot \pi^{h(PID_i \| U)}) \psi \times [(pub_o^{h(m_w, K)} \cdot K \cdot B^{PID_i}) \cdot (\gamma_G \cdot A_0 \cdot A_1^{PID_i})^{h(m_w, K)}]^{h(SID_i, PID_i, PerID_i, t)} \pmod{p}$ , if this equation is correct, the RSU uses  $SK$  to provide service; otherwise, the RSU terminates the procedure.

## 4 Security Analysis

In this section, we will analyze our proposed scheme with security requirements.

### 4.1 Mutual Authentication

In request service phase, after receiving message  $(Y_i, A_{i0}, A_{i1}, SID_i, PerID_i, VID_i, m_w)_{pub_o}$  from the  $V_i$ , the SP will use the  $V_i$  public key and  $A_{i0}, A_{i1}$  in message to verify whether  $g^{y^A}$  is correct. Then, when the  $V_i$  receives  $\sigma_i$  from the SP, the  $V_i$  will verify  $g^{\sigma_i}$  by using the SP's public key, if the equation  $g^{\sigma_i} = pub_o^{h(m_w, K)} \cdot K \cdot B^{PID_i} \pmod{p}$  is correct, the  $V_i$  and SP are both authenticated.

In verification phase, after receiving  $U, (y'_{R,i}, B_{R0}, B_{R1})_{SK}$  from the RSU, the  $V_i$  will verify that equation  $g^{y'_{R,i}} = B_{R0} \cdot B_{R1}^{h(PID_i \| U)} \pmod{p}$  is correct or not. The RSU receives  $(\gamma_i, SID_i, PerID_i, \psi, \pi, t)_{SK}$  from the  $V_i$ , verifying  $g^{y_i}$  by using the SP's public key and information in the RSU's database. If correct, both the  $V_i$  and RSU achieve mutual authentication.

### 4.2 Message Integrity

In verification phase, all messages between the  $V_i$  and the RSU are encrypted by session key  $SK$  and the RSU's public key. Therefore, anyone cannot modify the message content without knowledge of session key  $SK$  and the RSU's private key. Consequently, our proposed scheme can achieve message integrity.

### 4.3 Conditional Anonymity

In request service phase, the SP will set  $PID_i$  for the  $V_i$ . No one will know the identity of the user with knowledge of  $PID_i$ .  $PID_i$  will expire in a period; therefore, if the  $V_i$  want to request the same service after period end,  $PID_i$  will be changed by the SP.

### 4.4 Traceability and Revocability

If the RSU finds that  $PID_i$  is a malicious user, reporting this situation to the SP. The SP will use  $PID_i$  and check database to find matching  $VID_i$ . Therefore, the SP can trace the real user of  $PID_i$  and revoke his/her right to protect system security.

## 4.5 Comparisons

We will compare security requirements in this subsection. PAACP and SECSPP provide authentication for non-safety service. Although Choi et al.'s scheme is designed for road-safety messages, our scheme takes advantages of their method and makes securer protocol for value-added applications. The following is the comparison between our and related schemes.

**Table 1.** Comparison with related schemes

	Our scheme	PAACP [17]	SECSPP [13]	Choi et al.'s scheme [6]
Mutual authentication	Yes	Yes	Yes	Yes
Message integrity	Yes	Yes	Yes	Yes
Conditional privacy	Yes	NO	NO	Yes
Traceability and revocability	Yes	NO	NO	Yes
Scalability	Yes	Yes	NO	NO

## 5 Conclusion

In this article, we provide a useful protocol for value-added applications. No matter whether it is rush-hour or normal traffic, our scheme uses a proxy signature and it can protect service providers from bottlenecking. The basic security requirements that our scheme can achieve are mutual authentication, conditional anonymity, traceability and revocability. With the rapid growth of wireless networks, vehicular ad hoc networks not only provide road-safety messages but more and more value-added service to user. Therefore, our scheme can provide great scalability for VANETs.

## References

1. Hao, Y., Cheng, Y., Ren, K.: Distribute key management with protection against RSU compromise in group signature based VANETs. In: Proceedings of the IEEE GLOBECOM 2008, pp. 1–5 (2008)
2. Studer, A., Shi, E., Bai, F., Perrig, A.: TACKing together efficient authentication, revocation, and privacy in VANETs. In: Proceedings of the IEEE SECON 2009, pp. 1–9 (2009)
3. Pointcheval, D., Stern, J.: Security arguments for digital signature and blind signature. *Journal of Cryptography* 769 13(3), 361–396 (2000)
4. Lu, R., Lin, X., Zhu, H., Ho, P.H., Shen, X.: ECPP: efficient conditional privacy preservation protocol for secure vehicular communication. In: Proceedings of the 27th IEEE Communications Society Conference on Computer Communications (INFOCOM 2008), pp. 1903–1911 (2008)
5. Lu, R., Dong, X., Cao, Z.: Designing efficient proxy signature scheme for communication. *Science in China Series F*, vol. 51(2), pp. 183–195 (2008)

6. Choi, H.K., Kim, I.H., Yoo, J.C.: Secure and efficient protocol for vehicular ad hoc network with privacy preservation. *EURASIP Journal on Wireless Communication and Networking* 2011 (2011)
7. Chaum, D.: Blind signature systems. In: *Proceedings of Advances in Crypto 1983*, New York, USA, p. 153 (1983)
8. Leinmüller, T., Schoch, E., Maihöfer, C.: Security requirements and solution concepts in vehicular ad hoc networks. In: *Proceedings of the 4th Annual Conference on Wireless on Demand Network Systems and Services*, pp. 84–91 (2007)
9. Raya, M., Hubaux, J.P.: The security of vehicular ad hoc networks. In: *Proceedings of the 3rd ACM Workshop on Security of Ad Hoc and Sensor Networks*, Alexandria, USA, pp. 11–21 (2005)
10. Jundels, D., Raya, M., Papadimitratos, P., Aad, I., Hubaux, J.P.: Certificate revocation in vehicular ad hoc networks. *Technical LCA-Report-2006-006*, LCA (2006)
11. Tseng, Y.M., Jan, J.K.: ID-based cryptographic schemes using a non-interactive public-key distribution system. In: *Proceedings of the 14th Annual Computer Security Applications Conference (IEEE ACSAC 1998)*, Phoenix, Arizona, pp. 237–243 (1998)
12. Yang, C.C., Tang, Y.L., Wang, R.C., Yang, H.W.: A secure and efficient authentication protocol for anonymous channel in wireless communications. *Applied Mathematics and Computation* 169(2), 1431–1439 (2005)
13. Li, C.T., Hwang, M.S., Chu, Y.P.: A secure efficient communication scheme with authenticated key establishment and privacy preserving for vehicular ad hoc network. *Computer Communications* 31(12), 2803–2814 (2008)
14. Zhang, C., Lin, X., Lu, R., Ho, P.H., Shen, X.: An efficient message authentication scheme for vehicular communications. *IEEE Transactions on Vehicular Technology* 57(6), 3357–3368 (2008)
15. Wang, N.W., Huwang, Y.M., Chen, W.M.: A novel secure communication scheme in vehicular ad hoc networks. *Computer Communications* 31(12), 2827–2837 (2008)
16. Ren, K., Lout, W., Kim, K., Deng, R.: A novel privacy preserving authentication and access control scheme for pervasive computing environments. *IEEE Transactions on Vehicular Technology* 55(4), 1373–1384 (2006)
17. Yeh, L.Y., Chen, Y.C., Huang, J.L.: PAACP: A portable privacy-preserving authentication and access control protocol in vehicular ad hoc networks. *Computer Communications* 34(3), 447–456 (2011)
18. Hwang, M.S., Lin, I.C., Lu, E.J.L.: A secure nonrepudiable threshold proxy signature scheme with known signers. *Informatica* 11(2), 1–8 (2000)
19. Sun, H.M.: An efficient nonrepudiable threshold proxy signature scheme with known signers. *Computer Communications* 22(8), 717–722 (1999)

# The Computer-Aided System for Promoting Judgments of Umpires

Chin-Fa Chen<sup>1</sup>, Chia-Chun Wu<sup>3</sup>, Duc-Tai Dang<sup>2</sup>, and Iuon-Chang Lin<sup>2,4</sup>

<sup>1</sup> Graduate Institute of Sports and Health Management,  
National Chung Hsing University, Taichung, Taiwan

<sup>2</sup> Department of Management Information Systems,  
National Chung Hsing University, Taichung, Taiwan

<sup>3</sup> Department of Industrial Engineering and Management,  
Nation Quemoy University, Taiwan

<sup>4</sup> Department of Photonics and Communication Engineering,  
Asia University, Taichung, Taiwan

{fachen, iclin}@nchu.edu.tw, ccwu0918@gmail.com,  
g101029035@mail.nchu.edu.tw

**Abstract.** Basketball is one of the most popular sports, and it's also popular in Taiwan because of fierce competition among SBL, UBL and HBL. Umpires are the pivots who can let games to proceed smoothly. Basketball games are played in quick progress and changing all the time. As an umpire, being calm and just in such a tense situation can be difficult. Therefore, how to train an experienced umpire is especially important. But till now the check-up system without aid of computer is not suitable for training umpires. For this reason, using the new technique to help training and assessing is necessary. We want to propose a computer-aided system based on video database for promoting judgments of umpire in order not to influence the game results by erroneous judgments and misses of judgment. Training by simulation instead just by statement of situation can let umpires be more familiar to real situations and do better judgments in real games. Also, we will do reliability and validity tests to prove our scheme.

**Keywords:**Initials in Capitals; Separate with Semicolons.

## 1 Introduction

Owing to the development of computer prediction technique, using this technique to test the perceptual skill level of particular person is possible. Many research results point out that there is possibility of using computers to aid in training and testing of umpires.

According to interrelated researches of Adolphe, Vickers, & Laplante, 1997, Franks & Hanvey, 1997 and Tayler, Burwitz, & Davids, 1994, we can find that the training of perceptual skill can make athletes perform better in competitions. As an example, *the predictions of serving balls from badminton beginners* by Tayler,

Burwitz, & Davids, (1994) indicate that the beginners who have been trained by imaging simulation will have much better ability of prediction. Another example, the perceptual training of 12-yard penalty kicks prediction for goalkeepers by Franks & Hanvey, 1997 indicates that the goalkeepers who have been trained by film simulation will do a better job on correctness of predictions.

Besides, there are some differences between theory and reality. For right now, the check-up system is not compatible with computer-aided system. Therefore, how to improve the check-up system and make it compatible with computer science will be found out in this paper. The information system can help umpires to be familiar with the real competition situations, in order to advance umpires' level of perceptual skill and make competitions to be set fairly with no erroneous and miss judgment.

## 2 Methods

There are four steps in this study. In order to establish the computer-aided judging system with image database and conduct the reliability and validity test for this system.

### 2.1 Collect Judgments in Different Situations

The scope contains Super Basketball League in 2008, the preliminary contests of University Basketball League in 96th academic year. We set up two video cameras at each side of half course to record the competition. According to whistling time and situations in video, we can establish rules in non-separated responsibility area and eliminate sound of video in order not to let whistling influence users.

### 2.2 Establish Video Database of Computer-Aided System

By actual situations of judging, we separate judgments into different types. At first, we discuss films of competition with three experts in judging (There are two international umpires and a well experienced one in class A) and decide the films with different view in order to make the database including all types of film. Then, let the three experts check the films in this database with slow motion, to make sure that whether each judgment and its time is right or not.

By programming, we let each chosen user to test this system by mouse. They can select classified or non-classified film (Figure 1). During the playing of game film, the first left-click will be identified as the timing of whistling, and then the system will show gesture pictures to let users choose the first gesture of judgment (violation or foul, Figure 2). If the first choice is correct, then let them choose the second gesture to decide the type of violation or foul (There are 11 kinds of gesture, Figure3). The checking sequence is as following; first, the system will check the timing of whistle (The error time should not be more than 2 seconds), and then check the first gesture and second gesture are correct or not. The system will identify it as an wrong answer if there is any wrong choice in the sequence.

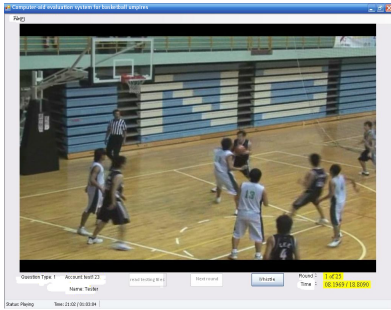


Fig. 1. Major view of the System

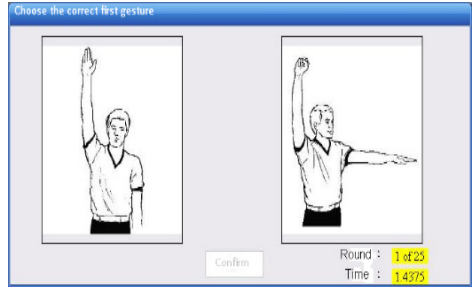


Fig. 2. Choices of first judging gesture

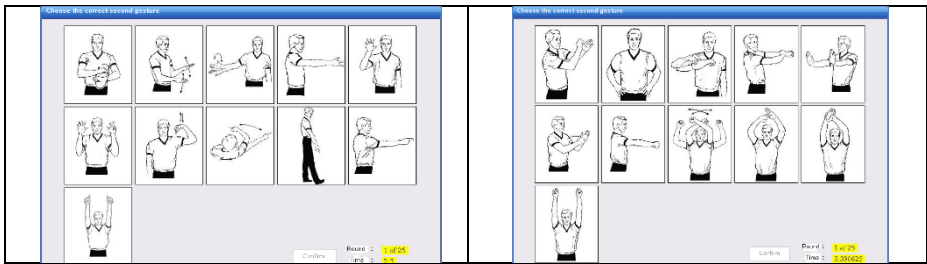


Fig. 3. Choices of second judging gesture (Types of violations and fouls)

Our system is constructed and tested by the platform of Intel Pentium4 4.2GHz CPU, 512MB RAM and Windows 2000/XP Professional operating system. We used Microsoft Visual Basic 6.0 to program the system and used Ulead VideoStudio 9 and Ulead PhotoImpact 10 to edit the competition video. We also need to use MediaPlayer components to control and play the video, so the installation of MediaPlayer Encoder is essential. The database is used to store user accounts, training and testing records in order to estimate and analyze interrelated data. And we chose Microsoft Office Access 2003 or Microsoft SQL Server 2003 to establish the database.

### 2.3 Completion of Computer-Aided Evaluation System for Basketball

By films from different type database collected before, we distributed the sampling rate for each database according to the appearance frequency of each type. Based on our observation of normal competitions; the happen rate of reaching foul, blocking foul, charging foul, holding foul and traveling violation is higher than others. For each unit composed of 25 judging cases, should be composed with 20 cases from the 5 types we mentioned above and other 5 cases sampled randomly.

The computer-aided system makes 25 judging cases to be a unit. It records users' (who are chosen to take the test) first mouse-click time which will be recognized as the whistling time, and then compare with the normal correct time. Besides, it will also record the validity of first gesture and second gesture. Through these acts, we can

comprehend validity of the whistling times and gestures. If the gestures of a judgment are correct, the column of it will be marked as "1". Otherwise, the column will be marked as "0". The system will calculate scores automatically. If user's first gesture and second gesture are all correct, he or she will be given 4 points as the score. Then if only first gesture is correct, scores they will get are reduced to 2 points, and if the first gesture is false, scores they will get will be 0 points, whether second gesture is correct or not. Because there are 25 judging cases in a test, the perfect score is 100 points.

## 2.4 Confidence and Effectiveness Testing of Computer-Aided Evaluation System

Randomly sample 32 domestic basketball umpires as testing object for computer-aided evaluation system. After the "computer-aided evaluation system for basketball umpires" is completed. (Testing was done between April and May in 2008.) We used the system we developed in this research. We installed the "computer-aided evaluation system for basketball umpires" on each computer. Every user (Umpire who is chose to be tested) uses each computer independently. We explain testing procedure and operating functions. Then we provide 3 practices before the testing starts in order to let users know how to proceed. After the test begins, the system will record users' click time and the choices they chose. After the first test is done, let users take 10 minutes break and then do the same test with identical content but in different sequence. Use "Pearson product-moment correlation" to find the coefficient of correlation  $r$ . After the second test is finished, also let users take 10 minutes break and then start a new test with another set of video in the same type. Use "Pearson product-moment correlation" to find the coefficient of correlation  $r$ .

We use "known group difference method" (Thomas and Nelson, 2001, p185.) to check validity of our system. Make the first testing scores as the measured value, and use "Analysis of variance" to see whether there is any difference between groups of different certificate level. If there is a good validity in the test, there will also be noticeable differences of group.

## 3 Testing Procedures

As we described in this paper before, using editing technology and 0.1 second as the time counting unit, we set up the video database for our "computer-aided evaluation system for basketball umpires". In this database, there are several judgment types about violations and fouls. There are totally 358 judging instances from 9 class of competition situations including 47 violations by ball holders (traveling, double dribble), 8 violations about time series (3, 5, 8, 24 seconds), 177 fouls by hands, 88 blocking fouls (offense and defense), 7 fouls by elbow, 8 fouls of charging and 9 fouls of others (double fouls, technical fouls and Unsportsmanlike fouls).

With the video database we set up, users can choose questions from group No. 1, 2 or 3. There are 25 judging instances in each group. After testing, the system records

results and shows total scores automatically in Excel files. Through the system, we can realize evaluation results of users to complete the computer-aided evaluation system.

We install the system on the computers in the college of social science and management. Then we randomly sampled 32 umpires in A, B and C class having umpire certification in Taichung, Chungghwa and Nantou. Let them do questions from group No. 1, 2 and 3 at the same time and take 10 minutes breaks between three groups of questions. There are 25 fixed instances generated from the evaluation system in question group No. 1. Question group No. 2 has the same content as No. 1, but the sequence of No. 2 is different from No. 1. There are another 25 instances generated from the evaluation system in group No. 3. Then we input the results of test that had finished to SPSS10.0 program and use the statics analysis which is called ANOVA to produce descriptive static results from different-level basketball umpires' scores (See Table 1) and the result from variance analysis (See Table 2).

We can find out there are significant differences between different class umpires' results in table 2 from static analysis. Then we did "Scheffe's post hoc tests", and there are the results in table 3.

**Table 1.** Static results from different-level basketball umpires' scores

Question Type	Class	Amount	Mean	Standard deviation	Standard error	95% Confidence Interval of mean		Minimum	Maximum
						Low-bound	High-bound		
1	A	8	85.7500	3.77018	1.33296	82.5980	88.9020	82.00	92.00
	B	11	80.0000	5.44059	1.64040	76.3450	83.6550	72.00	90.00
	C	13	69.6923	8.19944	2.27411	64.7374	74.6472	58.00	82.00
	Total	32	77.2500	9.16867	1.62081	73.9443	80.5557	58.00	92.00
2	A	8	87.0000	4.78091	1.69031	83.0031	90.9969	80.00	94.00
	B	11	82.9091	4.67877	1.41070	79.7658	86.0523	74.00	90.00
	C	13	74.6154	5.12410	1.42117	71.5189	77.7118	68.00	84.00
	Total	32	80.5625	7.06165	1.24834	78.0165	83.1085	68.00	94.00
3	A	8	85.2500	4.52769	1.60078	81.4648	89.0352	78.00	92.00
	B	11	79.6364	5.64398	1.70172	75.8447	83.4280	68.00	88.00
	C	13	68.3077	6.47282	1.79524	64.3962	72.2192	58.00	78.00
	Total	32	76.4375	9.08362	1.60577	73.1625	79.7125	58.00	92.00



**Table 2.** Variance analysis results from different-class basketball umpires' scores

		Sum Square	Freedom	Mean of Sum Square	F	Significant
QT No.1	Between	1403.731	2	701.865	16.930**	.000
	Within	1202.269	29	41.458		
	Total	2606.000	31			
QT No.2	Between	851.889	2	425.944	17.799**	.000
	Within	693.986	29	23.931		
	Total	1545.875	31			
QT No.3	Between	1593.060	2	796.530	23.942**	.000
	Within	964.815	29	33.269		
	Total	2557.875	31			

## 4 Testing Results

According to table 4, we can find out that there is a significant correlation between group No. 1, 2 and 3 from the results of Pearson product-moment correlation. It indicates that there are re-testing reliability and alternate reliability in our system. According to analytic results in table 2 and 3, we find umpires of class A and B significantly score more than umpires of class C, but there is no significant difference between class A and B. We infer that our system possesses great reliability and validity in separating class C and above class B. The lack of samples or indetermination of classification may result in no significant difference between class A and B. An umpire should at least have one year experience or over 20 games refereeing experience before he or she can attend a short-term training of the higher level class. They should be able to pass the test after the training is over. But it can't be avoided that identifiers may be subjective and emotional. There is also a lack of objectively quantifiable bases. We will continue the research to modify this problem. Another way to improve the results is to collect much larger amount of video films and to do reliability and validity tests for each judging instance instead of every 25 judging instances in order to reduce the error of scores.

**Table 3.** Results of Scheffe’s post hoc tests from different-class umpires’ scores (\*\*p < .01)

Question Type	Umpire Class		Standard deviation	Standard error	Significant	95% Confidence Interval	
						Low-bound	High-bound
QT No.1	A	B	5.7500	2.99183	.195	-1.9683	13.4683
		C	16.0577	2.89331	.000	8.5936	23.5218
	B	A	-5.7500	2.99183	.195	-13.4683	1.9683
		C	10.3077	2.63779	.003	3.5028	17.1126
	C	A	-16.0577	2.89331	.000	-23.5218	-8.5936
		B	-10.3077	2.63779	.003	-17.1126	-3.5028
QT No.2	A	B	4.0909	2.27306	.232	-1.7731	9.9549
		C	12.3846	2.19821	.000	6.7137	18.0555
	B	A	-4.0909	2.27306	.232	-9.9549	1.7731
		C	8.2937	2.00407	.001	3.1236	13.4638
	C	A	-12.3846	2.19821	.000	-18.0555	-6.7137
		B	-8.2937	2.00407	.001	-13.4638	-3.1236
QT No.3	A	B	5.6136	2.68015	.151	-1.3006	12.5278
		C	16.9423	2.59189	.000	10.2558	23.6288
	B	A	-5.6136	2.68015	.151	-12.5278	1.3006
		C	11.3287	2.36298	.000	5.2327	17.4247
	C	A	-16.9423	2.59189	.000	-23.6288	-10.2558
		B	-11.3287	2.36298	.000	-17.4247	-5.2327

**Table 4.** Pearson product-moment correlation from different groups of questions (\*\* $p < .01$ )

		QT No.1	QT No.2	QT No.3
QT No.1	Pearson Correlation	1	.820**	.782**
	Significant (two-tailed)	.	.000	.000
	Amount	32	32	32
QT No.2	Pearson Correlation	.820**	1	.829**
	Significant (two-tailed)	.000	.	.000
	Amount	32	32	32
QT No.3	Pearson Correlation	.782**	.829**	1
	Significant (two-tailed)	.000	.000	.
	Amount	32	32	32

## 5 Conclusions

By establishing video database and the computer-aided system, we can provide basketball umpires more practice opportunities to improve their judging skills. It also reduces the cost of each practice because we can simulate an actual game by games in past time. In accordance with results of testing, our system possesses great reliability and validity, so it could be really helpful to evaluate the judgment quality and review. We suggest our system should be added into present short-term training and classification-test. In the future, we will try to let the system connect to the internet, in order to provide evaluations and trainings for users immediately.

## References

1. Adolphe, R.M., Vickers, J.N., Laplante, G.: The effects of training visual attention on gaze behavior and accuracy: A pilot study. *International Journal of Sport Vision* 4, 28–33 (1997)
2. Allard, F., Graham, S., Paarsalu, M.E.: Perception in sport: basketball. *Journal of Sport Psychology* 2(1), 14–21 (1980)
3. Allard, F., Starkes, J.L.: Perception in sport: volleyball. *Journal of Sport Psychology* 2(1), 22–33 (1980)
4. Franks, J.M., Hanvey, T.: Cur for goalkeepers: High-tech methods used to measure penalty shot response. *Soccer Journal*, 30–38 (May/June 1997)
5. Kevin, L., et al.: An exploratory investigation of perceptions of anxiety among basketball officials before, during, and after the content. *Journal of Sport Behavior* 23, 11–19 (2001)
6. Rainey, D., Winterich, D.: Magnitude of stress reported by basketball referees. *Perceptual & Motor Skills* 81, 1241–1242 (1995)

7. Rainey, D., Hardy, L.: Ratings of stress by rugby referees. *Perceptual & Motor Skills* 84, 728–730 (1997)
8. Stewart, M.J., Ellery, P.J.: Amount of psychological stress reported by high school volleyball officials. *Perceptual & Motor Skills* 83, 337–338 (1996)
9. Tayler, M.A., Burwitz, L., Davids, K.: Coaching perceptual strategy in badminton. *Journal of Sport Science* 12, 213 (1994)
10. Taylor, A.H., Daniel, J.V., Leith, L., Burke, R.J.: Perceived stress, psychological burnout and paths to turnover intentions among sport officials. *Journal of Applied Sport Psychology* 2, 84–97 (1990)
11. Thomas, J.R., Nelson, J.K.: *Research methods in physical activity*. Human Kinetics, Champaign (2001)

# A Distributed Model to Analyzed QoS Parameters Performance Improvement for Fixed WiMAX Networks

A.L. Ibrahim<sup>1\*</sup>, A. Md. Said<sup>1</sup>, K. Nisar<sup>2</sup>, Peer Azmat Shah<sup>1</sup>,  
and Abubakar Aminu Mu'azu<sup>1</sup>

<sup>1</sup> Department of Computer & Information Sciences (CIS),  
Universiti Teknologi PETRONAS,  
Bandar Seri Iskanda, 31750,  
Tronoh, Perak, Malaysia

<sup>2</sup> School of Computing InterNet Works Research Laboratory  
UUM College of Arts and Sciences,  
Universiti Utara Malaysia,  
Sintok 06010 Kedah Malaysia  
{Ibrahim\_g01867, abuaminu\_g01797}@utp.edu.my  
abass@petronas.com.my,  
kashif@uum.edu.my,  
peer.azmat@comsats.edu.pk

**Abstract.** Quality of Service (QoS) is an important parameter that is use for evaluating Network performance. In this paper a Distributed model was developed to analyze QoS Parameters Performance Improvement for Fixed Worldwide Interoperability for Microwave Access (WiMAX) Network using parameters that include Delay, Throughput and Application Response Time to enhance the services that are provided to the end users. The designed has been evaluated using the simulation tool OPNET modeler 16.0 and compared with the Existing model which is centralized. The results obtained from the new Distributed model shows significant increase in the network throughput from 12,000bit/sec to 14,000bit/sec, and decrease in delay from 0.012sec to 0.009sec and also decrease in application response time from 0.94bit/sec to 0.69bit/sec in Scenario1. The result of the distributed model enhances the QoS performance.

**Keywords:** QoS parameters, fixed WiMAX, Master-Slave model, Client\_server, Base Station, Subscriber Station, OFDM, and OPNET Modeler

## 1 Introduction

WiMAX is a promising wireless technology based on the IEEE 802.16 standards that provide high-speed and reliable communications in large areas up to 30 miles (50 km). This technology has a target of data transfer rate exceeding 100 Mbps [1]. WiMAX supports various multimedia applications like VoIP, voice conference and Internet based gaming. The IEEE 802.16 technology (WiMAX) is the perfect substitute

---

\* Corresponding author.

for 3G or wireless LAN networks for providing last mile connectivity by radio link because of its high data rates, low cost of deployment and larger coverage area [1]. In view of that WiMAX can be one of the hottest wireless technologies around the world today [2]. WiMAX has two classes which include fixed and mobile WiMAX. Fixed WiMAX is designed to provide fixed and nomadic services, and mobile is designed to provide portable mobile connectivity [3]. Also according to [2] the fixed is for fixed wireless and falls under the IEEE 802.16-2004 standards and the mobile is generally for mobile applications Networks, that is under the 802.16e specification.

In this paper a distributed model was developed using Master-Slave BSs and client-server in BSs-SSs communication to analyze QoS Parameters Performance improvement in Fixed WiMAX Networks to enhance the services that are provided to the end users. The Nearest Neighborhood Algorithms was used for master-slave BSs selections and client-server was used in BSs and SSs communication as discussed in section 2 and 3. Addition of master BSs enhances the performance of QoS with respect to throughput, delay and application response time. Point-to-multipoint connection with multicast transmission based on OFDM techniques was used to propagate the network information. The proposed techniques were evaluated using the simulation tool, OPNET modeler 16.0 and compared with the existing centralize model. This approach could effectively enhance the QoS performance.

## 2 Proposed Distributed Model

OPNET Modeler 16.0 was used for the simulations; OPNET Modeler is a highly sophisticated simulation software package that enables to developed and model communications networks and distributed systems [7]. Also OFDM techniques were used to propagate the networks between central server, BSs and SSs, as OFDM is a multiple carrier transmission technique that is used for high speed bi-directional wireless data communication transmission [8]. The used of OFDM reduced the amount of bandwidth needed for data transmission by producing a compressed multiple modulated carrier [9].

### 2.1 Master-Slave Base Station Selection

Master BSs is selected in the MAC layer on the Central server using Nearest Neighborhood Algorithm as shown in Algorithm 1 below. Each BS from the set of BSs register with the central server, the nearest BSs with network information will be selected to provide network information to the closer slave BS that does not have. If there is any BS with network information closer to slave BS than any other previous nearest Neighbor BS then delete the farthest master BS in the set Nearest Neighbor. IF two or more base stations have the same distance and are in final Nearest Neighbor set then Select Master at random from base stations as describes in Algorithm 1.

### 2.1.1 Algorithm1 Process of Master-Slave BS Selection

```

1. START
2. Each BS  $BS_i$  from set of base stations  $BS = \{BS_1, BS_2, BS_3, \dots, BS_n\}$ 
3. Register with the central server
4. Let  $P = \{P_1, P_2, P_3, \dots, P_m\} \subseteq BS$  be the set of  $m$  Base Stations (Candidates for Master) and  $m < n$ 
5. Input  $Q$ , the Slave Base Station which don't have network information
6. Set  $i \leq K \leq n$ 
7. Set  $i = 1$ 
8. DO
9. {
10.   Compute distance from  $Q$  to  $P_i$ 
11.   IF ( $i \leq K$ ) THEN
12.     Include  $P_i$  in the set of Nearest Neighbors base stations
13.   ELSE
14.     { IF ( $P_i$  is closer to  $Q$  than any other previous Nearest Neighbor base station) THEN
15.       DELETE farthest base station  $P$  in the set Nearest Neighbor
16.     }
17.     Include  $P_i$  in the set of Nearest neighbor base stations
18.   }
19. }
20. WHILE (Nearest Neighbor base station with network information is found)
21. IF (two or more base stations have the same distance and are in final Nearest Neighbor set) THEN
22.   Select Master at random from base stations
23. ELSE
24.   Select the single value as Master from the set
25. END IF
26. END
    
```

### 2.2 Master-Slave Communication

Figure 1 shows the communication between Central Server and Master BSs, Master BSs and Slave BSs and also between BSs and SSs for getting network information, where a Central server will send an advertised existence message, master BS also send existence message, central sever send acknowledgement to master BSs, Master BSs sends network information request , central server sends authentication request, Master BS send authentication replay containing the authentication information, central server process the authentication if the authentication is verified then the central server will sends the network information to the master BS otherwise the network information will be denied . The same communication process will take place between Master BS and Slave BS. But the communication between the SSs and BSs is a client-server communication where a client SSs will send network information request then the other procedure will follow as in the server master BS communication as describes in Algorithm 2.

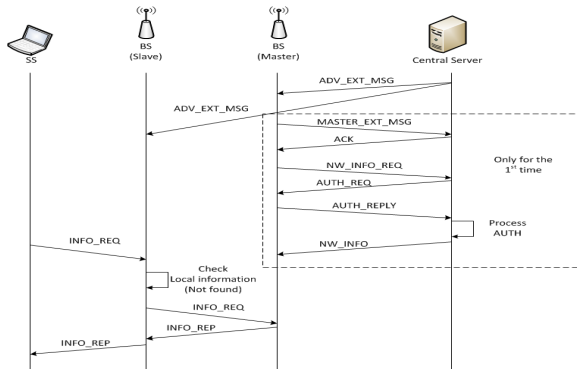


Fig. 1. SSs, Master-Slave BSs and Central Server Communication

### 2.2.1 Algorithm 2 the Process of Master-Slave Communication

```

1.  START
2.  Take input (Master Base Station  $M$ , Slave Base Station  $S_i$ ) from Algorithm_1
3.   $M$  has network information
4.   $M$  advertises its existence through MASTER_ADV message
5.  Each Slave BS  $S_i$  when receives MASTER_ADV message
6.  { IF  $S_i$  don't have network information earlier received Send NW_INF_REQ message to Master  $M$ 
7.  ELSE
8.  }
9.  END IF
10. Master  $M$  sends AUTH_REQ to validate Slave  $S_i$  from which got request
11. Slave  $S_i$  sends AUTH_REP containing the authentication information
12. IF (authentication information is verified) Master  $M$  sends NW_INF to Slave  $S_i$ 
13. ELSE
14.   Master  $M$  sends message that node  $S_i$  is not authenticated
15. END IF
16. END

```

---

## 3 Simulation Setup and Results

A detailed explanation of the simulated network model together with configured traffic that was developed for evaluating the performance of the QoS over fixed WiMAX is given below. Basic parameters associated with WiMAX Configuration attributes, Application Configuration, Application Profile, Task Definition, BSs configuration and Subscribers Station for the proposed Distributed model in fixed WiMAX are configured.

### 3.1 Scenarios for the Existing Model

In the scenario 1 of the existing model, 5 WiMAX BSs were developed with 150 SSs, 30 SSs around each BS. All BSs are connected with IP backbone using point-to-point protocol (ppp), without any master BS. All Basic parameters are configured, while in scenario 2, 10 WiMAX BSs were developed with 300 SSs, thirty SSs around each BS in the subnet without any master BSs. All other parameters are as in scenario\_1 as showed in Figure 2(a) and 2(b).

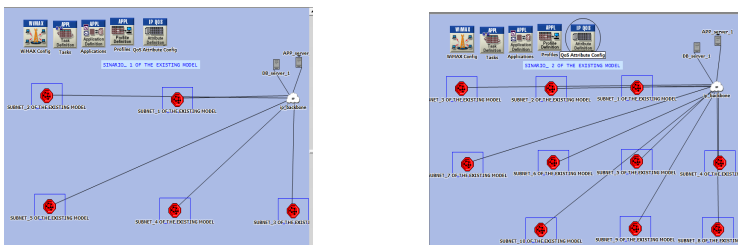


Fig. 2(a) Scenario\_1 for the Existing Model      Fig. 2(b) Scenario\_2 for the Existing Model

### 3.2 Scenarios for New Distributed Model

In scenario 1 of the proposed distributed model, 5 WiMAX BSs were developed with 150 SSs, 30 SSs around each BS in the subnets. All BSs are connected with IP



backbone (Internet) using point- to- point protocol (ppp), with BS in the subnet 1 and 4 as master BS selected by algorithm 1 and the remaining are slaves. Basic parameters are configured, while in scenario\_2, 10 WiMAX BSs were developed with 300 SSs, 30 SSs are around each BS in the subnets with BS in subnet1, 2, 5, 6 and 9 as masters selected by the algorithm1 and the remaining are slaves as shows in figure 3(a) and 3(b) . All other parameters are as in Scenario 1.

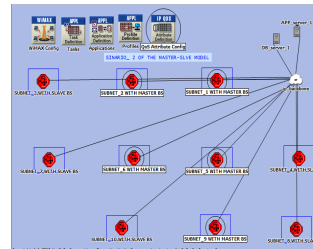
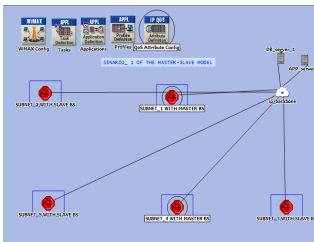


Fig. 3(a) Scenario\_1for the Distributed Model Fig. 3(b) Scenario\_2 for the Distributed Model

### 3.3 Result and Analysis

#### 3.3.1 Application Response Time for Existing and Distributed Model Results

When the design of scenario1 of the existing and distributed model, was simulated, the graph in figure 4(a) and 4(b) was obtained for the Application Response Time.

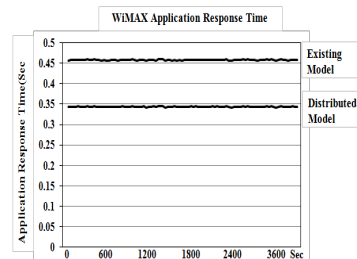
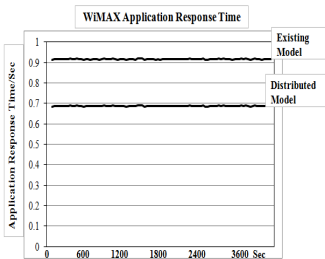
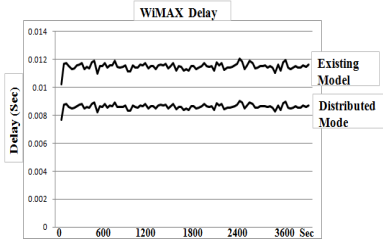


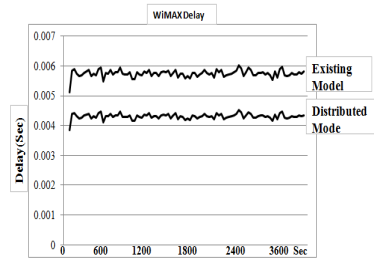
Fig. 4(a) Application Response Time for the Existing and Distributed model (scenario\_1) Fig. 4(b) Application Response Time for the Existing and Distributed model (scenario\_2)

#### 3.3.2 Fixed WiMAX Delay for Existing and Distributed Model Results

Figure 5(a) and 5(b) shows the graph obtained for the delay in the design of scenario\_1 and 2 of the existing and distributed model.



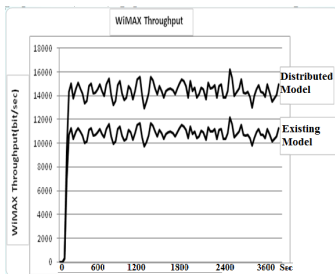
**Fig. 5(a)** Delay for the Existing and Distributed Model (scenario\_1)



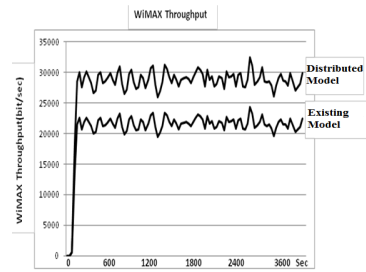
**Fig. 5(b)** Delay for the Existing and Distributed Model (scenario\_2)

**3.3.3 Throughput for Existing and Distributed Model Results**

When the design of scenario\_1 of the existing and distributed model, was simulated, the graph for the Throughput obtained as shown in figure 6(a) and 6(b)



**Fig. 6(a)** Throughput for the Existing and Distributed Model (scenario\_1)



**Fig.6(b)** Throughput for the Existing and Distributed Model (scenario\_2)

**3.4 Analysis of the Result**

If we compare the result obtained from Figures 4(a), 4(b), 5(a), 5(b), 6(a) and 6(b) we can observe that the result of the new distributed model has an improved performance as compared to the Existing model, because the results showed an increase in the network throughput with respect to time from 12,000bit/sec to 14,000bit/sec in the scenario 1 and 24,000bit/sec to 33,000bit/sec in scenario 2. Also the delay reduces from 0.012sec to 0.009sec in scenario 1 and 0.061sec to 0.0043sec in the scenario 2. Lastly the application response time drastic decrease from 0.94sec to 0.69 in scenario 1 and 0.46 to 0.35 in scenario 2 respectively with increasing number of BSs, SSs and the introduction of additional master BSs. Also the decrease occurred as a result on the network environment such as link and bandwidth allocation.

## 4 Conclusion and Future Work

Wireless communication mostly affected with the QoS performance problem, this paper proposed a distributed model to analyze QoS Parameters Performance improvement for Fixed WiMAX Networks to enhance the services that are provided to the end users. Addition of master BSs enhances the performance of QoS with respect to throughput, delay and application response time. The design has been evaluated using the simulation tool OPNET modeler 16.0 and compared with the Existing centralized model. The results obtained from the distributed model shows significant increase in the network throughput and decrease in network delay and application response time. Furthermore, this Model helped the Internet Service providers (ISPs) in terms of data delivery by not operating from one central server, and hence will reduce the cost of infrastructures development.

As future work, the proposed algorithms will be completed with some topology changes to increase the coverage area of WiMAX through distributed master BSs and consequently, may improve the WiMAX QoS performance.

## References

1. Zakhia, A., Yanlin, P., Morris, J.C.: WiMAX, The Emergence of Wireless Broadband. IT Pro. Magazine, 44–48 (2006)
2. IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for
3. Broadband Wireless Access Systems, IEEE Std., 2009 Deploying License-exempt WiMAX Solutions. Technical Report, Intel (2005)
4. Jeffrey, G.A., Arunabha, G., Riaz, M.: Fundamentals of Wimax: understanding broadband wireless networking. Prentice Hall Communications Engineering and Emerging Technologies Series (2007)
5. Yekanlu, E., Joshi, A.: Performance Evaluation of an Uplink Scheduling Algorithm in WiMAX kinge Institute of Technology Sweden (October 2009)
6. Ramachandran, S., Bostian, C.W., Midkiff, S.F.: Performance Evaluation of IEEE 802.16 for Broadband Wireless Access. Opnetwork (2012)
7. Ashjaei, M., Liu, M., Behnam, M., Mifdaoui, A., Almeida, L., Nolte, T.: Worst-Case Delay Analysis of Master-Slave Switched Ethernet Networks: Worst-Case Traversal Time (WCTT), Workshop (January 2012)
8. IEEE 802.16e: IEEE 802.16e Task Group (Mobile Wireless MAN) (2005), <http://www.ieee802.org/16/tge/>
9. Haidar, S., Farah, A.S.: A Policy-Based Trust-Aware Adaptive Monitoring Scheme to enhance WiMAX QoS. Journal of Computer Networks 55, 2465–2480 (2011)

# A Hybrid of Fixed-Size and Dynamic-Size Tile Algorithm for Panoramic View on Mobile Devices

Chen Kim Lim<sup>\*</sup>, Kian Lam Tan, and Abdullah Zawawi Talib

School of Computer Sciences, Universiti Sains Malaysia,  
11800 USM Penang, Malaysia  
chen-kim.lim@inria.fr,  
andrewtankianlam@gmail.com, azht@cs.usm.my

**Abstract.** Digital image is a numeric representation of a two-dimensional image. With the current digital camera technology, every mobile device is also equipped with a camera which is usually more than one megapixel in resolution. Besides, the size of the basic image is getting bigger, for example around 1.8MB to 3.4MB on iPhone4. However, the storage for mobile devices is limited. Basically, for a panoramic view with  $360^0$  horizontal representations and with its X-resolution of more than one thousand pixels, the size of the panoramic view is around 12MB without any compression. In this paper, we proposed a hybrid of fixed-size and dynamic-size tile algorithm that overcomes the storage restriction in mobile devices. In the proposed algorithm, the panoramic view is divided into tiles and only tiles of interest are shown on the mobile screen. Based on the evaluation, the amount of the time taken to load the panoramic view is faster and the display requires less bandwidth and memory.

**Keywords:** Tile algorithm, Mobile device, Panoramic view.

## 1 Introduction

Nowadays, the price of mobile devices such as smart-phones or tablets is rather cheap and can be very low. Besides, the number of mobile phone subscriptions worldwide is over 5 billion and the number is still increasing. Nonetheless, mobile devices have several limitations compared to other devices such as the limited number of processors and memory. Panoramic view also known as wide-angle photography is a technique of photography that uses a specialized equipment or software to capture images with elongated fields of view. Basically, the size of a panoramic view is huge and it takes more than 5MB to store a single panoramic view. Tile algorithm is a technique for generating large images in pieces (tiles). In the algorithm, large image files can be generated without allocating a full-sized image buffer in the main memory. This paper presents a tile algorithm that cuts large images in pieces and re-produces a cylindrical panoramic view on mobile screen using a strategy that uses both fixed-size tiles and dynamic-size tiles where the dynamic-size tiles and one of the fixed-size tiles are in

---

<sup>\*</sup> Corresponding author.

image buffers. These image buffers help to reduce the times taken to scroll to the left or right. The user can still navigate the view interactively through mobile devices as in conventional panoramic views but the method requires less bandwidth and memory compared to the conventional panoramic views and other tile algorithms. This algorithm is particularly useful since mobile devices have very limited resources in terms of the number of processors and storage.

## 2 Related Work

Hsiao *et al.* [1] proposed a primitive-hierarchy fitting algorithm which can handle any size and shape of the images. The can reduce the storage pressure, improve the building time and enhance the data locality if layer-based rendering is exploited. Kim *et al.* [2] proposed an effective tile-based rendering method for a multi-core based GP-GPU environment with the aim of reducing the amount of data transfer between the processor and external memory for resource-limited mobile environment. Oliver *et al.* [3] investigated the power implications of tile size selection for tile-based processors. The Granularity Indicator since it provides a novel way to encapsulate power-scaling factors when trying to meet performance targets with parallelism. Silpa *et al.* [4] proposed an accurate work-load estimation technique and two Dynamic Voltage and Frequency Scaling (DVFS) schemes: (i) tile-history based DVFS and (ii) tile-rank based DVFS for tiled-rendering architectures. The schemes are to more efficient in terms of power and performance than the frame level DVFS schemes.

Basically, tile-based rendering requires that the primitives (commonly triangles) are sorted into bins corresponding to the tiles. Antochi [5] described several algorithms for sorting the primitives into bins and evaluates their computational complexity and memory requirements. Dai *et al.* [6] developed a method to select the most efficient tile size for the best viewing quality. In this method, the panorama is divided into tiles and only the tiles implicated with the perspective view are transmitted and decoded. In summary, we can say that the purpose of using tile-based rendering technique is to reduce the amount of data transfer between the processor and storage since the size of a single panoramic view is usually more than 5MB. So, in this paper, an algorithm is proposed to overcome the limitation on the number of processors and storage for viewing panoramic view on mobile devices.

## 3 Proposed Algorithm

The algorithm begins with loading of the panoramic image based on the method by Lim *et al.* [7]. The length of the panoramic image is needed to determine whether the tile can be cut into equal pieces using the following calculation:

$$E = L \% a \quad (1)$$

where L is the length of the panoramic image and a is the size (length) of the tile. The calculation yields a Boolean value E. If E is “0” (false), it means that the panoramic image cannot be cut into equal sizes and we need to have both dynamic-size and

fixed-size tiles. On the other hand, if E is “1” (true), then the algorithm directly calculates the value of fixed-size (Fs) as follows:

$$F_s = \left\lfloor \frac{M}{a} \right\rfloor \quad (2)$$

where M is the length of the panoramic image and “a” is the size of the tile. In this case we only have fixed-size tiles.

For the case where E is “0” (false), we need to know the total number of tiles needed and define the size for the dynamic-size and fixed-size tiles. The total number of tiles, Na is calculated as follows:

$$N_a = \left\lfloor \frac{M}{a} \right\rfloor + 1 \quad (3)$$

where M is the length of the panoramic image and “a” is the size of the tile. Next, the proposed tile algorithm needs to cut the size of the image and calculate the size of the layer for the dynamic-size (Ds) as follows:

$$D_s = M - (F_s * a) \quad (4)$$

where M is the length of the panoramic image and Fs is the size of the layer for fixed-size tiles (Equation (2)). Layering is a technique where an image can be stacked on top of another image and a new image is reproduced after combining the two layers. Also in this case (if E is “0”), Fs is also required because the tiles consist of dynamic-size and fixed size tiles. Basically, a dynamic-size tile and one of the fixed-size tiles are in image buffer but invisible to the user since the former is located to the left and the latter is to the right of the screen. The purpose of having these tiles is to allow faster and smoother scrolling to the left or right. A conventional tiling algorithm usually needs to cut the image again and this process consumes a few seconds when the user scrolls to the left or right. If the user scrolls faster than the normal speed, then the image needs a few seconds to be loaded and will appear slowly on the screen. Thus, a mechanism is needed to ensure that the image buffer appears faster on the screen during scrolling. If we scroll to the left with more than the size of the image buffer, then the process will go back to start of the loop. The same scenario also happens when scrolling to the right.

## 4 Implementation

The proposed tile algorithm was implemented by using the Objective C programming language and Apple X-code 4.3 as the Integrated Development Environment (IDE). Quartz 2D library is needed because it is an advanced two-dimensional drawing engine available for the iOS application development.

In this algorithm, the image is cut into tiles using `CGImageCreateWithImageInRect`. It is cut according to the coordinate system based on dynamic-size and fixed-size tiles by using Equation (4) and Equation (2) respectively. The purpose of having image buffers for both dynamic-size and fixed-size tiles, is to make sure that the system responds instantaneously when the user scrolls to the left or right. The dynamic-size tile is invisible to users and appears in the image buffer to the left of the screen. The fixed-size tiles consist of two types. The first type is invisible to

users where it appears in the image buffer to the right of the screen and the second type is visible to the user and appears directly on the mobile screen. For example, Figure 1 shows an arrangement of the tiles consisting of six tiles. Fixed-sized tiles are image1, image2, image3, image4 and rightImage. The leftImage, referred to as dynamic-sized tile is of different size. Only image1, image2, image3 and image4 are displayed (visible) on the screen while leftImage and rightImage are temporarily held in an image buffer (invisible) while scrolling to the left or right.

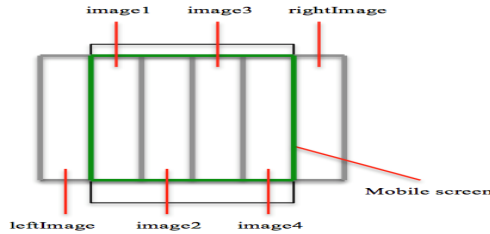


Fig. 1. The arrangement of the tiles

A layer has to be created and assigned to the image of both the dynamic-sized and fixed-sized tiles. CALayer (Core Animation Layer) is used create the layers for the tiles. The layer of the dynamic-size tile named leftImage is added to the leftImage view of the scrollLayer in order to allow the user to interact with the image. For fixed-sized tiles, layers have also to be created and assigned. The loadImage function is used to cut the image with fixed-size. The layers are then added to their respective image view of the scrollLayer in order to let the user interact with the image.

The user is allowed to scroll to the left. The loadImage function is used to cut the image and assign the image to the layer (the container to store the image) of image1 which is the leftmost-visible tile on the screen and the layer of leftImage which is the invisible tile to the left of the screen. The user is also allowed to scroll to the right. In this case, the loadImage function is used to cut the image and assign the image to the layer of Image-n which is the rightmost-visible tile on the screen and to the layer of rightImage which is the invisible tile to the right of the screen. Figure 2(a) shows an example of a display using the proposed tile algorithm on a mobile device and its corresponding panoramic image is shown in Figure 2(b) with the highlighted part corresponds to the display.

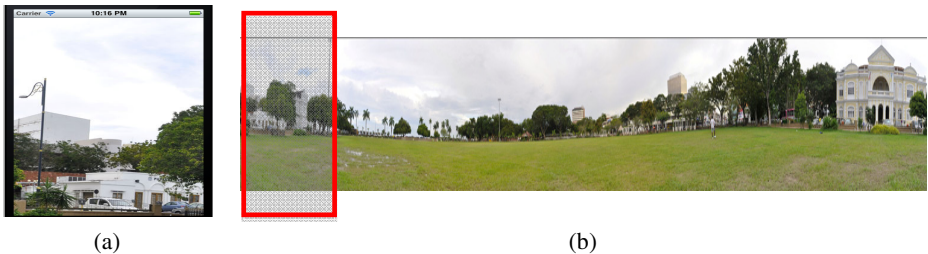


Fig. 2. (a) The output on mobile device and (b) the cylindrical panoramic view

## 5 Experimental Results

One set of cylindrical panoramic view as shown in Figure 2 was used in the experiment and the image size of this image is 6319 x 460, and to be displayed on iPhone 4. Besides, we set the tile size as 80 for the width and 460 for the height in this experiment. In addition, the default coordinate system for iPhone 4 is 320 x 480. In the experiment, six tiles are created as in Figure 1. The size of fixed-sized tiles (image1, image2, image3, image4 and rightImage) is 80 x 460. The leftImage (dynamic-sized) is of different size.

Table 1(a) presents the size of each tile when the image was first loaded by using the proposed algorithm. The total size of all the tiles when the image was loaded is 18999 bytes. Table 1(b) presents the size of each tile when the image is scrolled to the left by using the proposed algorithm. The total size of all the tiles loaded when the user scrolls to the left is 19484 bytes. The size of the panoramic image without using the proposed algorithm or any tile algorithm is 299864 bytes.

**Table 1.** The size (a) when the image was first loaded, (b) when scrolling to the left

(a)		(b)	
Tile	Size (bytes)	Tile	Size (bytes)
leftImage	3521	image1	3521
image1	3269	leftImage	3596
image2	2784	image2	3269
image3	2683	image3	2784
image4	3331	image4	2683
rightImage	3411	rightImage	3331

**Table 2.** Comparison between the cases

The total size in bytes	Without using any tile algorithm	Using the conventional tile algorithm	Using the proposed tile algorithm	Size reduction	Improvement in size
When first loaded	299864	21738	18999	2739	12.60%
When first loaded and scrolled to the left	299864	40817	38483	2334	5.72%
When first loaded and scrolled to the right	299864	39810	37709	2101	5.27%

In order to evaluate the efficiency of the algorithm, we have performed a comparison on the proposed algorithm with a conventional tile algorithm, and a method without using the algorithm or any tile algorithm based on the total size of the images in bytes. The normal tile algorithm used the CATiledLayer library for the implementation. Table 2 shows the results of the experiment. The first case is when it was first loaded and the second and third cases are when the image was first loaded and scrolled to the left and right respectively. For the first case, tremendous improvement



of the reduction in size that is 12.60% compared to the size using the conventional tile algorithm. For the second case, the proposed algorithm helps to reduce the size of the image from 40817 to 38483 bytes while for the third case it helps to reduce the size of the image from 39810 to 37709 bytes.

## 6 Conclusion and Future Work

We have shown that the proposed algorithm have great potential to be used in panoramic view because the viewing quality remains the same even through less consumption of the bandwidth and memory is required compared to conventional panoramic views especially on mobile devices where the capacity of the memory is limited. For future work, we plan to create a database to store the images inside the mobile devices so that the images can be reused if needed. The purpose of reusing the images is to reduce the total size of the images although the current result using the proposed algorithm already shows a marked improvement.

**Acknowledgments.** The authors would like to thank the support of Universiti Sains Malaysia (USM) under the Research University Grant: Grant no. 1001/PKOMP/817062.

## References

1. Hsiao, C.-C., Chung, C.-P., Yang, H.-C.: A Hierarchical Primitive Lists Structure for Tile-Based Rendering. *International Conference on Computational Science and Engineering (SCE) 2*, 408–413 (2011)
2. Kim, J.-S., Kim, D.-H., Lee, K.-Y., Kwon, Y.-S., Eum, N.-W., Lee, C.: A Hierarchical Tiling Algorithm for Tile Based Rendering with Global Scratch Counter under Multi Core Environment. In: *TENCON*, pp. 197–200 (2011)
3. Oliver, J.Y., Rao, R., Brown, M., Mankin, J., Franklin, D., Chong, F.T., Akella, V.: Tile Size Selection for Low-power Tile-based Architectures. In: *Conference on Computing Frontiers*, pp. 83–94 (2006)
4. Silpa, B.V.N., Krishnaiah, G., Panda, P.R.: Rank Based Dynamic Voltage and Frequency Scaling for Tiled Graphics Processors. In: *International Conference Hardware/Software Codesign and System Synthesis*, pp. 3–12 (2010)
5. Antochi, I., Juurlink, B., Vassiliadis, S., Liuha, P.: Scene Management Models and Overlap Tests for Tile-based Rendering. In: *Euromicro Symposium Digital System Design*, pp. 424–431 (2004)
6. Dai, F., Shen, Y., Zhang, Y., Lin, S.: The Most Efficient Tile Size in Tile-based Cylinder Panoramic Video Coding and its Selection under Restriction of Bandwidth. In: *International Conference on Multimedia and Expo*, pp. 1355–1358 (2007)
7. Lim, C.K., Tan, K.L., Talib, A.Z.: Low-cost Methods for Generating Panoramic View for a Mobile Virtual Heritage Application and its Application to the Heritage Zone of George Town Malaysia. *International Journal of E-Entrepreneurship and Innovation 1*, 58–73 (2011)

# A Relationship Strength-Aware Topic Model for Communities Discovery in Online Social Networks

Juan Bi<sup>1,\*</sup>, Jia Huang<sup>2</sup>, and Zhiguang Qin<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering  
University of Electronic Science and Technology of China  
No. 4, Section 2, North Jianshe Road, Chengdu, China

<sup>2</sup> China Science Publishing & Media Ltd. (Science Press),  
16 Donghuangchenggen North Street, Beijing 100717, China  
bijuan66@gmail.com, huangjia@mail.sciencep.com,  
qinzg@uestc.edu.cn

**Abstract.** Automatic discovering latent communities of users from observed textual content and their relationships is vital for understanding the cooperation and interaction patterns of users on large scale. In this paper, a novel probabilistic generative model was proposed to detect latent communities in a social network based on semantic information and the social relationships between users. In this model, it was assumed that users from the same community tend to share similar interests, and those who engage in common topics should be closely connected to each other on the topology structure of the social network. Users can have multiple interests and participate in multiple communities. Further, heterogeneous relationship strength was used in this paper to improve community discovery. The research indicated that the probabilistic generative model present in this paper has a good capability of discovering meaningful communities and topics on real-world data from Twitter.

**Keywords:** Community Discovery, LDA, Probabilistic Generative Model, Social Networks.

## 1 Introduction

The rapid development of online social networks has tremendously changed the way of people to communicate with each other. A lot of user-generated content is available on these online social networks. The rich textual content reveals crucial information about their interests and tastes. In traditional social network analysis, a social network is represented as a social graph in which nodes represent users and links correspond to relationships between the users. Incorporating both linkage structure and text informant can provide a unique ability of detecting latent social structure among group of users. In this paper, we address the problem of automatic discovering latent communities of users from observed textual content and their relationships. A common method of

---

\* Corresponding author.

community detection based on the graph partitioning of graph theory which entirely based on the topology information of the network links. However, such link-based methods ignore the content attribute of the nodes which crucial for uncovering and understanding the latent community in online social networks. By combining the textual information along with the linkage structure of underlying network, an appropriate community is considered as a group of users who are closely connected by social relationships and share similar interests. Moreover, the relationship is another one of crucial factor for community discovery. We consider that link formed via retweet (RT) or mention (@) interactions can be an appropriate representation of actual relationships.

In this paper, we propose a probabilistic topic model to detect latent communities in a social network based on semantic information and the social relationships between users. Our model naturally allows soft membership of users to communities, according to which user can belong to multiple community. Further, a user can be interested in more than one topic. Therefore, we assume that multiple topics can be interested by one community. In our work, we consider community and topic as different latent variable. The model cannot only discovery communities and topics simultaneously, but also enable them benefit each other. We also utilize the interaction intensity as a proxy for relationship quality which represents the strength of the relationship.

## 2 Related Works

In this section, we review related works in the areas of community discovery and topic modeling. The study of community structure in networks is primarily based on the graph partitioning algorithm [5] and probabilistic model. For graph partitioning algorithm, community discovery aims to divide the network nodes into closely connected components. The method presented in [4] is based on agglomerative algorithm where edges are removed from the network iteratively to split it into communities. These methods are purely based on graph partition algorithm they fail to account for other node attributes and communication content information. Another problematic aspect is the so-called hard membership which does not allow users to participate in multiple communities. On the other hand, the probabilistic generative models have been gained significant attention in recent years. Much of extensions of the Latent Dirichlet Allocation (LDA) model have focused on community detection. [9] Proposed an LDA-based hierarchical Bayesian algorithm called SSN-LDA where community is defined as a distribution over the social link space. LDA-G [3] simply adapts the original LDA model for community discovery in a social graph. Although these models build a generative process for links by introducing a mixture community membership, they merely consider the link structure in a graph.

There are several studies based on generative model that incorporate the link information and text content into a unified model. Such as Group-Topic (GT) model [6], Community-User-Topic (CUT) model [7]. The CUT model leverages the semantic content information to extract communities. However, the model extracts communities from just content information, ignoring the link structure in the graph. The GT model

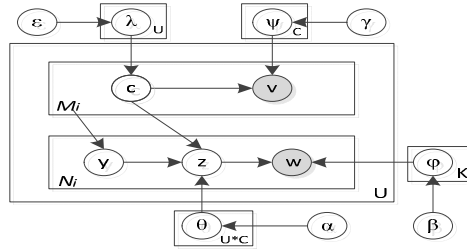


Fig. 1. Graphical notation of RCTM

simultaneously clusters entities to groups and clusters words into topics. Although the model works with both link structure and text attributes of the user, the assumption of links between users are binary has limitations. In our model, we utilize the interaction intensity as the representation of the relationship strength.

### 3 The Proposed Method

#### 3.1 Generative Progress

The graphic model representation of our model is shown in Fig 1. The proposed generative progress is as follows:

1. For each of community  $c \in C$ , draw a multinomial  $\vec{\psi}_c \sim \text{Dir}(\vec{\gamma})$
2. For each of topic  $z \in K$ , draw a multinomial  $\vec{\varphi}_z \sim \text{Dir}(\vec{\beta})$
3. For each of user  $i \in U$ :
  - (a) Draw a community distribution  $\vec{\lambda}_i \sim \text{Dir}(\vec{\epsilon})$
  - (b) For each of community  $c \in C$ , draw topic distribution  $\vec{\theta}_{i,c} \sim \text{Dir}(\vec{\alpha})$
  - (c) For each interaction  $v_{i,j} \in M_i$  associated with user  $i$ :
    - i. Draw a community  $c_{i,j} \sim \text{Mult}(\vec{\lambda}_i)$
    - ii. Draw a user  $v_{i,j} \sim \text{Mult}(\vec{\psi}_{c_{i,j}})$
  - (d) For each of words  $w_{i,j} \in N_i$  for user  $i$ :
    - iii. Draw a community index  $y_j \sim \text{Uniform}[1: M_i]$
    - iv. Draw topic  $z_{i,j} \sim \text{Mult}(\vec{\theta}_{i,c_{y_j}})$
    - v. Draw a word  $w_{i,j} \sim \text{Mult}(\vec{\varphi}_{z_{i,j}})$

The graphical model representation has shown in Fig 1. This generative model represents content information as a mixture of topics and link information as a mixture of communities. We first generate communities for each link structure for a specific user  $u$  from multinomial  $\lambda_u$ , and then generate the topic assignment of the user dependent on the communities that the user truly belongs to. We use the indexing variable  $y$  to indicate which community generates the corresponding topic.  $\theta_{u,c}$  representing the topic distribution for community  $c$  and user  $u$ . we generate topic assignments from multinomial  $\theta_{u,c}$  based on the assumption that users in the same

**Table 1.** Five topics selected from the Twitter dataset

Topic 6	Topic 7	Topic 10	Topic 17
student	Romney	Twitter	iPhone
classroom	Obama	social	apple
educational	president	Google	app
elementary	speech	Facebook	mobile
teaching	Clinton	online	android

community are likely to share the same topics. Formally, let  $Z$  and  $C$  be the set of latent topic and latent community respectively,  $W$  be the set of words in the corpus,  $V$  be the set of interactions that observed on the social graph. The joint probability on the texts, links and the latent variables for a given user is given by:

$$\begin{aligned}
 P(W, V, Z, C, \theta, \varphi, \lambda, \psi | \alpha, \beta, \varepsilon, \gamma) &= P(W|Z; \varphi)P(V|C; \psi)P(C|\lambda)P(Z|C; \theta)P(\theta|\alpha)P(\varphi|\beta)P(\lambda|\varepsilon)P(\psi|\gamma) \\
 &= \int \prod_{i=1}^U \prod_{n=1}^{N_i} p(w_{i,n} | \bar{\varphi}_{z_{i,n}}) \prod_{z=1}^K p(\bar{\varphi}_z | \bar{\beta}) d\varphi \times \int \prod_{i=1}^U \prod_{n=1}^{M_i} p(v_{i,n} | \bar{\psi}_{c_{i,n}}) \prod_{c=1}^C p(\bar{\psi}_c | \bar{\gamma}) d\psi \\
 &\times \int \prod_{i=1}^U \prod_{n=1}^{N_i} p(y_n | M_i) p(z_{i,n} | \bar{\theta}_{i,c_{y_n}}) \prod_{i=1}^U \prod_{c=1}^C p(\bar{\theta}_{i,c} | \bar{\alpha}) d\theta \times \int \prod_{i=1}^U \left( \prod_{n=1}^{M_i} p(c_{i,n} | \bar{\lambda}_i) p(\bar{\lambda}_i | \bar{\varepsilon}) \right) d\lambda
 \end{aligned}$$

### 3.2 Parameter Estimation

In order to estimate parameters  $\theta, \varphi, \lambda, \psi$ , we adopt the collapsed Gibbs sampling, a stochastic approach for approximate inference in high-dimensional models. In particular, the conditional distribution of the community assignment is given by:

$$\begin{aligned}
 p(c_i = c | Z, C_{-i}, V, W, \theta, \varphi, \lambda, \psi) &= p(c_i = c, z_i = z, v_i = v | Z_{-i}, C_{-i}, V_{-i}, \theta, \varphi, \lambda, \psi) \\
 &= \frac{p(Z, C, V, W, \theta, \varphi, \lambda, \psi)}{P(Z_{-i}, C_{-i}, V_{-i}, W, \theta, \varphi, \lambda, \psi)} \\
 &= \frac{n_{-i,c}^{(v)} + \gamma}{n_{-i,c}^{(c)} + E\gamma} \times \frac{n_{-i,(u,c)}^{(z)} + \alpha}{n_{-i,(u,c)}^{(z)} + K\alpha} \times \frac{n_{-i,u}^{(c)} + \varepsilon}{n_{-i,u}^{(c)} + C\varepsilon}
 \end{aligned}$$

Where  $n_{-i,c}^{(v)}$  is the number of times of user  $v$  assigned to community  $c$ , and not including the current user.  $n_{-i,c}^{(c)}$  is the total number of users assigned to community  $c$ , but not including the current one. Similarly,  $n_{-i,(u,c)}^{(z)}$  is the number of times of topic  $z$  is sampled from community  $c$  which user  $u$  belongs to, not including the current topic.  $n_{-i,u}^{(c)}$  is number of users who interact with user  $u$  that are assigned to community  $c$ , not including the current one. Further, the conditional distribution of a topic assignment is given by:

$$\begin{aligned}
 p(z_i = z | C, Z_{-i}, V, W, \theta, \varphi, \lambda, \psi) &= p(z_i = z, w_i = w | Z_{-i}, W_{-i}, \theta, \varphi, \lambda, \psi) \\
 &= \frac{p(Z, C, V, W, \theta, \varphi, \lambda, \psi)}{P(Z_{-i}, C, W_{-i}, V, \theta, \varphi, \lambda, \psi)} = \frac{n_{-i,z}^{(w)} + \beta}{n_{-i,z}^{(z)} + T\beta} \times \frac{n_{-i,(u,c)}^{(z)} + \alpha}{n_{-i,(u,c)}^{(z)} + K\alpha}
 \end{aligned}$$

Where  $n_{-i,z}^{(w)}$  is the number of times of word  $w$  is assigned to topic  $z$  excluding the current word  $i$ .  $n_{-i,z}^{(c)}$  is the total number of words assigned to topic  $z$  excluding current word  $i$ . Finally, the multinomial parameters can be computed as follows:

$$\lambda_{u,c} = p\{c|u\} = \frac{n_u^{(c)} + \epsilon}{n_u^{(c)} + C\epsilon} \quad \psi_{c,v} = p\{v|c\} = \frac{n_c^{(v)} + \gamma}{n_c^{(v)} + E\gamma}$$

$$\theta_{(u,c),z} = p\{z|c, u\} = \frac{n_{(u,c)}^{(z)} + \alpha}{n_{(u,c)}^{(z)} + Z\alpha} \quad \varphi_{z,w} = p\{w|z\} = \frac{n_z^{(w)} + \beta}{n_z^{(w)} + T\beta}$$

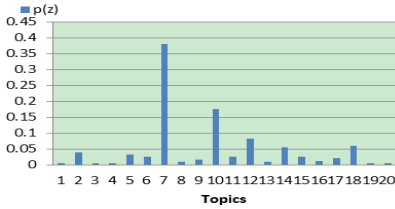


Fig. 2. Topic profiles for community 9

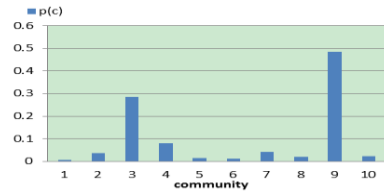


Fig. 3. Community profiles for a user

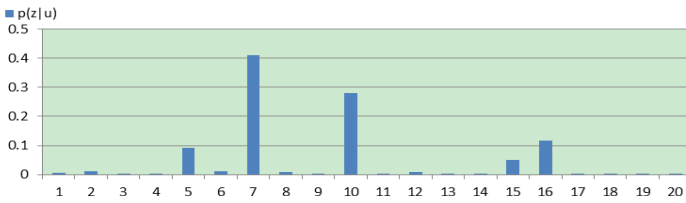


Fig. 4. Topic profiles for a user

## 4 Experiments

### 4.1 Dataset Description

Here, we present the data collected from Twitter. Since our goal is to explore the relationship between user’ interests and their interactions in the social network, we need to collect information about users, content and link information. The content in Twitter refers to tweets. And we connect two users only if there is an interaction took place between them via mention actions (@user name) or retweet action (RT), each link weighted by counting the number of times these actions has taken place between the two users. All the data is collected from Twitter API. We applied preprocessing to tweets content by removing non-English tweets, punctuations and stop words. Finally, our collection contains 1054 users, 5675 links and 17633 distinct words.

## 4.2 Experiment Results

In our experiment, we set the number of communities  $C$  at 10 and the number of topics  $Z$  at 20. All of the model hyper-parameters were set as value of 0.01 empirically, in our model, we run the Gibbs sampling process for 1000 iterations.

Table 1 shows a few topics discovered by our model in the Twitter dataset, we give the top 5 words for these interesting topics. We find our model reveals more interesting topics in Twitter. Topic 6 is about education topics; Topic 7 is about the US presidential election in 2012; Topic 10 is about social media; Topic 17 is about electronics products.

Next, Fig 2 illustrates the topic probabilities for a specific community 9. It can be found that topic 7 is the dominant topic in this community. The community is also interested in topic 10. This result accords with our assumption that one community can be related with multiple topics.

Fig 3 presents the community probabilities given a particular user 8. For example, user 8 has a high degree possibility for involving with community 9. Moreover, the result is indicated that the user also participates in community 4 to some extent.

Besides the community membership mining for each user, in Fig 4, we show the particular user 8 pays close attention to topic 7, topic 10, and topic 16. Recall the topic 7 is the dominate topic in community 9 which is the primary community for user 8. This analysis is useful in recognizing user interests and preferences. From these results presented, our model could discover topically meaningful communities in our real-world dataset.

## 5 Conclusion

In this paper, we have proposed a probabilistic generative model for community detection based on user interest extraction and social relationship analysis. The model extends prior works by combining the textual content and link information into a unified generative process to discover topically meaningful communities. The underlying assumption behind the model is that users in the same community are not only have strong relationship but also share common interest. Through extensive experiments on a Twitter dataset, we demonstrate that the model is able to extract well connected and topically meaningful communities.

## References

1. Cha, Y., Junghoo, C.: Social-network analysis using topic models. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM (2012)
2. Wasserman, S., Katherine, F.: Social network analysis: Methods and applications, vol. 8. Cambridge University Press (1994)
3. Henderson, K., Tina, E.: Applying latent dirichlet allocation to group discovery in large graphs. In: Proceedings of the 2009 ACM Symposium on Applied Computing. ACM (2009)
4. Fortunato, S.: Community detection in graphs. Physics Reports 486(3), 75–174 (2010)

5. Wang, C., David, M.: Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2011)
6. Wang, X., Natasha, M., Andrew, M.: Group and topic discovery from relations and their attributes. Massachusetts Univ. Amherst Dept of Computer Science (2006)
7. Zhou, D., et al.: Probabilistic models for discovering e-communities. In: Proceedings of the 15th International Conference on World Wide Web. ACM (2006)
8. Huang, H., Horng, Y.: Semantic Clustering-Based Community Detection in an Evolving Social Network. In: 2012 Sixth International Conference on Genetic and Evolutionary Computing (ICGEC). IEEE (2012)
9. Zhang, H., et al.: An LDA-based community structure discovery approach for large-scale social networks. In: 2007 IEEE Intelligence and Security Informatics. IEEE (2007)
10. Chang, J., David, M.: Hierarchical relational models for document networks. *The Annals of Applied Statistics* 4(1), 124–150 (2010)



# Replication in Data Grid: Determining Important Resources

Yuhanis Yusof

School of Computing, Universiti Utara Malaysia,  
Malaysia  
yuhanis@uum.edu.my

**Abstract.** Replication is an important activity in determining the availability of resources in data grid. Nevertheless, due to high computational and storage cost, having replicas for all existing resources may not be an efficient practice. Existing approach in data replication have been focusing on utilizing information on the resource itself or network capability in order to determine replication of resources. In this paper, we present the integration of three types of relationships for the mentioned purpose. The undertaken approach combines the viewpoint of user, file system and the grid itself in identifying important resource that requires replication. Experimental work has been done via OptorSim and evaluation is made based on the job execution time. Results suggested that the proposed strategy produces a better outcome compared to existing approaches.

**Keywords:** Grid computing, data grid, replication strategies, OptorSim.

## 1 Introduction

Over a number of recent years, the grid has become progressive information technology trend that enables high performance computing for scientific applications. Such a technology offers researchers the availability of powerful resources which allows them to broaden their simulations and experiments. As the grid infrastructure progress, issues are shifted towards resource management. This is realized in the Data grid where huge amount of data enables grid applications to share data files in a coordinated manner. Such an approach is seen to provide fast, reliable and transparent data access. Nevertheless, Data Grid creates a challenging problem in a grid environment because the volume of data to be shared is large despite the limited storage space and network bandwidth [1, 2]. Furthermore, resources involved are heterogeneous as they belong to different administrative domains in a distributed environment. Operationally, it is infeasible for various users to access the same data (e.g. a data file) from one single organization (e.g. site). Such situation would lead to the increase of data access latency.

Motivated by these considerations, a commonly strategy used in distributed system is also employed in Data Grid, that is replication. Experience from distributed system design shows that replication promotes high data availability, low bandwidth

consumption, increased fault tolerance, and improved scalability. In grid environment, replication is one of the major factors affecting performance of Data Grids [3]. With this, it is suggested that well-defined replication strategies will smooth data access, and reduce job execution cost [4]. However, replication is bounded by two factors: the size of storage available at different sites within the Data Grid and the bandwidth between these sites [5]. Furthermore, the files in a Data Grid are mostly large [6, 7]; so, replication to every site and hosting unlimited number of replicas would be unfeasible. Hence, we need to carefully decide which file that requires replication. We propose a relationship based replication that integrates the viewpoint of three parties; user, system and grid environment. Existing study either identify the required resource based solely on users' perspective, i.e. number of access on the file [8, 9], or based on system's perspective, i.e. storage cost and read cost of a file [10-12]. As a result, there will be an insufficient utilizing of storage resource space, which in turn will lead to less storage availability. According to [13], less storage availability would lead to longer job execution time and larger network usage because only fewer replicas can be accommodated in the Data Grid, and most files will be read remotely.

The rest of this paper is structured as follows. Section 2 provides a brief description on existing work in data replication in the data grid. We include the details of our proposed replication strategy in Section 3 and the performance evaluation is presented in Section 4. Finally, we summarize the study in Section 5.

## 2 Background

The replication algorithm proposed in [4] determines popularity of a file by analyzing data access history. The researcher believes that the popular data in the past will remain popular in the near future. Having analyzed data access history, the average number of access, NOA, is computed. Files with NOA's value that is greater than the computer average NOA will be replicated. Hence, the order of which files to be replicated depends on the NOA. The larger the NOA, the more popular the file is and will be given a higher priority during the replication process.

Nevertheless, such an approach did not consider time period of when the files were accessed. If a file was accessed for a number of times in the past, while none was made recently, the file would still be considered popular and hence will be replicated. The algorithm proposed in [8] called Last Access Largest Weight (LALW) tries to solve this problem. The key point of LALW is to give different weights to files having different age. The LALW algorithm is similar to other algorithms [4] by means of using information on access history to determine popularity of a file. But the innovation is included by adding a tag to each access history record of a file.

The work in [14] suggested a model that helps to determine number of replicas needed to maintain the desired availability in P2P communities. With this, each site within the Data Grid is authorized to create replicas for the files. The availability of a file depends on the failure rate of peers in the network. However such a model has its own disadvantage: the exact number of replicas is not determined; rather it depends on the location service accuracy which depends on the existing number of replicas.

The accuracy of the replica location service determines the percentage of accessible files, and thus if the location service is ineffective, more replicas are created to ensure data availability. On the other hand, the work discussed in [9] proposed a replication strategy that makes replication decisions whether to increase number of replicas to face the high volume of requests, or to reduce the number of replicas to save more storage space. Evidently, increasing the number of replicas will decrease the response time, but the storage cost will be increased accordingly [9].

### 3 Method

In a data grid, when a resource (e.g a data file) is required by a job and is not available on a local storage, it may either be replicated or read remotely. If a file has been replicated, in the future, when it is requested, any job can accessed it quickly and the job execution time can be reduced. Due to the limited storage capacity, replication decision should be made to conform users' needs so that high demanded files (popular replicas) are efficiently maintain and files that are rarely utilized are removed. Our strategy (known as Relationship-based Replication, RBR) is designed by utilizing three types of relationships:

- 1) File-to-user (F2U) [15] - behavior of a file being requested by users, and notes the change to this request(whether is a growth or decay change). The relationship is represented using the exponential model. The F2U provides us with the *FileLifetime* (FL),
- 2) File-to-file relationship (F2F) [15] - behavior of a file requesting other files and is noted by *FileWeight* (FW)
- 3) File-to-grid (F2G) - lifetime of a file in the grid system and is represented by *File Age* (FA).

Hence, the work presented in this study determines the importance of a resource (i.e data file) to three parties; users (F2U), file system (F2F) and the grid system (F2G). The integration of such information is represented by *File Value* (FV) , and is computed as using Eq. 1.

$$File\ Value(t, f) = \frac{FileLifetime(t, f) + FileWeight(t, f)}{File\ Age(t, f)} \quad (1)$$

The *FileLifetime*, *FileWeight* and *File Age* are used to compute the *File Value* (FV) that is used as an indicator for the importance of a resource. The larger the value of FV, the more important the resource is, hence, requiring for replications. The operation of *FileLifetime* and *FileWeight* are provided in the work reported by Madi [15] and can be obtained using the followings:

$$File\ Lifetime = N_f^t \times (1 + r)$$

Where  $N_f^t$  represents the number of access for file  $f$  at time  $t$ , and  $r$  is the growth or decay rate in number of access of a file in one time interval. The value of  $r$  can be obtained using:

$$r = (N_f^{t+1}/N_f^t) - 1$$

On the other hand, the *FileWeight* is calculated as follows:

$$File\ Weight = \sum_{i=1}^n FL_i \times DL_i$$

where, *n*: total number of files in a grid system, *FL*: File Lifetime, and *DL*: dependency level of other files on the underlying file, and if there is no dependency, DL is assumed to be zero.

In addition, the age of a resource can be calculated as the time of the resource is included in the grid until the current time. Hence, it is as follows:

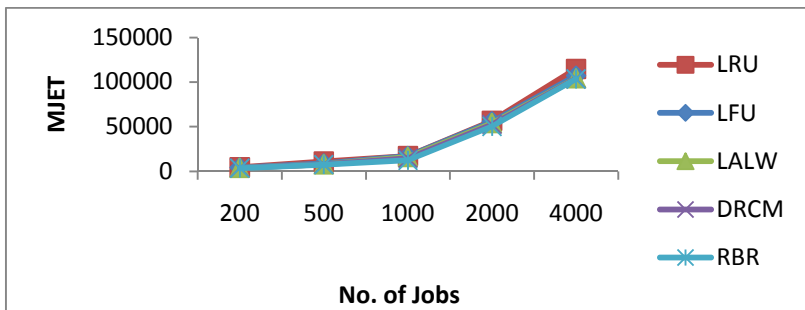
$$File\ Age = Time_{current} - Time_{attach}$$

### 4 Evaluations

In this research, the *OptorSim* [16-18] simulator was utilized to simulate the proposed RBR. The scalability of RBR is tested by the number of jobs running during the simulation. In this paper, number of jobs that is considered in our evaluation varies between 200 and 4000 jobs. The evaluation is later based on the mean job execution time (MJET)[8, 16, 17, 19-21]. We depict the obtained results in Table 1 and Figure 1 where comparison of MJET is made between the proposed RBR and four other replication strategies; LRU, LFU, LALW [8] and DRCM [15].

**Table 1.** Simulation Results of MJET

No. of Jobs	LRU	LFU	LALW	DRCM	RBR
200	4582	4398	3931	3792	<b>3545</b>
500	10911	8994	7839	7791	<b>7566</b>
1000	17108	17030	16241	14522	<b>12311</b>
2000	56567	55948	54133	52689	<b>50361</b>
4000	114652	106979	104129	103771	<b>103396</b>



**Fig. 1.** Line Chart of Simulation Results

The results show a linear increase in the MJET as the number of jobs on the grid increases. This is because, as more jobs are submitted, the queue at the sites increases. If the job submission rate is higher than the grid's job processing rate, this build-up of queues is inevitable. Hence, a preferred algorithm is an algorithm that has less MJET. As shown above, the RBR is the best among existing algorithms. Utilizing the RBR, the mean job execution time is reduced and is noted to improve by 5.12% over DRCM, 24.25% over LALW, and about 7 % over LRU and LFU.

## 5 Conclusion

In this paper, we presented the integration of three viewpoints in identifying important resource files that requires replication. A file is assumed to be important if its access grows exponentially and is required by other files. In the simulation experiments, even though different workload has been tested, it is noted that the undertaken approach produces better result (less job execution time) as compared to its competitors. Hence, this may suggest that the proposed Relationship based Replication, RBR, could be a possible approach in data replication.

**Acknowledgement.** The author would like to thank Mohammed Madi for his assistant in completing the experiment. Also, author expresses the gratitude to Universiti Utara Malaysia for providing the financial and management support under the High Impact Individual research grant (s/o 12155).

## References

1. Wilkinson, B.: Grid computing: techniques and applications. Chapman & Hall/CRC (2009)
2. Nicholson, C., Cameron, D.G., Doyle, A.T., Millar, A.P., Stockinger, K.: Dynamic data replication in lcg 2008. *Concurrency and Computation: Practice and Experience* 20, 1259–1271 (2008)
3. You, X., Chang, G., Chen, X., Tian, C., Zhu, C.: Utility-Based Replication Strategies in Data Grids. In: *Fifth International Conference on Grid and Cooperative Computing*, pp. 500–507 (2006)
4. Tang, M., Lee, B.S., Tang, X., Yeo, C.K.: The impact of data replication on job scheduling performance in the Data Grid. *Future Generation Computer Systems* 22, 254–268 (2006)
5. Venugopal, S., Buyya, R., Ramamohanarao, K.: A taxonomy of data grids for distributed data sharing, management, and processing. *ACM Computing Surveys (CSUR)* 38, 3 (2006)
6. Rahman, R.M., Barker, K., Alhaji, R.: Replica placement strategies in data grid. *Journal of Grid Computing* 6, 103–123 (2008)
7. Rahman, R.M., Barker, K., Alhaji, R.: Performance evaluation of different replica placement algorithms. *International Journal of Grid and Utility Computing* 1, 121–133 (2009)
8. Ruay-Shiung, C., Hui-Ping, C., Yun-Ting, W.: A dynamic weighted data replication strategy in data grids. In: *AICCSA 2008: Proceedings of IEEE/ACS International Conference on Computer Systems and Applications*, pp. 414–421 (2008)

9. Al Mistarihi, H.H.E., Yong, C.H.: Replica management in data grid. *International Journal of Computer Science and Network Security IJCSNS* 8, 22 (2008)
10. Yi-Fang, L., Pangfeng, L., Jan-Jan, W.: Optimal placement of replicas in data grid environments with locality assurance. In: 12th International Conference on Parallel and Distributed Systems, ICPADS 2006, p. 8 (2006)
11. Pangfeng, L., Jan-Jan, W.: Optimal replica placement strategy for hierarchical data grid systems. In: Sixth IEEE International Symposium on Cluster Computing and the Grid, CCGRID 2006, p. 4 (2006)
12. Mansouri, Y., Garmehi, M., Sargolzaei, M., Shadi, M.: Optimal Number of Replicas in Data Grid Environment. In: First International Conference on Distributed Framework and Applications, DFmA 2008, pp. 96–101 (2008)
13. Cameron, D.G.: Replica management and optimisation for data grids. PhD. Thesis, University of Glasgow (2005)
14. Ranganathan, K., Iamnitchi, A., Foster, I.: Improving data availability through dynamic model-driven replication in large peer-to-peer communities. In: Global and Peer-to-Peer Computing on Large Scale Distributed Systems Workshop, pp. 376–381 (2002)
15. Madi, M.: Replica Creation Algorithm for Data Grid. PhD, School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia, Sintok (2012)
16. Cameron, D.G., Millar, A.P., Nicholson, C., Carvajal-Schiaffino, R., Stockinger, K., Zini, F.: Analysis of scheduling and replica optimisation strategies for data grids using OptorSim. *Journal of Grid Computing* 2, 57–69 (2004)
17. Bell, W.H., Cameron, D.G., Millar, A.P., Capozza, L., Stockinger, K., Zini, F.: Optorsim: A grid simulator for studying dynamic data replication strategies. *International Journal of High Performance Computing Applications* 17, 403–416 (2003)
18. Cameron, D.G., Carvajal-Schiaffino, R., Millar, A.P., Nicholson, C., Stockinger, K., Zini, F.: Evaluating scheduling and replica optimisation strategies in OptorSim. *Journal of Grid Computing*, 57–69 (March 2004)
19. Bell, W.H., Cameron, D.G., Capozza, L., Millar, P., Stockinger, K., Zini, F.: Simulation of Dynamic Grid Replication Strategies in OptorSim. *Journal of High Performance Computing Applications* 17 (2003)
20. Ben Charrada, F., Ounelli, H., Chettaoui, H.: An Efficient Replication Strategy for Dynamic Data Grids. In: Proceedings of International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), pp. 50–54 (2010)
21. Cameron, D.G., Carvajal-Schiaffino, R., Millar, A.P., Nicholson, C., Stockinger, K., Zini, F.: UK grid simulation with OptorSim. In: Proceedings of UK e-Science All Hands Meeting, Nottingham, UK (2003)

# Backpropagation Neural Network for Sex Determination from Patella in Forensic Anthropology

Iis Afrianty<sup>1</sup>, Dewi Nasien<sup>1</sup>, Mohammed R.A. Kadir<sup>2</sup>, and Habibollah Haron<sup>1</sup>

<sup>1</sup>Faculty of Computing

Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

<sup>2</sup>Faculty of BioScience and Medical Engineering

Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia

afrianty\_iis@yahoo.com,

{dewinasien,habib}@utm.my, rafiq@fkm.utm.my

**Abstract.** Forensic anthropology is a discipline that concerned on postmortem identification from skeletal remains in sex determination. In sex determination, besides empirical techniques such as Discriminant Function Analysis (DFA), Artificial Intelligence techniques such as Artificial Neural Network (ANN) should be considered to get more accurate result. This paper proposes back propagation ANN model for sex determination. By using data and DFA result from previous work, this paper compares the result with the result of ANN model obtained from the experiment. A total sample data of 113 patellae has been generated based on statistics values of previous study. The data is divided into three groups of ages (young, middle, and old) and is measured using three parameters (width, height, and thickness). The ANN model produces average accuracy until 96.1% compared to 92.9% result from DFA technique. This concludes that ANN produces more accurate result in sex determination compared to DFA.

**Keywords:** Backpropagation neural network, Forensic anthropology, Patella, Sex determination.

## 1 Introduction

Forensic anthropology represents the application of knowledge and techniques that primarily concerned with the examination of material believed to be human to answer medico-legal questions including those related to identification [1]. Forensic anthropology is a sub-discipline within the subfield of physical anthropology and has been one of the fastest growing discipline [2, 3]. Forensic anthropology have goals namely to establish identity of skeletal remains of the decedent, analyze useful information for a case, make use of all available scientific resources for these purposes, and to help determine what happened to the remains, specific to the cause and the manner of death [4, 5]. Forensic anthropologists often bemoan about the major problem in forensic anthropology are identification of skeletal remains. The identity of skeletal remains is an essential part of post-mortem to recognize of biological profile of unknown remains such as sex determination. Sex determination is the first and essential

steps taken by forensic anthropologists in the identification of an unknown skeleton and consequently for positive identification of skeletal remains [6, 7]. Sex determination is the classification of an individual as either male or female. Knowledge of the sex of an unknown set of remains is essential to make a more accurate estimation of age [8].

Sex determination is more reliable if the complete skeleton is available, but in forensic cases skeletal remains are often with conditions incomplete, burned, and damaged, at times the only thing is the skeletal remains or even decomposed and amputated body fragments [9, 10]. In previous work, for identifying skeletal remains, forensic anthropologists used DNA analysis in the laboratory. DNA is largely widespread that DNA could lead to identification but it has lack. It cannot be extracted if skeleton in burned or damaged condition, thus cannot give data on some of the essential parameters of the biological profile so it is sometimes not evident whether it is of human or non-human bones [11]. Nevertheless, DNA cannot replace the anthropological analysis. Therefore, with lack of DNA analysis, forensic anthropology present with give contributes in identification, particular in sex determination.

Generally, sex determination process has two methods for measurement of data collection namely morphologic and metric method [12, 13]. The morphologic method is visual judgment of sex-dependent morphological traits on bones sample [12, 14]. This method has advantage namely the ability to obtain results quickly with high classification accuracy if the bone is available and the observer has enough experience[12]. Nevertheless, judging the sex can be problematic because the classification criteria for this method are rather subjective and some of the bone samples may indicate features between those of males and females[12]. The metric method is based on measurements and statistical techniques [15]. Metric measurements can be used in two ways for sex estimation. The first is to compare the measurement value obtained with the average values of each sex and the second is to classify the value by using methods [12]. Metric measurements were preferred due to their easy repeatability, high accuracy, and no requirement for special skill [16]. A metric method should be used in combination with the morphological method that is called morphometric.

Commonly, in previous studies, the forensic anthropologists using many parts of the skeleton for identification of a person, such as pelvic [17], skull [18], mandible [19, 20], clavicle [16, 21], and femur [22-25] have been used for identification. One of the skeletal elements drawing more attention recently is the patella [15]. This study has objectives for sex determination of patella using backpropagation neural network.

## 2 Materials and Methods

The patella is a small compact bone that articulates with the distal anterior end of the femur [6]. As the patella is very resistant to postmortem changes, the present study aims to estimate the sex of individuals on the basis of metric methods of patella bone. In previous study, researchers more use DFA for determination. A total sample data of 113 patellae has been generated based on statistics values of previous study, namely data of [15]. There are three parameters of patella that used in this study as the most useful sex differentiating parameters, namely height, width, and thickness.



In this paper, we propose Backpropagation Neural Network (BPNN) method for sex determination and will compare result of average accuracy of BPNN with DFA.

Artificial Intelligence (AI) is a branch of computer science concerned with ability of a machine or artifact to perform the same kind of functions that characterize human thought[26]. Neural network is one branch of AI. The neural network methodology is well known by its ability for generalization, its massive parallel processing power and its high nonlinearity, making it perfect for sex estimation [6]. Backpropagation (BP) is a systematic method of training multilayer artificial neural networks. BP consists of (1) an input layer with nodes representing input variables to the problem, (2) an output layer with nodes representing the dependent variables, and (3) one or more hidden layers containing nodes to help capture the nonlinearity in the data [27]. The neurons between layers can be fully or partially interconnected between layers with weight ( $w$ ).

In this paper for establishing that significant difference exists between male and female mean values for each group and analyzing backpropagation neural network were performed with MATLAB 2010 software.

### 3 The ANN Model

Architecture of neural network that is used of this cases consist of three inputs (height, width, and thickness), three architectures for hidden layer, namely [3, 4, 3] respectively, and 2 output layer (male and female). This study using 6000 epochs, multilayer networks may use the tan-sigmoid transfer function tansig, and learning rate is 0.1. The structure of cases is shown in Figure 1.

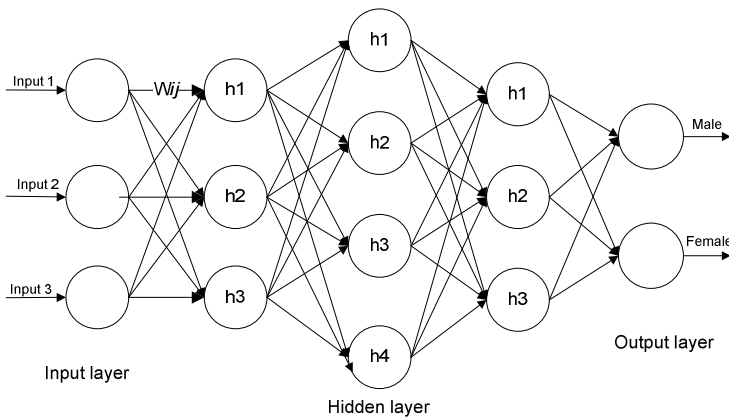


Fig. 1. The structure of backpropagation neural network of cases

### 4 Result and Discussion

There are three age groups of that cases, namely group A (young), group B (middle aged), and group C (old age). Group A had an age range of 19 to 39 years, group B

had age range of 40 to 64 years, and group C had an age range over 65 years. After measurement of patella (cm) between males and females, then average accuracy of sex determination using backpropagation neural network that has been developed in MATLAB 2010 is calculated. In determination of average accuracy of male and female can be visualized as a confusion matrix, where each column represents the predicted instances of a character, while each row represents the actual instances of an output. A confusion matrix is a visualization tool typically used in supervised learning. The data are generated based on statistics values of previous study [15]. Three age groups and parameters are shown in Table 1.

**Table 1.** The results of descriptive statistical analysis of patella measurements (cm) [15]

Patella parameters	Age group (years)	19-39			40-64			≥ 65		
		Male	Female	Total	Male	Female	Total	Male	Female	Total
Height	Mean	4.504	3.82	4.172	4.42	3.94	4.19	4.49	3.78	4.12
	Min.	4.1	3.5	3.5	3.88	3.5	3.5	3.98	3.42	3.42
	Max.	4.89	4.16	4.89	5.13	4.53	5.13	5	4.28	5
Width	Mean	4.61	3.96	4.29	4.59	4.1	4.36	4.47	3.99	4.22
	Min.	4.3	3.55	3.55	4.1	3.82	3.82	4.1	3.68	3.68
	Max.	5.04	4.25	5.04	5.1	4.5	5.1	4.9	4.42	4.9
Thickness	Mean	2.2.7	2.1	2.18	2.23	2.04	2.14	2.11	1.97	2.03
	Min.	2.08	1.7	1.7	1.76	1.6	1.6	1.82	1.78	1.78
	Max.	2.6	2.36	2.6	2.6	2.38	2.6	2.24	2.1	2.24

From descriptive statistical analysis of patella measurements of Table 1, data are generated as input values for BPNN. Data that has been generated, then it is calculated by MATLAB. The result of average accuracy that produced by BPNN is shown in Table 2.

**Table 2.** The result of average accuracy using BPNN

Age Groups	Average Accuracy (%) By BPNN		Average Accuracy (%) by Previous study	
	Males	Females	Males	Females
Group A: Young (19-39 years)	95.3 %	97.7 %	100%	89.7%
Group B: Middle aged (40-64 years)	96.72%	97.49%	88.2%	93.3%
Group C: Old age (≥65 years)	94.15%	95.22%	90.9%	95.8%
Total average accuracy for males and females	96.1%		92.9%	

In this paper, we obtained a classification rate of 96.1% in the sex determination of the patella, compared to 92.9% found in the literature [15]. This confirms that the use of neural network will improve the accuracy in determination of sex for forensic anthropology. The work by [15] DFA can be improved by applying artificial intelligence techniques.

**Acknowledgment.** This research has been supported by Malaysian Ministry of Education under Fundamental Research Grant Scheme with UTM Vote Number RJ130000.7828.4F115.

## References

- [1] Blau, S., Briggs, C.A.: The role of forensic anthropology in Disaster Victim Identification (DVI). *Forensic Sci. Int.* 205, 29–35 (2011)
- [2] Adebisi, S.S.: Forensic Anthropology In Perspective: The Current Trend. *Forensic Sci. Int.* 4, 2–2 (2009)
- [3] Iscan, M.Y., Olivera, H.E.S.: Forensic anthropology in Latin America. *Forensic Sci. Int.* 109, 15–30 (2000)
- [4] Schmitt, E.C.A.J.P.A.: Forensic Anthropology and Medicine: Complementary Sciences From Recovery To Cause of Death. *International Journal of Osteoarchaeology, Int. J. Osteoarchaeol.* 17, 434–436 (2007)
- [5] Vaz, M., Benfca, F.S.: The experience of the Forensic Anthropology Service of the Medical Examiner's Office in Porto Alegre, Brazil. *Forensic Sci. Int.* 179, e45–e49 (2008)
- [6] Mahfouz, M., Badawi, A., Merkl, B., Fatah, E.E., Pritchard, E., Kesler, K., Moore, M., Jantz, R., Jantz, L.: Patella sex determination by 3D statistical shape models and nonlinear classifiers. *Forensic Sci. Int.* 173, 161–170 (2007)
- [7] Gomez-Valdes, J.A., Quinto-Sanchez, M., Menendez Garmendia, A., Velemenska, J., Sanchez-Mejorada, G., Bruzek, J.: Comparison of methods to determine sex by evaluating the greater sciatic notch: Visual, angular and geometric morphometrics. *Forensic Sci. Int.* 221, 156 (2012)
- [8] Koçak, A., Özgür Aktas, E., Ertürk, S., Aktas, S., Yemisçigil, A.: Sex determination from the sternal end of the rib by osteometric analysis. *Legal Medicine* 5, 100–104 (2003)
- [9] Mostafa, E.M., El-Ellemi, A.H., El-Beblawy, M.A., Dawood, A.E.-W.A.: Adult sex identification using digital radiographs of the proximal epiphysis of the femur at Suez Canal University Hospital in Ismailia, Egypt. *Egyptian Journal of Forensic Sciences* 2, 81–88 (2012)
- [10] Akhlaghi, M., Sheikhzadi, A., Ebrahimnia, A., Hedayati, M., Nazparvar, B., Saberi Anary, S.H.: The value of radius bone in prediction of sex and height in the Iranian population. *J. Forensic Leg. Med.* 19, 219–222 (2012)
- [11] Cattaneo, C.: Forensic anthropology: developments of a classical discipline in the new millennium. *Forensic Sci. Int.* 165, 185–193 (2007)
- [12] Ogawa, Y., Imaizumi, K., Miyasaka, S., Yoshino, M.: Discriminant functions for sex estimation of modern Japanese skulls. *J. Forensic Leg. Med.* 20, 234–238 (2012)
- [13] Cattaneo, C., Porta, D., De Angelis, D., Gibelli, D., Poppa, P., Grandi, M.: Unidentified bodies and human remains: An Italian glimpse through a European problem. *Forensic Sci. Int.* 195, 167.e1–167.e6 (2010)

- [14] Akhlaghi, M., Hajibeygi, M., Zamani, N., Moradi, B.: Estimation of stature from upper limb anthropometry in Iranian population. *J. Forensic Leg. Med.* 19, 280–284 (2012)
- [15] Akhlaghi, M., Sheikhzadi, A., Naghsh, A., Dorvashi, G.: Identification of sex in Iranian population using patella dimensions. *J. Forensic Leg. Med.* 17, 150–155 (2010)
- [16] Akhlaghi, M., Moradi, B., Hajibeygi, M.: Sex determination using anthropometric dimensions of the clavicle in Iranian population. *J. Forensic Leg. Med.* 19, 381–385 (2012)
- [17] Dixit, S.G., Kakar, S., Agarwal, S., Choudhry, R.: Sexing of human hip bones of Indian origin by discriminant function analysis. *J. Forensic Leg. Med.* 14, 429–435 (2007)
- [18] Guyomarc'h, P., Bruzek, J.: Accuracy and reliability in sex determination from skulls: a comparison of Fordisc(R) 3.0 and the discriminant function analysis. *Forensic Sci. Int.* 208, 180 e1–186 e1 (2011)
- [19] Hu, K.S., Koh, K.S., Han, S.H., Shin, K.J., Kim, H.J.: Sex determination using nonmetric characteristics of the mandible in Koreans. *J. Forensic. Sci.* 51, 1376–1382 (2006)
- [20] Saini, V., Srivastava, R., Shamal, S.N., Singh, T.B., Pandey, A.K., Tripathi, S.K.: Sex determination using mandibular ramus flexure: a preliminary study on Indian population. *J. Forensic Leg. Med.* 18, 208–212 (2011)
- [21] Papaioannou, V.A., Kranioti, E.F., Joveneaux, P., Nathena, D., Michalodimitrakis, M.: Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: applications in forensic identification. *Forensic Sci. Int.* 217, 231 e1–237 e1 (2012)
- [22] du Jardin, P., Ponsaille, J., Alunni-Perret, V., Quatrehomme, G.: A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population. *Forensic Sci. Int.* 192, 127 e1–136 e1 (2009)
- [23] Harma, A., Karakas, H.M.: Determination of sex from the femur in Anatolian Caucasians: a digital radiological study. *J. Forensic Leg. Med.* 14, 190–194 (2007)
- [24] Purkait, R.: Triangle identified at the proximal end of femur: a new sex determinant. *Forensic Sci. Int.* 147, 135–139 (2005)
- [25] Rissech, C., Schaefer, M., Malgosa, A.: Development of the femur—implications for age and sex determination. *Forensic Sci. Int.* 180, 1–9 (2008)
- [26] Mellit, A., Kalogirou, S.A.: Artificial intelligence techniques for photovoltaic applications: A review. *Progress in Energy and Combustion Science* 34, 574–632 (2008)
- [27] Liang, L., Wu, D.: An application of pattern recognition on scoring Chinese corporations financial conditions based on backpropagation neural network. *J. Computers & Operation Research* 32, 1115–1129 (2005)

# Learning History Using Role-Playing Game (RPG) on Mobile Platform

Guo Haur Lee, Abdullah Zawawi Talib,  
Wan Mohd Nazmee Wan Zainon\*, and Chen Kim Lim

School of Computer Sciences, Universiti Sains Malaysia,  
11800 USM Penang, Malaysia  
{leeguohaur, chenkinlim}@gmail.com,  
{azht, nazmee}@cs.usm.my

**Abstract.** Mobile learning is one of the most convenient approaches in e-learning as it is accessible from virtually anywhere. This paper describes our experience in designing and implementing a framework for a mobile learning application that can assist students in understanding history lesson using role-playing game (RPG) approach. The implementation is based on the narrative of a Malaysian legendary warrior, *Merong Mahawangsa*. This application was developed on an iOS platform and a crowd simulation technique was used in the application in order to make it more interactive and realistic. The Dijkstra shortest path algorithm was used to search the shortest path for the avatar to move around. Initial investigations suggest that applying the RPG concept has indeed provided a much better learning environment especially in helping students in learning their history lessons.

**Keywords:** Mobile Learning, Role-playing Game, Mobile History Education.

## 1 Introduction

History is very important as it provides a cognitive process for us to recognize what had happened in the past on the planet that we are living now. It helps in inspiring values and shaping our life. We will be able to make better decisions with the knowledge that we acquired by learning history and live a better life by knowing history. However, history is a difficult subject to learn as it involves learning and memorizing of names, dates, events and so on including their significance. Sometimes history can be very interesting. However, it normally fails to attract the interest of many students. Therefore an alternative approach is needed to help these students by making the learning process to be more attractive and made available on mobile platform for better accessibility. Mobile learning and game-based learning are some of the approaches that can be used to solve such problem. Some researchers found that game-based learning can indeed motivate children [1]. Most of the current game-based learning products focus on language learning and many of them are also based on mobile platform [2].

---

\* Corresponding author.

In this paper, we present a framework and its implementation for a mobile learning application which was designed for learning history based on Role-Playing Game (RPG) that incorporates multimedia games. The implementation is based on the narrative of a Malaysian legendary warrior, *Merong Mahawangsa*. Using narrative in teaching history is a valuable means to move beyond just acquiring fragmented facts of historical figures and events. In addition, games are also recognized as a fruitful means for narrative learning environment that allow pupils to learn in a realistic way about on a certain topic [5]. Given the changes in society, and constructivist approaches towards learning, it becomes important that students not only learn by receiving knowledge, but also by searching and making knowledge. In line with this, an approach towards learning in which pupils do things that matter to them and to the society, games seem to be an appropriate means to allow learning in a more meaningful way [6].

## 2 Background and Related Work

Digital games have become one of the popular e-learning preferences as they provide a multimedia learning environment that can attract the students' attention and help them improve their study. The gap between digital games and education is getting much closer. Many research works have integrated their learning content into digital games. Some of them adopted the role-playing game (RPG) in their work. RPG allows player to play the game through a point of view of a character in order to maximize the user's experience of the game. RPG presents narrative experience and storytelling to a player during game play. RPG can be single or multiple players. The multiple-player RPG is call massively multiplayer online role-playing game (MMORPG) and these MMOPRG connects all the multiple players through the internet.

Many RPG-based learning applications had been developed on various platforma. This kind of game usually focuses on language learning. "Knuckles in China Land" for example, is a console-style RPG that helps the user in learning the Japanese language. During a battle, a picture or a word is shown on the screen and the user needs to type the correct word or spelling to defeat the enemy. This game also provides a vocabulary editor that allows the user to insert new vocabulary into the game. Math Quest is an RPG-based flash game for teaching Mathematics [7]. In this game, a player is given a mission to bring back the knowledge of Mathematics in order to save the world. The player will learn mathematics skills along the journey through the battle or apply their mathematics skills in solving the quest.

"The Romance of the Three Kingdoms 2" was developed on iPhone/iPod touch platform. Its storyline is based on a historical novel, the Romance of the Three Kingdoms and it also incorporates general information and history timeline of the story. This game is a turn-base strategy game and each player can perform some commands such as attack, internal administration and strategy planning. The players are expected to learn the history of China from the characters' description and storyline or storytelling that goes with the game.

### 3 Proposed Framework

In this paper, an application that can help students in learning history using game approach on mobile platform (M-History) was proposed. This application will be designed and developed on iOS platform. The framework of the application as shown in Fig. 1 consists of two main components: the RPG Approach and information corner. The RPG approach also deals with mini games and quiz. The information corner is used to display historical information.

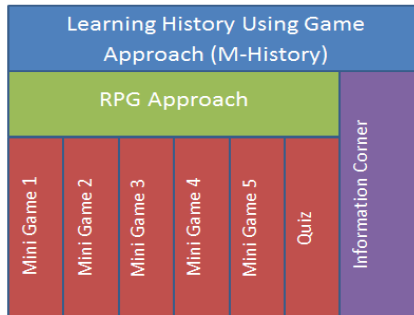


Fig. 1. The framework of M-History

### 4 Design and Implementation

The application is developed on iOS platform specifically for iPhone with iOS version 5 and above using Xcode 4.3.2 and Cocos2d game engine and tested on iPad mini and iOS simulator 6.0.

The application applies RPG concept that allows the user to control an avatar to complete the game. The RPG game map was designed using Tiled Map Editor. The tile map is shown in Fig. 2. The Dijkstra shortest path algorithm was used to connect each turning point together to improve the routing function and avoid collision with the object on the tile map. The algorithm is used to calculate the shortest path that need to be traversed by the avatars when they move to complete the mission. Besides, crowd simulation was also applied for a more realistic movement of the avatars.

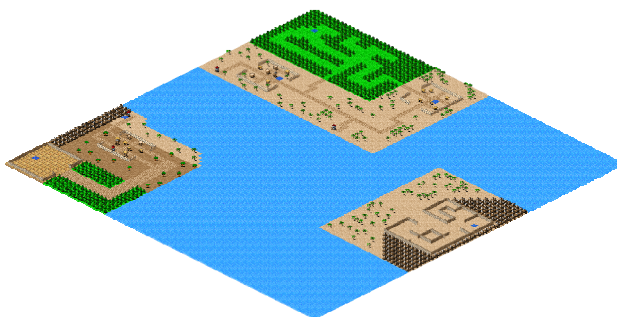
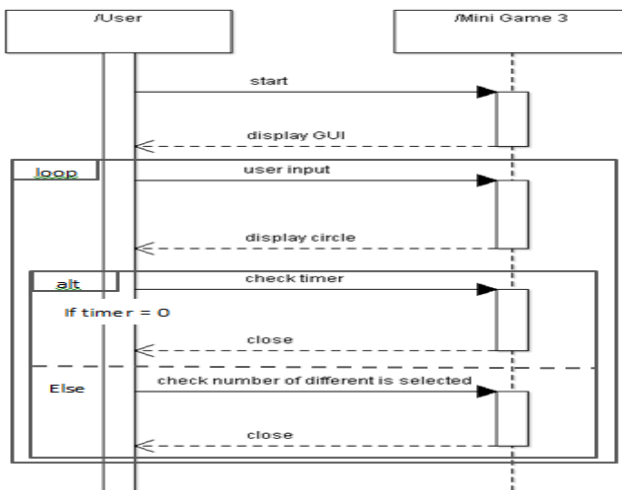


Fig. 2. RPG tile map

A summary of the game flow is given in Table 1. The game consists of five stages with a mini game for each stage. The game plot is based on the narrative history of *Merong Mahawangsa*. The mini games are used to attract the user’s attention to focus on the specific topics that he/she is supposed to learn. For example, the mini game 3 of this application requires the user to find the differences between two pictures in a certain time limit. The system matches the coordinates touched by the user with the coordinates of the differences in the database. Fig. 3 shows the sequence diagram of this mini game. If the coordinates are correct then a red circle will be displayed at the location on the picture. Fig. 4 shows the corresponding storyboard for this mini game.

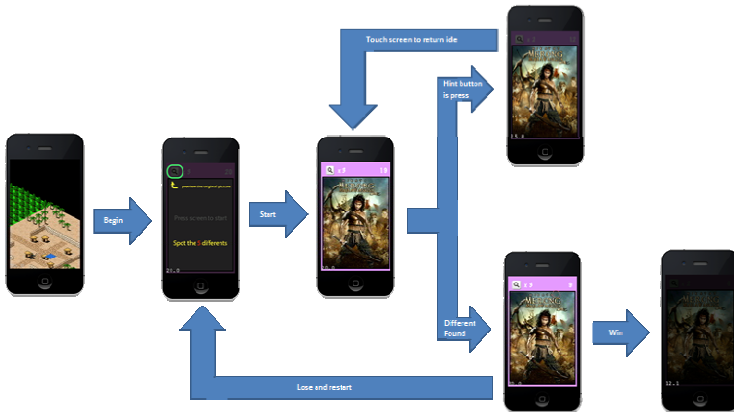
**Table 1.** Summary of the game flow

Stage	Description	Mini Games
1	The King of Rome orders Merong Mahawangsa to find the Prince of Rome, Marcus and bring him to the peninsula to meet a princess of the Han Dynasty, Meng Li Hua.	Scissor paper stone
2	Marcus and Merong Mahawangsa reach the peninsula and search for the location of the Han Dynasty army camp. They reach the camp and meet the princess.	Swap the picture
3	The princess runs away from the camp. Marcus and Merong Mahawangsa search for her in the jungle.	Spot the differences
4	The pirate, Garuda attacks them and kidnaps the princess. Merong Mahawangsa helps the Han army in defeating the enemy.	Defend the tower
5	Merong Mahawangsa goes to the Garuda’s base to save the princess and fights with Taji. Finally, Merong Mahawangsa stays on this land and creates an empire.	Whack-a-mole



**Fig. 3.** Sequence diagram of mini game 3





**Fig. 4.** Example of the proposed storyboard for mini game 3

The quiz module will pop out a question after each mini game was completed. The question is randomly picked from the database. All questions are related to the Merong Mahawangsa history. The last module of this application is the information corner. This module is an info sharing corner to help the users to acquire some knowledge the history of Merong Mahawangsa. All the information will be stored inside a SQLite database.

## 5 Evaluation and Discussion +

Evaluation of this application was gathered through a questionnaire from 13 participants, and by conducting semi-structured interview and analyzing the storyboards. The findings reveal that the perception of students about history has changed. The application is also believed to be able to help students to develop communication and interpersonal skills. The students and teachers also realized the potential of Mobile phones as one of the teaching and learning tools. Such activities engage students in thinking, reflecting and researching about ways of taking teaching of history beyond the pages of textbooks. The result of the evaluation has shown that by participating in the game, student interactivity can be increased which can eventually lead to motivation and greater interest in what is being learnt.

## 6 Conclusion

We have developed M-history which uses RPG concept that allows the user to learn and understand history in a more interesting manner. M-History was developed with nice graphical user interface and high user interactivity. Each stage contains a mini game and quiz to attract the user's attention on the subjects that are being learnt.

For future research it would be appropriate to explore more on whether it is possible to further combine the usual learning processes with mobile games in order to allow students to access education from virtually anywhere.

**Acknowledgments.** We thank Universiti Sains Malaysia (USM) for providing the funding (RUI Grant - no. 1001/PKOMP/817062) through which this article was produced.

## References

1. Adams, D.M., Mayer, R.E., MacNamara, A., Koenig, A., Wainess, R.: Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology* 104(1), 235–249 (2012)
2. Admiraal, W., Raessens, J., Van Zeijts, H.: Technology enhanced learning through mobile technology in secondary education. In: Cunningham, P., Cunningham, M. (eds.) *Expanding the Knowledge Economy. Issues, Applications, Case Studies (Pt. 2)*, pp. 1241–1248. IOS Press, Amsterdam (2007)
3. Winters, N.: What is mobile learning? In: Sharples, M. (ed.) *Big Issues in Mobile Learning*, pp. 7–12. LSRI, Nottingham (2007)
4. Hermans, H., Hermans-Jansen, E.: *Self-Narratives. The construction of meaning in psychotherapy*. The Guilford Press, London (1995)
5. Shaffer, D.W., Squire, K.R., Halverson, R., Gee, J.P.: Video games and the future of learning. WVER Working Paper No. 2005-4 (2005)
6. Burguillo, J.: Using Game Theory and Competition-Based Learning to Stimulate Student Motivation and Performance. *Computers & Education* 55, 566–576 (2010)
7. Ahmad, W.F., Shafie, A.B., Latif, M.H.: Role-playing game-based learning in mathematics. *The Electronic Journal of Mathematics and Technology* 4(2), 185–196 (2010)
8. Jenkins, H.: Game Design as Narrative Architecture. In: Wardrip-Fruin, N., Harrigan, P. (eds.) *First Person. New Media as Story, Performance, and Game*, pp. 118–130. The MIT Press, Cambridge (2004)
9. Gee, J.P.: Learning by design: Good video games as learning machines. *E-Learning* 2(1), 5–16 (2005)
10. Li, K., Huang, J., Heh, J., Chen, C., Wang, H., Yeh, S.: Designing Game-Based Learning Framework - A Motivation-Driven Approach. In: *10th Proceedings of the IEEE International Conference on Advanced Learning Technologies*, pp. 215–216. IEEE Computer Society, Washington DC (2010)

# Experiment on Modified Artificial Bee Colony for Better Global Optimisation

Muhammad Shahrizan Shahrudin and M. Mahmuddin\*

Computer Science Department, School of Computing,  
Universiti Utara Malaysia  
06010, UUM Sintok, Kedah, Malaysia  
m.shahrizan@ymail.com,  
ady@uum.edu.my

**Abstract.** Although there are many works have been done in Artificial Bee Colony (ABC) algorithm, yet, there still an issue for faster convergence for this algorithm. This paper will present a modified ABC algorithm to find optimum value for optimisation function. This hybrid ABC algorithm will adapt the modification especially in searching mechanism and probability function. The modified algorithm is tested on general global optimisation functions. The result demonstrates that the proposed algorithm produces better convergence compared to the original algorithm.

**Keywords:** Optimisation, Artificial Bee Colony, Convergence.

## 1 Introduction

Recent years, swarms intelligent are becoming a trend to academicians in solving many complex problems. Swarm intelligent basically mimics the behaviour of natural swarm insects (such as ants, bees, and wasps) or other animal flocking (such as birds, and fish). Artificial Bee Colony (ABC) algorithm is one of these techniques that inspired from the bees' foraging activities. In this work, ABC algorithm is selected and adapts several modifications and then will undergo simulation to determine better value for optimisation. This algorithm was introduced by Karaboga in 2005 [1] gains so much attentions and popularity. As a result, numerous works have used the ABC algorithm in many optimisation problems.

ABC algorithm is one of optimisation algorithm and capable of achieving result equal to or better than classical optimisation algorithm [2][3] such as genetic algorithm. The algorithm also can solve optimisation problem that classical optimisation algorithm incapable of doing it [2][4]. However, one of the setbacks of ABC is the long convergence to produce the optimal answer. This work will propose a modified version of the ABC algorithm to enhance the existing ABC algorithm.

---

\* Corresponding author.

## 2 Modified Artificial Bee Colony

### 2.1 Artificial Bee Colony

There are two main stages. The first stage is the initialisation of the algorithm. In this stage, the initial value for the solution will be obtained by using Eq. 1 including the colony size, cycle, limit of the iteration and runtime. The second stage is the “bee” phase where the colony bee is divided into three groups employed bee (half of the colony), onlooker bee (another half) and scout bee. An employed bee is the bee that forage the food source. In the optimisation problem, this group of bee will be the seekers of the solution [5]. The area covered by employed bee is the area that it has visited before and the area around it. Employed bee will return after finding food location and tell the other bee with a waggle dance [6]. The dance is the method use by employed bee to inform the other bee about the food location. The onlooker bee is the bee that waits in the hive for employed bee. They follow the employed bee after the information is passed. The last group of bee is the scout bee where this bee does not have information about the food location but still searching randomly for the food [7]. The scout bee will only exist when employed bee’s solution cannot be improved and will be abandoned [2]. The best solution in a cycle will be remembered and evaluate in the next cycle for comparison. This iteration or cycle will repeat until the limit set for a solution is reached.

There are three main equations in general ABC Algorithm. The first equation is to initialize solution population:

$$x_{i,j} = lower_j + rand(0,1)(upper_j - lower_j) \quad (1)$$

where  $x_{i,j}$  is the solution with  $i = 1..NS$  is the possible solution which is equal to employed bee,  $j = 1..D$  is the dimension of solution space and  $rand(0,1)$  is a random number between 0 and 1. Lower and upper bound of  $x_{i,j}$  is represented by  $lower_j$  and  $upper_j$ . These lower and upper bound is the range that optimum value will be obtained. The next equation is use to determine neighborhood food source:

$$v_{i,j} = x_{i,j} + \varphi_{i,j}(x_{i,j} - x_{k,j}) \quad (2)$$

where  $v_{i,j}$  is the neighborhood food source or new solution of  $x_{i,j}$ ,  $\varphi_{i,j}$  is random number ranging from -1 to 1, and  $x_{k,j}$  is solution of neighboring food source. Greedy selection will be used to compare  $v_{i,j}$  and  $x_{i,j}$  for the best solution. Basically, Greedy selection will always select the next solution that offers obvious and immediate benefit. The probability of the solution being chosen as in Eq. 3:

$$P_i = \frac{fitness_i}{\sum_{i=1}^{NS} fitness_i} \quad (3)$$

where  $P_i$  is the probability of the food source being chosen,  $fitness_i$  is the fitness or objective function value of food source  $i$ . Fitness is calculated using Eq. 4:

$$fitness_i = \begin{cases} \frac{1}{1 + f_i}, & f_i \geq 0 \\ 1 + abs(f_i), & f_i < 0 \end{cases} \quad (4)$$

where  $f_i$  is the food source or objective function of  $x_i$  and  $P_i$  will be normalized to  $[0,1]$ . Algorithm 1 depicted the overall procedure of the algorithm.

---

### Algorithm 1: Artificial Bee Colony

---

Step 1: Initialize population using  $x_{i,j} = lower_j + rand(0,1)(upper_j - lower_j)$

Step 2: Evaluate population

Step 3: Set cycle = 1

Step 4: Repeat this

Step 4.1: For each employed bee

Step 4.1.1: Produce new solution using  $v_{i,j} = x_{i,j} + \varphi_{i,j}(x_{i,j} - x_{k,j})$

Step 4.1.2: Determine fitness of food source

Step 4.1.3: Compare  $v_i$  and  $x_i$  using Greedy Selection

Step 4.1.4: Determine probability using  $P_i = \frac{fitness_i}{\sum_{i=1}^{NS} fitness_i}$

Step 4.2: For each onlooker bee

Step 4.2.1: Choose solution  $x_i$  based on  $p_i$

Step 4.2.2: Produce new  $v_i$

Step 4.2.3: Determine fitness of food source

Step 4.2.4: Compare  $v_i$  and  $x_i$  using Greedy Selection

Step 4.3: If abandoned solution or food source exist, replace with new random produced solution  $x_i$  for scout using  $v_{i,j} = x_{i,j} + \varphi_{i,j}(x_{i,j} - x_{k,j})$

Step 4.4: Memorize current best solution

Step 4.5: Increased cycle by 1, cycle=cycle+1

Step 5: Until cycle reach number of maximum cycle

---

## 2.2 Modified Artificial Bee Colony

Enhancement and modification to this standard algorithm has been made by many researchers since it was first introduced. The modification and enhancement were made so that the result of algorithm specifically the mean value, fitness, and probability will give better result. Many researchers make modification or enhancement on the searching mechanism and probability value. The original theory is when fitness level is high; the probability of that fitness being chosen also will be high. This is because based on Eq. 3 probability is proportional to fitness value. Other than probability, many studies also make modification on searching mechanism. The reason of modification and enhancement for searching mechanism is that original and standard ABC algorithm does not work best on exploitation [4][8].

The proposed modified ABC algorithm are adopted from other researchers including from [2][4][8][9]. All result of hybrid algorithm will be compared to the standard ABC algorithm. For the search mechanism, the current global best solution to replace  $x_{i,j}$  in Eq. 2 as in [2],[4],[8], and [9]. Combining and modifying those in [1], [7], [8], and [9], the standard searching mechanism of Eq. 4 below:

$$v_{i,j} = x_{i,j} + \phi_{i,j}(x_{i,j} - x_{k,j}) + \varphi_{i,j}(y_j - x_{i,j}) \tag{5}$$

where  $\phi_{i,j}$  and  $\varphi_{i,j}$  are uniformly distributed random numbers such  $\phi_{i,j} \in [-1,1]$  and  $\varphi_{i,j} \in [0,1.5]$  while  $y_j$  is  $j$  th element of global best solution. This equation is inspired from the particle swarm optimisation[10] searching mechanism. This equation improved the convergence of the algorithm and recognized as HABC1.

Besides searching mechanism, probability equation also is changed. The standard probability equation as shown in Eq. 3 will be change to:

$$P_i = \exp\left(-\frac{1}{\rho} * fitness_i\right) \tag{6}$$

where  $fitness_i$  is the fitness value and  $\rho = 2.5$  as explained in [4] is called as HABC2. The final combination is to apply Eq. 5 and Eq. 6 in the algorithm and named as HABC3.

### 3 Simulation and Experimental Results

#### 3.1 Simulation Setting

In evaluating the algorithm ABC and HABC, fourwidely known global optimisation (GO) benchmark test function have been tested as listed in Table 2,namely Rosenbroock [11], sphere [12], Rastrigin [12] and Griewank [12].All simulation uses the same controlled parameters as simulation are tabulated as in the Table 1.

**Table 1.** Parameters of the simulation

Parameter	Value
Colony	50
Dimension	50
Limit	100
Cycle	3300
Runtime	10

#### 3.2 Simulation Result

Every test function has its own area range for the best optimum. They also have their own global minimum that is the  $x_i$  value. This  $x_i$  value will change according to the solution of the algorithm. There are more test function in GO but majority of

**Table 2.** Global optimisation test function

Function	Range	Equation
Sphere	[-5.12, 5.12]	$f(x) = \sum_{i=1}^n x_i^2$
Rosenbrock	[-2.048, 2.048]	$f(x) = \sum_{i=1}^{n-1} [100(x_{i-1} - x_i^2)^2 + (x_i - 1)^2]$
Rastrigin	[-5.12, 5.12]	$f(x) = 10n \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i))$
Griewank	[-600, 600]	$f(x) = \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$

**Table 3.** Best mean value of the tested functions

	Sphere	Rosenbrock	Rastrigin	Griewank
<b>Original</b>	1.787E-15	<b>2.637E+01</b>	9.620E-11	7.149E-10
<b>HABC1</b>	<b>1.550E-15</b>	3.522E+01	8.574E-11	3.752E-10
<b>HABC2</b>	1.643E-15	3.467E+01	5.064E-08	5.467E-10
<b>HABC3</b>	1.684E-15	3.207E+01	<b>1.755E-11</b>	<b>3.721E-10</b>

studies of ABC algorithm uses these four test function as the main test function. The result of simulation will determine which algorithm can produce optimum value in shorter time using the proposed hybrid ABC algorithm. Simulation will produce best mean value for the tested function. The simulation result for all four test functions is as in Table 3. The table shows the original ABC algorithm work best on the Rosenbrock test function. While as HABC3 is work best on Rastrigin and Griewank test function.

## 4 Discussion

The original version of the algorithm is still valid for optimisation problem. By adding more parameters to the original algorithm could lengthen the time taken to get the optimum value even though the additional parameters are meant to make the searching space larger so that the exploitation of the onlooker bee is better. The change of probability equation could not give optimum value on all test function as this was shown by the HABC2. But if we compared by test function, the best mean value can

be obtained using Sphere function for the probability equation changes. Finally, the result of HABC3 shows some confirmation that the proposed modified ABC algorithm by adapting and changing Eq. 2 and Eq. 3 at the same time can give better optimum result.

## 5 Conclusion and Future Work

In this paper, a hybrid version of ABC algorithm is proposed and named as HABC3. This hybrid algorithm applied changes to onlooker bee searching mechanism and to the probability equation. By plot, the results show better performance in HABC compared to the Original algorithm. This proposed algorithm has been tested in four general GO test functions for optimisation. From the simulation, we can say that whatever values were assigned to the parameters in Table 1, the results always in favour of HABC. This means that the original ABC could give better result if the searching mechanism include larger searching space. The probability equation also have effect to the result of the optimisation hence a proper changes to this equation could give better result compared to the original ABC. For the future work, the tuneable parameters will be the focus and tested to get the desired optimum value in term of performance and some other parameters that only give almost none or only small effect to the result will be reduce in values.

**Acknowledgement.** This work partly supported by PIBT research grant of Malaysian Ministry of Education under code S/O 12159.

## References

1. Karaboga, D., Gorkemli, B.: A Quick Artificial Bee Colony -qABC- Algorithm for Optimization Problems. Paper presented at the Innovations in Intelligent System and Applications (INISTA), pp. 1–5 (2012)
2. Karaboga, D., et al.: Cluster Based Wireless Sensor Network Routings using Artificial Bee Colony Algorithm. Paper presented at the International Conference on Autonomous and Intelligent System (AIS), pp. 1–5 (2010)
3. Pham, D.T., et al.: The Bees Algorithm - A Novel Tool for Complex Optimisation Problems. In: Proceedings of the 2nd International Virtual Conference on Intelligent Production Machines and Systems (IPROMS 2006), pp. 454–459 (2006)
4. Alam, M., Islam, S., Monirul, M.: Artificial Bee Colony Algorithm with Self-Adaptive Mutation: A Novel Approach for Numeric Optimization. Paper presented at the TENCON 2011 - 2011 IEEE Region 10 Conference, pp. 49–52 (2011)
5. Baykasoglu, A., et al.: Artificial Bee Colony Algorithm and Its Application to Generalized Assignment Problem. In: Chan, F.T.S., Tiwari, M.K. (eds.) *Swarm Intelligence, Focus on Ant and Particle Swarm Optimization*, pp. 532–564 (2007)
6. Karaboga, D.: An Idea Based on Honey Bee Swarm for Numerical Optimization. Erciyes University. Technical Report-TR06 (2005)



7. Babayigit, B., Ozdemir, R.: A modified artificial bee colony algorithm for numerical function optimization. Paper presented at the IEEE Symposium on Computers and Communications (ISCC), pp. 245–249 (2012)
8. Kong, X., Liu, S., Wang, Z.: A new hybrid artificial bee colony algorithm for global optimization. *International Journal of Computer Science Issues* 10(1), 287–301 (2013)
9. Pei, W.S., et al.: Enhanced Artificial Bee Colony Optimization. *International Journal of Innovative Computing, Information and Control* 5(12), 1–12 (2009)
10. Rosenbrock, H.H.: An automatic method for finding the greatest or least value of a function. *The Computer Journal* 3, 175–184 (1960)
11. Yang, S., et al.: Investigation of Neural Networks for Function Approximation. In: *Procedia Computer Science 17 Information Technology and Quantitative Management (ITQM 2013)*, pp. 586–594 (2013)
12. Kennedy, J., Eberhart, R.: Particle swarm optimization. Paper presented in the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)

# Ontology-Supported Development for Drug Analysis Laboratory Corresponding to the ISO/IEC 17025 Standard

Orapan Apirakkan<sup>1</sup>, Wanna Sirisangtragul<sup>2</sup>, and Pusadee Seresangtakul<sup>1,\*</sup>

<sup>1</sup>Department of Computer Science

<sup>2</sup>Department of Forensic Science, Faculty of Science

Khonkaen University, 123 Mittraphap Road, Muang, Khonkaen, 40002, Thailand

aaorapan@gmail.com,

{wannas, pusadee}@kku.ac.th

**Abstract.** This paper introduces the development of an ontology that will support the work procedures of a drug analysis laboratory for the accreditation according to the ISO/IEC 17025 standard. This is an international standard for the competence of testing and calibration laboratories. The ontology conformed to both management and technical requirements and it represents scientific observational data and management workflow. The benefits of this ontology are that it will formalize the experimental description, allow for the sharing of data between the analysts, and allow data interoperability within the framework of the ISO/IEC 17025 standard.

**Keywords:** ISO/IEC 17025, Ontology development, Accreditation, Knowledge representation.

## 1 Introduction

A drug analysis laboratory is a testing laboratory that provides testing solutions and guarantees the quality of finished drug products. Therefore, quality control for quality management systems, administrative and technical operations is a requisite. An accreditation to International Organization for Standardization / International Electrotechnical Commission 17025 (ISO/IEC 17025), which is an international standard for the competence of testing and calibration laboratories [1], is a requirement. The monetary benefits of accreditation are increased customer satisfaction, improved reliability of results and staff qualifications, a decrease in equipment malfunctions, international recognition between countries of testing and calibration results. As part of the laboratory accrediting, the committee will also determine the laboratory's testing process in compliance with the requirements of the ISO/IEC 17025 guidelines. These assessments include evaluation of laboratory quality systems and management systems. The data from testing, calibrating,

---

\* Corresponding author.

validating methods, chemical suppliers, environmental conditions etc. should be collected for audit trial, data analysis and traceability of result. Observational data naturally appears in heterogeneity format for many reasons, e.g. individual analysts, institutions, collection methods or format requirements of analysis. This shows that many data sets should be controlled, collected and/or exchanged in order to improve the quality of the laboratory's management workflow.

Ontology is a representational model of the explicit formal terms in any domain knowledge that identifies relationship between concepts in domain. To share common understanding of the structure of information among human or software agents is one of the more common goals in developing ontologies (Musen 1992; Gruber 1993) and it is a basic tool for adding semantic remarks to scientific observational data[8,9]. With regard to semantic interoperability, ontology is a useful technique for support the laboratory accredit, create uniform structure and unify access of observational data. This technique advocates a comparative testing between laboratories, which is one of the requirements in the standardization process. The most correct method for developing ontologies is to have them developed by experts in their field[6,7,8].

As mention above, the researchers tried to develop ontology of drug analysis's domain knowledge for support the quality workflow of laboratory toward the ISO/IEC 17025 standard. We used an ontology editor tool called "Hozo" to develop. The Hozo is an effective ontology editor tool based on a role-concept [10] and is open source. The paper is organized as below: Section 2 is a description of an important theory of quality management system of the ISO/IEC 17025, which was used to build the ontology and literature reviews of related works. Section 3, we introduce the framework of the Drug analysis's ontology building. Section 4 covers the experimental results and discussion. Finally, the conclusions and future work will be informed in section 5.

## 2 Related Works

The ISO/IEC 17025 standard is an international standard for testing and calibrating laboratories. The main purpose of cooperation for the accreditation system is the harmonization of international trade for the acceptance of products that are tested by an accreditation laboratory. In order to receive certification to the ISO/IEC 17025 standard, test results procured from that laboratory have to be confirmed as reliable [2, 3, 4]. The ISO/IEC 17025 standard consists of five parts [1], namely, scope, normative references, terms and definitions, management requirements and technical requirements. The management and technical requirements are the main contents of the standard. The management and technical requirements consist of 15 and 10 items, respectively. Inês Hexsel and Carla Schwengber [5] identified activities that related to 16 testing laboratory processes, which correspond to items of the ISO/IEC 17025 standard.

Inês Hexsel and Carla Schwengber [5] introduced quality management systems (QMS) for the accrediting laboratories in the university. During implementation of the

QMS, the authors analyzed the laboratory’s processes that met the requirements of the 17025’s standard and identified their input, output, stakeholders and related activities. The processes were made up of 16 sub-processes, which were grouped into five macro processes. These processes are referred in this paper as a template for creating our ontology.

L.N. Soldatova and R.D. King [8] developed the general ontology of scientific experiment, EXPO, which was used to formalize knowledge about scientific experiments. The EXPO is a part of general ontology of science; it can represent all fields of science. In order to test the EXPO effectiveness, they applied EXPO to the different experiments of scientific domain and found that EXPO’s classes still cover the essential parts of the experiments.

### 3 The Frameworks of an Ontology Building

The ontology-supported drug analysis laboratory corresponding to the ISO/IEC 17025 standard development is composed of four steps, which are shown in Figure 1.

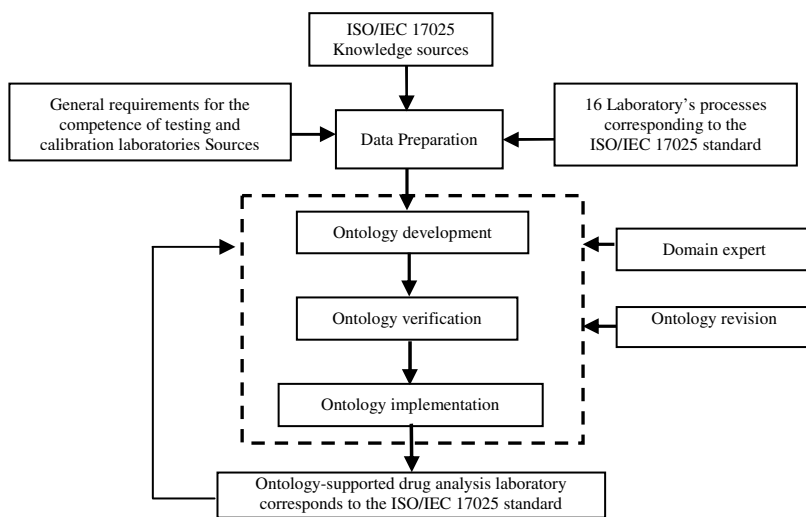


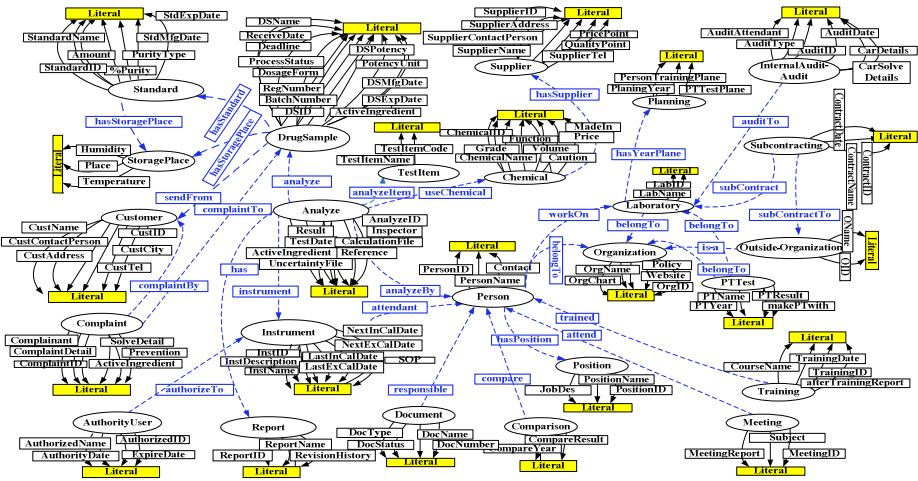
Fig. 1. The framework of ontological building

The data preparation step is the concepts extraction process. This data set is extracted from processes and activities of testing laboratory that correspond to the items of the ISO/IEC 17025 standard as defined by Inês Hexsel and Carla Schwengber [5]. The concepts received from 16 processes are specific to a drug analysis laboratory. As a result of the data preparation process, we extract the process into 25 concepts, which are shown in Table 1.

At the ontology development and verification stage, basic properties were added to core concepts, where the properties are composed of part-of relation, attribute-of relation and is-a relation (hierarchical relation). Following this, the laboratory’s analyst and quality manager for the 17025’s standard checked the ontology model; the whole relations are shown in Figure 2.

### 4 Experimental and Results

After developing the Ontology-supported drug analysis laboratory, our ontology contains 25 concepts, 119 data type properties and 30 object properties as shown in Figure 2. All of the concepts and properties are covered by the ISO/IEC 17025 standard requirements as shown in table 1.



**Fig. 2.** The ontology-supported drug analysis laboratory, which is corresponds to the ISO/IEC 17025 standard, where dashed line is a part-of relationship and normal line is an attribute-of relationship

In order to test the effectiveness of the proposed ontology, an actual drug analysis experiment was carried out and monitored in the laboratory. The selected drug, Glipizide 5 mg, was is in tablet form and it was this drug that the laboratory had introduced into the ISO/IEC 17025 standard process.

From the experiment, an annotated ontology of the data analysis can make the data more explicit. The ontology classes can be grouped into 3 parts based on the link properties between classes. The 3 parts are: 1) Receive and store sample 2) Analyze sample and 3) Report sample and stakeholder. The inclusions of ontology over the process of drug analysis are shown in table 2.

**Table 1.** The list of Concepts created from macro process of the ISO/IEC 17025 requirements

Macro process	Process	The 17025's Standard items	Concepts create
	Activities		
Product realization	<b>1. Review of requests, tenders and contracts</b>		
	- Contact client - Specify requirement, agreement, method, price of drug analysis	4.4 4.4	Customer DrugSample, TestItem, Analyze
	<b>2. Testing</b>		
	- Collect, receive, identify, handle, store sample - Subcontract the test	5.1, 5.8, 5.7 4.5	DrugSample, StoragePlace, Standard, Subcontracting, Outside-organization
System management	<b>3. Reporting the results</b>		
	- Analyze data, write, protect, send report	5.10	Analyze, Report
	<b>4. Management responsibility</b>		
	- Meet organization requirements - Establish policy and objectives - Plan, execute and record manage review meeting	4.1 4.2 4.10, 4.15	Organization Laboratory, Organization Planning, Meeting
Resource management	<b>5. Information management</b>		
	- Issue, approve, distribute and manage document - Identify, collect, index, store, dispose records	4.3, 4.13, 5.4, 5.10	Document Document
	<b>6. Purchase</b>		
	- Evaluate suppliers, specify requirement for purchase, accomplish purchase and check compliance with specifications	4.6	Supplier, Chemical
	<b>7. Personnel</b>		
	- Hire, train, evaluate, and authorize staff - Describe personnel functions	4.1, 5.1, 5.2	Person, Position, Training, AuthorityUser, Planning
	<b>8. Infrastructure</b>		
	- Monitor, control and record environment conditions - Control the access to areas	5.1, 5.3	Laboratory
	<b>9. Methods</b>		
	- Acquire, create, validate, implement and use testing methods - Estimate measurement uncertainties	5.1, 5.4	Analyze
Effectiveness of quality management system and its improvement	<b>10. Equipment</b>		
	- Acquire, identify, monitor, maintain and calibrate equipment and instruments - Elaborate procedures for safe handling, transport, storage, use and planned maintenance of measuring equipment	5.1, 5.5	Instrument AuthorityUser, Instrument
	<b>11. Customer service</b>		
	- Provide reasonable access to laboratory, provide a guide to the preparation, packaging and dispatch of test items - Maintain communication throughout the process, seek feedback through customer satisfaction	4.7	Laboratory, Document Customer
	<b>12. Complaints</b>		
	- Record and solve complaints	4.8	Complaint
Quality assurance of tests	<b>13. Nonconformities, corrective and preventive actions</b>		
	- Record and analyze nonconformities - Plan and execute corrections - Ensure corrective and preventive actions	4.9, 4.10, 4.11, 4.12	Complaint
	<b>14. Audits</b>		
	- Plan, conduct and record internal audits	4.10, 4.14	Planning, InternalAudit-Audit
Quality assurance of tests	<b>15. External QC</b>		
	- Establish program and procedure for the calibration of equipment and acquisition of standards - Participate in interlaboratory comparison or proficiency testing programs and analyze the laboratory performance	5.1, 5.6, 5.9	Planning, Instrument PTTest
	<b>16. Internal QC</b>		
- Provide intermediate checks to maintain confidence in the calibration status of equipment and reference standards - Develop intralaboratory comparisons - Establish quality control procedures as regular use of CRM - Replicate testing or retesting of retained items among others - Analyze obtained data	5.5, 5.6, 5.9	Instrument, Comparison, Analyze	

**Table 2.** The ontology formalization of drug analysis process (laboratory analysis part)

Receive and storage sample		Storage	DessicatorNo.1	Testitem	Assay
<b>DrugSample</b>	Glipizide	<b>humidity</b>	NMT 80%RH	<b>testItemName</b>	assay
<b>activeIngredient</b>	glipizide	<b>temperature</b>	+25 °C	<b>Chemical</b>	Methanol
<b>custID</b>	customer001	<b>place</b>	dessicatorNo.1	<b>chemicalID</b>	met003
<b>storagePlace</b>	dessicatorNo.1	<b>Analyze sample</b>		<b>chemicalName</b>	methanol
<b>dsName</b>	AAAA	<b>Analyze</b>	Glipizide	<b>grade</b>	HPLC
<b>receiveDate</b>	07dec2012	<b>analyzeID</b>	sample001	<b>function</b>	mobilephase
<b>deadline</b>	07Jan2013	<b>analyzeDrug</b>	glipizide	<b>volume</b>	500ml
<b>processStatus</b>	in-process	<b>activeingredient</b>	glipizide	<b>madeIn</b>	Thailand
<b>dosageForm</b>	tablet	<b>instrument</b>	HPLC-UV detector	<b>price</b>	650bath
<b>dsPotency</b>	tablet	<b>testIitem</b>	assay	<b>supplier</b>	KKcompany
<b>potencyUnit</b>	5	<b>byPerson</b>	MS.CCC	<b>caution</b>	methanolMSDS.doc
<b>mfgDate</b>	Mg	<b>testDate</b>	10Dec2012	<b>Report sample and stakeholder</b>	
<b>expDate</b>	10Sep2012	<b>result</b>	pass	<b>Person</b>	MS.CCC
<b>regNumber</b>	10Sep2017	<b>reference</b>	USP 35 NF 30	<b>contact</b>	087-5546544
<b>batchNumber</b>	AA/AAAA	<b>chemical</b>	methanol	<b>workOn</b>	drug analysis lab.
<b>Standard</b>	AA001	<b>calculationFile</b>	glipizide001.xls	<b>position</b>	analyst
<b>standardName</b>	Glipizide	<b>inspector</b>	MS.DDD	<b>positionName</b>	analyst
<b>amount</b>	standard	<b>uncertaintyFile</b>	uncer_glipizide001.doc	<b>jobDescription</b>	jdAnalyst.doc
<b>%purity</b>	glipizide ws	<b>report</b>	report001	<b>Customer</b>	Customer001
<b>purityType</b>	200 mg	<b>Instrument</b>	HPLC UV-Detector	<b>custName</b>	AAAFactory
<b>stdMfgDate</b>	99.99	<b>attendant</b>	No.10	<b>contactPerson</b>	Mr.AAA
<b>stdExpDate</b>	as it is	<b>nextInCal</b>	Mr.EEE	<b>address</b>	BBB
<b>storagePlace</b>	02Feb2010	<b>lastInCal</b>	16Aug2013	<b>city</b>	Bangkok
	02Feb2014	<b>nextExCal</b>	16Aug2012	<b>tel</b>	02-xxxxxxx
	Refrig.No.1 2-	<b>lastExCal</b>	15Feb2014		
	5°C		15Feb2013		

## 5 Conclusions and Future Works

The purpose of this paper was to develop ontology to support the workflow of drug analysis laboratory corresponding to the ISO/IEC 17025 standard. The ontology covers laboratory analysis and laboratory management that conforms to the 17025's standard requirements. Future work will map this ontology to the database for use in a semantic search system and searching for the ontology's instances.

## References

1. ISO/IEC 17025 (2005) General requirements for the competence of testing and calibration laboratories. International Organization for Standardization (ISO), 2nd edn., Geneva (May 2005)
2. Cortez, L.: The implementation of accreditation in a chemical laboratory. Trends Anal. Chem. 18, 638–643 (1999)
3. International Laboratory Accreditation Cooperation (ILAC). The ILAC mutual recognition arrangement (2013), [http://www.ilac.org/documents/Bro\\_english/ILAC\\_MRA\\_English.pdf](http://www.ilac.org/documents/Bro_english/ILAC_MRA_English.pdf) (accessed May 23, 2013)
4. Bureau of Laboratory Accreditation (2013), <http://www.dss.go.th/ewdss/index.php/2013-05-21-05-47-59/eng-bla-background> (accessed May 25, 2013)

5. Hessel, G.I., Schwengber, T.C.C.: A process approach to ISO/IEC 17025 in the implementation of a quality management system in testing laboratories. *J. Accreditation and Quality Assurance* 17(5), 519–527 (2012)
6. Brinkman, R., Courtot, M., Derom, D., Fostel, J.M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Soldatova, L.N., Stoeckert Jr., C.J., Turner, J., Zheng, J.: OBI consortium, Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics* 1(suppl. 1), S7 (2010)
7. Heeptaisong, T., Srivihok, A.: Ontology Development for Searching Soil Knowledge. In: *The 9th International Conference on eBusiness (iNCEB 2010)*, Bangkok, Thailand, November 18-19 (2010)
8. Soldatova, L.N., King, R.D.: An ontology of scientific experiments. *J. of the Royal Society, Interface* 3(11), 795–803 (2006)
9. Bowers, S.: Scientific workflow, provenance, and data modeling challenges and approaches. *Journal on Data Semantics* 1(1), 19–30 (2012)
10. Mizoguchi, R., Sunagawa, E., Kozaki, K., Kitamura, Y.: The model of roles within an ontology development tool: Hozo. *J. of Applied Ontology* 2, 159–179 (2007)



# Modeling the Natural Capital Investment on Tourism Industry Using a Predator-Prey Model

Zhonghua Cai<sup>1,\*</sup>, Qing Wang<sup>1</sup>, and Guangqing Liu<sup>2</sup>

<sup>1</sup> College of Economic Management

<sup>2</sup> Biomass Energy and Environmental Engineering Research Center  
Beijing University of Chemical Technology

No.15 North 3<sup>rd</sup> Ring Road East, Chaoyang District, Beijing, China  
caizh@mail.buct.edu.cn,  
wqing90@126.com, guangqing.liu@gmail.com

**Abstract.** Based on the predator - prey model and the correlation between natural capital and physical capital, this paper explores the natural capital investment efficiency in tourism. By determining the effect of the natural capital investment rate on tourism development, the paper shows the impact of physical capital on tourist attraction and environmental degradation through the model simulation. Furthermore, we use the real data of Guangxi Guilin city and the Anhui Huangshan for model validation. The simulation and empirical results show that by appropriate proportion of natural capital investment, it will take a long-term growth, and achieve large number of tourists as far as possible.

**Keywords:** Natural Capital, Predator-Prey, Tourism, Effect of Investment.

## 1 Introduction

Tourism industry which developed rapidly made use of natural resources directly and renewably, namely natural capital possessed of the environmental services and their external effects provide recreational and cultural services. So the tourism industry should be a minimum conflict with environment protection and ecological civilization construction. However, due to human utility, system design limitations, non-identity of thought and action and other reasons [1], there has been some alienation in the development process of tourism with various contradictions and conflicts. Tourist overloaded many tourist attractions seriously, not only affected the quality of tourism, but also impact the scenic environment and bring environmental pollution, resource degradation. Domestic tourism market is a mess, and the healthy development of the tourism industry in urgent need of our correct attitude.

However, because ecological environment is a complex whole, there are linear and nonlinear interactions between each part. In order to achieve sustainable development of ecological, economic and social together, the theory of system dynamics is needed. Predator - Prey model has been widely studied, since Lotka and Volterra constructed the classic model in the 1930s and the publication of *Silent Spring* [2]. It prompted

---

\* Corresponding author.

scholars have started to use predator - prey model study on tourism development, with the development of Predator - Prey model and gradually attention to tourism. The analysis of Casagrandi etc. [3] is purely theoretical and based on very simple and general assumptions about the interactions between the three main components of the system: the tourists, the environment, and the capital. Juan M. Hernandez and Carmelo J. Leon (2007, 2013) [4-5] presented a model of the interactions between natural resources and physical capital in the evolution of a tourist destination. Obtain different patterns of the post-stagnation phase, depending on the impacts of physical capital on both demand and environmental degradation.

Based on previous studies, this paper supplemented the interrelationship between natural capital and physical capital in tourism industry, and put forward the benefits of natural capital investment, to explore the sustainability of the tourism economy development. The structure is as follows, the second part is the model; the third part is the simulation; the fourth part is empirical example; finally is the conclusion.

## 2 Model

The model is based on the achievement from Juan M. Hernandez and Carmelo J. Leon (2007) [4]. Tourist attraction can be determined by two aggregated elements which are considered: physical capital, which includes hotels, transport, services and facilities, and natural capital, which is the environmental value of the area. Fig.1 shows the correlation between physical capital and natural capital.

### 2.1 The Model of Physical Capital

The upper half of Fig. 1 shows the flow diagram for the physical capital. Physical capital grows through the reinvestment of revenues, which are determined by the product between the "Revenue rate", and the number of "Visitors" at any period of time. The number of tourists attracted depends on the combination of both types of capital, which evolve along the lifecycle.

Both capitals are combined by means of a multiplicative function where parameter "ShareT" indicates the role of physical capital in tourist attraction. Thus, if K represents physical capital, X natural capital,  $\epsilon$  is the "Exploitation rate" and  $\alpha$  "ShareT", the "Visitors" (V) function follows the expression, where h stands for the "Rest of factors".

$$V = hK^\alpha (\epsilon X)^{1-\alpha} \quad (1)$$

The exponent  $\alpha$  determines the degree of dependence of the tourist activity from physical capital. With the increased awareness of environmental protection, the natural capital investment process has been raised, and assumed p "natural capital reinvestment". Then r and (1-p) symbolize "Revenue rate" and "Physical capital reinvestment proportion", respectively, the dynamic evolution of the physical capital K follows the equation

$$\dot{K} = (1-p)rhK^\alpha (\epsilon X)^{1-\alpha} - \delta_K K \quad (2)$$

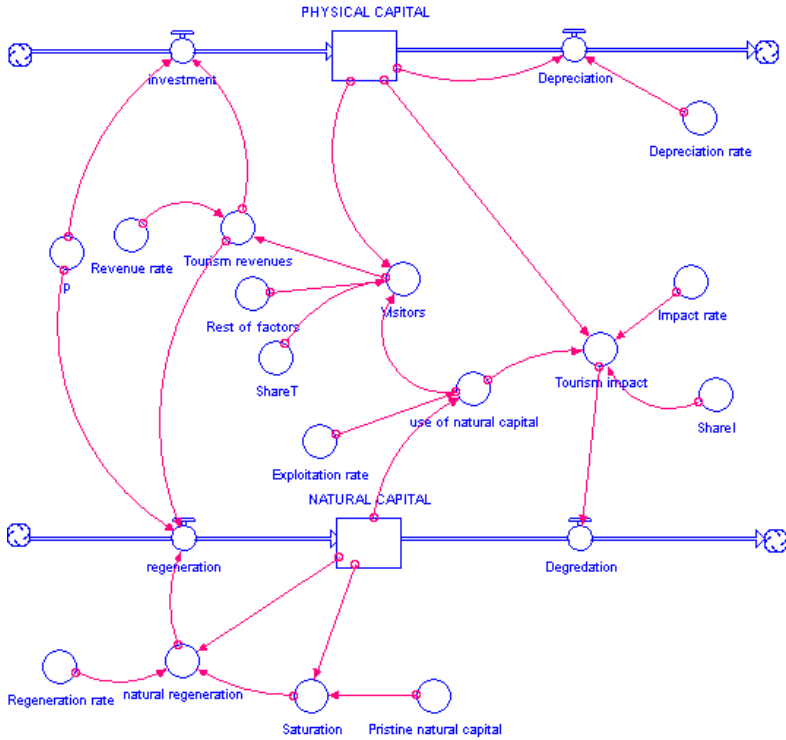


Fig. 1. The relationship between natural capital and physical capital

### 2.2 The Model for the Natural Capital

The dynamic evolution of the natural capital is represented in the lower part of Fig. 1. The stock of the natural resource grows through the inflow “Natural regeneration”. This flow represents the capacity of the environment to absorb pollution or degradation. The maximum level for the “Natural capital” (X) is indicated by the “Pristine natural capital” ( $\hat{X}$ ).

Accordingly, a logistic regeneration structure for natural capital is presented. Thus, the natural capital grows through a “Regeneration rate” ( $\delta_x$ ). However, this explosive loop is limited by a negative feedback loop through “Saturation”, which represents the distance of the actual level of natural capital to its maximum level. The “Natural regeneration” (R) can be described by the function

$$R = \delta_x X(\hat{X} - X) \tag{3}$$

The two capitals are connected through the “Degradation” factor. The natural capital is negatively affected by the existence of physical capital. However, the “Use of natural capital” also influences the environmental value. Therefore, natural and physical capitals determine the “Tourism impact”. Parameter “ShareI” indicating the relevance

of the physical capital in the depletion of the natural resource, and determines the type of tourism. Thus,

$$I = dK^\beta (\epsilon X)^{1-\beta} \tag{4}$$

where  $\beta$  is “ShareI”,  $I$  is the “Tourism\_impact”, and  $d$  is the “Impact\_rate” which transforms the combination of physical and natural capital in terms of natural damage. The dynamic evolution of the natural capital with investment to natural capital is,

$$\dot{X} = prhK^\alpha (\epsilon X)^{1-\alpha} + \delta_x X(\hat{X} - X) - dK^\beta (\epsilon X)^{1-\beta} \tag{5}$$

The dynamical system (2) and (5) represents a typical predator–prey structure. The physical capital, the predator, needs of the natural capital, the prey, for its growth.

### 3 Simulation

We should check the model’s validity by simulation and empirical data, because of the abstraction and theoretical property of the model. Compared with the qualitatively different development scenarios discussed by Butler in his work on life cycle (Butler 1980) [6] is concluded, which has been widely accepted in a variety of empirical contexts.

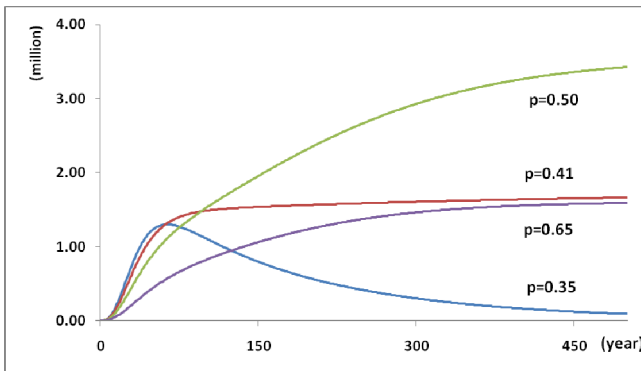


Fig. 2. Simulations of tourists depending on p

Fig. 2 shows the output of the model for the number of tourists visiting a destination according to four different assumptions regarding the “natural capital investment proportion” parameter  $p$ , to analysis the result of natural capital investment.

As can be seen in the simulated results, all trajectories for the number of tourists show an initial logistic shape, successfully replicating the tourist product lifecycle. However, the assumptions about parameter  $p$  lead to different results.

In a short time, the higher  $p$  value set, the lower growth rate of tourists is, i.e. we can forecast the number of tourists by  $p$  value in the beginning stage.

In a long-term, the number of tourists rises rapidly at the beginning, and a deep decline reached lower but stable values than the peak of the trajectory when  $p < 0.5$ . The lower  $p$  value set, the lower tourists stable is at last. When  $p > 0.5$ , the growth rate of tourists decrease slowly so that tourists slow-grew in a long time and stable at last.

### 4 The Empirical Examples

In order to compare the performance of the simulation model against empirical data, the case of Guilin and Huangshan was considered. China has enforced the opening to the outside world since 1978, with many tourism attractions developed, likewise Guilin in Guangxi province and Huangshan in Anhui province.

The data of 28 years from statistical yearbook of Guangxi and the number of foreign visitors to Huangshan during the periods of 21 years from statistical yearbook of Anhui were matched simulation with STELLA. Fig. 3 shows the real data and simulated data for the case of Guilin and Anhui. And the simulation model has been calibrated to represent the growth patterns of the two tourism cities.

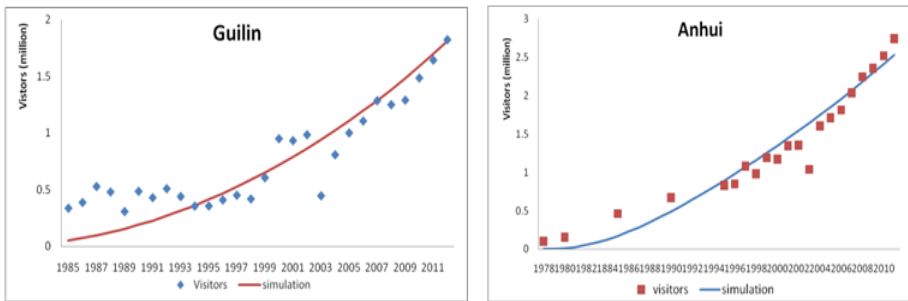


Fig. 3. Real data and simulated data for the case of Guilin and Anhui

A nonparametric statistical validation has been recommended by system dynamics modelers (Barlas, 1989) [7] for the comparison between simulated and real data, as is shown in Table 1. The Theil's inequality coefficient (U) is used to measure the gap between two series, which is combined with the coefficient of determination ( $R^2$ ) and the root mean square error (RMSE). The latter can be split in three percentage components: bias, variation and covariant, indicating where the error takes place.

Table 1. Statistical for the comparison between real and simulated data

	Guilin	Huangshan
	$(\alpha=0.69, \beta=0.64, p=0.64)$	$(\alpha=0.63, \beta=0.47, p=0.45)$
$R^2$	0.91	0.97
RMSE	190629.96	142082.24
Theil coefficient	0.106	0.046

The calibration of the impact parameters classified both of the tourism cities. For Guilin, the best representation was obtained with high values of the impact of physical capital on both tourist attractions and environmental degradation ( $\alpha=0.69$ ,  $\beta=0.64$ ). However, Huangshan presents a high value of the impact of physical capital on tourist attraction, but a lower value of the impact of physical capital on environmental degradation ( $\alpha=0.63$ ,  $\beta=0.47$ ).

## 5 Conclusion

Due to the direct and regenerative use of natural resources by tourism, it is very good for environmental protection in the process of tourism development, by studies on the role and correlations of natural capital and physical capital in tourism. Tourisms take full use of both capitals, while the overusing will lead to degradation of natural capital. The incomes of tourism should feedback to natural capital to steady the sustainable development of tourism system.

Both the natural capital and physical capital attract tourists, and the number of tourists impacts both capitals. Appropriate but not excess tourists deserve recommendation for natural resource protection and tourism sustainable development. It is contributed to environment protection in tourism development to analyze the relationships and the results of natural capital investment in tourism.

It will minimize the instability of tourists, take a long-term growth, and achieve large number of tourists as far as possible, to determine the appropriate proportion of natural capital investment according to the model.

## References

1. Xia, H., Liu, X., Cao, S.: Analysis of tourism development and ecological civilization construction. *Research and Development* 165(2), 73–76 (2013)
2. Rachel, C.: *Silent Spring*. Translated by Ruilan L., Changsheng L. Science (1979)
3. Casagrandi, R., Rinaldi, S.: A theoretical approach to tourism sustainability. *Conserv. Ecol. Rinaldi* 6(1), 13 (2002)
4. Hernandez, J.M., Leon, C.J.: The interactions between natural and physical capitals in the tourist lifecycle model. *Ecological Economics* 62, 184–193 (2007)
5. Juan, M.H., Carmelo, J.L.: Welfare and environmental degradation in a tourism-based economy. *Tourism Economics* 19(1), 5–35 (2013)
6. Butler, R.: The concept of a tourist area cycle of evolution: implications for management of resources. *Can. Geogr.* 24, 5–12 (1980)
7. Barlas, Y.: Multiple test for validation of system dynamics type of simulation models. *Eur. J. Oper. Res.* 42, 59–87 (1989)

# Study on Tracking Derivative Based Method for DC System Grounding Fault Detection

Wang Peng<sup>1,\*</sup>, Li Huashun<sup>2</sup>, Wang Ning<sup>3</sup>, Qu Linkui<sup>3</sup>, and Yu Yin<sup>4</sup>

<sup>1</sup> Electric Engineering School of Northeast Dianli University, Jilin 132012, China

<sup>2</sup> Jilin Power Supply Company, Jilin 132011, China

<sup>3</sup> Tieling Power Company, Tieling 112000, China

<sup>4</sup> China Southern Power Grid EHV Power Transmission Company,  
Liuzhou 545006, China

wangpeng8776@163.com

**Abstract.** Low-frequency signal injection and leakage current measurement are two main methods for DC ground fault detection. The system capacitance has great impact on the ground fault detection accuracy in low-frequency signal injection method, and the leakage current measurement base on the absolute-value detecting also have influence by temperature drift, zero drift and the residual magnetism. For the deficiencies in wavelet detection method, a tracking derivative based method for dc system grounding fault detection is proposed, which avoids the influence by ground capacitance and temperature drift, zero drift and the residual magnetism in current sensor. The simulation and application results show that, this method has rapid detection, little calculation and accurate result, the operational status of each branch are fully reflected, it is a new ground fault detection method for DC system.

**Keywords:** Tracking Derivative based Method, DC System, Fault Detection.

## 1 Introduction

DC system in power plants and substation is a very large and complex multi-branch power supply network, its ground fault has been ranked first in the rate of electrical faults, ground fault variety, the situation is complex, it is a very serious potential danger for the safe operation of the power system. DC system provide reliable uninterrupted power supply for relay protection, automatic device, the signal circuit and other secondary devices in power plant and substation, which directly affects the normal operation of primary and secondary equipment in power system[1-3], the branch grounding of DC system is the most common faults, if the DC system ground fault occurs at one place without treated, which can cause maloperation or miss trip in device, as well as damage to the equipment, widespread power outages, and even serious collapse of the system consequences.

---

\* Corresponding author.

At present, the low-frequency signal injection method detection method is the low frequency signal injection method [4-5] (usually 10 ~ 30 Hz), since the high-order harmonic and fundamental frequency component interference signal in DC system submerge the low-frequency signals when the capacitance to ground greater than  $2\mu\text{f}$ . It difficult to accurately measure the value of ground resistance and ground capacitance, and the low-frequency signal injection method can not detect because the capacitance to ground is often greater than  $2\mu\text{f}$ . Leakage current measurement using a sensor absolute measurement, its resolve technical on the sensor temperature drift, zero drift and the residual magnetism is still very difficult.

Domestic scholars have done a lot of research on DC system ground fault detection, advanced signal processing methods have been applied to the detection of branch grounding resistance [6-8], especially, wavelet-based time-frequency analysis method has a wide range of applications, which improves the detection accuracy, but in the case of large capacitance to ground, its detection accuracy should be improved.

A neural network detection system based on the wavelet analysis is given in [6], in which, wavelet entropy is used as the characteristic parameter to achieve the intelligent DC system grounding fault identification. The ground fault detection device which embedded ARM microprocessor and based on wavelet transform is addressed in [7]. Based on wavelet transform and fractal theory, concavo-convex parameter and fractal dimension are used to detect grounding fault of looped network is proposed in [8].

Wavelet-based detection method can better suppress the noise, but in the case of low SNR signal distortion is more serious, and it is not easy to select the wavelet function, the lack of large computation [9-11]. Dimensional linear tracker is an infinite impulse response filter, which is based on the principle of minimum variance with the gradient descent method, and then obtained by rotation transform [13-15], It has amplitude-frequency characteristic and phase-frequency characteristics, can extract a specific frequency of the input signal, and obtain the estimates of signal's amplitude and phase.

The tracking derivative based method for DC system grounding fault detection is proposed, which collected disturbance voltage and current signal in order to obtain the voltage and current differential signal to calculate the value of insulation resistance, by using the simple method of least squares linear fit to calculate the DC system grounding resistance value and avoid the adverse effects of the ground capacitance. The simulation and application results show that, this method is simple, has high precision, strong anti-interference ability and little calculation, the operational status of each branch are fully reflected.

The tracking derivative based method for DC system grounding fault detection is proposed, which collects disturbance voltage and current signal in order to obtain the voltage and current differential signal to calculate the value of insulation resistance, by using the simple method of least squares linear fit to calculate the DC system grounding resistance value and avoid the adverse effects of the ground capacitance.



The simulation and application results show that, this method is simple, has high precision, strong anti-interference ability and little calculation, the operational status of each branch are fully reflected.

The ground capacitance of DC system in substation is often greater than  $2\mu\text{f}$ , it is the parallel values of all distributed capacitance in each branches, and the distributed capacitance of each branch is small. Under this condition, using transient component of current signal to detect the ground fault in DC system is proposed. When put sense resistor on the positive and negative bus, the drop in insulation fault slip sensor will decrease as the degree of insulation, resulting in different sizes of transient current process.

## 2 Physical Characteristics and Expression for Derivative Tracking Method

Now the request of grid intelligent, automated signal during the execution can not analysis process data by statistical probability, such information can only be used as a reference to human judgment. Intelligent, automated execution signal should ensure to be true value.

In electrical theory: Voltage  $U$ —voltage of positive and negative bus to ground ,  
Current  $I$ —sensor current,  
 $R$  —insulation resistance.

$$R = \frac{U}{I} \quad (1)$$

This expression gives the relationship of absolute value measured, which is influenced by zero-regulator circuit values and adjust the magnification (full scale) circuit values.

The Derivative tracking method applied a disturbance signal on the bus and measured the responding:  $\Delta U$ -- the change in bus voltage to ground,  $\Delta I$ —the change in Sensor current, then the insulation resistance can be expressed as function (2)

$$R = \frac{\Delta U}{\Delta I} \quad (2)$$

The expression above gives the relationship of relative value measured, which does not need the zero-regulator circuit and adjusts the magnification circuit. So it uses the disturbance signal basic point as a reference point, which gives the solution on temperature drift, zero drift and the residual magnetism in absolute value measurement. The insulation resistance can be obtained through this method.

Basic theory of Derivative tracking method is shown above, Digital simulation and measurement confirmed its feasibility. The formula is simple, practical, only use simple filtering, without any complex process and wavelet analysis, can get the results quickly and accurately.

### 3 Relationship between the Resistance and Capacitance to Ground

Due to the big capacitance to ground in the DC system will affect the detection accuracy of grounding resistance, and easily lead to misjudgment. The capacitance to ground of the DC system is composed with multi-branch capacitors in parallel, the device filter capacitor, etc., which results the large capacitance to ground. The capacitance to ground has some influence on the increase of the disturbance voltage signal rate of increase, which can be determined by the delay time and rise rate of the measured voltage.

By using the least squares fit, the branch resistance  $R$  and capacitance  $C$  to ground can be obtained, relationship between the insulation resistance and capacitor is expressed as function (3), where  $I_i$ —Leakage current in the  $i$  branch,  $C_i$ —Capacitance for  $i$  branch.

$$I_i = \frac{\Delta U}{R} + C_i \Delta U \quad (3)$$

The simulation and experimental results show that the capacitance in each branch is small and the transient process in the sensor current signal is much less than the transient process in bus, so the branch capacitor has no effect on the current measurement.

### 4 Principle of DC Ground Fault Detection

The process of detect DC system insulation, first should monitor the bus insulation resistance and choose branch detection by the insulation situation of DC system, according to the current detection techniques the relevant standard is should detect the value of DC system insulation is less than 25k $\Omega$ , the requirements of technical regulations design is that the positive and negative bus to ground insulation should not be less than 100k $\Omega$ , to maintain the insulation indicators of design procedures, it need to alarm when the positive and negative bus to ground insulation is less than 100k $\Omega$ ; it should alarm when the slip insulation is less than 100 k $\Omega$  not is the positive and negative bus to ground insulation in our design, whose detection target is to distinguish the level of slip insulation. It is easy to measure insulation parameters of high resistance, but very hard to measure slip insulation of high resistance, which is related to the sensor disturbed by 10KV strong space electromagnetic fields and sensor resolution, temperature drift, zero drift, residual magnetism factors, the signal of leakage current in branch is weak to microampere level. By using unbalanced bridge detection method, combined with switch 2, switch 3, switch 4, and the development of high-precision and removable DC weak current sensor as the insulation detection( the resolution is up to 0.01mA and can automatically zeroing before detection), which can accurately measure a high insulation value  $R_+$ ,  $R_-$ . The DC current sensor is not very sensitive to AC signal. Therefore, we use unbalanced resistance detection method to accurately detect the DC bus insulation condition. The middle section is shown below.

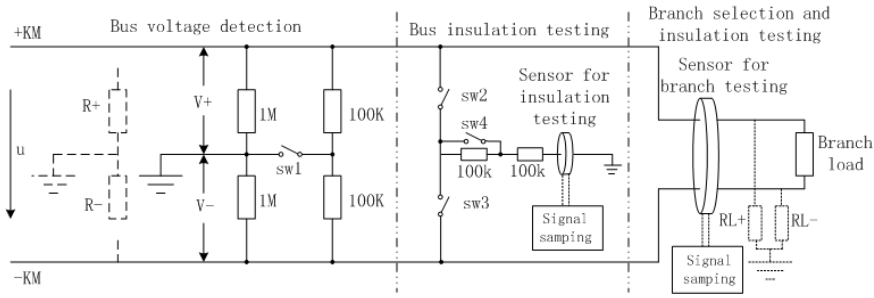


Fig. 1. Schematic diagram of DC ground fault detection

## 5 Conclusion

Now the request of grid intelligent, automated signal during the execution can not analysis process data by statistical probability. Intelligent and automated execution signal should ensure to be true value, it is one or zero. The automated execution information in a control system may not only be one data, so the related information should ensure to be true value.

A tracking derivative based method for DC system grounding fault detection is proposed in this paper, can directly and accurately measure the fault signal, then the combine measurements of grounding resistance and capacitance to ground is completed, the operational status of each branch are fully reflected, which provides a new idea for DC ground fault detection. Compare to wavelet-based detection method, this method is easy to detect, it has strong anti-interference ability, high precision and no need to preprocess the signal in the environment of complex frequency components and strong noise. The matlab simulation and application results show that, this method has little calculation, faster response and better practical value.

## References

1. Ciheng, Z.: Problem existing on DC System grounding fault monitoring device. *Electric Power* 27(7), 64–66 (1996)
2. Wang, R., Liu, F., Xiong, X., et al.: On-line insulation monitoring of substation DC power supply. *Electric Power Automation Equipment* 12, 65–67 (2009)
3. Song, R., Wang, X., Su, C.: Design of apparatus for monitoring DC system ground faults. *High Voltage Engineering* 36(4), 975–979 (2010)
4. Fei, W., Zhang, Y., Wu, Z.: New measuring principle of DC system grounding resistance value in electric power system. *Automation of Electric Power Systems* 25(6), 54–56 (2001)
5. Li, D., Shi, L.: An overall scheme to detect grounding fault in DC system of power plants and substations. *Power System Technology* 29(1), 56–59 (2005)
6. Li, D., Bo, W., Ma, Y.: Grounding fault detection based on wavelet entropy and neural network for loop net of DC system. *Electric Power Automation Equipment* 28(3), 51–54 (2008)

7. Jie, S., Hongwei, S.: DC grounding fault detection device implemented by ARM. *Journal of Shenyang University of Technology* 31(3), 342–344 (2009)
8. Li, D., Xu, J.: A new method to detect grounding fault of looped network in DC power system. *Power System Technology* 30(2), 57–60 (2006)
9. Lin, Z., Xia, X., Wan, Y., et al.: Harmonic detection based on wavelet transform. *Transactions of China Electrotechnical Society* 21(9), 67–74 (2006)
10. Zhang, D., Liu, Z., Zhang, Y.: Complex wavelet transform review and its applications in power system. *Automation of Electric Power Systems* 30(17), 97–104 (2006)
11. Chen, X., Liu, B., Bo, Z.: Harr wavelet-based modified phasor to detect and quantify power quality disturbance. *Proceedings of the CSEE* 26(24), 37–42 (2006)
12. Li, T., Fan, X., Wang, J., et al.: A method to detect dielectric loss factor based on wavelet filtering and tracking differentiator. *Power System Technology* 35(4), 221–226 (2011)
13. Chu, Z., Zhang, C., Feng, X.: Adaptive notch filter-based frequency and amplitude estimation. *Acta Automatica Sinica* 36(1), 60–66 (2010)
14. Chu, Z., Zhang, C., Feng, X.: A new linear band-adjustable algorithm for real-time signal analysis in power system. *Proceedings of the CSEE* 29(1), 94–99 (2009)
15. Ziarani, A.K., Konrad, A.: A method of extraction of nonstationary sinusoids. *Signal Process* 84(8), 1323–1346 (2004)

# Towards a VLIW Architecture for the 32-Bit Digital Signal Processor Core

Khoi-Nguyen Le-Huu<sup>1</sup>, Thanh T. Vu<sup>2</sup>, Diem N. Ho<sup>1</sup>, and Anh-Vu Dinh-Duc<sup>1</sup>

<sup>1</sup>University of Information Technology – VNU-HCM

<sup>2</sup>Ho Chi Minh City University of Technology

Ho Chi Minh City, Vietnam

{khoinguyen, diemhn, anhvu}@uit.edu.vn

thanhv@tse.hcmut.edu.vn

**Abstract.** Digital Signal Processors have been developed for ages due to the great effectiveness in digital signal processing algorithms such as digital filtering and Fourier analysis which cannot be achieved in general-purpose processors. This work is to propose a VLIW architecture for Digital Signal Processor core including the top-level design of the data path. RTL implementation of the proposed architecture will be carried out in the future to verify the micro-architecture as well as instruction set of the DSP core.

**Keywords:** Digital Signal Processor; VLIW Architecture.

## 1 Introduction

Digital signal processing is increasingly important for applications in real life such as communications [1], medical imaging [2], radar & sonar [3], high fidelity music reproduction [4], oil prospecting [5], etc. A digital signal processor is a specialized microprocessor with an architecture optimized for operational needs of digital signal. Although DSP processors have a comprehensive change in the past few decades, there are still common features in most DSP processors today. DSP processors need multiple memory banks with independent buses, specialized instruction sets, addressing modes, control and peripherals. Modern DSP architectures can be divided into 3 or 4 categories (generations) [6].

For the conventional DSP processors such as Texas Instruments' TMS320C2xx family, Analog Devices' ADSP-21xx family, one instruction is issued and executed in one clock cycle. These processors typically include a single multiplier (MAC unit) and an ALU, but few additional execution units. DSP processors like the Motorola DSP563xx and Texas Instruments TMS320C54x operate at higher clock speeds and often include a modest amount of additional hardware (barrel shifter, instruction cache), to improve performance in common DSP algorithms. These processors also tend to have deeper pipelines.

Another DSP generation was built by expanding conventional DSP architectures, for instance, adding parallel execution units, i.e. a second multiplier and adder.

Multi-issue processors use very simple instructions that typically encode a single operation. These processors achieve a high level of parallelism by issuing and executing instructions in parallel groups rather than one at a time. Using simple instructions simplifies instruction decoding and execution, allowing multi-issue processors to execute at higher clock rates than conventional or enhanced conventional DSP processors. The two sub-categories of implementation of this architecture that execute multiple instructions in parallel are VLIW (Very Long Instruction Word) and superscalar. The biggest difference between them is how instructions are grouped for parallel execution.

This paper is structured as follows. Section 2 describes the proposed architecture for a 32-bit fixed-point DSP core. Conclusion as well as future work discussion will be presented in section 3.

## 2 Proposed Architecture

In this section, we will propose a general architecture for the 32-bit Fixed-Point Digital Signal Processor-DSP core. Note that, the design of instruction set is still under construction, thus, the components that mainly relate to the instruction set architecture will not be presented in this work.

### 2.1 Top-Level Architecture Design

DSPs are typically based upon the Harvard architecture, or upon modified versions of it, such as the Super-Harvard architecture as shown in Fig. 1. In the Harvard architecture, there are separate memories for data and instructions, and two separate buses connect them to the DSP core. The Harvard architecture can be improved by adding to the DSP core a small bank of fast memory, called ‘instruction cache’, and allowing data to be stored in the program memory.

In general, digital signal processors usually consist of DSP core, peripheral controller, external memory controller, power management as well as acceleration hardware such as FFT core (Fast Fourier Transform), DCT core (Discrete Cosine Transform), DMA units (Direct Memory Access). They are all depicted in Fig. 2.

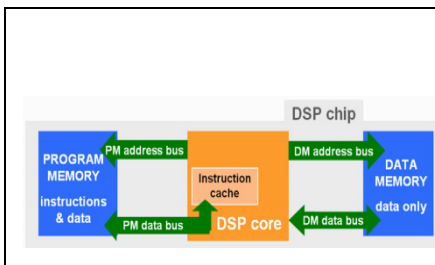


Fig. 1. Super-Harvard architecture

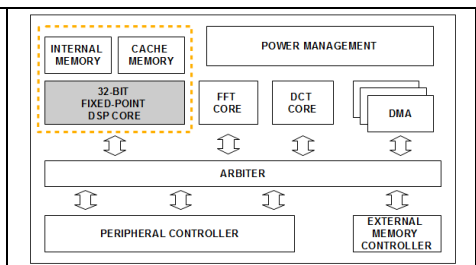


Fig. 2. Architecture of DSP processor

In this work, we only focus on the DSP core. The proposed DSP core architecture consists of CP (Control Path) and DP (Data Path) as in Fig. 3. On-chip memory subsystem is located out of the core, containing the Program memory and two Data memories (DMX and DMY). Besides, there is an instruction cache integrated to the processor core due to its advantages as mentioned above. The DP includes the Program Memory, Cache, Instruction Fetch, Instruction Decoder and Execution Units. The Execution Units on the whole are composed of four functional units: FALU (Arithmetic Logic Unit for Fixed-point computation), MDMAC (Multiplier/Divider and Multiplication and Accumulation unit), BALU (Arithmetic Logic Unit for Branching computation), and LSU (Loading/Storing Unit). Besides, the DP also includes RF (Register File) with 32 32-bit registers, Register Bus and support for acceleration instructions.

## 2.2 Data Path (DP) Design

The DP of DSP core can perform arithmetic operations (addition, subtraction, multiplication, division), logic operations (and, or, not, etc.), compare operations (<, >, =, etc.) and consists of 32 32-bit general-purpose registers (A0-A15, B0-B15). Every CPU clock cycle, the program fetch and instruction decode units can deliver up to four instructions to the functional units.

### 2.2.1 FALU and BALU Unit

In order to increase the parallel capacity of processor, we design two separate ALUs for different purposes. In fact, FALU is aimed to operations relating to fixed-point computation (cosine, sine, natural logarithm, exponential function, etc.), while BALU is dedicated for bit-oriented as well as branching computation. Although, they might have the same top-level architecture, they differ from detailed designs with extended logic blocks to support their own purposes. Consequently, we can utilize both of the FALU and BALU for different objectives, simultaneously. In our top-level architecture design, the ALU performs all of the logic and arithmetic operations, even shifting and rotation computations as depicted in Fig. 4. The most interesting issue in our design is the conjunction between barrel shifter and adder instead of separate components in order to perform shifting before computing operation in only one instruction.

Operands after preprocessing stage are then sent to the kernel that contains the adder, shifting and logic units. After execution, result is sent to the post processing unit for selecting the output. Besides, ALU writes to flags that are used for determining processor status, especially in conditional execution instructions. Moreover, it is a common practice to support saturation of final result. As illustrated in Fig. 4, the ALU consists of following units:

#### 2.2.1.1 Shifter

The shifter can perform arithmetic and logical shift. While the arithmetic shift replicates the MSB bit to keep the sign of the operand, the logical shift only appends a '0' and hence is more preferable for unsigned binary operands.

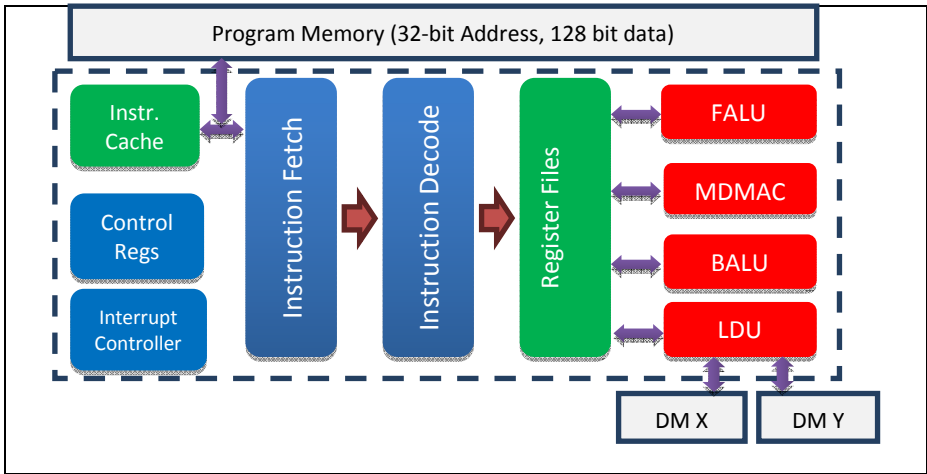


Fig. 3. Top-level architecture of DSP core

2.2.1.2 Logic Unit

The DSP processor can perform AND/OR/NOT/XOR instructions as logic operations with both of operand A and B as inputs. Of course, there might be a control signal to indicate exactly what kind of logic operations to be used.

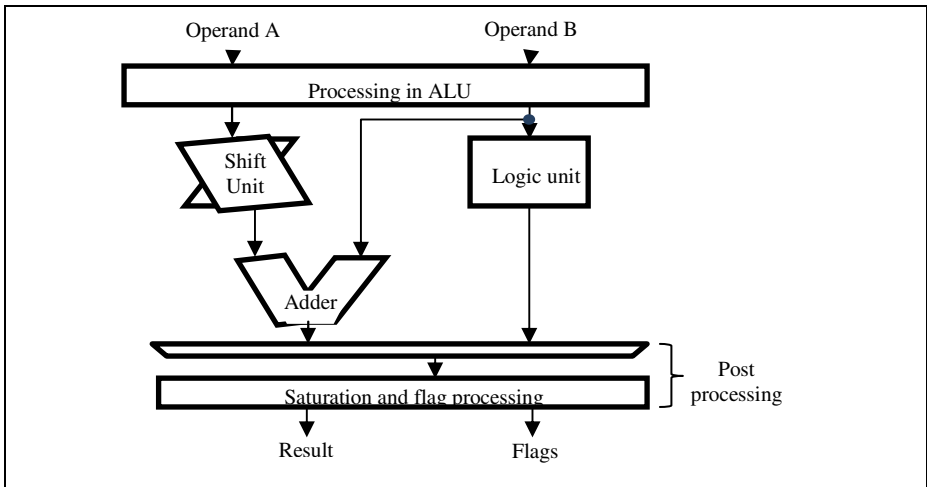


Fig. 4. Top-level architecture of ALU in DSP core

2.2.1.3 Saturation Unit

This unit is in charge of saturating the result whenever the guard bit is changed and noticed by the overflow flag, it means that the current result cannot be expressed in the native word length.



#### 2.2.1.4 Adder and MAX/MIN Operation

One of the common ALU operations is MAX/MIN of two operands. It can be simply implemented by taking advantage of the adder to calculate subtraction. Consequently, the MSB bit of adder output will be utilized as an input of the multiplexer to determine which operand is greater than the other.

#### 2.2.1.5 Leading Count Operation

In some digital signal processing applications, when we want to count how many bits are equal to MSB, LSB or a specific bit of the operand, we need to implement a small piece of code. The LED box is an available hardware that can perform the same function and consequently, it reduces the programming complexity.

### 2.2.2 MDMAC Unit

The MDMAC consists of two main components including MAC and Divider. Absolutely, MAC is the most important components of the DSP core DP as it performs iterative operations such as convolution and supports double precision as well as implements the hardware for auto-correlation, filtering and transform functions. Also, the guard bits are necessary for iterative computing to keep their sign value. For moving the data from the MAC to the general purpose registers (native word extract) the accumulator value needs to be rounded, scaled and saturated. The most interesting point in this design is the emergence of barrel shifter that allows to multiply by  $2^x$  number without using heavy multiplier. Moreover, the divider helps us perform division operations separately from MAC unit.

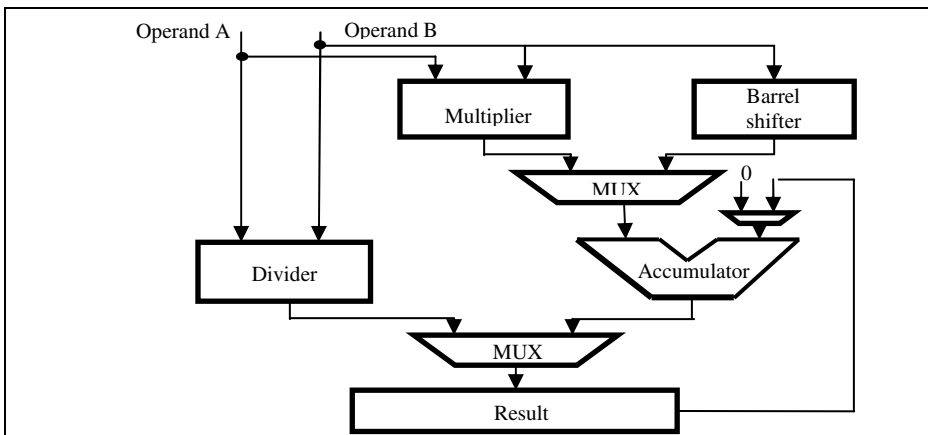


Fig. 5. Top-level architecture of MAC in DSP core

### 2.3 LSU Design

The LSU generates the address required for the two Data Memories such as DMX and DMY. Unlike general-purpose processors, DSPs include address generator blocks,

which control the address generation for specialized addressing modes such as circular buffers, and bit-reversal addressing. The circular buffer addressing mode allows easier performing of the convolution, one of the most common DSP instructions executed in the MAC. The equation of convolution is described as  $y(n) = \sum_{i=0}^{m-1} x(n-i) * c(i)$ . As seen in the equation, the computation needs data which are multiplied with coefficients. To implement it, a circular buffer would be required. Circular buffers are limited storage regions where data are stored in a First-In First-Out (FIFO) way. Besides, the bit-reversal addressing mode supports the Fast Fourier Transform (FFT). Furthermore, the greatest advantage of LSU is the capability to load/store data from/to two data memories simultaneously. Consequently, it improves the performance of processor significantly.

### 3 Conclusion

In this work, we have proposed a top-level design for a 32-bit VLIW digital signal processor. However, this architecture has not been proven to function correctly and the instruction set is still under construction and analysis. Surely, RTL implementation will be carried out in the future to verify the micro-architecture as well as instruction set of the DSP core.

**Acknowledgement.** This work was granted under Project 39/2013/HĐ-SKHCN by the Department of Science and Technology of Ho Chi Minh City.

### References

1. Nguyen, X.-T., Do, Q.-D., Tran, H.-D., Huynh, H.-T., Pham, C.-K.: A PCIe-based FFT Implementation for High-speed Spectrum Analysis. In: Proc. 3rd IEICE Int. Conf. Integrated Circuits and Devices in Vietnam, Danang, Vietnam, August 13-15, pp. 126–131 (2011)
2. Yagi, M., Shibata, T.: An Image Representation Algorithm Compatible with Neural-Associative-Processor-Based Hardware Recognition Systems. *IEEE Trans. Neural Networks* 14(5), 1144–1161 (2003)
3. Titlebaum, E.L.: Frequency- and time-hop coded signals for use in radar and sonar systems and multiple access communications systems. In: Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers. Dept. of Electr. Eng., Rochester Univ., NY (1993)
4. Olswang, B.S.: Separation of Audio Signals Into Direct and Diffuse Soundfields for Surround Sound. In: Cvetkovic, Z. (ed.) *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*. LOUD Technol. Inc., Wodenville (2006)
5. Mottl, V.: Pattern recognition in spatial data: a new method of seismic explorations for oil and gas in crystalline basement rocks. In: Dvoenko, S., Levyant, V., Muchnik, I. (eds.) *Proc. of 15th International Conference on Pattern Recognition*. Tula State Univ., Russia (2000)
6. Tan, E.J., Heinzelman, W.B.: DSP architectures: past, present and futures. *SIGARCH Comput. Archit. News* 31(3), 6–19 (2003)

# Application and Effectiveness of Cooperative Teaching Strategies in Entrepreneurship Education

Chien-Hua Shen<sup>1,\*</sup> Chun-Mei Chou<sup>2</sup>, and Hsi-Chi Hsiao<sup>3</sup>

<sup>1</sup> Department of Business Administration, TransWorld University Yunlin, Taiwan

<sup>2</sup> Institute of Vocational and Technological Education,  
National Yunlin University of Science & Technology, Yunlin, Taiwan

<sup>3</sup> Graduate Institute of Business and Administration,  
Cheng Shiu University Kaohsiung, Taiwan  
shen17@ms51.hinet.net, choucm@yuntech.edu.tw,  
hchsiao@cc.ncue.edu.tw

**Abstract.** The main purpose of this study was to understand the application and effectiveness of cooperative teaching strategies in entrepreneurship education. This study invited specialist faculty and school teachers to form a cooperative teaching group to present lectures in an entrepreneurial management module in the case school. After the implementation of cooperative teaching for one semester, a questionnaire survey and a general survey were conducted to understand the students' learning satisfaction. A total of 135 students were investigated, and the results indicated that the mean of the students' overall satisfaction in a Likert 5-point scale was above 4.0.

**Keywords:** Cooperative Teaching, Entrepreneurship Education.

## 1 Introduction

Entrepreneurship education refers to developing and improving students' basic entrepreneurial qualities, as well as the education for developing their ambition, initiative, pioneering spirit and spirit of adventure to engage in the planning of a certain career, enterprise or business. The United Nations Educational, Scientific and Cultural Organization (UNESCO) regarded the entrepreneurship passport as "the third passport to learning," other than academic and professional passports (UNESCO, 1989). For technical and vocational education, one of the important issues of having an innovative spirit is to implement innovation in entrepreneurship. Entrepreneurial abilities will become a basic essential ability of students in technical and vocational schools in the future.

In addition to including the Best Practices Guidelines for Entrepreneurship and Start-up Companies proposed by Taiwan into its formal documents, the 2001 APEC Ministerial Meeting also included it as a reference document for leadership conferences. Therefore, Taiwan should extend this issue to further establish the Asia

---

\* Corresponding author.

Pacific Entrepreneurship Development Center as the integrator and locomotive to promote the economic development of the Asia-Pacific region. Previous studies have indicated that entrepreneurship courses are beneficial to students' analytical skills, organizational skills, judgment skills, communication skills and team work skills (Chou & Shen, 2003), and that these are critical development skills (Chou & Shen, 2002). However, current students in the technical and vocational education system are deficient in such skills.

Technical and vocational education should attach importance to the opportunities for students to practice or implement what they have learned on campus to enhance their employability and reduce the gap between what they have learned in school and what they put into practice after graduation, in order to fulfill the objective of technical and vocational education. Consequently, this study intended to investigate the approaches and effectiveness of applying the teaching strategies of cooperative learning to entrepreneurship education. The results could serve as a reference for promoting entrepreneurship education in technical and vocational education.

## **2 Connotations of Cooperative Learning Teaching Strategies**

The main objective of cooperative learning is to enable heterogeneous groups of students to learn together, encourage one another, share their perspectives with one another, provide the results of information sharing, and criticize mutual perspectives (Liao, 2009). In a cooperative learning group, each student is responsible for his/her group and partners in the same group. Students have to be mutually dependent and use their social skills for mutual coordination to jointly complete assignments. Cooperative learning is a structural and systemic teaching strategy that is suitable for various grades and various academic fields. In cooperative learning, teachers assign students into a heterogeneous group according to their abilities, gender, ethnic background, etc. to encourage them to help one another, improve individual learning effectiveness and achieve the objective of the group. Therefore, cooperative learning establishes a learning environment that increases students' interactions, leads to more cooperative behaviors, improves students' learning achievement and expands the development of social skills.

From the perspective of cooperative learning, the teaching of entrepreneurship education should attach importance to the students' active participation in learning and spontaneous thinking during the learning process. Entrepreneurship education is different from traditional education, as it attaches great importance to the inspiration of students' creativity and critical and proactive thinking. Moreover, students learn to face and solve problems through inter-group mutual inspiration (Nilson, 2003). In other words, from the perspective of cooperative learning, entrepreneurship education should attach importance to students' active participation in the learning of entrepreneurship activities, which not only enables learners to absorb textbook knowledge but also further develops students' creativity and critical and proactive thinking (Jones & English, 2004).

### 3 Application and Analysis of Learning Satisfaction

#### 3.1 Application

This study introduced the concepts of cooperative teaching using entrepreneurship lectures in the entrepreneurial management module of the case school. The teachers in the school and specialist faculty cooperated with one another to develop teaching guidelines and teaching programs to implement lecture-style learning. The explanations on the main connotations of the subject are given as follows:

Name of the subject: [entrepreneurship lecture] subject

1. Relevant teaching theories: Situated learning and cooperative learning
2. Teaching objectives: To share the entrepreneurship processes of success, failure and second undertakings, to provide students with entrepreneurship benchmarks for learning and developing their spirit of entrepreneurship, and to further strengthen their entrepreneurial motivation and entrepreneurial abilities.
3. Implementation duration and frequency: One semester, one lecture per week
4. Implementation method
  - (1) Subjects: All the students in the school who were interested in the subject, or students who were taking relevant courses.
  - (2) Activities: Lectures and forums
  - (3) Speakers: The convener of the course or the person who opened the course invited entrepreneurs (including successful entrepreneurs, early entrepreneurs, failed entrepreneurs and entrepreneurs who started a second undertaking) to share their entrepreneurship process in school.
  - (4) Teaching assessment
    - (a) Students had to write down a summary of the lecture content and their feedback on every lecture. In addition, they had to collect relevant information and submit their learning files at the end of the semester for the assessment of their performance.
    - (b) The frequency and content of the interactions (dialogues) with the entrepreneurs during every lecture and the students' attendance records were included in the regular assessment.

[Implementation of entrepreneurship] subject – Entrepreneurial practice

1. Relevant teaching theories: Situated learning, cooperative learning and scaffolding theory.
2. Teaching objectives: To enable students to participate in an entrepreneurial internship program, in order to understand and experience the work connotations and further amend an entrepreneurship plan so as to start an undertaking successfully.
3. Implementation period and frequency: Half of one semester; eight hours per week
4. Implementation method

- (1) Subjects: Students who were about to graduate or students who were taking relevant courses and had participated in the teaching program of [micro-enterprise business practices].
- (2) Activities: The students were divided into groups according to their entrepreneurship plan and the business or products they expected to focus on. The teacher opening the course assisted in finding enterprises with which to implement internship programs, and the students looked for enterprises as well. However, the consent of the teacher had to be obtained before implementing entrepreneur observations and workplace internships.
- (3) Location of internship: The location was the business site of the specialist faculty. The specialist faculty also acted as counselors. The students had to attend the internship program for at least eight hours per week.
- (4) Teaching assessment
  - (a) The students had to submit a weekly internship report that contained a detailed explanation about their internship content and feedback.
  - (b) The students had to submit a weekly entrepreneur observation record that indicated how they had observed the entrepreneurs handling business-related events.
  - (c) The specialist faculty offered written comments in the middle and at the end of the semester, and the students' scores were assessed according to their attendance status, work status and learning performance.
  - (d) The students had to amend and complete their entrepreneurship plan according to the feedback to the internship program. In addition, they had to participate in a planning result forum, in which the specialist faculty offered their comments on the amendments and score assessments.
  - (e) In regards to the calculation of the students' grades (the internship accounted for 50% of the entrepreneurial internship and practices), the submission status and content of various written reports accounted for 50% of the grade and the specialist faculty's assessment of the score accounted for 50% (the internship score accounted for 35%, and the entrepreneurship plan score accounted for 15%).

### 3.2 Analysis of Learning Satisfaction

As shown in Table 3, after the implementation of the program, the item with the highest satisfaction was "As a whole, the instruction of the experts (specialist faculty) has a positive effect on my practice and learning" ( $M=4.33$ ), followed by "I am satisfied with the overall teaching attitude of the specialist faculty" ( $M=4.29$ ) and "The instruction content of the specialist faculty meets the learning needs of the course" ( $M=4.25$ ). As a whole, the mean of the students' satisfaction after the implementation of the program was above 4.0.

**Table 1.** Summary of the Students' Satisfaction with the Specialist Faculty's Cooperative Teaching Program N=1591

Item	SD	Mean	Order
● As a whole, the instruction of the specialist faculty has a positive effect on my practice and learning.	.782	4.33	1
● I am satisfied with the specialist faculty's teaching attitude.	.806	4.29	2
● The instruction content of the specialist faculty meets the learning needs of the course.	.808	4.25	3
● The instruction of the specialist faculty is beneficial to improving my understanding of the industrial environment.	.804	4.20	4
● The instruction of the specialist faculty is beneficial to improving the application of my professional skills into practice.	.797	4.18	5
● I am satisfied with the professional and instructional skills of the specialist faculty.	.755	4.18	5
● The specialist faculty is willing to answer students' questions in class or out of class.	.710	4.16	7
● Compared to the instruction of general courses, I am more satisfied with the practical content of the cooperative instruction provided by the specialist faculty.	.730	4.14	8
● The specialist faculty attaches importance to teaching interactions and encourages students to ask questions or express their comments.	.694	4.11	8
● The specialist faculty can grasp the teaching atmosphere in class.	.738	4.09	10

#### 4 Conclusion

The teaching findings after the implementation of the subject for one semester are summarized as follows.

In general, the teachers and students agreed that the cooperative teaching program was beneficial to the students' professional development. The teachers and students both indicated that the instruction from the specialist faculty in the classroom met the teaching needs and expectations for the increase in practical and professional experiences. In terms of the cognitive apprenticeship learning model, the specialist faculty acted like mentors. Providing adequate teaching approaches for them to express their abundant practical experiences would be beneficial to the professional growth of the teachers and students in the school (Woolley & Jarvis, 2007). Because the case school

had established a three-level review system for the selection of specialist faculty for various departments, the specialist faculty possessed practical and abundant experiences and sufficient ability to express their ideas. Consequently, the mean of satisfaction of the teachers and students was above 4.0.

Entrepreneurship education is a student learning-centered educational program that integrates the human resources of enterprises with school faculty. At present, technical and vocational schools are aggressively promoting industry-academia collaboration to introduce the resources of business units onto their campuses. However, in general, the primary collaborative model is one in which the business units provide funds and the schools provide techniques. The secondary model is one in which students attend internship programs in workplaces and teachers visit enterprises for learning. Entrepreneurship education programs focus on the integration of the business units' human resources with teaching, which can substantially improve the practice of the teaching content. With the implementation of entrepreneurship programs, human resources that meet the needs of practical teaching in schools can be integrated into teaching. Therefore, they are the most effective education programs for resource integration.

Cooperative teaching strategies can effectively trigger students' motivation. This subject was developed based on experiential learning and the educational idea of learning through practice and practice through learning. The students had to attend on-site visits, internship programs, and practices after school. In addition, under the instruction of the specialist faculty, the students were immersed in the atmosphere of entrepreneurs and became more satisfied with their learning. Therefore, their learning motivation could be effectively triggered.

The perfection of the development of entrepreneurship education will rely on more action studies. The existing relevant studies in Taiwan have mainly focused on the investigation of entrepreneurial behaviors and their relevant factors. There is a lack of studies on teaching content or teaching approaches. Although the teaching program of this subject could be provided as a reference, if the objective of entrepreneurship education is to develop micro enterprises, teaching content needs to be developed, and the engagement of more researchers is required.

## References

1. Shen, C.H.: A Study on the Technical and Vocational Education System's Approaches for Promoting Entrepreneurial Industry-academia Collaborative Education. *Educational Research* 148, 81–97 (2006)
2. Chou, C.M., Shen, C.H.: From Employment to Entrepreneurship – How the U.S. Entrepreneurship Education Inspires the Improvement of Employability of Students in Technical and Vocational Schools in Taiwan. *Business Education Quarterly* 88, 46–50 (2003)
3. Chou, C.M., Shen, C.H.: A Study on the Technical and Vocational Schools' Approaches for Promoting Entrepreneurship Education. Paper presented at the Annual Meeting for the Challenges of Human Resources under the Environmental Changes in 2003—Symposium of Educational Training, July 6-7. SME and Public Policies, Taiwan (2003)



# Isarn Dharma Alphabets to Thai Language Translation by ATNs

Phoemporn Lakkhawannakun and Pusadee Seresangtakul

Natural Language and Speech Processing Laboratory (NLSP),  
Department of Computer Science, Faculty of Science,  
Khon Kaen University, Khon Kaen, Thailand  
phoemporn@hotmail.com, pusadee@kku.ac.th

**Abstract.** This paper presents a method of translating Isarn Dharma alphabets into the Thai language. In order to develop the system, an Isarn to Thai dictionary was constructed. The Isarn Dharma input text was segmented into a sequence of Isarn words using the longest matching algorithm. In this study, we proposed a hybrid of dictionary and Augmented Translation Networks (ATNs) to translate the Isarn Dharma to Thai text. In order to evaluate the efficiency of the system, the Buddha foretell, Jataka legend, Stone inscription, Isarn foretell and common sentences were used to test the system. The experimental results showed that the correctness of the system is 61.81%.

**Keywords:** Isarn Dharma Alphabets, Augmented Transition Networks(ATNs), Translation.

## 1 Introduction

The Isarn Dharma alphabet is an ancient alphabet that was used in northeastern Thailand during the 22<sup>nd</sup> – 24<sup>th</sup> Buddhist century. It was used to record histories, traditions, rituals and remedies. This information was inscribed onto palm leaves or “Bailan”. The existence of these characters is not widely known with only some Buddhist monks and elder people being able to read and write them today. In order to preserve the remaining information, computerized systems must be used in order to translate the inscribed script into the more current Thai script. Hence, this research proposed the development of Isarn Dharma Alphabet to Thai translation system.

Isarn Dharma is a tonal language. The difficulty with using natural language processing on this language is that it does not have tone markers. Same words may have different tones and different meanings depending on their surrounding words [1], [2]. For example a word “ $\text{๕}$ ” ( $/p\bar{u}:/$ ) means a crab (mid tone), a grandfather (low tone) and unsharpened (falling tone). In order to choose the words to be translated correctly, the reader must have experience in reading.

In this paper, we will present Isarn Dharma to Thai translation based on a hybrid of dictionary and Augmented Transition Networks (ATNs) that can analyze the syntactic grammar according to Thai and Isarn grammar rule to merge the similar grammar with representing a more compact rule.

## 2 Background and Related Work

There are many multilingual translation systems available today [3-5] but none of them support the Isarn Dharma Alphabet and there is very little research related to the Isarn Dharma Alphabet [6].

N. Phaiboon and P. Seresangtakul [6] proposed an Isarn Dharma phoneme transcription. In their work, the Isarn Dharma Alphabet-Thai dictionary was constructed using trie structure. As mentioned before, Isarn Dharma words do not have tone markers and same words may have different tones. The problem with the previously mentioned dictionary is that it could not serve words that had many tones. Therefore, our work modified the trie structure in order to solve this problem. A more detailed explanation of this can be found in section 3.

N. Yapom and P. Seresangtakul [7], proposed Lanna to Thai language using dictionary and ATNs. The study focused on only some Thai phrase structures, namely noun phrase, verb phrase, location adverb phrase and time adverb phrase. In order to apply this to our system, more Thai phrase structures must be analysed.

## 3 System Overview

In order to develop the Isarn Dharma alphabet to Thai translation system, the first step was to consider the bilingual dictionary. In this study, we modified the Isarn Dharma – Thai dictionary proposed by N. Phaiboon and P. Seresangtakul [6]. The dictionary was constructed using the trie structure [8], which is a fast and compact double arrays structure. Our research modified the leaf node of the trie in order to serve words that had more than one meaning as well as words in different tones. Fig. 1 shows examples of the trie structure. The dictionary consists of 8,000 Isarn words. The dictionary contents consist of the Isarn word and its corresponding Thai words, phoneme, word type, Thai meaning [9], the sub part of speech tagging [10] that applied from the orchid corpus, special characteristics [11], English meaning and Thai description. The architecture of the developed system is given in Fig. 2. The system consists of four main modules: 1) the Isarn Dharma text pre-processing 2) sentence analysis and 3) transcription and 4) translation modules.

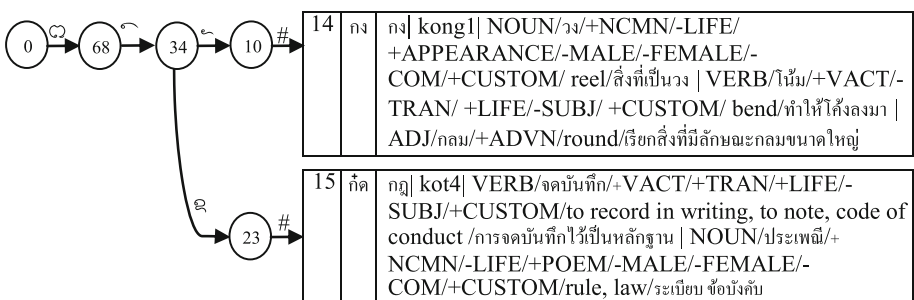


Fig. 1. Words in the trie data structure

### 3.1 Pre-processing

Isarn Dharma Alphabet is a non-segmented language. It rarely uses white space between words. Therefore, the input text will tokenize into a sequence of words using the forward longest matching algorithm [12]. This algorithm will scan a sentence

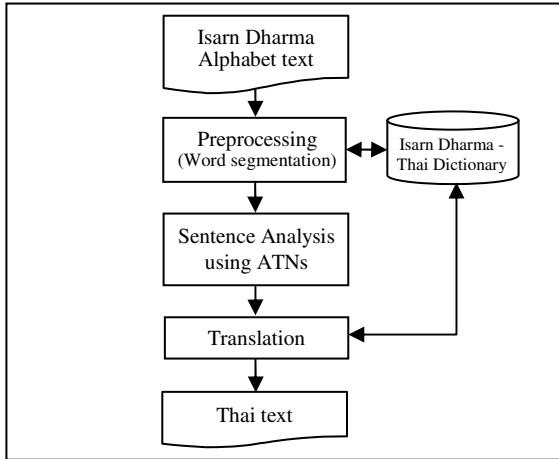


Fig. 2. Overview of the Isarn Dharma to Thai Translation System Architecture

```

S = {c1, c2, .., cN}; #set of character in a sentence
D = {d1, d2, .., dM}; #set of word in the dictionary
W = {w1, w2, .., wn}; #Set of output word
N = length(S); #N is number of character in S
#Initialize variable
i ← 1; #i is position of wi
j ← 1; #j is current starting position marker
LP ← N; #LP is last position maker
TN ← N; #TN is length of remaining sentence
DO WHILE (TN > 0) {
  IF (S[j to LP] ∈ D) THEN {
    W[i] ← S[j to LP]; i++;
    j ← LP+1;
    S ← S[cj to cN]
    TN ← length(S); LP ← N
  } ELSE {
    IF (LP == j) {
      W[i] ← S[cj]; i++;
      j ← LP+1;
      S ← S[cj to cN]
      TN ← length(S); LP ← N
    } ELSE { LP ← TN-1; }
  }
}
ENDDO

```

Fig. 3. The Longest Matching Algorithm

from the leftmost longest word to compare with words in the dictionary. If words are not found in the dictionary, the algorithm will backtrack to search for the next longest word. Fig. 3 shows the longest matching algorithm.

### 3.2 Sentence Analysis

The Isarn Dharma phrase/sentence structure is almost the same as the Thai structure. In this study, the ATNs [13, 14] are used to examine the syntactic structure of the input sentence. The ATNs [13] are a form of augmented pushdown store automata that developed by Woods [13, 14]. The ATNs consist of nodes and arcs. The nodes apprise the state of process, which indicates the process name. The arcs are the parts of speech (POS) values that are sent between node and node in order to analyse structure of a sentence. The start node is represented by “-” and the end node is represented by “+”. Fig.4. shows an example of the Isarn Dharma sentence “ບຸ້ນຊາ”. The symbols S, NP, VP, VERB represent sentence, noun phrase, verb phrase and verb, respectively. The subject of a sentence is “ບຸ້ນ” (/pù:/), which is noun. Verb is “ຊາ” (/pay/) and objective is “ຊາ” (/na:/).



Fig. 4. An example of ATN grammar

#### Isarn Dharma and Thai Phrase Structure

There are five main Thai and Isarn Dharma phrase structures [15-17]: noun phrase, verb phrase, extra adverbial phrase, time adverbial phrase and location adverbial phrase. The following are brief of the Isarn Dharma and Thai phrase structures.

- 1) *A Noun Phrase (NP)* consists of noun and nounal modifier. The nounal modifier may be an adjective, classifier or determinative.
- 2) *A Verb Phrase (VP)* is composed of a verb and a verbal modifier. The verbal modifier may be an auxiliary verb, adverb, preposition or ending.
- 3) *Extra Adverbial Phrase (EAP)* is a supplemental sentence. The EAP consists of a combination of noun, adjective, verb, adverbs and auxiliary verb.
- 4) *Time Adverbial Phrase (TAP)* indicates the time. The TAP consists of a noun, classifier, auxiliary verb, adverbs, determinative and ending.
- 5) *Location Adverbial Phrase (LAP)* indicates the place. The LAP consists of a preposition or a conjunction followed by a noun.

#### Isarn Dharma and Thai Sentence Structure

Isarn Dharma and Thai sentence structures usually consist of elementary and complementary parts. All sentences must have an elementary part, whereas the complementary part is optional.

1) *Elementary part*

The elementary part may be a noun or a verb. The noun officiates as subject, a direct object, an indirect object and single noun. The verb officiates as intransitive, transitive, and ditransitive.

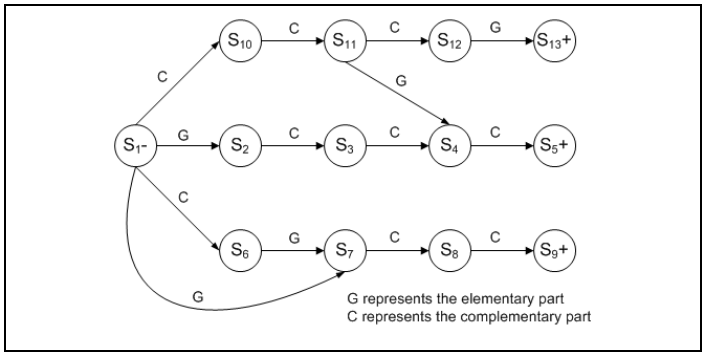
2) *The complementary part*

The complementary part is the supplemental sentence such as Extra Adverbial Phrase (EAP), Time Adverbial Phrase (TAP) and Location Adverbial Phrase (LAP). This part may appear in front of the sentence or change to another place in the sentence. The structure changing does not change the meaning of the sentence.

From the relationship of the elementary and complementary parts [16], the ATN notation of the Thai sentence structure can be represented by Fig. 5.

**3.3 The Sentence Translated Using ATNs**

In order to translate Isarn Dharma to Thai, the input sentence was segmented in to a sequence of words. Each Isarn Dharma word looks for the corresponding Thai words and their indication POS tag in the dictionary. By using ATNs graph, all POS and words can be represented in the ATNs graph to generate all possible patterns. The most suitable Thai grammatical pattern will be chosen as the output sentence. Table 1 shows an example of the translation process.



**Fig. 5.** ATNs of Thai sentence structure

An example sentence is “*นอนที่เขียงที่บ้าน*” (นอนที่เขียงที่บ้าน, /nɔ:n - t<sup>h</sup>i:am - k<sup>h</sup>â:ŋ - t<sup>h</sup>i: - bâ:n/). This example sentence can be represented with Fig.5 according to the elementary part (G) and the complementary part (C) respectively.

**Table 1.** Example of choosing the correct POS tag “ငွေဝေ့ဝေ့ဝိပူ”

**Vocabulary:**  
 ငွေ (/ဂ:န/) = VERB/นอน  
 ဝေ့ (/t<sup>h</sup>i:am/) = NOUN/กระเทียม, ADV/เท่า, ADJ/เคียง  
 ဝေ့ (/k<sup>h</sup>á:ŋ/) = NOUN/ข้าง  
 ဝိ (/t<sup>h</sup>í:/) = PREP/ที่  
 ပူ (/bá:n/) = NOUN/บ้าน

Step	Word	ATNs	Method	Description
1			First word go to VP	
2	ငွေ	VP	VP ← နဝ	
3	ဝေ့	NP	ADJ ← ကေး	using Noun Phrase
4	ဝေ့	NP	NP ← ခံ	
5	ဝိ	LAP	LAP ← ကဲ	using Location
6	ပူ	LAP	LAP ← ဃံ	Adverbial Phrase (LAP)
7	-	-	VP ← နဝ, NP ← ကေးခံ LAP ← ကဲဃံ	END

### 4 Experimental and Results

In order to evaluate the efficiency of the proposed system, sentences from the Buddha foretell, Jataka legend, stone inscription, Isarn foretell and generated sentences were translated by this system. The efficiency of the system is determined by the percentage of the translation correctness as shown in the following equation.

$$Efficiency (\%) = \frac{X}{N} \times 100 \tag{1}$$

Where *X* is the number of correctly translated sentences and *N* is the number of total sample sentences. The experimental results are shown in table 2.

**Table 2.** The efficiency of the translation system

Document Type	All sentences	Correct sentences	Results (%)
Buddha foretell	404	241	59.65
Jataka legend	35	21	60.00
Stone inscription	63	46	73.02
Isarn foretell	20	12	60.00
generated sentence	20	15	75.00
<b>Average Accuracy</b>	<b>542</b>	<b>335</b>	<b>61.81</b>

### 5 Conclusions

This paper presented an Isarn Dharma to Thai translation system. The proposed system, first extracted the Isarn Dharma input text into a sequence of words using the longest matching algorithm. ATNs was applied to these sequences to translate the Isarn Dharma to Thai text. By testing the proposed system using the Isarn Dharma

text, the experimental performance results obtained a 61.81% correct translation rate, which is a quite low accuracy rate. The main reason for this is that although the Isarn Dharma is a tonal language its writing style does not have tone markers. The difficulty in predicting word tones applies to both machine and human. A second reason is the testing data was mostly literature structures, which are quite different from normal writing structures.

In order to improve the efficiency of the translation system, we need more vocabulary and corpus. Future work will focus on corpus based and machine learning in order to improve the translation performance.

## References

1. Booncherm, S.: Ancient alphabets textbook, 2nd edn. Moraduk Isan, Ubon Ratchathani (2005)
2. Weerawong, S.: Isarn Dharma Alphabet (January 7, 2012), <http://www.esansawang.in.th/tham/thamhome.htm> (accessed)
3. Bouzit, F., Laskri, M.T.: Multi-Lingual Machine Translation System Based On a Linguistic Approach. Paper presented at the Fourth International Symposium on Innovation in Information & Communication Technology (ISIICT 2011), pp. 155–158. Faculty of Information Technology, Philadelphia University, Amman (2011)
4. Hung, V.T., Fafiotte, G.: UVDict - A Machine Translation Dictionary for Vietnamese Language in UNL System. Paper presented at the Fifth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS 2011), pp. 310–314. Korean Bible University, Seoul (2011)
5. Zeman, D.: Data issues of the multilingual translation matrix. Paper presented at Proceedings of the Seventh Workshop on Statistical Machine Translation (WMT 2012), Montreal, Canada, pp. 395–400 (2012)
6. Phaiboon, N., Seresangtakul, P.: Isarn Dharma Alphabet Dictionary. Paper presented at Proceeding of the 13th National on Computer Science and Engineering Conference (NCSEC 2009), pp. 287–292. King Mongkut's University of Technology, Thonburi (2009)
7. Yaprom, N., Seresangtakul, P.: Lanna to Thai language translation. Paper presented at Proceeding of the 11th National Computer Science and Engineering Conference (NCSEC 2007), pp. 180–189. Sripatum University, Bangkok (2007)
8. Morita, K., Atlam, E.-S., Fuketa, M., Tsuda, K., Aoe, J.-I.: Fast and compact updating algorithms of a double-array structure. *Information Sciences* 159, 53–67 (2004)
9. Pinthong, P.: Isarn – Thai – English dictionary. Siritham opset, Ubon Ratchathani (1989)
10. Sornlertlamvanich, V., Charoenporn, T., Isahara, H.: ORCHID: Thai Part-of-Speech Tagged Corpus. NECTEC technical report, 1, 5–19 (1997)
11. Ruengchotwit, P.: Analysis the meaning of words. Chiang Mai University, Chiang Mai (1992)
12. Poowarawan, Y.: Dictionary-based Thai Syllable Separation. Paper presented at Proceedings of the Ninth Electronics Engineering Conference (EECON 1986), Thailand, pp. 409–418 (1986)
13. Woods, W.A.: Cascaded ATN Grammars. *American Journal of Computational Linguistics* 6(1), 1–12 (1980)

14. Woods, W.A.: Transition Network Grammars for Natural Language Analysis. *Computational Linguistics: Communications of the ACM* 13(10), 591–606 (1970)
15. Pankhuenkhat, R.: *Thai Linguistics*. Mahachulalongkornrajavidyalaya University, Bangkok (1997)
16. Phanupong, V.: Grammar structure: sentences and phrases. In: *Thai Language* 3, pp. 127–172. Sukhothai Thammathirat Open University, Bangkok (2004)
17. Sayangkun, P.: Grammatical words in Northeastern Thai inscriptions from B.E.1957 to B.E. 2466. Silpakorn University, Bangkok (2005)



# Simulation and Analysis of Space Station Redocking

Rui Chen<sup>1\*</sup>, Shengjing Tang<sup>1</sup>, and Guojiang Sun<sup>2</sup>

<sup>1</sup> School of Aerospace Engineering, Beijing Institute of Technology, China

<sup>2</sup> China Academy of Space Technology, China  
chenrui\_bit@163.com, tangsj@bit.edu.cn

**Abstract.** Massive module redocking is a key technology of space station on-orbit assembly, due to test condition limit, it is difficult to perform effective experiment on the ground. In order to simulate the assembly process, here taking a simplified four rigid bodies space station model as an example, first, its dynamic equation is established based on Kane method, then numerical simulation compared to other methods is provided to verify model validation.

**Keywords:** Space Station, Redocking Process, Attitude Dynamics.

## 1 Introduction

China aims to finish own space station by 2020, its configuration is similar to Mir's multi-modular structure, one core module is connected to several experimental modules through a nodal module. Experimental module first dock with nodal module in axial direction, before next experimental module docking, the previous one need to shift to lateral dock port, this process is called redocking.

Due to microgravity environment of space, large size and mass of space station, and test condition limit, it is difficult to perform effective experiment on the ground to simulate the redocking process. So theory modeling and numerical simulation are main means to analyze dynamic characteristics of space station assembly. We simplify actual system into four rigid bodies model consisted of core module (body  $A$ ), one experimental module (body  $D$ ) and two robotic arms (body  $B$  &  $C$ ), its dynamic model is established by Kane method.

## 2 Dynamic Modeling

### 2.1 System Description

First of all, let's introducing four type coordinate system used in this paper: geocentric inertia frame  $E - XYZ$ , orbit frame  $O - X_O Y_O Z_O$ , arbitrary body frame  $I - x_I y_I z_I$  ( $I = A, B, C, D$ ) and entire system body frame  $S - x_S y_S z_S$  (Fig. 1.).

---

\* Corresponding author.

Specific definition of these coordinate can refer to [1]. Unit vectors  $\vec{e}_k, \vec{o}_k, \vec{i}_k, \vec{s}_k$  ( $k=1,2,3$ ) are fixed in corresponding frame.

Generalized coordinates  $R_1, R_2, R_3, \psi, \phi, \lambda, \alpha, \beta, \delta, \gamma, \eta$  are introduced as follows. Vector  $\vec{R}_{ES} = [R_1, R_2, R_3]^T$  is position coordinate of space station in frame  $E$ , and pitch angle  $\psi$ , roll angle  $\phi$ , yaw angle  $\lambda$  are used to describe the attitude of frame  $S$  in frame  $O$ , rotation sequence is 2-1-3<sup>[2]</sup>.  $\alpha, \beta$  are rotation angles of body  $B$  relative to body  $A$ , rotation sequence are 2-3;  $\delta$  is rotation angle of body  $C$  relative to body  $B$ , rotation axis is  $Z_C$ ;  $\gamma, \eta$  are rotation angles of body  $D$  relative to body  $C$ , rotating sequence are 3-2.

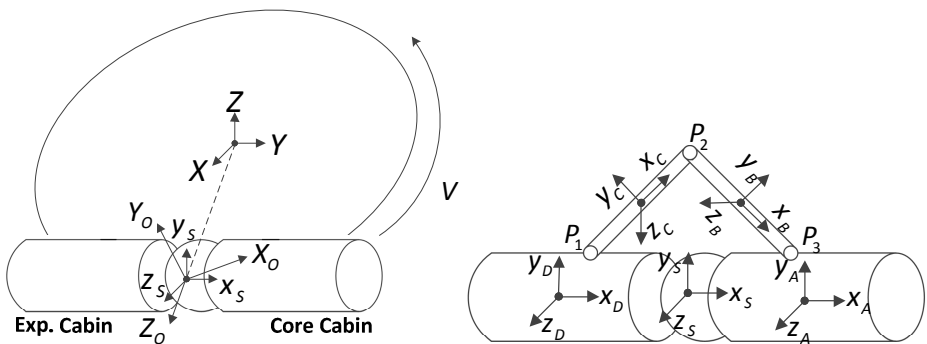


Fig. 1. Illustration of coordinate system

In order to facilitate the description of redocking, the lateral port of nodal module are labeled as port 1, 2, 3, 4, in which port 1 and port 3 are at the positive and negative direction of  $Z_A$  axis respectively, and port 2 and port 4 are at the positive and negative direction of  $Y_A$  axis respectively.

### 2.2 Generalized Speed

Generalized speed is a key concept in Kane method, selecting appropriate generalized speeds can greatly simplify the system dynamic equations. Refer to a novel generalized speed selection guideline proposed by Paul C. Mitiguy<sup>[3]</sup>, the generalized speeds of system is defined as follow:

$$\begin{aligned}
 u_i &= {}^E \vec{v}^S \cdot \vec{e}_i \quad (i=1,2,3) \quad , \quad u_i = {}^E \vec{\omega}^A \cdot \vec{a}_i \quad (i=4,5,6) \\
 u_7 &= ({}^E \vec{\omega}^B - \dot{\beta} \vec{b}_3) \cdot \vec{a}_2 \quad , \quad u_8 = {}^E \vec{\omega}^B \cdot \vec{b}_3 \quad , \quad u_9 = {}^E \vec{\omega}^C \cdot \vec{c}_3 \\
 u_{10} &= ({}^E \vec{\omega}^D - \dot{\eta} \vec{d}_2) \cdot \vec{c}_3 \quad , \quad u_{11} = {}^E \vec{\omega}^D \cdot \vec{d}_2
 \end{aligned}
 \tag{1}$$

Angular velocity of each body can be expressed as:

$$\begin{aligned}
 {}^E\vec{\omega}^S &= u_4\vec{s}_1 + u_5\vec{s}_2 + u_6\vec{s}_3, & {}^E\vec{\omega}^A &= u_4\vec{a}_1 + u_5\vec{a}_2 + u_6\vec{a}_3 \\
 {}^E\vec{\omega}^B &= u_4\vec{a}_1 + u_7\vec{a}_2 + u_6\vec{a}_3 + [u_8 - \sin(\alpha)u_4 - \cos(\alpha)u_6]\vec{b}_3 \\
 {}^E\vec{\omega}^C &= u_4\vec{a}_1 + u_7\vec{a}_2 + u_6\vec{a}_3 + (u_9 - u_8)\vec{c}_3 \\
 &\quad + [u_8 - \sin(\alpha)u_4 - \cos(\alpha)u_6]\vec{b}_3 \\
 {}^E\vec{\omega}^D &= u_4\vec{a}_1 + u_7\vec{a}_2 + u_6\vec{a}_3 + (u_{10} - u_8)\vec{c}_3 + [u_8 - \sin(\alpha)u_4 \\
 &\quad - \cos(\alpha)u_6]\vec{b}_3 + [u_{11} + \cos(\alpha)\sin(\beta + \delta + \gamma)u_4 \\
 &\quad - \sin(\alpha)\sin(\beta + \delta + \gamma)u_6 - \cos(\beta + \delta + \gamma)u_7]\vec{d}_2
 \end{aligned} \tag{2}$$

Velocity and acceleration of center of mass, angular acceleration can be obtained by differentiating corresponding vector. For example:  ${}^E\mathbf{v}^A = \dot{\vec{R}}_{EA}$ ,  ${}^E\boldsymbol{\sigma}^A = \ddot{\vec{R}}_{EA}$ ,  ${}^E\vec{\alpha}^A = \dot{{}^E\vec{\omega}^A}$ . The explicit expression of such variable is tedious, it isn't listed here.

Absolute angular velocity  ${}^E\vec{\omega}^I$  and velocity  ${}^E\vec{v}^I$  are functions of generalized speeds  $u_r$  ( $r = 1, 2, \dots, 11$ ), according to the definition of partial angular velocity and partial velocity:

$$\vec{v}_r^I = \frac{\partial {}^E\vec{v}^I}{\partial u_r}, \quad \vec{\omega}_r^I = \frac{\partial {}^E\vec{\omega}^I}{\partial u_r} \quad (I = A, B, C, D, S; r = 1, 2, \dots, 11) \tag{3}$$

### 2.3 Kinematics Equation

Relation between generalized coordinates and generalized speeds can be obtained by kinematic analysis. According to  ${}^E\vec{v}^S = \dot{\vec{R}}_{ES}$ , it can be easily concluded that:

$$\dot{R}_i = u_i \quad (i = 1, 2, 3) \tag{4}$$

Angular velocity of frame  $O$  in frame  $E$  is equal to  ${}^E\vec{\omega}^O = \dot{\theta}\vec{o}_2$ , and angular velocity of frame  $S$  in frame  $O$  is:

$$\begin{aligned}
 {}^O\vec{\omega}^S &= [\dot{\psi}\cos(\phi)\sin(\lambda) + \dot{\phi}\cos(\lambda)]\vec{s}_1 + [\dot{\psi}\cos(\phi)\cos(\lambda) \\
 &\quad - \dot{\phi}\sin(\lambda)]\vec{s}_2 + [-\dot{\psi}\sin(\phi) + \dot{\lambda}]\vec{s}_3
 \end{aligned} \tag{5}$$

Base on superposition principle of angular velocity:  ${}^E\vec{\omega}^S = {}^E\vec{\omega}^O + {}^O\vec{\omega}^S$ , substituting  ${}^E\vec{\omega}^O$ ,  ${}^O\vec{\omega}^S$ ,  ${}^E\vec{\omega}^S$  into above equation, we can get:

$$\begin{aligned}
 \dot{\psi} &= (\sin(\lambda)u_4 + \cos(\lambda)u_5) / \cos(\phi) - \dot{\theta}, \quad \dot{\phi} = \cos(\lambda)u_4 - \sin(\lambda)u_5 \\
 \dot{\lambda} &= u_6 + \tan(\phi)(\sin(\lambda)u_4 + \cos(\lambda)u_5)
 \end{aligned} \tag{6}$$

Similarly,

$$\dot{\alpha} = u_7 - u_5, \quad \dot{\beta} = u_8 - \sin(\alpha)u_4 - \cos(\alpha)u_6 \tag{7a}$$

$$\dot{\delta} = u_9 - u_8, \quad \dot{\gamma} = u_{10} - u_9$$

$$\begin{aligned} \dot{\eta} = & u_{11} + \cos(\alpha)\sin(\beta + \delta + \gamma)u_4 - \cos(\beta + \delta + \gamma)u_7 \\ & - \sin(\alpha)\sin(\beta + \delta + \gamma)u_6 \end{aligned} \tag{7b}$$

## 2.4 Generalized Active Force and Generalized Inertia Force

Ignoring small disturbance such as rarefied atmospherical drag, magnetic force, solar radiation pressure and so on, active forces and torques act on space station are gravity  $\vec{F}_S$ , gravity gradient torque  $\vec{T}_S$ , control force  $\vec{F}_{AC}$  and torque  $\vec{T}_{AC}$  act on core module, and control torque  $\vec{T}_{A/B}$ ,  $\vec{T}_{B/C}$ ,  $\vec{T}_{C/D}$  act on robotic arms.

Using two point gravity model, total gravity force and gravity gradient torque act on system are equivalent to resultant force and the moment acting on the instantaneous CM of system, then:

$$\vec{F}_S = -\frac{\mu M}{|\vec{R}_{ES}|^2} \vec{u}_{ES}, \quad \vec{T}_S = \frac{3\mu}{|\vec{R}_{ES}|^3} \vec{u}_{ES} \times \vec{J}_S \cdot \vec{u}_{ES} \tag{8}$$

In which  $\mu$  is gravity constant,  $\vec{J}_S$  is the moment inertia tensor about instantaneous CM of system, and  $\vec{u}_{ES} = \vec{R}_{ES} / |\vec{R}_{ES}|$ .

Control force and torque act on core module are:

$$\vec{T}_{AC} = \sum_{i=1}^3 T_{ACi} \vec{a}_i, \quad \vec{F}_{AC} = \sum_{i=1}^3 F_{ACi} \vec{a}_i \tag{9}$$

And control force act on robotic arms are:

$$\vec{T}_{A/B} = T_{AB1} \vec{a}_2 + T_{AB2} \vec{b}_3, \quad \vec{T}_{B/C} = T_{BC} \vec{c}_3, \quad \vec{T}_{C/D} = T_{CD1} \vec{c}_3 + T_{CD2} \vec{d}_2 \tag{10}$$

Therefore, generalized active force can be obtained by:

$$(\vec{F}_r)_S = \sum_I (\vec{\omega}'_r \cdot \vec{T}'_I + \vec{v}'_r \cdot \vec{F}'_I) \quad (I = A, B, C, D, S; r = 1, 2, \dots, 11) \tag{11}$$

Inertia force and torque of each body are:

$$\vec{F}_I^* = -M_I {}^E \vec{a}'_I, \quad \vec{T}_I^* = -{}^E \vec{\alpha}'_I \cdot \vec{J}_I - {}^E \vec{\omega}'_I \times \vec{J}_I \cdot {}^E \vec{\omega}'_I \quad (I = A, B, C, D) \tag{12}$$

Generalized inertia force can be expressed as:

$$(\vec{F}_r^*)_S = \sum_I (\vec{\omega}'_r \cdot \vec{T}_I^* + \vec{v}'_r \cdot \vec{F}_I^*) \quad (r = 1, 2, \dots, 11; I = A, B, C, D) \tag{13}$$

### 2.5 Kane Dynamic Equation

Kane dynamic equation is first-order differential equation of  $\dot{u}_r (r = 1, 2, \dots, 11)$ , it is the combination of generalized active force and generalized inertia force:

$$(F_r)_s + (F_r^*)_s = 0 \quad (r = 1, 2, \dots, 11) \tag{14}$$

Dynamic equations (14) together with kinematic equations (4), (6), (7) constitute the closed form motion equations of system. In order to save computing time and improve simulation accuracy, processing the motion equation with nondimensional length  $R = R_E$  and nondimensional time  $T = 2\pi\sqrt{R_E^3 / \mu}$ , in which  $R_E$  is the average radius of Earth.

### 3 Numerical Simulation

To verify the Kane dynamic equation, numerical simulation compared to Lagrange and Newton-Euler methods is provided. However, dynamic equation in literature [1][4] are based on simplified two rigid model, let  $M_B = M_C = 0, L_B = L_C = 0$ , then four rigid body model used in this paper reduce to two rigid body model.

Suppose that space station is moving in a counterclockwise equatorial orbit which is 500km height away from ground, ‘Mir’ module data [5] is selected as reference value of body A and B, core module takes no attitude control during redocking process, and simulation time is an orbit period.

At the beginning of redocking, frame S coincide with frame O, relative angular velocity and acceleration of this two frames are zero, namely  $\psi = \phi = \lambda = 0 \text{deg}$ ,  $\dot{\psi} = \dot{\phi} = \dot{\lambda} = 0 \text{deg/s}$ ,  $\ddot{\psi} = \ddot{\phi} = \ddot{\lambda} = 0 \text{deg/s}^2$ . By analysis of geometrical

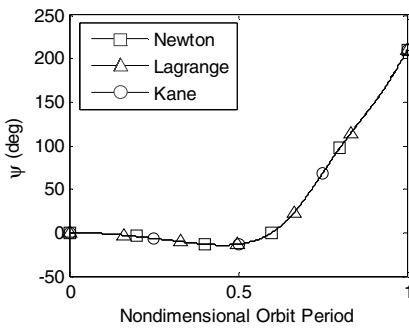


Fig. 2. Comparison of different method

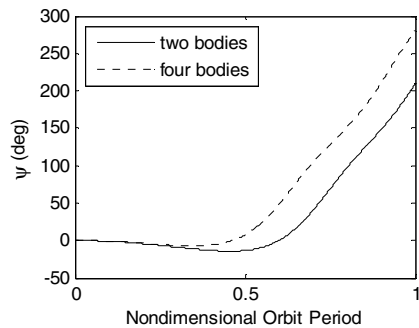


Fig. 3. Comparison different model

relation of system at the start and end moment, the initial and final value of  $\alpha, \beta, \delta, \gamma, \eta$  can be determined, the by using method of undetermined coefficient, time history function of these angle can be determined. For example, the sin form function of  $\alpha$  is:

$$\alpha(t) = (\alpha_f - \alpha_0) \left( 1 + \sin\left(\frac{\pi t}{\tau} - \frac{\pi}{2}\right) \right), \quad \tau = t_f - t_0$$

Take redocking to port 1 as example. Under this situation, experimental module always moves in orbit plane, so value of roll angle  $\phi$  and yaw angle  $\lambda$  always keep zero. Fig 2 is the time history of pitch angle  $\psi$ , it can be seen that the simulation results of Kane method is consistent with Lagrange and Newton-Euler method, it shows the correctness of Kane dynamics equation. Comparison of pitch angle time history between two body and four body model is shown in fig 3, there is a little difference at first half period, however, as the geometrical asymmetry of system becomes significant (Fig. 4 and Fig. 5), the difference gets larger.

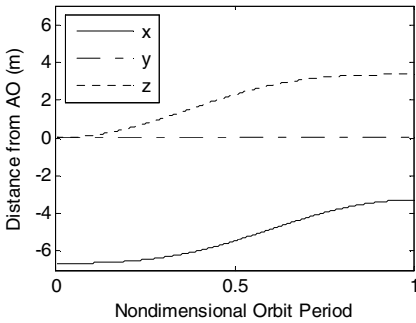


Fig. 4. Displacement of CM in frame A

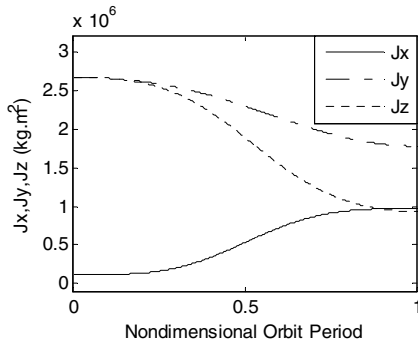


Fig. 5. Variation of moment of inertia

## 4 Conclusion

In this paper, motion equations of a simplified space station model based on Kane method is established, the derivation is much simpler than Lagrange and Newton-Euler method. Simulation results indicated that effective control scheme must be taken to overcome the great attitude movement caused by geometrical asymmetry during redocking. The motion equations proposed here is an appropriate choice for attitude control system design.

## References

1. Mah, H.W.: On the Dynamics of Spacecraft with A Slewing Appendage. The University of British Columbia (1986)
2. Kane, T.R., Likins, P.W., Levinson, D.A.: Spacecraft Dynamics. McGraw-Hill Book Company (1983)
3. Mitiguy, P.C., Kane, T.R.: Motion Variables Leading to Efficient Equations of Motion. The International Journal of Robotics Research 15(5), 522–532 (1996)
4. Syromiatnikov, V.S.: Manipulator system for module redocking on the Mir Orbital Complex. In: Proceeding of the 1992 IEEE International Conference on Robotics and Automation, Nice, France, vol. 1, pp. 913–918 (1992)
5. Fehse, W.: Automated Rendezvous and Docking of Spacecraft. Cambridge University Press (2003)

# Campus Mobile Navigation System Based on Shortest-Path Algorithm and Users Collaborations

Shir Ni Ler and Wan Mohd Nazmee Wan Zainon\*

School of Computer Sciences, Universiti Sains Malaysia,  
11800 USM, Penang, Malaysia  
noblesn@live.com, nazmee@cs.usm.my

**Abstract.** The main purpose of this paper is to present a research and development of an intelligent system for mobile phones that can help users finding the shortest path from a point-of-interest to another point-of-interest within Universiti Sains Malaysia (USM) main campus. Research is done to acquire an optimal navigation solution which primary goal is to provide shortest-path calculation. A mobile application based on Android that allows user to find a location quickly with minimum effort and collaborate with other users to enhance the experience of route-planning has been developed. Preliminary evaluation suggests that the proposed methods and the mobile application are helpful for the users in finding their point of interest within shortest time possible.

**Keywords:** Campus Navigation, Shortest Path Algorithm, Flexible Map Representation.

## 1 Introduction

USM main campus spans across a land area of 591.72 acres (240.13 hectares), which is approximately the dimension of 296 football fields. It has a total of 4 entrances, 25 blocks of in-campus accommodations, 5 administrative centers, 18 cafeterias, 6 campus service centers, 24 centers of excellence, 29 halls, 16 schools, 8 sport facilities, a mosque, and a museum [1]. Every year, thousands of new students enroll in this university. These students either take a campus commuter or walk around to get familiar with campus compound. Visitors in USM might have a hard time searching for a particular location in the campus. Each and every day, uncountable numbers of students, staffs, and visitors move around campus compound to perform tasks by means of walking, cycling, driving, or riding campus commuters.

The main intention of this research is to provide an optimal navigation solution through mobile application that are able to show a shortest-path calculation that benefits all USM residents, prospective students and staffs, as well as visitors in terms of location-search and route-planning. The research aims at building an architecture

---

\* Corresponding author.



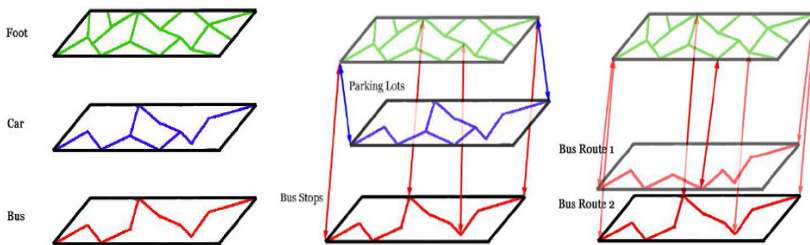
for a mobile campus navigator that finds the shortest path between any two points. The objectives of this research are:

- [1] to study on a flexible map representation method
- [2] to choose a suitable shortest-path algorithm with collision detection
- [3] to develop a prototype mobile application that can navigate a specific maps based on users collaborations

## 2 Background and Related Work

Mobile phones nowadays are embedded with Global Positioning System (GPS), Wireless-Fidelity (Wi-Fi), Bluetooth, and so forth. In line with the rapid growth of smart phone market, demand for location-based service (LBS) is sky-rocketing. Many tertiary institutions around the world have taken the initiative to research and develop a navigation system for the campus. Each navigation system proposed or developed has different implementations from others ranging for simple web-based navigation system with existing map application programming interface (API) plug-in to complex three-dimensional (3D) navigation system running on mobile phones. Although the project goal is to develop a campus navigation application for mobile phone, the research covers also web-based navigation applications as web-based campus navigation applications are still prevalent. Many useful concepts and theories of navigation system development were discussed in terms of web-based navigation application.

Boyd B., et al. [9] from Texas A&M University (TAMU) has built a web-based Campus Navigator on top of Google Map API that comprises three core units i.e. the web server, the database, and the query server. Multiple transportation modes are supported. Roadmap is denoted by graphs in which an edge represents a path and a vertex represents a point-of-interest. In order to cater multiple transportation modes, the authors implemented layering in building roadmap. A feature is added into Campus Navigator to allow administrator to add, edit, or remove edges at anytime.

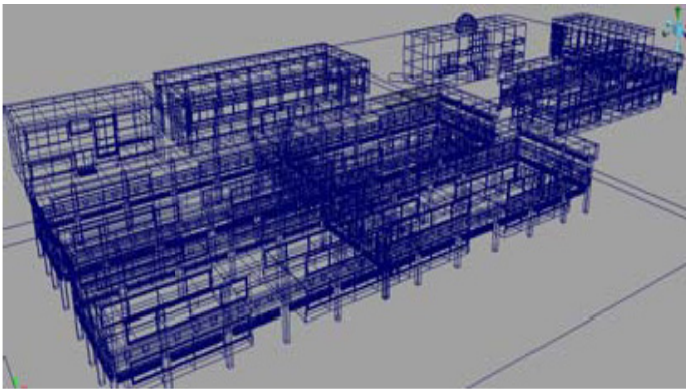


**Fig. 1.** Roadmap Layering, Sub-layering, and Layers Mapping

This Campus Navigator is a rather simple navigation system to implement. Basically, it makes use of existing map API. As virtual reality (VR) becomes a hotspot of research, it opens up a possibility of virtual campus navigation.

Wang J. et al. [3] proposed a web-based 3D Campus Navigation System for Asian Institute of Technology Campus Navigator (AITCN). They used two concepts i.e. spatial environment and nodal graph. Spatial environment is specified as a collection of buildings. A nodal graph is formed to represent buildings by vertexes and roads by edges. AITCN comprises four modules namely the spatial environment module, the nodal graph module, the animation module and the behavior module. The spatial environment module is an independent module that creates 3D models of each buildings in spatial environment. Decisions concerning polygons rendering, processing time and resource capabilities must be made in this module. The major goal of nodal graph module is to create a graph based on spatial environment and propose a shortest-path algorithm. Each building is assigned exactly one node near the main entrance to that building. The animation module has two components. 3D animation that shows user's avatar moving from origin point to destination. 2D mini map that gives an overview of the entire terrain, and a red box moving along with the avatar to indicates the location that the avatar is currently staying. The behavior module handles user-system interactions.

This 3D campus navigation system is more favorable and easy to use. However there are a few challenges facing 3D modeling of campus buildings (see Fig. 2). First of all, rendering of 3D model is time and resources consuming. Besides, changes of the structure of a building requires remodeling. In order to produce a better visual effect, models should be rendered with more polygons which means a significant increase of computation time and resources consumption.



**Fig. 2.** Models of 3D Nodes Example Created in Maya

Gallagher T. J. et al. [5] proposed campus navigation system that aims at providing students and staff with software able to guide them between any two locations on the university campus, either indoors or outdoors. The system is broken down into 3 main components i.e. client, server, and database. GPS signals are used for outdoor positioning. Wi-Fi signals are used for indoor positioning. Fingerprinting is used to build a database of signal strengths. The switch between indoor and outdoor modes when the user requests his or her location has to be made manually on the device.

### 3 Proposed Work

A collection of points-of-interest can be represented by a collection of vertices on a nodal graph whereas paths connecting points-of-interest are represented by edges on the graph. Vertices and edges can be added into, modified on and removed from the graph at any time according to the change of USM main campus topology. The chosen shortest-path algorithm (discussed in later section) does not detect collisions. To eliminate collisions, additional vertices will be added at every road junctions and near to the entrances of each buildings.

Dijkstra's shortest-path algorithm will be used to calculate shortest path connecting any two points-of-interest. Dijkstra's algorithm starts from a source (s) and computes each vertex v in set V the cost d(v) of the shortest path found so far between s and v. Initially, cost for vertex s is set to 0,  $d(s) = 0$ . Costs are set as infinity for all other vertices,  $d(v) = \infty$  for every v in V as we do not know any path leading to those vertices. At the end of the algorithm, d(v) should be the cost of the shortest path from vertex s to vertex v. If no such path exists,  $d(v) = \infty$ . To find the shortest path between vertex s and vertex v, edge relaxation is applied iteratively until shortest path is found. The algorithm maintains two sets of vertices S and Q. Set S contains all vertices for which we know that the value d(v) is already the cost of the shortest path. Set Q contains all other vertices. Set S starts empty. In each iteration, one vertex is moved from Q to S. This vertex is chosen as the vertex with lowest d(u). When a vertex u is moved to S, the algorithm relaxes every outgoing edge (u,v).

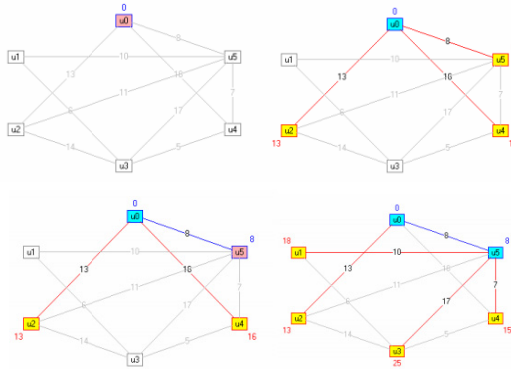


Fig. 3. Illustration of 4 Dijkstra's Shortest-Path Algorithm Iterations

Both GPS and Wi-Fi are suitable for building a campus navigation system as they are available and accessible in USM main campus. Besides, most mobile phones are GPS and Wi-Fi enabled. GPS is reliable for outdoor positioning but not indoor positioning. Wi-Fi is an alternative of GPS in indoor-environment. However, switching between GPS and Wi-Fi when going from outdoor to indoor or the other way round has to be done manually. Accuracy of Wi-Fi positioning greatly relies on density of Wi-Fi APs in the campus and the stability of Wi-Fi SS in every buildings.

In this project, only GPS positioning will be used due to the low stability of Wi-Fi SS at certain areas of USM main campus. Integration of Wi-Fi positioning will be listed as suggestion for future work.

A server having a database component must exist to store the vertices, edges, and weights of edges of the nodal graph. In addition, user collaborations are broadcasted through server.

## 4 Mobile System Descriptions and Functions

A mobile application that runs on Android platform will be created to solve the problem mentioned in Section 1. The mobile application comprises the following features:

- Function to select/search/input initial position (Point A) and final destination (Point B)  
*Scenario:* A student wants to go Library Hamzah Sendut II from School of Computer Sciences. He sets "School of Computer Sciences" as Point A and "Library Hamzah Sendut II" as Point B.
- Calculates and displays shortest path between Point A and Point B  
*Scenario:* The application finds all possible paths and ultimately, the shortest path among all possible paths that direct to Library Hamzah Sendut II from School of Computer Sciences. It displays the shortest path on map.
- Function that enables user to select points-of-interest (Point C, Point D,...,Point N) to stop by before reaching final destination  
*Scenario:* Let says the student who wants to go Library Hamzah Sendut II from School of Computer Sciences also wants to withdraw money from the nearest ATM machine (Point C) and meets a friend in front of DTSP (Point D) before heading to his final destination, Library Hamzah Sendut II. The application calculates the shortest path that connects Point A and Point B through Point C and Point D.
- Collaborations of users to enhance the efficiency of route-planning  
*Scenario:* A staff of USM, who is also an user of the application noticed a construction work is under progress and blocks off one of the road connecting School of Computer Science and Library Hamzah Sendut II, he marks the road as "blocked" so that the road will be automatically omitted while calculating other users' shortest path.

## 5 Discussion and Conclusion

Based on this study, we found out that both GPS and Wi-Fi has high availability and accessibility. GPS is suitable for outdoor positioning. Wi-Fi is suitable for indoor positioning. A hybrid of GPS and Wi-Fi positioning method is proper for building

navigation system for a campus. However, extensive research must be made to solve the issue of manually switching between GPS mode and Wi-Fi mode. 3D map representation consumes more time and computing resources than 2D map representation. LOD technology can be employed to lower rendering time. Most widely-used shortest-path algorithm is Dijkstra's shortest-path algorithm. Additional vertices can be added to avoid collisions.

Initial investigation suggest that users can find any buildings in USM with ease via the mobile application installed on their smart phones. Besides, they are able to go any points-of-interest within the shortest time possible by following the calculated shortest path. An extensive user acceptance studies will be conducted later this year to validate the proposed research and study the effectiveness of using the E-community collaborations features that can further enhance the process of route-planning. At the end, it is hoped that this mobile application prevents user from getting lost in USM as well as helping in avoiding unnecessary detour.

## References

- [1] Dzul kifli, A.R., et al. (eds.): Transforming Higher Education For A Sustainable Tommorow 2010: Laying The Foundation. Universiti Sains Malaysia (2011)
- [2] Chou, T.S., et al.: CaNPAs: A Campus Navigation and Parking Assistant System. In: IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan, October 8-11, vol. 1, pp. 631–638 (2006)
- [3] Wang, J., Lin, Z.: A General 3D Campus Navigator System. In: 2009 8th IEEE/ACIS International Conference on Computer and Information Science, pp. 1074–1078 (2009)
- [4] Wang, C.S., Hong, C.Y.: A Location and Interest Based Peer to Peer Virtual Navigation System. In: 2nd International Conference on the Applications of Digital Information and Web Technologies 2009, pp. 676–681 (2009)
- [5] Gallagher, T.J., Li, B., Andrew, G.D., Rizos, C.: A sector-based campus-wide indoor positioning system. In: 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Zurich, Switzerland, September 15-17, pp. 1–8 (2010)
- [6] Luo, D., Tan, G.: A Study for 3D Virtual Campus Navigation System Based on GIS. In: WiCOM 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1–5 (2008)
- [7] Zirari, S., Canalda, P., Spies, F.: WiFi GPS Based Combined Positioning Algorithm. In: 2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), pp. 684–688 (2010)
- [8] Ching, W., Teh, R.J., Li, B., Rizos, C.: Uniwide WiFi Based Positioning System. In: 2010 IEEE International Symposium on Technology and Society (ISTAS), pp. 180–189 (2010)
- [9] Boyd, B., Haigler, D.: The Campus Navigator,  
<https://parasol.tamu.edu/people/darlah/finalPaper.pdf>
- [10] Soares, M.G., Malheiro, B.: An Internet DGPS Service for Precise Outdoor Navigation. In: Proceedings of IEEE Conference Emerging Technologies and Factory Automation, vol. 1, pp. 512–518 (2003)

- [11] Han, D., Lee, M., Chang, L., Yang, H.: Open Radio Based Indoor Navigation System. In: Proceedings of 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 844–846 (2010)
- [12] Manabe, T., Yamashita, S., Hasegawa, T.: On the M-CubITS pedestrian navigation system. In: IEEE Intelligent Transportation Systems Conference, ITSC 2006, pp. 793–798 (2006)
- [13] Liu, W., Liu, Z., Zhang, Y.: Interactive Mobile Campus Based on Position Perception. In: Proceedings of 2011 Fourth International Symposium on Computational Intelligence and Design (ISCID), vol. 1, pp. 147–150 (2011)

# Tracking Multiple Fish in a Single Tank Using an Improved Particle Filter

Wong Poh Lee<sup>1</sup>, Mohd Azam Osman<sup>1</sup>, Abdullah Zawawi Talib<sup>1</sup>,  
Jean-Marc Ogier<sup>2</sup>, and Khairun Yahya<sup>3</sup>

<sup>1</sup>School of Computer Sciences, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia

<sup>2</sup>Laboratoire L3i, Pôle Sciences & Technologies, Université de La Rochelle, France

<sup>3</sup>Centre For Marine & Coastal Studies (CEMACS), Universiti Sains Malaysia,  
11800 Penang, Malaysia

wongpohlee@hotmail.com,

{azam, azht}@cs.usm.my,

jean-marc.ogier@univ-lr.fr, khairun@usm.my

**Abstract.** Studies on tracking fishes have become a popular research endeavour in recent years. Many methods have been used to track fishes by integrating microchips in fishes, using infra-red cameras, image processing and motion sensor. The use of particle filter in the process of tracking has been widely used by researchers. Particle filters are used to track people, fluid movement and animals. In this paper, the particle filter algorithm is improved to track multiple fish in a fish tank. The aim is to identify every fish trajectory and fish target location for further analysis. The main challenge is to ensure that the correct fish are tracked and the algorithm manages to identify specific fish even if they overlap with each other. The objective of the study is to improve the existing particle filter to track multiple fish in a single fish tank. The improved algorithm contains an additional cache which stores the object's position to estimate the next potential move of the fish. The result is evaluated by comparing existing algorithm without the enhancement with the improved algorithm. Besides, suggestions in improving the particle filter will also be discussed in this paper.

**Keywords:** particle filter; fish tracking; tracking system.

## 1 Introduction

Tracking multiple fish using computational methods have become a research endeavour among researchers. Different concepts have been proposed such as installing water sensors and video cameras to identify movement speed, colours, shapes and swimming patterns of the fish [1][2]. In this research, an existing particle filter algorithm is further improved. This algorithm is improved by considering multiple fish in the water and identifying the trajectory patterns of the fish. A cultured fish tank installed with water sensors to monitor water pH, dissolved oxygen and water temperature is set up together with a network camera for the case study. Koi fish are chosen due to their active swimming behaviour, variety of colours and easy-to-adapt habitat

in the water. A real-time prototype system which models the fish swimming pattern of the enhanced algorithm was developed for the purpose of experimenting with the proposed enhanced algorithm of this research.

In our case study, a network camera is placed above the fish tank to monitor fish swimming patterns on a single screen. The particle filter is applied to detect the change in speed or vector movement (trajectory) of the fish in the viewing field. This method allows the fish to be tracked down digitally.

## 2 Background Study

### 2.1 Particle Filter

Particle filter algorithm uses the weighted distribution approach to determine the trajectory of the tracked object by taking the particle with the highest weight or the weighted mean of the particle set at each time step[3][4].

Particle filter algorithm is suitable for tracking the fish movement in the water as the algorithm is also able to track multiple objects (fish) and not limited to just a single object in a frame[5].

In a particle filter algorithm, there are 3 steps in the implementation process which are the initialization step (prediction), sampling step (update) and selection step (re-sample) as shown below [6].

1. Initialisation,  $t = 0$ .
  - For  $i = 1, \dots, N$ , sample  $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$  and set  $t = 1$ .
2. Importance sampling step
  - For  $i = 1, \dots, N$ , sample  $\tilde{\mathbf{x}}_t^{(i)} \sim p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)})$  and set  $\tilde{\mathbf{x}}_{0:t}^{(i)} = (\mathbf{x}_{0:t-1}^{(i)}, \tilde{\mathbf{x}}_t^{(i)})$ .
  - For  $i = 1, \dots, N$ , evaluate the importance weights
 
$$\tilde{w}_t^{(i)} = p(\mathbf{y}_t | \tilde{\mathbf{x}}_t^{(i)}). \quad (1.8)$$
    - Normalise the importance weights.
3. Selection step
  - Resample with replacement  $N$  particles  $(\mathbf{x}_{0:t}^{(i)}; i = 1, \dots, N)$  from the set  $(\tilde{\mathbf{x}}_{0:t}^{(i)}; i = 1, \dots, N)$  according to the importance weights.
  - Set  $t \leftarrow t + 1$  and go to step 2.

Particles are initially scattered across the screen. For every time frame, the object that moves will be identified based on motion model or colour model. The particles around the moving object will be given higher weight. As the object continues to move, the particles will be accumulated around the object. This enables the tracking of the object's trajectory patterns and location based on the highest weight at each time step.



## 2.2 Condensation Algorithm

Condensation algorithm was proposed by Isard et. al[6] using a probabilistic algorithm to detect and track the contour of an object. This algorithm was originally improved from the existing particle filter algorithm.

Condensation algorithm is able to track objects which are represented by parameterized spline curves, in substantial clutter at video frame rate. Besides, condensation algorithm is capable of running in real-time on modest hardware. This algorithm has been applied to tracking multiple of objects and often used in traffic scenarios as vehicles come and go[7,8].

Condensation algorithm works in a similar manner as particle filter. However, the unique thing about condensation algorithm is that it does not compute every pixel of an image frame. Pixels are chosen randomly.

In another paper, Yui et al. [9] used an adaptive approach to define condensation algorithm in modelling a face tracking problem.

## 3 Implementation

### 3.1 Setting Up the Fish Tank

A fish tank is set up with a camera placed above the fish tank. The camera is able to capture the view of the whole tank. The video is recorded in real-time to enable continuous monitoring of the fish. A total of 5 koi fish with an average of 12cm in length are taken as the case study. The reason koi fish are used is due to the active swimming behaviour and the in colours which enable different fish to be distinguished. A set of fish which consist of fish with nearly the same colour is also placed into the tank to identify if the use of the proposed algorithm is able to track and differentiate between the fish.

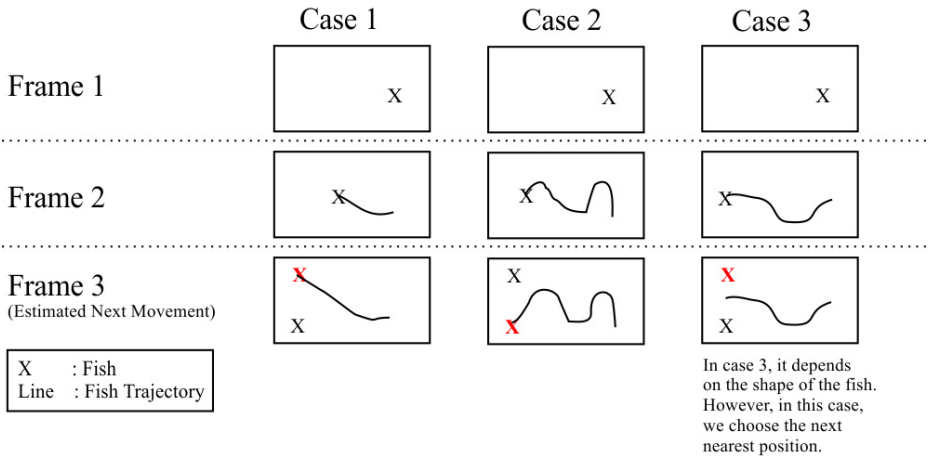
### 3.2 The Experiment and Enhanced Algorithm

In our experiment, we have recorded and cropped a video with duration of 4 minutes. The resolution of the video is 1280 x 720 pixels. The video is capable of displaying 30 frames per second which is sufficient to track the fish movement by comparing frames for every second.

The same video is used to evaluate the algorithm which has been improved from the existing algorithm. We improved the particle filter algorithm as proposed in [6]. The pseudo code of the improved algorithm is as follows:

- Particles are randomly scattered across the video frame.
- For every time frame, the object which moves will be identified based on a motion model.

- An additional cache storing the object’s position is proposed here. This cache will estimate the next potential move of the fish in order to continue tracking the trajectory pattern of the fish which has been lost due to overlapping or mis-track (as shown in Figure 1). The fish are swimming upwards (Case 1) and downwards (Case 2). Therefore, the possible next location of the fish depends on the previous direction of the fish.
- The particles surrounding the moving object are given higher weight.

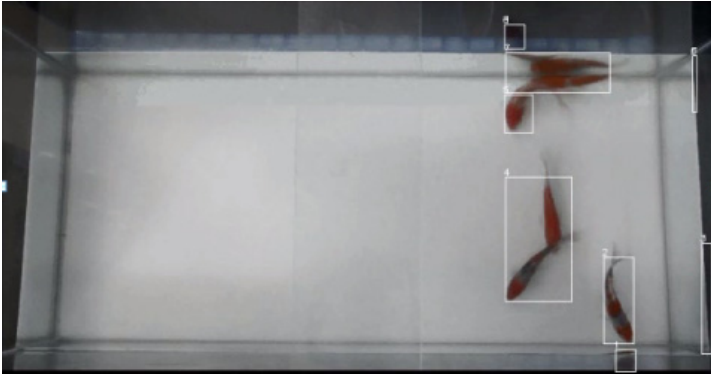


**Fig. 1.** Estimated Next Movement of Fish Target Location Based on 3 Frames

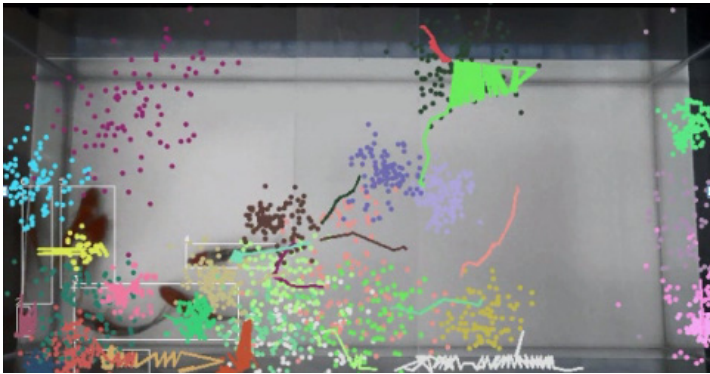
## 4 Results

Based on the experimental work carried out using the proposed algorithm, it is found that the algorithm manages to track the trajectory patterns and the locations of the fish (Figure 2). The lines signify the trajectory patterns of the fish in the water (Figure 3). However, some lines ends after a few seconds and the same fish is identified as a different fish. This is because some fish are overlap with each other and some colour changes affect the tracking of the fish. Besides, the quality of the video also affects the tracking of the fish. By adding a cache that stores the fish direction and moving patterns, the algorithm improves by displaying a longer trajectory pattern instead of short partial lines from the fish movement.

The experiment shows a non-consistent tracking environment where the percentage of accuracy cannot be evaluated. This is because every fish shows a non-consistent way of swimming. Other factors such as lighting, bubbles and swimming depth of the fish affect the accuracy of the tracker.



**Fig. 2.** Blob Images of Tracked Fish



**Fig. 3.** Trajectory Patterns of Tracked Fish

## 5 Conclusion

The improved tracking system performs well when fewer numbers of fish are tracked in the fish tank. By implementing a cache in the existing algorithm, the position and direction of the fish can be identified. This allows the algorithm to identify mis-tracked fish and to track back the same fish again. However, when more fish are placed, the algorithm fails to distinguish between the same and different fish.

In conclusion, further studies need to be conducted to improve the tracking system by considering the colour and shape of the fish. In addition, the swimming depth parameter of the fish is important as additional information in the tracking process. The swimming depth parameter allows the algorithm to identify if two or more fish are touching each other or swimming above one another in a fish tank.

## References

1. Wong, P.L., Osman, M.A., Talib, A.Z., Yahya, K.: Modelling of Fish Swimming Patterns Using an Enhanced Object Tracking Algorithm. *Frontiers in Computer Education*, 585–592 (2012)
2. Muizz, W.A., Osman, M.A., Talib, A.Z., Yahya, K.: Framework For Modelling of Fish Behaviour Through Fish Swimming Patterns. In: *Ocean & Coastal Observation: Sensors and Systems (OCOSS 2010)*. Congress Center Le Quartz, Brest (2010)
3. Mariño, C., Ortega, M., Barreira, N., Penedo, M.G., Carreira, M.J., González, F.: A Simple Implementation of the Condensation Algorithm. *Computer Methods and Programs in Biomedicine* (2011)
4. Monteiro, J.B.O., de Andrade Silva, J., Machado, B.B., Pistori, H., Odakura, V.: Multiple Mice Tracking using a Combination of Particle Filter and K-Means. In: *Computer Graphics and Image Processing, SIBGRAPI 2007*, Minas Gerais, pp. 173–178 (2007)
5. Williams, R., Purser, J.: Application of the Particle Filter to Tracking of Fish in Aquaculture Research. In: *Digital Image Computing: Techniques and Applications (DICTA)*, Sch. of Comput. & Inf. Syst., Univ. of Tasmania, Hobart, TAS, pp. 457–464 (2008)
6. Isard, M., Blake, A.: CONDENSATION – conditional density propagation for visual tracking. *International Journal on Computer Vision* 29 (1998)
7. Meier, E.B., Ade, F.: Tracking cars in range images using the CONDENSATION algorithm. In: *Proceedings of the 1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems*, pp. 129–134 (1999)
8. Koller-Meier, E., Ade, F.: Tracking Multiple Objects Using the Condensation Algorithm. *Journal of Robotics and Autonomous Systems* 34, 93–105 (2001)
9. Yui, M.L., Ross, B.J., Darrell, W.L.: A Novel Appearance Model and Adaptive Condensation Algorithm for Human Face Tracking. *Colorado State University, Fort Collins, CO 80523* (2008)

# Efficient 3D Model Visualization System of Design Drawing Based on Mobile Augmented Reality

Yeon-Jae Oh and Eung-Kon Kim\*

Department of Computer Engineering, Sunchon National University  
413 Jungangno, Sunchon-si, Jeollanam-do, Republic of Korea  
{Oksug10, kek}@sunchon.ac.kr

**Abstract.** In this paper, marker-less mobile augmented reality based 3D model visualization of efficient design was implemented by applying augmented reality that is recently getting the most brilliant spotlight in the shipbuilding industry. This system helps even an amateur understand relevant individual object with ease by visualizing ship design being progressed based on existing 2D individual object as realistic 3D ship individual object model. This 3D visualization would be helpful for activating negotiation for new shipbuilding order as it is easy to understand virtual ship design.

**Keywords:** Augmented Reality, Ship Design, Virtual Design, Design Visualization, Digital Shipbuilding.

## 1 Introduction

Recently, R & D on mobile augmented reality by using portable mobile terminal together with ubiquitous computing technology is being intensified. The reason of this trend is that as technology of augmented reality has been developed day by day, diversified research on augmented reality technology by using mobile based on merit of its mobility and popularity is under progress.

Mobile augmented reality as a technology of grafting virtual information in actual world by using smart phone is being utilized in various fields. As smart phone has a base through which platform based program is operable, diversified augmented reality application program such as GPS based LBS service and touch screen is able to be operated by building such function internally.

While in the past, augmented reality as an area of virtual reality focused on synthesis of actual image information with 3D virtual object, recent augmented reality is being evolved to the direction of diversified internet service, mash-up and information provision based on technology of camera, network, GPS information and directional sensor. Along with this trend, in the shipbuilding industry also, augmented reality is booming. However, as shipbuilding industry that is still unexplored field requires development of new technology, this industry needs graft of diversified IT technologies[1]. In other

---

\* Corresponding author.

words, in order to ensure next generation high value added shipbuilding industry, a key research that may support the best program in the field of shipbuilding off-shore facilities is required.

In this paper, in case of visualizing 2D drawing as 3D model combining it with actual world by using mobile augmented reality, understanding for each design document could be enhanced and exact and efficient design and manufacturing process control are enabled.

This paper comprises Section 2 (Relevant research), Section 3 (Mobile augmented reality based 3D visualized system of efficient design document), Section 4 (Implementation of mobile augmented reality based 3D visualized system of efficient design document) and Section 5(Conclusion).

## 2 Related Works

**Mobile Augmented Reality:** Recently, as distribution of smart phone is increased, concern over augmented reality has been rapidly emerged[2]. Mobile is furnished with conditions favorable for implementing augmented reality as it is individualized and miniaturized owing to its rapid development[3-5]. Mobile of today of which function is not just limited to simple text service has a close relation with our daily life by incorporating diversified functions such as entertainment, game and ubiquitous function. Ubiquitous interactive function in real time by diversified function of mobile device is called Mobile augmented reality.

diversified expectation psychology including increased patent application works in this forecast as well. Most typical augmented reality includes platform of Ovjat[6], Layar[7], Wikitude[8] and Junaio[9].

**Necessity and Status of AR Shipbuilding Industry:** Form, scale and compartment arrangement being required at the time of ship contract are required to be proposed based on the requirement of the client but existing ship exclusive 2D CAD design document was hard to be understood by ordinary people as ship design information and relevant data are diffused and complicated and it has a disadvantage of requiring long design time due to its repeated work. In order to solve this problem, an attempt of performing basic ship design by utilizing virtual reality and augmented reality technology is required.

Computer technologies being rapidly developed from the beginning of the end of 20th century presented a new area of virtual space and it has been developed so that all the life cycle from product design to ship operation as well as system composition could be controlled under virtual environment. In particular, nature of shipbuilding industry is not only labor intensive but also technology intensive as highly value added industry and recently, various researches and projects for enhancing quality through shortening of shipbuilding process and reducing process error rate by grafting IT convergence technology are under progress.

In addition, as AR technology visualizes shipbuilding process as well as design field more realistically by combining actual environment with virtual environment

and enables existing knowledge or information adequate for production site situation, AR technology is expected to improve productivity and quality level innovatively in existing ship design and shipbuilding process.

### 3 System Design

This system has been implemented in order to obtain industrial benefit by applying it to off-shore industry based on 3D model visualization technique of mobile AR based efficient design. The objective of this study is to develop 3D model algorithm of CAD design and 2D drawing through analysis of the ship owner’s requirement by using virtual simulator of ship operation for design verification.

For proposed implementation environment system, a smart phone Galaxy Note that employs 8 Million pixel Camera was used. The system was implemented in Android NDK and native code was implemented through JNI by using OpenCV library.

Figure 1 shows structure chart of design drawing 3D model visualization system based on mobile AR. By using mobile phone in the real world, 2D design drawing is recognized. Design drawing is tracked by tracking module through matching with design document of DB. 3D model is displayed on mobile phone under application of user interface module including adjustment of parameters, matching position and point through rendering process as virtual object after being imported from DB of design document.

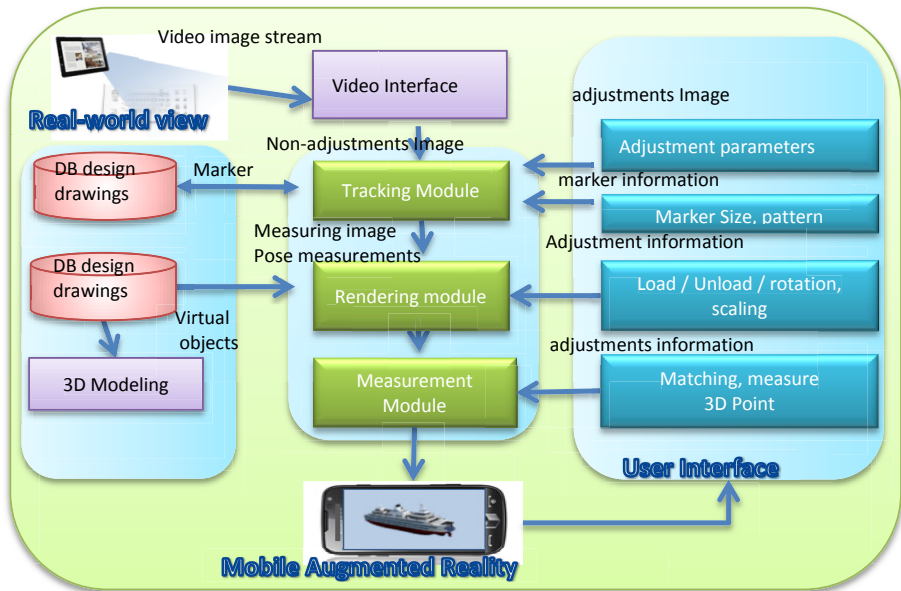


Fig. 1. Structure chart of design drawing 3D model visualization system

## 4 System Implementation

3D model visualization technique implementation of mobile AR based efficient design is classified into 5 stages; manufacturing 3D object model, algorithm recognizing ship drawing, algorithm of geometric transformation, development of AR browser for design document visualization and its visualization system.

**Manufacturing 3D Object Model:** In order to implement 3D ship system by using mobile AR, 3D MAX was used. By sampling 10 types of ship that are required to be directly built by shipbuilding company, it was manufactured by 3D model. As the most typical ship, car-ferry boat as a transport means that could connect each island was manufactured. Model being manufactured as car ferry boat was sampled as .OBJ and .3DS file.

**Image Marker of Ship Drawing:** In using marker for tracking object, there are artificial marker and marker-less. In this study, a mode of marker-less was adopted. In other words, by extracting a feature point in order to recognize 2D design document as marker, it was applied to AR application program. As storing characteristic of relevant drawing by extracting it, it may be used together with marker of a specific pattern being defined in advance. In this system, design document was recognized by using mobile phone camera and in case of recognized ship drawing, its edge point detection, boundary tracking algorithm were used by utilizing boundary edge detection of ship drawing. ROI (Region of Interest) of marker is extracted by selecting outline information after extracting all the edge points of marker inserted in ship design document.

**Extraction, Matching of Feature Point and Recognition of Design Document:** In order to recognize drawing without damaging contents of existing ship drawing, drawing contents is recognized by marker-less based image that allows drawing contents to be recognized as one marker.

At this time, as a matching algorithm of corresponding point, SURF (speeded up robust features) [10] is used. Matching is performed by extracting a few numbers including distribution of sampled pixel value and even though angle or distance of camera object at the time of shooting ship design document is different, same object is searched through a process of rotation and deformation. The most typical matching algorithm of corresponding point is SIFT (Scale- Invariant Feature Transform) and SURF but SURF algorithm is used as calculation volume of SIFT is large and it has limitation in its commercial use.

**Image Matching Module:** 3D visualized system of ship design document infers posture by extracting marker and image information in 2D design document image inputted from camera. In matching of 2D drawing and 3D model, geometric transformation is used for exact matching depending on variation of distance and angle between drawing and camera. Image matching is a process where two images are exactly coincided when actual drawing and 3D model are shown in overlapped condition. It shares data such as graphic image by receiving video image data in frame



buffer of graphic system. In order to express virtual object being displayed at present more realistically, create natural image and to show an effect of being covered by shadow or other object, an effect of natural light was used. Figure 2 is a screen showing position setting for image matching.

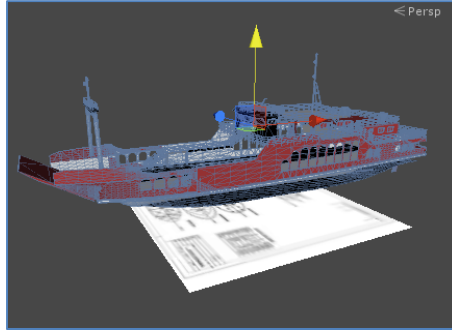


Fig. 2. Position setting for image matching

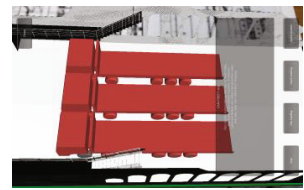
**AR Browser Development and Visualization System for Visualizing Design Drawing:** AR browser for recognizing design drawing is displayed on mobile phone by synthesizing virtual 3D model on design drawing that is a real image. In other words, it is visualized by augmenting virtual ship model comprising 3D model in smart phone or tablet PC. In this system, as more accurate mutual communication among the client, designer, construction engineer and managers is enabled, any communication problem could be reduced. Table 1 shows a screen implementing this system in mobile phone. When implementing mobile AR, most main screen was arranged to express total ship view together with detailed information by each floor. In addition, by leaving a window where comment of the client could be entered and reflected, a function of two-way communication for ensuring confirmation by the contractor, designer and construction engineer was supplemented. Figure 3 is a screen implementing this system.



(a) Overall ship view



(b) Car loading status



(c) Truck loading status

Fig. 3. Screen implementing this system

## 5 Conclusion

Mobile AR provides great change in the global market as it has been developed day by day for its excellent mobility and portability. Responding to this trend, shipbuilding industry is also trying to apply mobile AR. However, function of already applied AR system was limited to that of displaying simplified text or pictures on screen by using just simple POI.

In this paper, ship contract awarding system being progressed based on existing 2D system is intended to be progressed by applying 3D model. The client is able to place an engineering and emotional order while watching 3D applied mobile AR system rather than 2D based ship design drawing.

In addition, the contractor is able to confirm its own requirement by using virtual system and additional requirement could be reflected. As a future study, this system is intended to be changed by matching with opinion of the contractor and the client through application of actual environment. In addition, this system is scheduled to be utilized as an education program for the students who wish to study actual shipbuilding engineering and basic data of marketing for activating new shipbuilding contract negotiation.

**Acknowledgement.** This research was financially supported by the Ministry of Science, Ict & future Planning(MSIP) and National Research Foundation of Korea(NRF) and Jeonnam Science & Technology Promotion Center (JNSP) through the research & development project of Jeonnam Science Park.

## References

1. Oh, Y.J., Cho, O.-H., Kim, E.-K.: Design of 3D Display System using Android. The Journal of The Korea Institute of Electronic Communication Sciences 7(5), 1011–1016 (2012)
2. Paucher, R., Turk, M.: Location-Based Augmented Reality on Mobile Phones. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 9–16 (2010)
3. Michelis, G.D., Loregian, M., Martini, P.: Directional Interaction with Large Displays Using Mobile Phones. In: 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (2006)
4. Fails, J.A., Druin, A., Guha, M.L.: Mobile Collaboration: Collaboratively Reading and Creating Children's Stories on Mobile Devices. In: Proceedings of the 9th International Conference on Interaction Design and Children (2010)
5. Paucher, R., Turk, M.: Location-Based Augmented Reality on Mobile Phones. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 9–16 (2010)
6. <http://www.ovjet.com>
7. Layar, <http://www.layar.com>
8. Wikitude, <http://www.wikitude.com>
9. Junaio, <http://www.wJunaio.com>
10. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding 110, 346–359 (2008)

# A New Co-evolutionary Immune Algorithm for Flow Shop with Zero Wait

Zhenhao Xu<sup>\*</sup>, Shun Zhang, and Xingsheng Gu

Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China  
{xuzhenhao, xsgu}@ecust.edu.cn,  
zhangshun\_21@163.com

**Abstract.** The flow shop scheduling problems with zero wait is considered as one of the most challenging problems in the field of scheduling. This paper deals with the problem considering the makespan minimization as the objective. In order to solve the corresponding model, a new co-evolutionary immune algorithm (NCIA) is proposed based on the '80/20' rule and a designed local search strategy. Moreover, a global crossover operator is presented to preserve excellent genes fragment and increase the diversity of population during the process of evolution. to confirm the performance of the proposed algorithm, in terms of solution quality, the algorithm is applied to various test instances of benchmark problems available in the literature. The computational results show that the proposed algorithm outperforms other algorithms.

**Keywords:** Co-evolutionary Immune Algorithm; Global Crossing; Flow Shop; Zero Wait.

## 1 Introduction

In many flow shop situations, jobs need to pass through a sequence of processors without any delay between the successive operations of a job. It is such zero wait or non-delay permutation flow shop scheduling problems. To tackle this problem, many approaches have been proposed. Sapkal and Laha [1] presented an efficient heuristic method to minimize total flow time in no-wait flow shop scheduling. Nagano et al. [2] proposed a new hybrid metaheuristic Genetic Algorithm-Cluster Search to solve no-wait flow shop problem where the set-up time of a job is separated from its processing time. Hsieh et al. [3] applied an immune algorithm to solve the multiple-machine no-wait flow shop scheduling. Rabiee et al. [4] proposed a robust meta-heuristic algorithm named the adapted imperialist competitive algorithm for a no-wait two-machine flow shop. Ruiz and Allahverdi [5] proposed several heuristics and local search methods based on GA and iterated greedy procedures for the general m-machine no wait case. Bertolissi [6] presented a heuristic algorithm for no-wait flow-shop problem, and which goal is to minimize the sum of the total flow-times.

---

<sup>\*</sup> Corresponding author.

In this paper, a flow shop scheduling problem with zero wait is considered. The objective function of the problem is to minimize the makespan. A New Co-evolutionary Immune Algorithm (NCIA) is presented to solve this problem. The rest of this paper is organized as follows: Section 2 gives the problem definition. Section 3 designs an effective crossover operator. In Section 4, the proposed NCIA algorithm is given. The experimental results are provided in Section 5. Finally, Section 6 consists of conclusions.

## 2 Model Formulation

We take the same assumptions regarding the multiproduct process from the reference [7]. We formulate the scheduling problem as follows. There are  $N$  ( $i=1,2,\dots,N$ ) products for processing on  $M$  ( $j=1,2,\dots,M$ ) units. The processing time of products  $i$  on unit  $j$  is  $T_{i_k j}$ , and  $S_{i_k j}$  represents the starting time of product  $i$  on unit  $j$ , where  $i_k$  is the order of product  $i$  in processing procedure. Accordingly,  $S_{i_k e}$  and  $T_{i_k e}$  mean the starting time and the finishing time of the last operation of product  $i$ . The scheduling criterion is *makespan*, that is  $\min\{Z = \max(S_{i_k e} + T_{i_k e})\}$ . We used the procedure of calculating delays of the start of product on the first unit presented by Reddi and Ramamurthy [8]. Let  $d_{ij}$  define the delay between products  $i$  and  $j$ , when product  $j$  follows product  $i$  in the sequence. It is given by

$$d_{ij} = \max_{m=2,M} \left\{ 0, \sum_{k=2}^m O_{ij} - \sum_{k=1}^{m-1} O_{ij} \right\}.$$

## 3 New Crossover Operator

### 3.1 Typical Crossover Operators

Crossover operator is used to assemble new individuals and implement the effective search in the solution space. The individuals are selected randomly for crossover operations according to the crossover rate in the population. And it is a good method to increase the diversity of the offspring through the intersection between the parent individuals. There is a wide range of different crossover operators in the related research literatures. And the single point crossover and the local crossover are two typical operations.

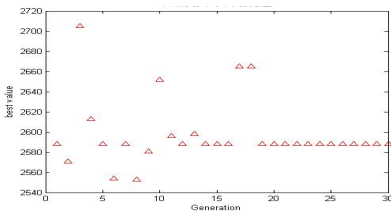
### 3.2 Global Crossover Operator

As we know, the single point crossover can restrain the diversity of effective individuals of the new population. And the local crossover will destroy the good genes fragment after the cut-point. So we design a new crossover operator for this

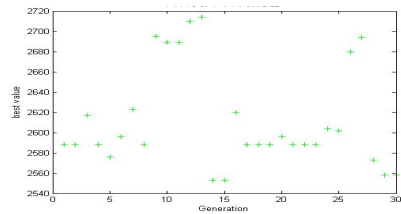
kind of scheduling problem, which is called global crossover operator. The global crossover has divided into two steps. In the first step, one group chromosome with two parent individuals is chosen to be intersected to produce two new groups with four offspring individuals by preserving the gene fragments before and after the cross-point separately. Obviously, two scenarios will emerge. One is that the two new groups are worse than the parents group, and the other is that they are better than the parents group. The former indicates that the degeneration appeared during the crossover. If it is happened in the process of evolution, the new groups will be ignored. While, the latter shows that the new offspring are excellent. And the new groups will be intersected to construct one new group, which is the second step of the global crossover.

### 3.3 Performance Analysis of Three Crossover Operators

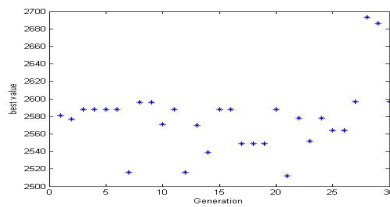
In the simulation, three crossover operators are all used in the improved co-evolutionary immune algorithm for various scales instances. The scatter diagrams of different crossover operators are Fig.1 – Fig.3.



**Fig. 1.** Distribution of the optimal values with single point crossover



**Fig. 2.** Distribution of the optimal values with local crossover



**Fig. 3.** Distribution of the optimal values with global crossover

As we can see from these figures, the single point crossover operator has less damage to good solutions, but it can lead to the local optimum which is shown clearly in Fig.1. In Fig.2, the local crossover operator can produce various individuals in the evolution, yet which is bad for the convergence of the algorithm. However, the global crossover operator can combine the advantages of both. It can keep the diversity of the population, which is demonstrated in Fig.3; besides, it can make the algorithm evolve in the direction of convergence.

## 4 A New Co-evolutionary Immune Algorithm

### 4.1 Extraction and Injection of the Vaccine

As we all know, the quality of the vaccine will be greatly affected by the validity of the prior knowledge. And it will be uncertain. Then we proposed a novel means of vaccine extraction to improve the validity of vaccine in this paper. In the evolution, the best individual of the current generation need to be found. And the local search around the best individual is to be conducted. The solution with well objective value found by the local search will be chosen as the vaccine. When performing the vaccination each time, the excellent gene segment will be injected into a certain number of individuals.

### 4.2 A New Mechanism of Population Selection – “80/20” Rule

Pareto’s 80/20-law is the rule given by Pareto who observed that 20 percent of the population owned 80 percent of the usable land [9]. During the evolution, the good individuals are always expected to play a vital role than the worse. On the basis of 80/20 rule, the mechanism of population reproduction is as follows. Firstly, evaluate each individual in current population, and rank them by the corresponding fitness value; secondly, select the top 20% individuals of the current population into the next generation; finally, select the other 80% individuals of the current population into the next generation according to the probability  $P = 20\% / 80\% = 0.25$ . At the same time, in order to keep the same number of population, extract randomly the top 20% individuals to replace the eliminated ones.

### 4.3 A Local Search Strategy

Although the speed of convergence can be increased by applying the ‘80/20’ rule at the early stage, it maybe leads to slow convergence in the later. Hence, the local search is introduced to improve the performance of the algorithm. It is showed as follows. Step1. Select one individual and compute its objective as  $f(x^{best})$ . Step 2. Search another individual  $x^{now}$  in the neighbor field of  $x^{best}$ . If  $f(x^{now}) < f(x^{best})$ ,  $x^{now} = x^{best}$ . Step 3. Check the stopping criterion. If not stop then go to Step 1.

## 5 Simulation Results and Analysis

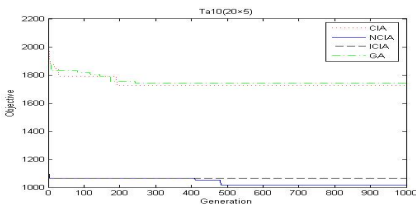
All the algorithms were tested over a range of different scales problems, and programs were coded and run on a PC with Intel Pentium CPU 2.66GHz and 1G of memory. The parameters of NCIA set as follows:  $num=2$ ,  $popsiz=200$ ,  $Pm=0.6$ ,  $Pv=0.8$ ,  $maxgeneration=1000$ ,  $To=500$ , and  $Pm=0.8$ . In the simulation, the four algorithms were performed on some instances from the benchmark problems, which are

GA, the general Co-evolutionary Immune Algorithm (CIA), the Improved Co-evolutionary Immune Algorithm (ICIA), and the New Co-evolutionary Immune Algorithm (NCIA) with global crossover operator. All results are summarized on Table I. It shows the name of the instance, the minimum (Min), average(Avg), maximum(Max) and variance deviation (Var) after 10 executions.

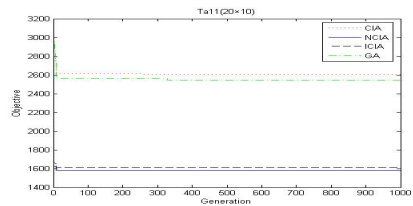
**Table 1.** Simulation results of benchmarks for algorithms

Instances	Algorithms	Objective	Var
		Min/Avg/Max	
Ta04(20x5)	GA	1890.0/1935.0/1978.0	91.20
	CIA	1930.0/1959.2/2000.0	81.56
	ICIA	1345.0/1345.4/1347.0	0.64
	NCIA	1345.0/1345.0/1345.0	0.00
Ta11(20x10)	GA	2484.0/2528.4/2548.0	530.64
	CIA	2508.0/2528.8/2567.0	434.96
	ICIA	1592.0/1607.0/1622.0	122.4
	NCIA	1579.0/1589.6/1595.0	42.24
Ta33(50x5)	GA	4531.0/4563.0/4587.0	482.0
	CIA	4520.0/4542.3/4569.0	266.4
	ICIA	2504.0/2523.4/2533.0	13.04
	NCIA	2499.0/2501.0/2504.0	6.00
Ta39(50x5)	GA	4307.0/4334.3/4405.0	541.43
	CIA	4361.0/4351.1/4417.0	344.44
	ICIA	2654.0/2684.4/2704.0	246.23
	NCIA	2654.0/2674.0/2684.0	89.73

In Table 1, the optimal value achieved by GA and CIA is worse than it of ICIA and NCIA for the same scale problems. Compared with the GA and CIA, the performance of ICIA and NCIA are improved greatly. According to the attribution of the global crossover operator, we believe in the superiority of the NCIA over the ICIA based on the effective protection of good genes by applying the global crossover operator. Moreover, the variance deviations of NCIA are general lower than those of other algorithms, which indicated that the NCIA has shown good stability and adaptability. Fig.4 ~ Fig.7 are the evolution curves of different scale problems. We can know that along with the iteration, the convergence rate of NCIA is faster than others, and the objective value is better than others.



**Fig. 4.** Evolution of Ta10 problem



**Fig. 5.** Evolution of Ta11 problem

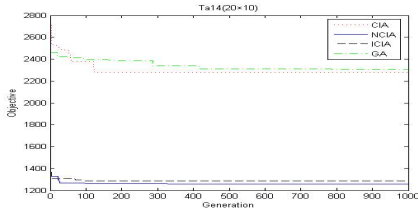


Fig. 6. Evolution of Ta14 problem

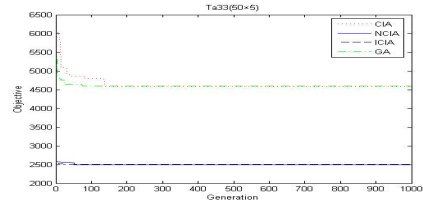


Fig. 7. Evolution of Ta33 problem

## 6 Conclusion

In this paper, we study the flow shop scheduling problems with zero wait. In order to search for optimal and near-optimal solutions, we proposed a new co-evolutionary immune algorithm, which can use the immune algorithm to expand search scope and cooperative evolutionary algorithm to improve search speed. In view of the poor search ability of the immune algorithm in the late evolution, we design a new population selection mechanism—"80/20" rule. In addition, a local search strategy is developed to improve the convergence. At the same time, a global crossover operator is presented to preserve excellent genes fragment and increase the diversity of population. Based on the analysis and comparison of results of benchmark problems, it is shown that the NCI A has better optimization performance beyond other three algorithms.

**Acknowledgments.** This work was supported by National Natural Science Foundation of China (Grant No. 61104178, 61174040).

## References

1. Sapkal, S.U., Laha, D.: A Heuristic for No-wait Flow Shop Scheduling. *International Journal of Advanced Manufacturing Technology*, 1–12 (2003)
2. Nagano, M.S., Da Silva, A.A., Lorena, L.A.N.: A New Evolutionary Clustering Search for a No-wait Flow Shop Problem with Set-up Times. *Engineering Applications of Artificial Intelligence* 25, 1114–1120 (2012)
3. Hsieh, Y.-C., Lee, Y.-C., You, P.-S., Chen, T.-C.: An Effective Immune Based Approach for the No-wait Flow Shop Scheduling Problems with Multiple Machines. *Advanced Materials Research* 97, 2432–2435 (2010)
4. Rabiee, M., Zandieh, M., Jafarian, A.: Scheduling of a No-wait Two-machine Flow Shop with Sequence-dependent Setup Times and Probable Rework Using Robust Meta-heuristics. *International Journal of Production Research* 50, 7428–7446 (2012)
5. Ruiz, R., Allahverdi, A.: New Heuristics for No-wait Flow Shops with a Linear Combination of Makespan and Maximum Lateness. *International Journal of Production Research* 47, 5717–5738 (2009)
6. Bertolissi, E.: Heuristic Algorithm for Scheduling in the No-wait Flow-shop. *Journal of Materials Processing Technology* 107, 459–665 (2000)



7. Ku, H.M., Karimi, I.: Completion Time Algorithms for Serial Multiproduct Batch Processes with Shared Storage. *Computers & Chemical Engineering* 14, 49–69 (1990)
8. Reddi, S.S., Ramamoorthy, C.V.: On the Flow-shop Sequencing Problem with No Wait in the Process. *Operational Research Quarterly* 23, 323–331 (1972)
9. Koch, R.: The 80/20 Principle. In: *Living the 80/20 Way*. Nicholas Brealey Publication, Melbourne (2004)
10. Zobolas, G.I., Tarantilis, C.D., Ioannou, G.: Minimizing Makespan in Permutation Flow Shop Scheduling Problems Using a Hybrid Metaheuristic Algorithm. *Computers & Operations Research* 36, 1249–1267 (2009)
11. Aldowaisan, T., Allahverdi, A.: A New Heuristic for M-machine No-wait Flow Shop to Minimize Total Completion Time. *Omega* 32, 345–352 (2004)

# Design and Implementation of Safety Verification for Civil Aviation Processing System Based on BPMN

Lisong Wang\* and Qian Zhang

Department of Computer Science and Technology,  
Nanjing University of Aeronautics and Astronautics  
Nanjing, Jiangsu, China  
wangls@nuaa.edu.com, zhangqiannow@gmail.com

**Abstract.** The civil aviation information system should be formally verified because it is a safety-critical system. Formal verification method is a vital approach to ensure the system safety. In this paper the civil aviation business system is modeled using business process modeling notation (BPMN) language, in order to satisfy the safety verification requirements of the civil business logic in China. Six safety properties are then defined by analyzing the built model, and a forward reasoning algorithm for safety verification is designed and implemented. Finally, an example of civil aviation ticketing system is demonstrated to verify the complete verification procedure and support the validity of the present method.

**Keywords:** Civil Aviation, Business Process, Safety Verification, BPMN.

## 1 Introduction

Several civil accidents have been reported in recent years, such as successful generation of massive exceptional free tickets, man-made cancellation of return flight, and the crippled civil cell system due to network failure [1]. Therefore, models and their consistencies should be verified for guaranteeing the safety and reliability. However, there have been so far no perfect technical solutions. Several researchers have been addressed on the issue regarding safety verification of the civil aviation information system in recent years. Formal verification methods were applied to Pi-calculus [2], visualization UML [3], Petri nets [4], fault tree analysis [5]-[6] and graph reduction technique [7]-[9]. Especially PVS [10] has been specifically presented aiming at civil aviation ticketing system. However, these methods do show some limitations. Most of them could not fully define safety properties and some unsafe factors were contained in their models. Russell [11] proposed the importance of considering the data model; however he did not show the concrete approach.

In this paper we analyze the civil aircraft business model and its interaction with the data model, and present a new verification method. BPMN is used to model the

---

\* Corresponding author.

process and define six safe properties. The task parallelism is discussed and a specialized technique for the verification is proposed to reduce the size of verification. A forward reasoning algorithm is deliberately designed to locate the unsafety with matrix operation. The verification process of a civil aviation ticketing system is outlined as an example to demonstrate the validity of the present algorithm.

## 2 Analysis in Civil Aviation

Business systems and processes should be formally modeled before analyzing the safety of the civil aviation business system. We use BPMN 2.0 [12] as the modeling language. The civil aviation ticketing system is here taken as an example to represent the model based on BPMN and the safety analysis method.

The system consists of one parent process shown in Fig.1 and two sub-processes (only one is shown in Fig.2 owing to space limitations). In Fig. 2 the parallel branches in the sub-processes are gathered in a XOR gateway generating lack of synchronization. The following nodes will be multiply activated. This constitutes a reason that leads to successful generation of massive exceptional free tickets.

The civil aviation ticketing system requires real-time data involving maintenance and safe access of database. The improper operation will severely influence the system safety. It is thus imperative to analyze the consistency between the process model and the data model. As seen in Fig. 1 the task “order revocation” relies on the data object “credits info”, but the data object “credits info” is created in the parallel branch. The task ‘order revocation’ may be abnormally executed because of the indeterminacy of the execution sequence of parallel tasks. In the parent process, the task “booking information recording” creates the data object “book info” with the dependent data object “fare info” created by the task “fare calculation” in the parallel branch. The data object “book info” may be abnormally created because of the indeterminacy of the execution sequence of the parallel task.

**Definition 1.** Define the set of tasks as  $T: T=\{t_1, t_2, \dots, t_n\}$  ( $n \geq 1$ ), is the set of the task nodes, where  $t_i$  is a task.

**Definition 2.** Define the set of data as  $D: D=\{d_1, d_2, \dots, d_m\}$ , ( $m \geq 0$ ), is the set of data object nodes, where  $d_i$  is a data object.

**Definition 3.** Define the set of all kinds of start nodes as  $ST: ST=\{start_1, start_2, \dots, start_p\}$ , ( $p \geq 1$ ), is the set of start nodes, where  $start_i$  is a start node.

**Definition 4.** Define the set of all kinds of end nodes as  $END: END=\{end_1, end_2, \dots, end_q\}$ , ( $q \geq 1$ ), is the set of end nodes, where  $end_i$  is a start node.

**Definition 5.** Define the set of gateway nodes as  $GW: GW=\{G_{andspl}, G_{orspl}, G_{xorspl}, G_{andjoin}, G_{orjoin}, G_{xorjoin}\}$ , ( $r \geq 0$ ), is the set of gateways nodes consisting of andsplit, orsplit, xorsplit, andjoin, orjoin and xorjoin.

**Definition 6.** Define the set of the relations of sequence as  $S_{sequence}: S_{sequence} \subseteq (T \cup G \cup ST \cup END) \times (T \cup G \cup ST \cup END)$ , is the sequence between the predecessor node and the successor node.

**Definition 7.** Define the set of the relations of creation as  $C_{create}: C_{create} \subseteq T \times D$ , which denotes that the task  $t_i$  creates the data object  $d_i$ .

**Definition 8.** Define the set of the relations of deletion as  $DL_{delete}: DL_{delete} \subseteq T \times D$ , which denotes that the task  $t_i$  deletes the data object  $d_i$ .

**Definition 9.** Define the set of the relations of precondition as  $P_{precondition}: P_{precondition} \subseteq T \times D$ , which denotes the task  $t_i$  relies on the data object  $d_i$  when executing.

**Definition 10.** Define the set of the relations of dependence as  $DP_{depend}: DP_{depend} \subseteq T \times D \times D$ , which denotes that the task  $t_i$  relies on the data object when creating  $d_i$ .

**Definition 11.** Define the business model with the data model as  $G = \langle V, E \rangle$ , and:

$$V = \{T, D, ST, END, GW\};$$

$$E = \{S_{sequence}, C_{create}, DL_{delete}, P_{precondition}, DP_{depend}\}.$$

We define predicates in Table 1 to better define safety:

**Table 1.** Predicates Definitions

Predicate	Explanation	Predicate	Explanation
IsST(x)	$x \in ST$	IsEND(x)	$x \in END$
IsD(x)	$x \in D$	IsAndSP(x)	$x \in G_{andsp}$
IsOrSP(x)	$x \in G_{orsp}$	IsXorSP(x)	$x \in G_{xorsp}$
IsAndJO(x)	$x \in G_{andjo}$	IsOrJO(x)	$x \in G_{orjo}$
IsXorJO(x)	$x \in G_{xorio}$	Sequence( $x_1, x_2$ )	$\langle x_1, x_2 \rangle \in S_{sequence}$
Create( $x_1, x_2$ )	$\langle x_1, x_2 \rangle \in C_{create}$	Depend( $x_1, x_2$ )	$\langle x_1, x_2 \rangle \in DP_{depend}$
Precondition( $x_1, x_2$ )	$\langle x_1, x_2 \rangle \in P_{precondition}$	GateW( $x_1, x_2$ )	$x_1$ and $x_2$ is a pair of split-join gateway

**Property 1.**  $\forall x(\neg IsST(x) \vee IsD(x)) \rightarrow \exists y(Sequence(y, x) \wedge \neg(IsEND(y) \vee IsD(y)))$ , which means each node in the case of neither a start task node nor a data object node has an input sequential edge.

**Property 2.**  $\forall x(\neg IsEND(x) \vee IsD(x)) \rightarrow \exists y(Sequence(x, y) \wedge \neg(IsST(y) \vee IsD(y)))$ , which means each node in the case of neither a end task node nor a data object node has an output sequential edge.

**Property 3.**  $\neg \exists x(IsAndSP(x) \wedge GateW(x, y)) \rightarrow \exists y(IsOrJO(y) \vee IsXorJO(y))$ , which means branches start from the gateway “andsplit” and end up with the gateway “and-join”.

**Property 4.**  $\neg \exists x((IsOrSP(x) \wedge GateW(x, y)) \rightarrow \exists y(IsANDJO(y)))$ , which means branches starting from the gateway “orsplit” do not end up with the gateway “andjoin”.

**Definition 12.** Data object state. When the task  $t_i$  ( $0 < i < n$ ) finishes, the state of the data object  $d_i$  is  $s_j$ , and  $s_j \in \{c, d, u\}$ .  $c$  stands for the created data,  $d$  denotes the deleted data and  $u$  represents the unknown state.

**Definition 13.** Predicates for data object state determination.  $S_{create}(j):s_j=c$ ;  $S_{delete}(j):s_j=d$ ;  $S_{unknown}(j):s_j=u$ .

**Property 5.**  $\forall x(IsT(x) \wedge Precondition(x,y) \rightarrow \exists y(IsD(y) \wedge S_{create}(y)))$ , which means if the task  $x$  relies on the data  $y$ ,  $y$  should have been created when executing the task  $x$ .

**Property 6.**  $\forall x(IsT(x) \wedge Create(x,y) \wedge Depend(y,z) \rightarrow \exists z(IsD(z) \wedge S_{create}(z)))$ , which means if the task  $x$  creates the data  $y$  and  $y$  relies on  $z$ ,  $z$  should have been created when executing the task  $x$ .

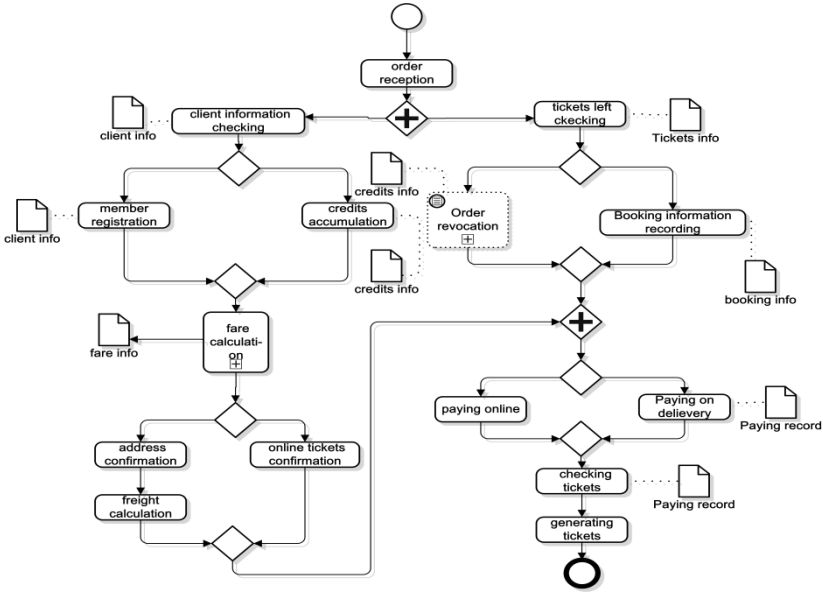


Fig. 1. Parent process model for civil aviation ticketing system

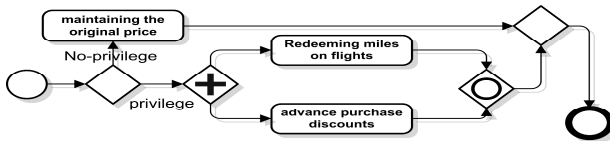


Fig. 2. Fare calculationsubprocess model for mivil aviation ticketing system

### 3 The Safety Verification Algorithm for Civil Aviation

#### 3.1 Calculated Corresponding Relationship among Gateways

The sequential flows split at the “split” gateway and join in at the “join” gateway. Therefore, the correct corresponding relationships among gateways should be guaranteed. The algorithm is shown as follows:

**Table 2.** Algorithm for Calculating the Corresponding Relationship among Gateways

<b>Algorithm 1</b>	
<b>Input :</b>	process model
<b>Output :</b>	Join gateway nodes corresponding to split gateway nodes
1.	traverse nodes in sequential flows in breadth-first order, 2. <b>for</b> each node $v_i$
3.	<b>if</b> $v_i \in (\text{andsplit} \cup \text{orsplit} \cup \text{xorsplit})$
4.	traverse along one sequential path from the $v_i$ back to the start and record the number of split gateways and join gateways, denoted as $a, b$ ;
5.	$c = a - b$ ;
6.	<b>while</b> $c \neq 0$
7.	traverse down the path from $v_i$
8.	<b>if</b> meeting a split gateway, then $c + +$ ;
9.	<b>if</b> meeting a join gateway, then $c - -$ ;
10.	<b>end while</b>

### 3.2 Parallel Analysis

The actual model of the aviation information system is complex. Therefore, verification procedure has to be simplified. The model is here parallel analyzed and simplified.

**Definition 14.** Set of parallel nodes  $PAR_i, \forall v_i \in (TUG), \forall v_j \in (TUG)$ , if  $v_i$  and  $v_j$  are neither in different branches which start from a xorsplit gateway nor in a sequential path from the start node to the end,  $v_j \in PAR_i$ .

**Theorem 1.** If  $v_j$  is the only direct predecessor for  $v_i, PAR_i = PAR_j$ .

Proof.  $v_j$  is the only direct predecessor for  $v_i$ . So each path passing  $v_i$  must pass  $v_j$ .

**Theorem 2.** If  $v_j \in (\text{andjoin} \cup \text{orjoin} \cup \text{xorjoin})$ ,  $PAR_i = \text{inter\_set}PAR_i$ , which the  $\text{inter\_set}PAR_i$  is the intersection of parallel nodes of all direct predecessors.

Proof. Each path passing the predecessors must pass  $v_i$ , then  $\text{inter\_set}PAR_i \subseteq PAR_i$  and  $PAR_i \subseteq \text{inter\_set}PAR_i$ .

So  $PAR_i = \text{inter\_set}PAR_i$

### 3.3 Matrix of Data Objects State

**Definition 15.**  $M$  is the matrix of data state, which consists of  $m$  rows and  $n$  columns.  $M_{[i][k]}$  is the  $j$ -th data object state when the  $k$ -th task completes. The state space of  $M_{[i][k]}$  is  $(c, d, u)$ .

**Theorem 3.**  $\forall v_i \in T$ , the set of predecessor nodes for  $v_i$  is  $PRE_i, M_{[i][k]} = c$ , iff  $Create(v_i, d_k) \vee (No\text{-}op(v_i) \wedge \forall v_j (v_j \in PRE_i \rightarrow S_{create}(j))) \wedge \neg \forall v_l (v_l \in PAR_i \rightarrow Delete(v_l, d_k))$ , where  $No\text{-}op$  denotes  $v_i$  has no operation on data  $d_k$ ;  $M_{[i][k]} = d$ , iff  $Delete(v_i, d_k) \vee (No\text{-}op(v_i) \wedge \forall v_j (v_j \in PRE_i \rightarrow S_{delete}(j))) \wedge \neg \forall v_l (v_l \in PAR_i \rightarrow Create(v_l, d_k))$ ;  $M_{[i][k]} = c$  under other conditions.

**Table 3.** Algorithm for Calculating the Set of Parallel Nodes

<b>Algorithm 2</b>
<b>Input</b> :process model
<b>Output</b> :set of parallel nodes
<ol style="list-style-type: none"> <li>1. traverse nodes in sequential flows in depth-first order</li> <li>2. <b>for</b> each node <math>v_i</math></li> <li>3.     <b>if</b> <math>v_j</math> is the only direct predecessor</li> <li>4.         <b>if</b> <math>v_j \notin (andsplit \cup orsplit \cup xorsplit)</math></li> <li>5.             <math>PAR_i = PAR_j</math>;</li> <li>6.         <b>else</b> <math>PAR_i</math> is the set of all nodes in sequential flows</li> <li>7.             traverse from <math>v_i</math> back to the start node;</li> <li>8.             delete the passing nodes in <math>PAR_i</math> ;</li> <li>9.             <b>if</b> <math>v_j \in xorjoin</math></li> <li>10.                 delete all nodes in the braches starting from the xorjoin;</li> <li>11.             traverse from <math>v_i</math> down to the end;</li> <li>12.             delete the passing nodes in <math>PAR_i</math> ;</li> <li>13.         <b>if</b> <math>v_j \in (andjoin \cup orjoin \cup xorjoin)</math></li> <li>14.             <math>PAR_i = inter\_setPAR_j</math>;</li> <li>15.         now <math>PAR_i</math> is the set of parallel nodes;</li> </ol>

**Proof.** If  $v_i$  creates or deletes  $d_k$ , the value of  $M_{[i][k]}$  can be deduced by definition 12. If  $v_i$  has no operation on data  $d_k$ , the value of  $M_{[i][k]}$  is effected by the tasks between  $v_i$  and  $PRE_i$ . The union set of these tasks is  $PAR_i$ . If  $\exists v_l \in PAR_i$  and  $v_l$  operates  $d_k$ , the state of  $M_{[i][k]}$  is unknown because the execution sequence between  $v_i$  and  $v_l$  cannot be determined.

### 3.4 Safety Verification

In this paper safety is verified based on the matrix of data object state.

**Table 4.** Algorithm for Safety Verification

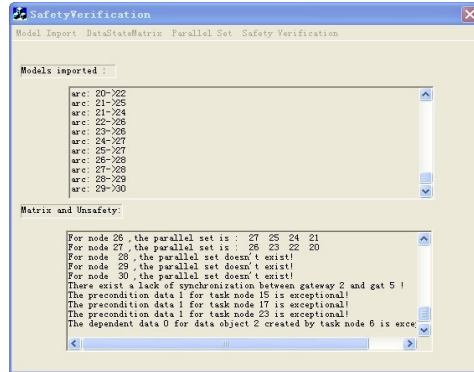
<b>Algorithm 3</b>
<b>Input</b> :process model, data model
<b>Output</b> :exceptions in models
<ol style="list-style-type: none"> <li>1. the models are transformed into a graph G;</li> <li>2. the corresponding relationship among join gateways are deduced by Algorithm 1;</li> <li>3. the set of parallel nodes are calculated by Algorithm 2;</li> <li>4. <b>for</b> each node <math>v_i</math>, nodes are traversed in sequential flows in depth-first order,</li> <li>5.     <b>if</b> (<math>v_i \notin ST</math> &amp;&amp; <math>v_i</math> has no predecessor)</li> <li>6.         output “<math>v_i</math> does not safety property 1 ”;</li> <li>7.     <b>if</b> (<math>v_i \notin END</math> &amp;&amp; <math>v_i</math> has no predecessor)</li> <li>8.         output “<math>v_i</math> does not safety property 2 ”;</li> </ol>

**Table 4.** (continued)

9.     **if** ( $v_i \in \text{andsplit}$  && the corresponding join gateway is not andjoin)
10.    output “ $v_i$  does not safety property 3 ”;
11.    **if** ( $v_i \in \text{orsplit}$  && the corresponding join gateway is andjoin)
12.    output “ $v_i$  does not safety property 4 ”;
13.    **if**  $v_i \in T$
14.    **if** (Precondition( $v_i, d_k$ ))
15.      **if**  $M_{[i][k]} \neq c$
16.        output “ $v_i$  does not safety property 5 ”;
16.    **if** (Create( $v_i, d_k$ ))
17.      **if** (Depend( $d_k, d_t$ ))
18.        **if**  $M_{[i][t]} \neq c$
19.        output “ $v_i$  does not safety property 6 ”;

### 3.5 Experiment and Analysis

The algorithm has been implemented in the environment of VC6.0. And the models are imported by XML files. The system interface is shown in Fig. 3. The experiment has been performed on a computer with an Inter Core I5 CPU clocked at 2.50 GHz.

**Fig. 3.** The interface for safety verification

Comparing to the algorithm in [7], which did not take the data flow into consideration, the algorithm in this paper can verify more properties. We have performed an experiment based on the civil aviation process system with 7 subprocesses, 231 tasks, 49 pairs of gateways and 58 data objects. The results are shown in Table 5.

Where  $tnum$  denotes the total number of exceptional nodes for each property,  $dnum$  denotes the number the number nodes being detected, and  $r$  denotes the detection rate.



**Table 5.** Results and Analysis

Properties	tnum	The graph reduction algorithm		The algorithm in the paper	
		<i>dnum</i>	<i>r</i>	<i>dum</i>	<i>r</i>
Property 1	2	0	0	2	100
Property 2	3	0	0	3	100
Property 3	15	14	93.3	14	93.3
Property 4	10	10	100	10	100
Property 5	29	0	0	27	93.1
Property 6	37	0	0	32	86.4

## 4 Conclusion

This paper provides systematic definitions of safety for civil aviation business process. We took both process models and data models into consideration when analyzing the six defined safe properties and designed the models based on BPMN. The algorithm was implemented by the verification tool written in C language and the nodes in models are imported by XML files. As being shown in the last section, the algorithm was demonstrated to be a more complete and reliable method of safety verification for the civil aviation processing system.

In the future we will try to analyze and define more safe properties and also attempt to build an integrated tool-chain from mapping the models based on BPMN automatically to verify the safeties.

**Acknowledgements.** This work was supported by Aeronautical Science Foundation of China granted no.20115552030.

## References

1. Wang, W.: The Importance of Information Technology in Civil Aviation. Guide to Business 14 (2011)
2. Liao, J., Tan, H., Liu, J.-D.: Describing and Verifying Web Service Using Pi-Calculus. Chinese Journal of Computers 28 (2005)
3. Fu, M., Wang, Y.: Research on Visual UML Model Verification Environment. Journal of University of Electronic Science and Technology of China 39, 289–293 (2010)
4. van der Aalst, W.M.P.: Workflow verification: Finding control-flow errors using petri-net-based techniques. In: van der Aalst, W.M.P., Desel, J., Oberweis, A. (eds.) Business Process Management. LNCS, vol. 1806, pp. 161–183. Springer, Heidelberg (2000)
5. Yang, Y., Keller, P., Livnat, Y.: Improving Safety-Critical Systems by Visual Analysis. In: Garth, C., Middel, A., Hagen, H. (eds.) VLUDS, pp. 43–58 (2011)
6. Conforti, R., Fortino, G., La Rosa, M., ter Hofstede, A.H.M.: History-aware, real-time risk detection in business processes. In: Meersman, R., et al. (eds.) OTM 2011, Part I. LNCS, vol. 7044, pp. 100–118. Springer, Heidelberg (2011)

7. Song, B., Wang, J.-Y., Yu, G.: Verification Method for Process Model Based on Graph Spreading and Graph Reduction. *Mini Micro* 26, 1073–1078 (2005)
8. Sadiq, W., Orłowska, M.E.: Analyzing process models using graph reduction techniques. *Information Systems* 25, 117–134 (2000)
9. Sadiq, W., Orłowska, M.E.: Applying graph reduction techniques for identifying structural conflicts in process models. In: Jarke, M., Oberweis, A. (eds.) *CAiSE 1999*. LNCS, vol. 1626, p. 195. Springer, Heidelberg (1999)
10. Yang, H., Liu, J., Han, J.-G.: Formal Specification and Verification of Air Ticket Reservation System Using PVS. *Journal of Xi An Institute of Postsand Telecommunications* 6 (2001)
11. Russell, N., ter Hofstede, A.H.M., Edmond, D., van der Aalst, W.M.P.: Workflow data patterns: Identification, representation and tool support. In: Delcambre, L.M.L., Kop, C., Mayr, H.C., Mylopoulos, J., Pastor, Ó. (eds.) *ER 2005*. LNCS, vol. 3716, pp. 353–368. Springer, Heidelberg (2005)
12. *OMG.BPMN2.0 (Standards style)*, BPMN Standard (2011)

# A Survey on Ontology Mapping Techniques

Yew Kwang Hooi<sup>1,\*</sup>, M. Fadzil Hassan<sup>1</sup>, and Azmi M. Shariff<sup>2</sup>

<sup>1</sup> Computer and Information Sciences Department  
Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Perak, Malaysia

<sup>2</sup> Chemical Engineering Department  
Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Perak, Malaysia  
yewkwanghooi@gmail.com,  
{m\_fadzil, azmish}@petronas.com.my

**Abstract.** This paper surveys existing ontology mapping techniques towards data interoperability. Existing matcher algorithms and strategies are discussed generally and the research gaps are highlighted. The study concludes that semantic mapping has the biggest share of unresolved problems. Bridging the gap in semantic mapping may help improving mapping of ostensibly different domain knowledge and disparate data sources.

**Keywords:** Enterprise interoperability, ontology mapping, basic matcher, semantics.

## 1 Introduction

Software systems can achieve more sophisticated abilities by accessing ontologies that provide computer-processable representations of a knowledge domain. The use of Knowledge Work Systems (KWS) to convert facts and knowledge into ontologies is already a growing trend in enterprises. Ontologies provide a layer of abstraction on top data, taxonomy or database schema [1] and offer richer semantics through concept mapping [2]. Various literature [2-10] indicate the potential of ontology mapping for data interoperability between heterogeneous sources.

## 2 Mapping Ontologies in Enterprises

Mapping multiple ontologies is envisaged a future necessity to support complex analysis, decision making and collaborative information systems to achieve a common business objective. Concepts [11] from separate database schemas [5], XML, and other data sources can be represented by ontologies. Future systems using ontologies may share and merge ontologies. However, this interoperability is not easy due to different vocabularies and granularities of ontologies [5, 12]. Aligning ontologies is the focus of a variety of works originating from diverse communities over the years. [2, 13].

---

\* Corresponding author.

Ontology mapping can resolve complex information exchange through a consensual understanding between concepts [14, 15]. This is achieved by identifying the exchange points between different representations [15]. The exchange points may share a common layer [4], a mediated schema that facilitates sharing of heterogeneous sources [5, 12] and tackles incomplete data issue in Knowledge Systems [16]. This leads to integration of evolving context, information or data [6] for inferencing new knowledge and accurate search [1, 17].

### 3 Literature Review

#### 3.1 Definitions

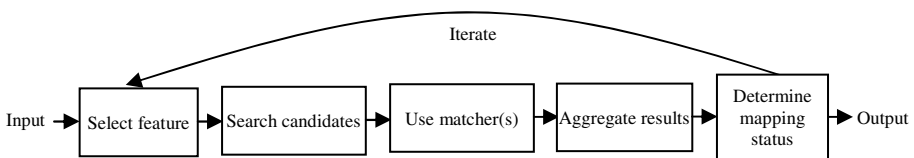
Ontology can be described as a pair of Signature and Axiom sets  $O = (S, A)$  where  $O$  is the ontology,  $S$  is ontological signature, and  $A$  is a set of ontological axioms, for restricting the meaning of the terms in the signature [4, 8].

Ontology mapping finds correspondences between entities of multiple ontologies [15]. Given ontologies,  $O_1 = (S_1, A_1)$  and  $O_2 = (S_2, A_2)$ , ontological mapping is a morphism  $f: S_1 \rightarrow S_2$ , such that  $A_2 \models f(A_1)$  where all correspondences that satisfy  $O_2$ 's axioms also satisfy  $O_1$ 's "translated axiom" [4]. A correspondence is a function that assigns symbol of one vocabulary to the symbol of another vocabulary [2] [1]. A set of such correspondences between a pair of ontologies is an alignment. Mapping is a directed alignment [3].

Partial ontology mapping, is defined as having a sub-ontology  $O'_1 = (S'_1, A'_1)$  where  $S'_1 \subseteq S_1, A'_1 \subseteq A_1$  such that there is a total mapping from  $O'_1$  to  $O_2$  [18]

#### 3.2 Mapping Framework

A high level view of mapping process can be simplistically depicted in Figure 1 [3, 15]:



**Fig. 1.** A simple high level view of a mapping process

The core component of mapping is the matcher. Matcher builds correspondences between ontologies by first selecting a suitable attribute or feature (entity label, structural description of concepts, range for relations, instantiated attributes or extensional descriptions) from the ontologies [3]. The feature selection transform the resource into a light-weight ontology.

Common mapping approaches [4] [3] link candidate ontologies to a common ontology using anchors. Anchors are entities which are declared to be equivalent (based on identity and user input).

Finally, matchers evaluate the similarity criteria of both ontologies. Often, functions based on heuristic similarity instead of exact logical similarity are used to avoid a costly exhaustive search. Multiple results from multiple matchers for the same entity pair can be aggregated. Recent advances introduce autonomous combination of multiple matchers. The tie-breaker determining the mapping of each pair may use threshold method, relaxation labeling or combining structural and similarity criteria using learning algorithms or/and user input. Iteration stops at predefined loops or when there is no more new mapping proposal [3, 8].

### 3.3 Categories of Matchers

Basic matcher is a similarity function of a pair of entities,  $\sigma : O \times O \rightarrow R$  where  $R$  is  $[0, 1]$ . In point-to-point approach, matching uses lexical or structural similarity of labels or instances. Various techniques have been developed and can generally be categorized as follows [4]:-

**Terminological Mapping.** Mapping uses token analysis to reduce a word to a common format and establish its importance through weighting of relations and comparison of paths. Various techniques specializing on features of concepts analyzed are:

- *String-based.* The analysis quantifies edit distance by counting and normalizing the required editing operations (insertion, deletion or substitution of affixes or substrings between two words) to transform the first word to the second.
- *Language-based.* The analysis tokenizes string using punctuations and cases; then uses lemmatization to find the possible basic forms of the base word.
- *Linguistic resource.* The analysis refers to an extrinsic source such as WordNet [3] for linguistic knowledge to interpret strings. Sense-based approach determines relationships of a word as hyponym, hypernym, synonym or antonym; whereas gloss-based approach counts the same words in a pair of phrases or sentences.

Terminological mapping face difficulty in processing word variations in the same ontology or across ontologies [3].

**Structural Mapping [3].** Structural mapping looks at the relationship (adjacency and path sharing) between concepts within the ontology structure. Two approaches are:

- *Taxonomy mapping:* The mapping uses super-concept rule and bounded-paths. Super-concept matching assumes similarity of actual concepts if both share the same parent concept. Bounded-paths compares two paths to identify similar concepts along the paths.
- *Tree-based mapping:* Similarity is based on the analysis of the positions within the graphs. Neighbour nodes are assumed to be somehow similar if two nodes from two ontologies are similar.

Similar to terminological, structural mapping faces difficulty in processing many kinds of variations that occur in ontologies [3].

**Semantic Technique.** Semantic mapping is the most challenging area and a key research area [8, 9]. The key feature of semantic mapping is the use of model theoretic semantics to define well-formed-formula (wff) to express the meaning of anything without ambiguity. Its advantage is the deductive methods for amplifying or cropping the mappings in anchored ontologies to ensure mapping completeness and to eliminate bad correspondences [3]. Semantic technique is dependent on anchored ontologies which contains mapping candidates serving as a common ground for comparison [3, 19]. Anchoring works by matching ontologies to the background ontology to extract meanings for concepts using a domain knowledge [3]. The two approaches are using external ontologies and deductive techniques.

*External Ontologies.* Semantic technique uses a mediated approach. An intermediate reference ontology can provide general concepts and axioms for clarifying the meaning of domain concepts and the relations. The intermediate reference ontology can be an external ontology or using using a hidden intermediate reference ontology that is built on the fly using lexicons, as proposed by Kotis [1]. The user of external ontology is more common. An external ontology is a general, top-level and formal ontology for conceptual modelling. Examples are General Formal Ontology (GFO), WordNet [20], Cyc, Suggested Upper Merged Ontology (SUMO) [21] and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [22]. It has a higher probability of finding a result by exploiting existing mappings but possibly at a lower accuracy [3]. A formal ontology provides reasoning and deduction methods.

*Deductive Techniques.* These techniques merge two ontologies and search correspondences through subsumption relation. Subsumption tests mappings and discard mappings that fares poorly in satisfiability test. Three types of satisfiability techniques are Propositional Satisfiability (SAT), modal satisfiability and Description Logic (DL). [3].

SAT technique builds a theory (the domain knowledge or a group of axioms using matchers referring to external sources) as a premise to establish a relationship between concepts, such as described by the following:

Axioms  $\rightarrow r(c, c')$  where  $c$  and  $c'$  are a pair of concepts and  $r$  is a relation {equivalent, subsumption, subclass, not equal}

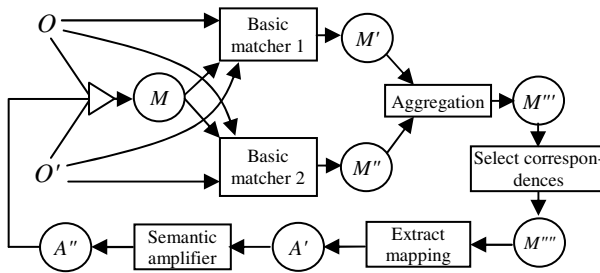
Validation is done by an exhaustive check to ensure that there is no possible negation of the formula using instances.

DL technique is a pure terminological formal knowledge representation technique. It uses subsumption reasoning to establish relations between different ontologies in a pure semantic manner [3, 5]. Two ontologies are merged and the pair of concepts and roles are tested for subsumption. The relationships are expressed in minimal description logics syntax. Relation inference is conducted on the description logics to see if the subsumption rules of the components are consistent.

### 3.4 Aggregating Matchers

As has been mentioned earlier, ontology mapping often compounds multiple basic matchers. The two generic compositions are [1]:-

- **Sequential Composition** - Matchers are arranged sequentially. The initial matchers in the order focus on smaller granularity such as linguistic matching. The subsequent matchers use higher granularity matching, such as structure matching.
- **Parallel Composition** - Both matchers evaluate similar set of ontologies simultaneously and generate a result respectively, which are compounded into one.
- **Mixed Composition** - The composition of matchers can be a mixture of sequential and parallel compositions, see Figure 2. AUTOMS for example, use sequential arrangement for multiple passes of mapping and parallel composition to aggregate results of multiple methods. The multiple passes include simple and iterative structural, instance-based or property-based matching; lexical matching; and semantic matching. Aggregation includes result matrices of concepts, properties and relations [23]. [3]



Note:  $M$  is a matrix of concepts, properties and/or relations

Fig. 2. A global matching system (adapted from [3])

## 4 Research Gap

According to Euzenat, semantics "provides the rules for interpreting the syntax which do not provide the meaning directly but constrains the possible interpretations of what is declared." [24]. Very few semantic techniques have been developed for ontology mapping despite the potentials of semantic techniques. A tall barrier is the difficulty of combining semantics' deductive technique with ontology's inductive structure.

Most matcher designs focus on a specific application domain or ontology type (DTD, relation schemas, OWL), hence reducing its reusability potential. Reasons are that ontology itself is designed for a specific application [3] due to intricacy of data in niche domains. Because rewriting a generic matcher for a new domain is inconvenient, it is therefore desirable to have modularized and general matcher designs. Few mapping techniques designed to handle general ontologies are S-Match, Similarity Flooding, COMA++ and Cupid.

**Table 1.** Summary of Ontology Matching Techniques

Techniques	Objective	Limitations and gaps	Ref.
<b>String-based</b> <i>Synonyms, Homonyms, Normalization, String equality, Substring test, Edit distance, Token-based distance, Path comparison.</i>	Simple matching effective when very similar strings are used on the schema name to denote the same concept.	Unable to distinct synonyms or homonyms effectively.	P.84 [3] [23]
<b>Language-based</b> 1. <i>Intrinsic methods / Linguistic normalization</i> Reduce words to standardized form using tokenization, lemmatization, term extraction or stop-word elimination  2. <i>Extrinsic methods</i> Uses external resources.	To improve interpretation and apprehension of terms used using natural language processing (NLP).	Very dependent on linguistic resources such as Stemmers, Part-of-speech taggers, Lexicons, and Thesauri Effectiveness hindered by presence of a foreign languages and syntactic variations of the same word (spellings, abbreviations, prefixes, suffixes). It does not take into account the structure of ontology entities to find the most coherent match.	P.92 [3] [20]
<b>Structure-based/ Constraint-based (Internal structure)</b> Keys are the most useful identifier. Works by comparing structure and the properties of entities.	To match schema to determine if the classes are equivalent. It is often used to quickly find possible matches with shallow accuracy.	Lacking accuracy. Ineffective when two equivalent entities has different data types for its properties. It may be possible that different entities have similar properties.	p.92 [3]
<b>Structure-based (Relational structure)</b> <i>Wu-Palmer similarity</i> <i>Upward cotopic similarity</i> Compare structure of entities using relations. Similarity is based on similar counting of edges in the graphs.	To match concepts in taxonomy, formal ontologies and semantic networks.	Difficulty in detecting (using iterative algorithm) and inability to handle mutual influence between related parts. Using edge count is inconclusive semantically as there is possibility that the same class hierarchy can be summarized as a short alternative. More work has been done on taxonomical form of relation but less on mereology (part-of) relation.	p.98 [3]
<b>Extensional</b> <i>Hamming distance</i> <i>Jaccard similarity</i> <i>Formal Concept Analysis (FCA)</i> <i>Concept Lattice</i> <i>Instance identification</i> <i>Disjoint extension</i> <i>Statistical-approach</i> <i>Similarity-based</i> <i>Matching-based</i>	Relatively accurate matching of classes when both ontologies are using a set of common individuals/instances; or tangible and non-changing indices.	When instance information is not available.	p.110 [3]
<b>Semantics</b> <i>External ontology</i> <i>Deductive</i> <i>Propositional</i> <i>Description logic</i>	Used to find all inconsistent correspondences to complete mappings and generating justification for mapping result.	Ontology requires inductive inputs but semantics technique is deductive in nature. Currently, there is a lack of interoperability between inductive technique and deductive semantic techniques.	p.110 [3]



Other areas noted for more research are:

- Relationship between entities are mostly quantitative, expressed in confidence range [0 1]. Matcher that determines qualitative relationship between entities using logical relations (e.g. equivalence or subsumption) may be more expressive.
- Most matcher analyzes schema despite the abundance of data repository and instances. The very few instance-based solutions use Naive Bayesian classifier and common value patterns.
- Only a few matchers (DCM, Wise-Integrator) handle more than one pair of ontologies. The results are usually one-to-one mapping although it may be possible to find one-to-many or many-to-many. Most matchers process ontologies with tree structure. Few matchers (COMA++, Cupid and OLA) process graphs.

Table 1 is a non-exhaustive list of existing basic mapping techniques.

## 5 Conclusion

This paper investigates the advancements made in technologies and theoretical foundations of ontology mapping. Matchers find correspondences between ontology entities from narrow perspectives. Therefore, results of multiple matchers are combined for better accuracy and mapping probability. Semantic mapping extends existing matching techniques through enrichment and formalism of anchored ontologies for higher accuracy. Future challenges are to enhance deductive-inductive operability, generalization and modularity of semantic mapping methods.

## References

1. Kotis, K.: *Ontology Matching*, Ai-Lab, ICS Eng. University of the Aegean (2007)
2. Kalfoglou, Y.A.M.S.: *Ontology mapping: the state of the art*. *The Knowledge Engineering Review* 18(1), 1–31 (2008)
3. Euzenat, J., Shvaiko, P.: *Ontology Matching*. Springer, Heidelberg (2007)
4. Kalfoglou, Y., Schorlemmer, M.: *Ontology mapping: the state of the art*. *The Knowledge Engineering Review* 18(1), 1–31 (2003)
5. Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: *Ontology-Based Integration of Information - A Survey of Existing Approaches*, pp. 108–117 (2001)
6. Kasabov, N.: *ECOS: The Knowledge Engineering Approach*. In: ICANN, Porto
7. Cui, L., Zhao, J., Zhang, R.: *The Integration of HAZOP Expert System and Piping and Instrumentation Diagrams*. *Process Safety and Environmental Protection* 88, 327–334 (2010)
8. Kotis, K., Vouros, G.A., Stergiou, K.: *Towards automatic merging of domain ontologies: The HCONE-merge approach*. *Web Semant.* 4(1), 60–79 (2006)
9. Uschold, M.: *Creating, Integrating and Maintaining Local and Global Ontologies*. In: ECAI 2000, Berlin, Germany (2000)
10. Beneventano, D., et al.: *Ontology-driven Semantic Mapping*. In: Mertins, K., et al. (eds.) *Enterprise Interoperability III*, pp. 329–341. Springer, London (2008)

11. Gottgroy, P.: Uilding Evolving Ontology Maps for Data Mining and Knowledge Discovery in Biomedical Informatics. In: BIOMAT - Brazilian Symposium of Mathematical and Computational Biology, Rio de Janeiro (2003)
12. Goh, C.H.: Representing and reasoning about semantic conflicts in heterogeneous information systems (1997)
13. Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: the state of the art. *The Knowledge Engineering Review* 18(1), 1–31 (2003)
14. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: *Ontological Engineering*. In: Wu, X., Jain, L. (eds.). Springer-Verlag London Limited, London (2004)
15. Davies, J., Studer, R., Warren, P.: *Semantic Web Technologies - Trends and Research in Ontology-based Systems*. The Atrium, Sothern Gate, Chichester, West Sussex. John Wiley and Sons Ltd. (2006)
16. Smith, R.G.: *Knowledge-Based Systems - Concepts, Techniques, Examples*. Canadian High Technology Show, Lansdowne Park, Ottawa (1985)
17. Kuo, Y.-T., et al.: Domain ontology driven data mining: a medical case study. In: *Proceedings of the 2007 International Workshop on Domain Driven Data Mining*. ACM, San Jose (2007)
18. Meseguer, J.: General logics. *Logic Colloquium* 87, 275–329 (1989)
19. Giunchiglia, F., Shvaiko, P., Yatskevich, M.: Discovering missing background knowledge in ontology matching. In: *Proceedings of ECAI* (2006)
20. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of semantic distance. *Computational Linguistics* 32(1), 13–47 (2006)
21. Niles, I., Pease, A.: Towards a standard upper ontology. In: *Proceedings of the International Conference on Formal Ontology in Information Systems*, vol. 2001, pp. 2–9. ACM, Ogunquit (2001)
22. Gangemi, A., et al.: Sweetening WORDNET with DOLCE. *AI Mag.* 24(3), 13–24 (2003)
23. Lasek, I.: *Ontology Matching Techniques* (2011)
24. *The Knowledge-based Economy*. General Distribution. Organization for Economic Cooperation and Development, Paris (1996)

# The Parametric Design and Automatic Assembly of Hydrostatic Rotary Table Based on Pro/Engineer

Qiang Cheng<sup>1</sup>, Can Wu<sup>1</sup>, Zhifeng Liu<sup>1</sup>, Jianhua Wang<sup>1</sup>, and Peihua Gu<sup>2</sup>

<sup>1</sup>College of Mechanical Engineering and Applied Electronics Technology,  
Beijing University of Technology, Beijing, China, 100124

<sup>2</sup>Department of Mechatronics Engineering, Shantou University, Shantou,  
Guangdong, China 515063

{chengqiang, lzf, wjh}@bjut.edu.cn,  
wucan@emails.bjut.edu.cn, cqaaok@163.com

**Abstract.** In order to achieve a reasonable and rapid design of hydrostatic rotary table which is used in heavy machine tools, the design flow of hydrostatic rotary table was proposed, and the bottom-up modeling method and database system were used to ensure the consistency between the design parameters. The rapid and parametric design and automatic assembly of hydrostatic rotary table based on Pro/E was implemented by taking Visual C++ 2008 as a development platform and Pro/Toolkit as a secondary development tool and adopting object-oriented programming techniques and visualization technique.

**Keywords:** Automatic assembly, Hydrostatic rotary table, Parametric Design, Pro/Toolkit.

## 1 Introduction

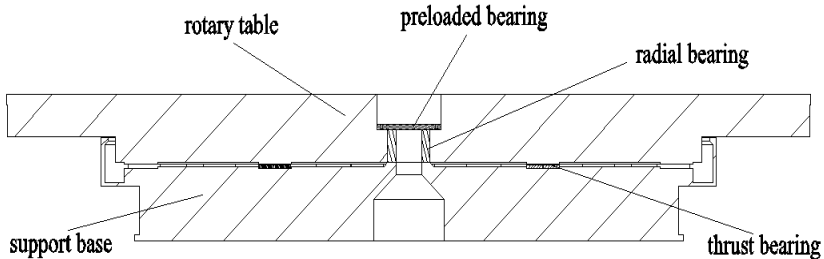
With the rapid development of computer technology and aided design, virtual manufacturing, digital prototyping, and other digital engineering techniques were widely used, the technical performance of enterprise product was greatly improved, at the same time, the product replacement cycle was shortened [1]. However, when the product designers complete a product design task with a common commercial three-dimensional computer aid design software, additional program by the product designers themselves is needed to complete the related computing tasks, and geometry dysfunctional relationship problems may caused during the process of modify the geometry dimensions. The cumbersome operating environment greatly restricted the use of the common commercial three-dimensional computer aid design software [2~4]. Meanwhile, Hydrostatic rotary table as a key component of heavy machine tools, is directly related to the machining ability and level of accuracy [5, 6], So a reasonable and rapid design tool will be helpful to designers when they asked to design a hydrostatic rotary table. For the above, this paper take hydrostatic rotary table as an example for parametric design, and using the bottom-up modeling method combined with database system to ensure the consistency between the design parameters.

## 2 Secondary Development Technology Based on Pro/Engineer

Pro/Engineer as a mature commercial three-dimensional computer aid design software is the first one proposed and supported for parametric design. At the same time, it provides Pro/Toolkit, a secondary development tool for Pro/Engineer, to third-party developers to achieve seamless connection of third party applications with Pro/Engineer. Pro/Toolkit provides a method of programming for each feature, but it requires various characteristic attributes were set in the element tree, and this is often a very complicated and tedious task. When it comes to adaptive design and variant design, simply changing some dimensions or parameters in the part can be achieved. The method of reading previously established parametric models and modifying the parameter values by Pro/Toolkit function has more advantages than the method of setting a variety of characteristic attributes in the element tree. Actually, approximately 70% of the design work is adaptive design and variant design, this paper adopted the method of reading previously established parametric models and modifying the parameter values by Pro/Toolkit function.

## 3 Proposed Design Flow of Hydrostatic Rotary Table

Hydrostatic rotary table can be divided into five parts: rotary table, support system, fuel system, drive system and support base. The infrastructures of hydrostatic rotary table which were mentioned in this article only include rotary table, support system and support base, wherein the support system comprises thrust bearing, preload bearing and radial bearing. The assembly as shown in the figure below:



**Fig. 1.** The assembly of infrastructures of hydrostatic rotary table

When starting a hydrostatic rotary table design, the weight and load carrying capacity of the rotary table directly affect other parts, particularly the thrust bearing. The weight and load carrying capacity of the rotary table is needed to calculate and check the pressure of the oil cavity of the thrust bearing, which will appear repeatedly modify the parameters between the rotary table and the thrust bearing. The design flow of the rotary table and thrust bearing was proposed as follows, to ensure the design of the hydrostatic rotary table properly.

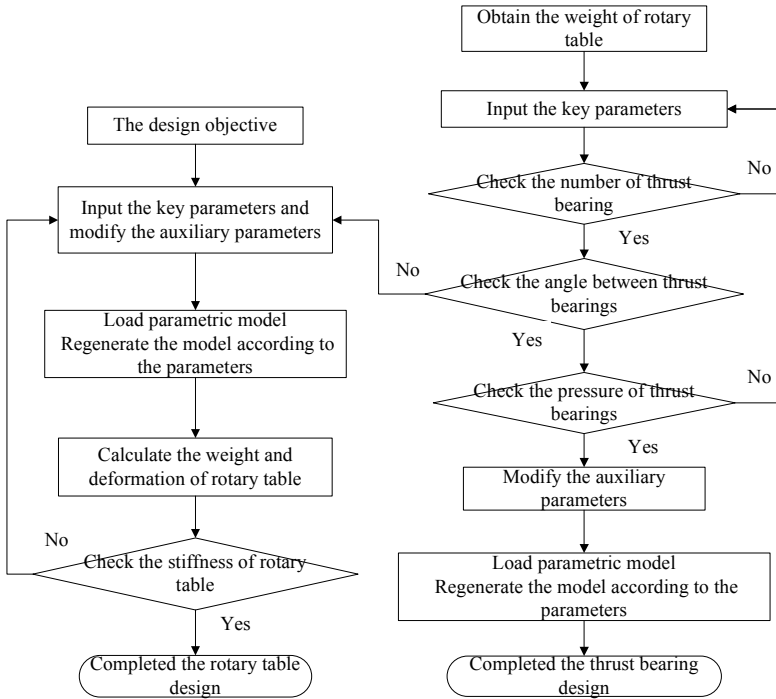


Fig. 2. The design flow of rotary table and thrust bearing

## 4 Design Example

### 4.1 Establish Parametric Models Library

The purpose of parametric design is obtain a new model by modify some dimensions or structure parameters defined of the model, and this are supported by the current commercial three-dimensional computer aid design software.

Rotary table is placement platform for workpiece, and drive them when it works. The sufficient stiffness of rotary table and rails is needed to reduce the structural deformation caused by the force form the oil chamber of thrust bearing. The main design parameters affecting the stiffness of rotary table are the diameter of rotary table, the height of rotary table, the laps of rails, the radius of rail and the width of rails. Thrust bearing is fixed on support base by bolts, and bear the axial load of rotary table. The main design parameters impact the pressure of oil chamber and the stiffness of thrust bearing are the diameter of thrust bearing, the width of rim and the number of thrust bearing.

Preload bearing is fixed on the spindle by bolts, with the thrust bearing to form a closed support structure to improve the axial stiffness of the rotary table. The main design parameters are the diameter of preload bearing and the height of preload

bearing. Radial bearing bear the centrifugal force when hydrostatic rotary table was working. The main design parameters are the diameter of radial bearing, the angle of oil chamber of preload bearing and the depth of oil chamber. Support base is mainly to provide support for other parts of hydrostatic rotary table and the oil pipelines and drive motor mounted on it. To make sure the support base has sufficient stiffness, the height of support base worked as a main design parameter. One parametric model can only describe one kind of structure, and make sure all kinds of structure have a parametric model when creating the parametric models library. When you create a parametric model, you should pay enough attention to the following:

- a) Using English characters instead of Chinese characters named with a parameter to avoid Pro/Toolkit does not support.
- b) Using geometric constraints as much as possible in the sketch to reduce the number of dimensions constraints.
- c) Adding relations between dimensions in the sketch-level rather than in the part-level to facilitate maintenance of the parametric models later. Good interface design can not only help designers understand the design parameters, but also to help users make better use of past design data.

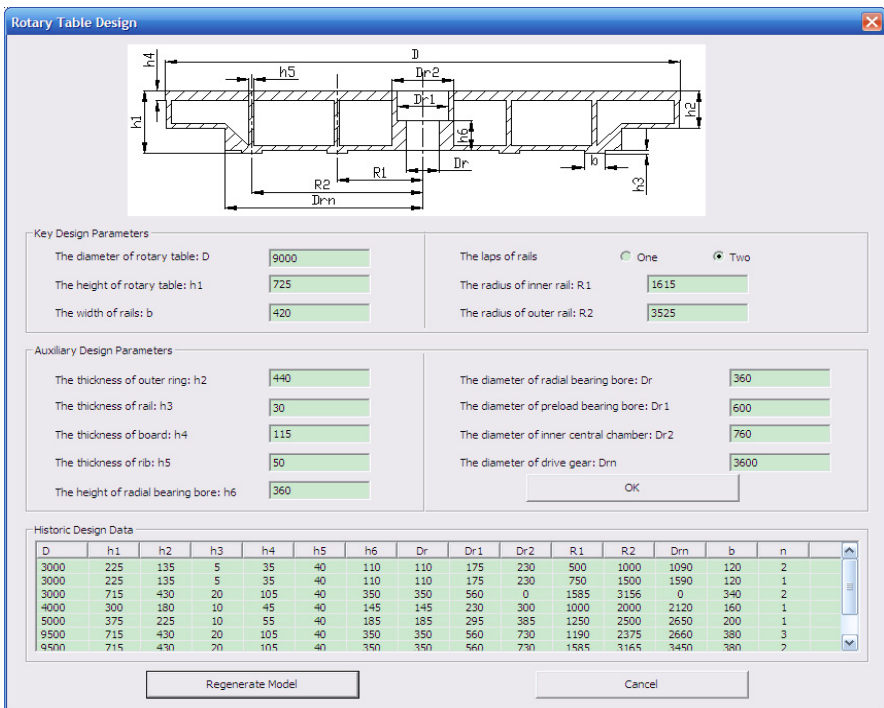


Fig. 3. The user interface of rotary table

## 4.2 Parts Design User Interface

After created parametric models library, user interface is needed to help designers understand the various parameters of the model and to facilitate the users to use the application. Below is the rotary table designed interface, the user can see the meaning of the various parameters visually and can read the historic design data. The user interfaces of other parts of hydrostatic rotary table have the same style.

## 4.3 Automatic Assembly

Mechanical assembly relations between parts can be divided into three categories: completely constraint, hole constraint and axis constraint plus plane alignment constraints [7].

In this paper, the research objects are rotary parts, and an axis-aligned constraint and a face-aligned constraint can constraint the position of part in assembly. When creating a parametric models library, the reference plane and the reference axis should created with a specified name during the modeling which can easily be find by Pro/Toolkit functions. The user only should point out the location of the parts respectively to finish the assembly of hydrostatic rotary table. The user interface of automatic assembly is showed below.

## 5 Conclusions

Pro/Toolkit and VC++2008 were used to establish the interactive parametric design system for hydrostatic rotary table, and the database were used to storage design data to ensure synchronization between the associated modified dimensions and further realization of the hydrostatic rotary table automatic Assembly. The hydrostatic rotary table can be easily finished by inputs design parameters on the user interface. The proposed system provides the function of historical records preservation and query which provides reference design data and automatic assembly function which avoid repeating simple work.

The proposed parametric design system can significantly shorten the design cycle of hydrostatic rotary table and can ensure design accuracy and significantly reduces design. It can greatly increasing design efficiency and reduce the costs.

**Acknowledgement.** The authors are most grateful to the National Natural Science Foundation of China (No.51005003), National Science and Technology Great Special Program (No. 2012ZX04010-011), the Leading Talent Project of Guangdong Province, Rixin Talent Project of Beijing University of Technology for supporting the research presented in this paper.

## References

- [1] Wei, L., Lu, Y.: The digital and virtual manufacturing of forklift truck mast based on Pro/E secondary development. *Hoisting and Conveying Machinery* (3), 30–33 (2010)
- [2] Liu, Q., Xi, J.: Case-based parametric design system for test turntable. *Expert Systems with Applications* (38), 6508–6516 (2011)
- [3] Zhou, Y., Huang, Z.: Rapid design system for down hole tools based on Pro/Engineer. *Guangxi Journal of Light Industry* (7), 69–70 (2011)
- [4] Wu, L., Chen, B.: The foundation of Pro/ENGINEER secondary development technology. Electronic Industry Press (2006)
- [5] Zhou, K., Xiong, W., Lv, L.: The overview of hydrostatic rotary table technology. *Manufacturing Technology & Machine Tool* 36(4), 29–32 (2011)
- [6] Chen, Y.: Theory and design of hydrostatic bearing. National Defence Industry Press, Beijing (1980) (in Chinese)
- [7] Xie, M., He, L., Xu, J.: A automated assembly method based on Pro/E secondary development. *Modern Machinery* (1), 46–48 (2006)



# Modeling and Simulation of China's Competition Strategy for Highly Educated Talents Based on System Dynamic

Wen Guo<sup>1,\*</sup> and Feng Dai<sup>2</sup>

Information Engineering University  
Zhengzhou, China

152183325@163.com, 152183325@qq.com

**Abstract.** As China is facing more and more fierce competitions for international highly educated talents resources. Highly educated talents competition system model is established based on the system dynamics and the trend is simulated and analyzed in this paper. By analyzing the competition system, elements in the system are divided into three categories. Vensim showed different trend while some exogenous variables in the model were adjusted. The simulation results show that, the density of highly educated talents in China has an close relationship with human development index, governmental efficiency index, the enterprise registration cost, etc. Chinese competitiveness in the international market of highly educated talents will be enhanced in the next 10 years with the development of highly educated talent density, and things will get better by increasing education investment, improving efficiency of government and perfecting social security system.

**Keywords:** Highly Educated Talents, Competition strategy, System Dynamics, Modeling and Simulation.

## 1 Introduction

With the advent of the era of knowledge economy, highly educated talents who can promote social progress with professional skills and creative labor have become a strategic resource. Little country can get a sustainable development of high-speed without them. In addition, the convenience of the flowing among different countries virtually increased the competition for highly educated talent resources in a globalized world. For variety of reasons, many of highly educated talents could not render a service for national construction. In recent years, with the improving of economy, the Chinese government has put forward "the national medium - and long-term talent development plan outline (2010-2020)" and Chinese position in the international talent competition is significantly improved.

---

\* Corresponding author.

Scholars have made a lot of research on the method and management of talent. Alvaro Aguiar (2009) [1] discussed the impact of monetary policy to the talent market. He find that, accommodating monetary policy will lead to positive social benefits and help bring positive welfare effects of the reform to the fore. Clayton Glen (2007) [2] analyzed the factors which have influence on talent strategy, such as leadership behavior and corporate brand. Wang Yao-hui[3] and Duan Li[4] summarized the related research domestic. Chang Xiao-yong (2007) [5] studied the influences of globalization on Chinese talent strategy. Additionally, talent strategy of developed countries have been compared by Wang Ming-jie(2005)[6]. Yu Yang (2006) [7] researched the talents development trend of the traffic in China and its impact on the transportation sector.

However, most of these researches are based on logic analysis without models. That is why the highly educated talents competition system model is establish though Vensim in this paper. As we know, Vensim is one of the common tools in System Dynamics modeling. The data which get from system dynamic model in this paper has strongly supported some points of the researches above.

## **2 Analysis of the Elements in International Talent Resource Competition**

### **2.1 Economic Elements**

Solid economic basis is the premise to implement the strategy of talent competition. Economically developed countries or regions can provide highly educated talents with high income jobs and comfortable living conditions, and provide economic support for talent continues to pursue advanced studies. These have attracted a large number of talents working for them. Science and technology innovation of highly educated talents linked to the spread of technology and new technology. According to the research(Griliches Zvi 1986[9]), the greater the country's technology innovation investment share of GDP, the faster its economic growth. Since the beginning of 21st century, most of the countries in the world are trying to improve its economic environment, increase funding for education and scientific research investment, improve its competitiveness in the international talent market. America, for example with the strongest talent attraction, its investment used for research and development accounts for about 3% of the gross national product (GNP), and its investment used for universal education increasing at least tens of billions of dollars every year. In order to increase enrollment, most of investment is used in university, especially famous universities' infrastructure rehabilitation and expansion (Li Ping-you 2012[10]). At the same time, the United States also set up Career program, Young Investigator program, Outstanding Junior Investigator program, Pathway to Independence, Independent Scientist Award, Early Independence Award and other projects to develop youth

talent to grow into independent researchers as soon as possible. The investment American military industry used in continue education accounted for 16% of the worker total wages(Wang Tong-xun 2005[11]).

## 2.2 Environmental Elements

Talent effectiveness can't be operated without a certain conditions, such as policy condition and culture condition. Due to the liquidity of international talent, no country can get advantage position in the international talent competition just rely on high salary treatment. System of talent cultivation and using, attitude to scientific research and innovation culture are also key factors to the competition for human resources.

National talent policy in general is divided into two aspects: one is the training policy, including a series of related factors such as basic education system, students dispatching system, talent international communication system and talent training for key project. In 2007, the south Korean government introduced a "study in South Korea plan" which set the ability of attracting foreign students as an important part of the assessment of diplomats (Wang Yao-hui 2011[12]). The second one is talent using policy, including talent selection and introduction, international scientific research platform construction and the related legal system. As a result of the government's high attention and the formulation of a series of laws and regulations such as "the special act of the national and public universities with foreign faculty", "the study exchange promotion act" and "the foreign technical personnel recruitment system", foreigners(83 person) account for 52% of all principle investigators(PI)(160 person) in National Institute for Materials Science WPI International Center for Materials Nanoarchitectonics(MANA) of Japan which has 1 local postdoctoral and 58 foreign postdoctoral(Wuyunqiige 2010[13]).

Culture is the source of creative and one of the most important factors which influence the talent competition strategy. People have different view of values, outlook on life and the way of thinking in different cultural environment. These factors affect the formation of self-confidence, autonomy, self-reliance spirit. The family values, work values and way of life determine people's employment tendency (Mobley 1979[14]).

Mature modern society can provide people with more superior development environment, including the probity of the administrative system, the identification of the immigrants, stable social order, etc. Through the statistics, Registering an enterprise took an average of 6 crossing in mature social system, while took 11 crossing in backward social system; Registering an enterprise needs 27 days under the mature social system while needs 59 days under backward social system; Registering an enterprise costs 8% of annual per capita income under the mature social system while costs 122% under backward social system.

### 3 Construction of Talent Competition Strategy Model

#### 3.1 Construction of Causality Diagram

According to the analysis of talent competition system mentioned above, several important variables affecting the operation of the system can be established based on the principle of system dynamics, e.g., economic development level, education investment, high-paying jobs, human development index, governmental efficiency, environmental pollution level, and prospects of personal development. The pros and cons of China highly educated talent competition strategy will be reflected by highly educated talents quantity. “Highly educated talents” appears in the diagram include: foreign talents, highly educated talents with local degree and returning students. The causality diagram is shown in figure 1.

The feedback loop in the system mainly includes: amount of talent—profits of large-sized enterprises—R&D spending—opportunity to get high salary job—number of foreign talents—amount of talent; amount of talent—human development index—higher education quality—number of foreign students study in China—number of foreign talents—amount of talent.

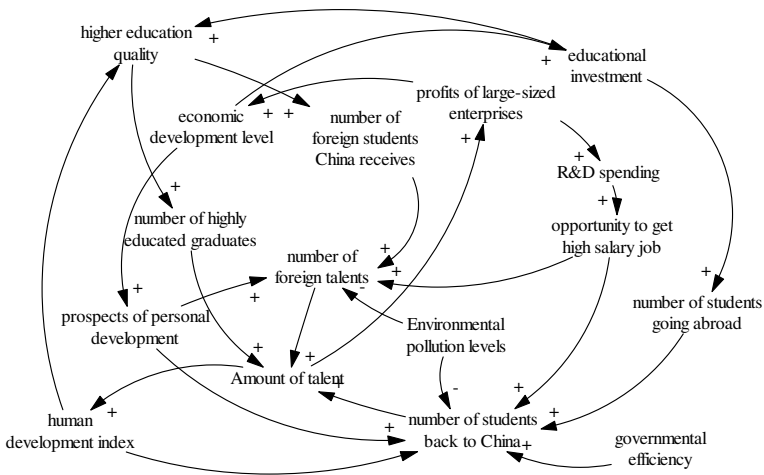


Fig. 1. Causal loop diagram of highly educated talents competition system

#### 3.2 Construction of the Flow Chart of System

After the difference between stock and flow rate in the system and the auxiliary variables and exogenous variables is distinguish, flow diagram of talent competition system (see figure 2) can be constructed based on the analysis of variables in the causality diagram and the distinction of the stock and flow rate, auxiliary variables and exogenous variables in the system. The relationship between variables and the basic elements is shown in table 1.

**Table 1.** Table of elements and variables

<i>Category</i>	<i>Elements</i>	<i>Variables</i>
Economic elements	economic development level	GDP GDP growth rate growth in GDP
	educational investment	number of education spending increase proportion of education funds account for GDP costs of higher education
	R&D spending	proportion of Research and development funds account for GDP corporate spending on new product development
	profits of large-sized enterprises	number of listed companies increased number of listed companies
	opportunity to get high salary job	talent gap for high-paying jobs
Environmental elements	Policy environment	proportion of business and financial regulation funds account for GDP investment of public and social security proportion of public service and social security spending account for GDP growth rate of Chinese government scholarships for international students study in China
		number of the enterprise registered per year Environmental pollution index medical costs per capita
	Cultural environment	human development index
	Social environment	efficiency of the government office enterprise registration cost

## 4 System Simulation and Prediction

### 4.1 Calculation Formula of Important Index

Most of the data used in this system come from statistics data released by the national bureau. The rest part of the data come from the related research scholars domestic and foreign have done. Environmental pollution index is one of them, it is expressed as a number of proportion of days when air quality is lower than the secondary data in capital share days all the year round; Human development index (HDI) is an specified indicator defined by United Nations development program (UNDP) which related to life expectancy (LE), education level (EI) and quality of life, its computation formula is:

$$\text{Life Expectancy Index(LEI)} = (\text{LE} - 20) / (83.2 - 20)$$

$$EI = \sqrt[3]{\frac{\text{Mean Years of School Education Index(MYSI)} \times \text{Expected Years of School Education Index(EYSI)} - 0}{(0.951 - 0)}} \quad (1)$$

$$\text{MYSI} = (\text{Mean Years of School Education (MYS)} - 0) / (13.2 - 0) \quad (2)$$

$$\text{EYSI} = (\text{Expected Years of School Education (EYS)} - 0) / (20.6 - 0) \quad (3)$$

$$\text{Income Index (II)} = (\ln(\text{Gross National Income Per Capita (GNIPC)}) - \ln(163)) / (\ln(108211) - \ln(163)) \quad (4)$$

$$\text{HDI} = \sqrt[3]{\text{LEI} \cdot \text{EI} \cdot \text{II}} \quad (5)$$

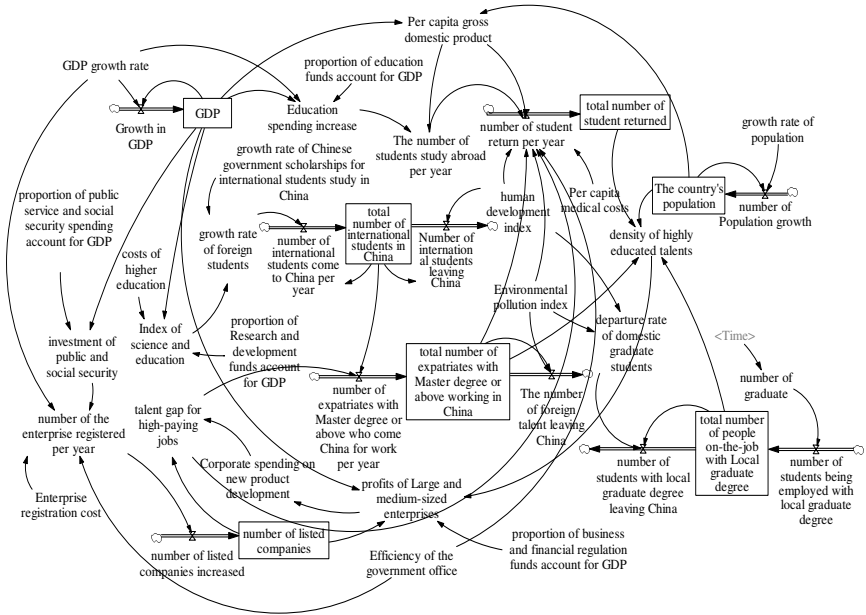


Fig. 2. System Dynamic of talent competition system

4.2 System Dynamics Equation

The system dynamics equation could not listed one by one due to the limitation of space, the main dynamic equation is shown as follows:

$$GDP = \text{INTEG}(\text{growth in GDP}, 1.83617e+009)$$

$$\text{investment of public and social security} = \text{proportion of public service and social security spending account for GDP} * GDP$$

$$\text{corporate spending on new product development} = \text{profits of large and medium-sized enterprises} * 0.2$$

$$\text{total number of international students in China} = \text{INTEG}(\text{number of international students come to China per year} - \text{Number of international students leaving China}, 10)$$

$$\text{total number of people on-the-job with Local graduate degree} = \text{INTEG}(\text{number of students being employed with local graduate degree} - \text{number of students with local graduate degree leaving China}, 12)$$

$$\text{total number of expatriates with Master degree or above working in China} = \text{INTEG}(\text{number of expatriates with Master degree or above who come China for work per year} - \text{The number of foreign talent leaving China}, 6.2754)$$

$$\text{density of highly educated talents} = (\text{total number of student returned} + \text{total number of expatriates with Master degree or above working in China} + \text{total number of people on-the-job with Local graduate degree}) / \text{The country's population}$$

### 4.3 Analysis of Simulation Results

To check the authenticity of the model, we set 2005 as the beginning of the simulation time and run the model with initial data (Specific process will be given in another paper due to the limitation of length). Through the simulation, the trend of Chinese higher education talents in the current economic and environmental conditions before 2025 is shown in figure 3. From the data, we can see that the Chinese highly educated talents density increasing smoothly from 2005 to 2011. This trend is consistent with reality. Predicted results also showed that, the density of highly educated personnel in China will keep the current momentum of development before 2018, and its growth rate will be faster after 2019 in current environment.

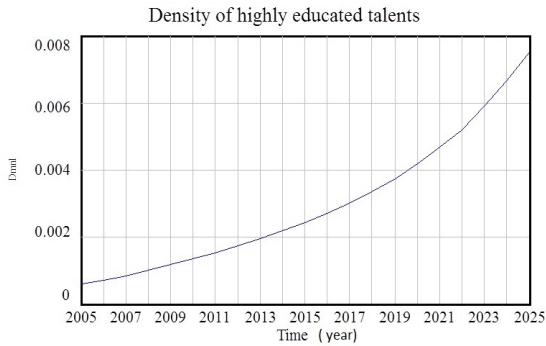


Fig. 3. Trend of the density of highly educated talents under current condition

According to the analysis of the development trend of Chinese highly educated talents in the future, three kinds of adjustment scheme are put forward in this paper. Solution 1: improve the culture and social environment; Solution 2: improve the policy environment; Solution 3: increase education investment. By changing the exogenous variable data, the simulation results are shown in figure 4.

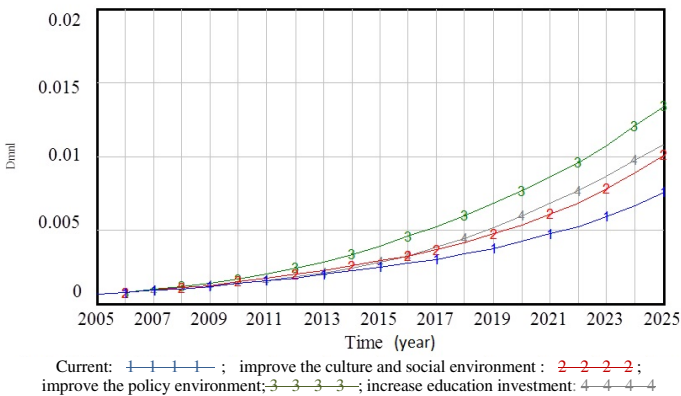


Fig. 4. Trend of the density of highly educated talents under different condition

The simulation results show that:

By changing some exogenous variables in the model such as the Chinese human development index, governmental efficiency index and the enterprise registration cost, the result of solution 1 is shown through Vensim. More students studying abroad will return to China to find chances if start-up costs is reduced, cultural environment and social environment is improved; Encourage innovative cultural environment and efficient government office efficiency will attract more foreign highly educated talents come to China for opportunities; Social identity of foreign immigration will also increase the sense of belonging of overseas students and improve the proportion of foreign students in China after their graduation.

Talent policy needs to be compatible with national conditions, the one which can not be implemented may backfire. By adjusting proportion of public service and social security spending account for GDP, proportion of business and financial regulation funds account for GDP, environmental pollution index and medical costs per capita and other parameters in the system, the result of solution 2 shows that the attention government paid on the pollution of the environment, reform of medical system, the improvement of the social security system, have a significant influence on the executive ability of the society.

Education funds in China accounts for only 0.88% of the central government fiscal spending compared with 6.03% in Japan, 16.01% in South Korea and 18.65% in Singapore (IMF 2012 [15]). There are a lot of things to do for China to increase its investment in education. By adjusting the proportion of education funds in GDP and costs of higher education, the simulation results show that the cost of higher education will be reduced in step with improvement of the higher education in China by increasing the education investment. By increasing international communication opportunity, making full use of the international education resource, cultivating highly educated talents, the proportion of local highly educated talents will be improved if the government increase international communication opportunity and make full use of the international education resource. By this means, the security of China's talent resources can be ensured. Moreover, China will win a good position in the international competition of higher education talent.

## 5 Conclusion

Through the simulation of China's highly educated talent competition strategy, we found: nearly 10 years will be needed for China if he want to enter high-speed development stage, but the density of highly educated talents in China will be inching higher in next 7 or 8 years.

Compared with the developed countries, China should pay more attention to education, healthcare, public and social security if it wants to get the initiative in international competition of higher education talent. A healthy and comfortable living environment with good personal development prospects will attract more talents to serve the country. Only by this way, talents will be "inspired, retained and used well".



The model in this paper still needs to be improved because the structure of talent competition system is so complicated. In order to make our suggestion more desirable, the model will be improved constantly in the follow-up work.

## References

1. Aguiar, A., Ribeiro, A.P.: Monetary policy and the transition costs of a labor market reform. *Journal of Macroeconomics* 31(4), 547–560 (2009)
2. Glen, C.: Fostering talent opportunity: getting past first-base. *Strategic Direction* 23(10), 3–5 (2007)
3. Huiyao, W.: China's New Talent Strategy: Impact on China's Development and its Global Exchanges. *The SAIS Review of International Affairs* 31(2), 49–64 (2011)
4. Li, D.: Summary of the Research of our Country's Talent Strategy. *Theory and Reform* 5, 41–44 (2007)
5. Cang, X.: Relationship Between Economic Globalization and China International Talent Competition Strategy, p. 11. Capital Normal University, Beijing (2007)
6. Mingjie, W.: Comparison of Talent Strategies in the Developed Countries. *Chinese Public Administration* (8), 77–79 (2005)
7. Yu, Y., Liu, S.: Analysis and Forecasting of Traffic Talents Based on System Dynamics. *China Water Transport (Theory Edition)* 4(8), 190–191 (2006)
8. Mineer, J.: Human capital and economic growth. *Economics of Education Review* 4(3), 195–205 (1984)
9. Zvi, G.: Productivity, R&D, and Basic Research at the Firm Level in the 1970's. *The American Economic Review* 76(10), 141–154 (1986)
10. Li, P.: Talent Competition, How Foreign Countries Attracting Talent. *Western China* (12), 60–63 (2012)
11. Tongxun, W.: Extraction of Talent Resource Based on the strategy of reinvigorating China through human resource development. *Chinese Talents Magazine* (3), 18–20 (2005)
12. Yaohui, W.: The international talent competition strategy of Japan, Korea and Singapore. *The First Resource* (2), 174–188 (2011)
13. Wuyunqigige: Trends in International Contest for Science & Technology Talents and China's Countermeasures. *Bulletin of Chinese Academy of Science* 25(6), 595–601 (2010)
14. Mobley, W.H., Griffeth, R.W., Hand, H.H., Meglino, B.: Review and conceptual analysis of the employee turnover process. *Psychological Bulletin* 86, 493–522 (1979)
15. International Monetary Fund. Statistics Dept. Government Finance Statistics Yearbook (2011) (March 13, 2012)

# Kernelized-QEMU: A Study of System-Level Virtual Layer in Linux Kernel\*

Yongxuan Xu, Xianglan Chen \*\*, Hua-Ping Chen, and Huang Wang

School of Computer Science and Technology, University of Science and Technology of China  
No. 443, Huangshan Road, Baohe District, Hefei City, Anhui Province, PRC  
{jyoeiken,ustc}@mail.ustc.edu.cn,  
{xlanchen,hpchen}@ustc.edu.cn

**Abstract.** In this paper, we propose a way to implement the cross-platform system-level virtual layer in Linux kernel to achieve a higher performance, because, the virtual layer can be supported directly by the operating system and can directly interact with the real hardware without the performance cost of virtual device and user-space. As an example, QEMU<sup>1</sup> is embedded into the Linux kernel. The tests show that the performance of kernelized-QEMU is better than QEMU.

**Keywords:** system-level virtual layer, QEMU, Linux kernel.

## 1 Introduction

Code migration problem [2] becomes increasingly important with the continuous development of computer architecture. In general, if a new architecture cannot be supported widely by software, it will be difficult to get opportunity to survive. Nowadays, an effective method is generally used to solve this problem by building a virtual layer above the architecture through virtualization to allow the common software running smoothly on it [3].

Actually, system-level virtualization research is not a new technology. In the early 60s and 70s of last century, this technology was used to achieve the high-level software to share the hardware resources by IBM mainframe system [4]. At that time, the goal of the virtualization was to make full use of expensive computer resources to allow more people to use the computer through terminal devices. But nowadays, by system-level virtualization, it is possible to unlock the tight coupling between the specific hardware architecture and software system. The virtualization can organize a variety of computing resources flexibly. It's also possible to achieve scalable transparent computing system architecture through system-level virtualization. In addition, it is possible to improve the use efficiency of the computing resources, and provide the personalized and pervasive computing resources usage environment to the users [5].

---

\* Supported by “the Fundamental Research Funds for the Central Universities” (WK0110000016).

\*\* Corresponding author.

Currently, through dynamic binary translation techniques, many system-level virtualization applications such as VMware and QEMU, simulate the full system to construct the system-level virtual layer. Those applications mainly execute in user space with many other applications also executing in the host computer. To the author's knowledge, it may have the following drawbacks:

- The virtual layer cannot take full advantage of performance of the hardware in host platform including the CPUs and peripherals.
- Needs multilayer calls when interacting with the hardware and frequent switches between user mode and kernel mode, therefore, performance loss is very serious.
- It costs more to use virtual devices than the real device at runtime.
- The performance of virtual layer is deeply influenced by the host environment for different kinds of host applications.

Although KVM [6] runs in kernel space, its virtualization requires the support of hardware, such as Intel VT or AMD V technology. Moreover, KVM is based on the homogeneous platform, which limits the possibility of the integration of heterogeneous computing resources.

Based on the above analysis, we proposed an idea that embeds the virtual layer that supports the cross-platform and does not depend on the hardware into the operating system. It is to make the operating system directly support the virtual layer and the virtual layer occupy the hardware exclusively.

In order to prove the idea is feasible, QEMU was embedded into the Linux kernel, which is called kernelized-QEMU. This paper will present how to construct the kernelized-QEMU, focus on solving the interfaces problem and device emulation problem, and the execution performance evaluation of the virtual layer constructed by kernelized-QEMU.

## 2 QEMU and the Design of Kernelized-QEMU

### 2.1 A Short Introduction to QEMU

QEMU is a multi-source multi-objective binary translation system, which supports process-level and system-level virtualization operating modes, with the characteristics of high-speed, cross-platform, open source, easy to transplant, etc [1]. The system-level virtualization refers to QEMU emulates the entire computer system, including one or more processors and a variety of peripheral devices.

QEMU translation system consists of front-end decoder, analysis optimizer, back-end translator and control core [1], which is good enough to support the heterogeneous platform. The front-end decoder is responsible to translate the binary code of the source platform into the intermediate code TCG (Tiny Code Generator). The analysis optimizer uses activity analysis method to remove the partial redundant instruction of TCG code. The back-end translator translates the TCG code into binary code of the target platform, so that the code can be run on the host computer. And the control core is responsible for the control flow of the entire system. The four modules work

together which makes QEMU virtual machine can simulate a heterogeneous virtual layer on the host computer.

### 2.2 The Design of Kernelized-QEMU

Until recently, there is some lack of knowledge about embedding the virtual machine into the operating system kernel to construct the kernel based virtual layer. In this occasion, we proposed to integrate the QEMU source code to the Linux kernel code and only execute the QEMU to build a system-level virtual layer with a target operating system running over the virtual machine when the kernel completes the initialization. Therefore, the host Linux becomes a tool that is only used to provide full service for the virtual layer. To this purpose, many kernel functions are unwanted and should be removed from the kernel. Also, since there is no host application running to share the resources, the QEMU can use the hardware exclusively. Figure 1 is the schematic diagram of ordinary virtual system and kernelized-QEMU system.

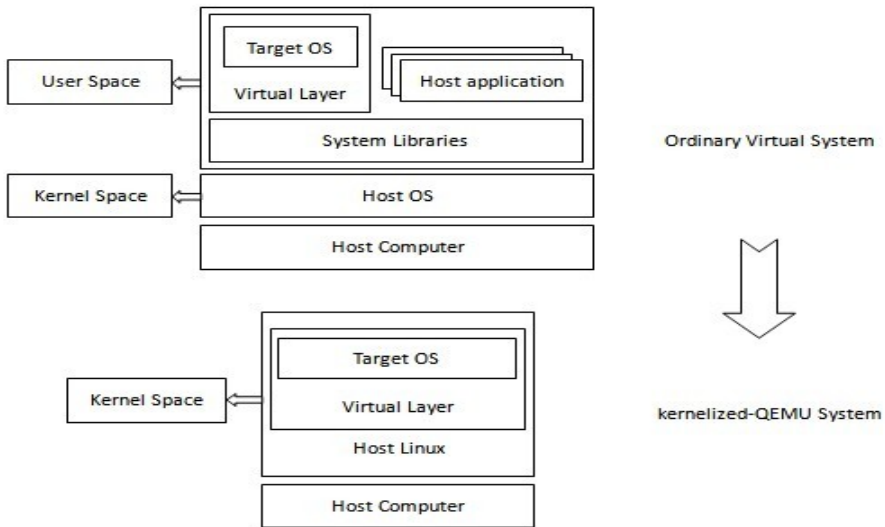


Fig. 1. Ordinary virtual system and kernelized-QEMU system

To make the QEMU embedded into the Linux kernel, there are three major problems: the interfaces problem, device emulation problem and floating point problem. Because the Linux kernel cannot support floating point operation perfectly, we simulated the floating point arithmetic by using fixed point instead of floating point after studied floating point format. The following two chapters will explain in detail about how to solve the interfaces problem and device simulation problem.

### 3 Interfaces Problem

QEMU in user space needs to call the C library functions, such as `open()`, `close()`, `read()` or `write()`. Multiple system calls in this process result in frequent switching between user mode and kernel mode, increasing the system overhead. How to get rid of the shackles of the interfaces and various libraries became a problem when the QEMU was embedded into the Linux kernel. After investigation, we proposed that change the interfaces of the library functions that relied by QEMU to the functions implemented by the Linux kernel, and when necessary, implement a part of C library functions in the kernel.

In this paper, the entire modified interfaces are divided into three major categories and analysis in the following three sections.

#### 3.1 Simple Interface

This part of the interfaces almost didn't need to be modified, such as `access()`, `read()`, `fclose()`, `fseek()`, `mprotect()`, `free()`. Most of these interfaces were file operation functions and memory handing functions, because the Linux kernel also need these functions to handle files and memory. The parameters and return values of the interfaces were same as the parameters and return values of user space interfaces.

#### 3.2 Partially Modified Interface

Part of the interfaces could not correspond to the kernel functions. In order to implement the interfaces, some related kernel functions were modified and combined. For example, the kernel function `sys_write()` and `sys_fsync()` were combined to construct the interface `write()`, because the `sys_write()` does not have the function to make the memory and hard disk in sync; And according to the meaning of the function, `sys_read()` was used to achieve the interface `fgets()`. A part of the interfaces were much more complicated to implement, such as `open()`, `fcntl()`, `fprintf()`, `pread()`, `pwrite()`.

Taking an example of function `pwrite()`: First, the kernel function `sys_write()` and `sys_fsync()` were used to complete basic write function; Then, `sys_lseek()` was used to change the position of the file pointer; Last, the atomic of `pwrite` operation had to be guaranteed.

#### 3.3 Rewritten Interface

There were parts of interfaces had to be rewritten because the kernel does not have any related functions. Such interfaces were very few, mainly involved the character conversion functions and the time functions.

## 4 Device Simulation Problems

Device simulation problems were the key to the experiment. The virtual layer can use the hardware exclusively, so the core idea of the solution was to make the virtual layer interact with the hardware directly.

### 4.1 Directly Operable I/O

QEMU simulates VGA to control the output of the virtual layer. And there are two main aspects to operate VGA: VGA port that explored to the user and VGA video memory (memory address: 0xA0000 to 0xC0000). The virtual registers of the simulated VGA (not the real VGA) are first changed when the upper operating system wants to change the VGA mode in the ordinary QEMU system. In order to make kernelized-QEMU can operate VGA directly, the two aspects were both modified.

The modification of VGA port operation was associated with the hardware, in the experiment. And the computer of X86 architecture was used as the experiment platform. In such cases, the kernelized-QEMU functions that registered to read and write the VGA ports are as follows:

```
register_ioport_write(0x3c0, 16, 1, cirrus_vga_ioport_write, s);
register_ioport_read(0x3c0, 16, 1, cirrus_vga_ioport_read, s);
```

Therefore, `cirrus_vga_ioport_read()` and `cirrus_vga_ioport_write()` were rewritten to achieve the function that operating VGA directly. The function “static void `cirrus_vga_ioport_write(void *opaque, uint32_t addr, uint32_t val)`” of QEMU writes the value of “`val`” to the address of “`addr`”. According to the meaning of the function, the parameter “`addr`” was checked and converted to the real address, and then the “`val`” was written to the address in kernelized-QEMU. The function `lq_inb()` was added into the function `cirrus_vga_ioport_write()` to write the port of VGA. And the modification of the function `cirrus_vga_ioport_read()` was similar to `cirrus_vga_ioport_write()`, function `lq_outb()` was added to read the value of the port. The method of Linux kernel was used for reference to accomplish the specific methods to read or write VGA port.

The virtual layer constructed by kernelized-QEMU is run in kernel space, so it only can access the memory address that above 0xC0000000 (1 GB kernel memory, 32-bit Linux kernel) [8] The address of VGA video memory had to plus the offset of kernel address, otherwise, the program could not access the memory.

The modification of VGA video memory operation was associated with the hardware too. Function groups `cirrus_vga_mem_read` and `cirrus_vga_mem_write` were registered to read or write VGA video memory in the experiment. And after a series of calls, `cirrus_vga_mem_readb()` and `cirrus_vga_mem_writeb()` were modified to read or write the video memory address.

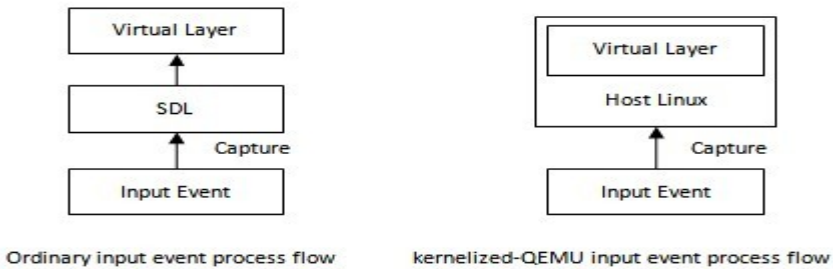
The pseudocode of function `cirrus_vga_mem_writeb()`:

Begin

```
char * kernel_vga_offset = (char *) 0xc00a0000
if addr >= 0x0000 AND addr < 0x20000
    return writeb(mem_value, kernel_vga_offset+addr);
else
    ERROR
```

End

The Linux kernel can detect the input event and stores it to the structure `input_event`. And the traditional virtual layer constructed by QEMU requires the support of specific libraries, such as SDL (Simple DirectMedia Layer), to handle the keyboard events. In order to remove the dependencies of libraries, the way to obtain keyboard events was changed to the host Linux and the keyboard handling function of QEMU was used to process the input through the keycode.



**Fig. 2.** Ordinary input event process flow and kernelized-QEMU method

Figure 2 is the schematic diagram of the ordinary input event process flow and the improved kernelized-QEMU method. The host Linux processing operation of the input event was truncated and the keycode was delivered to the keyboard event processing interface which provided by kernelized-QEMU. Each keycode had its own state to represent the three states: press, hold or release, and the parameter “value” was used to show the state in kernelized-QEMU. The key was pressed when “value” equaled 1; the key was held when “value” equaled 2; and the key was released when “value” equaled 0. Due to the only keycode was received by the keyboard event processing interface of kernelized-QEMU, an intermediary layer was built to deal with the state. The keycode was delivered to the interface immediately when the key was pressed or held ( $\text{value} > 0$ ); otherwise, the keycode was plus 128 to represent the key was released before delivering.

## 4.2 DMA Simulation

Asynchronous input and output process was used to simulate DMA in QEMU. And it was initialized when the virtual layer opened the hard disk image at the first time. The

interface `paio_submit()` was used to submit the asynchronous I/O request to read or write the hard disk when DMA request came. Then, a thread was created to respond to specific request. When the thread had completed the task, a custom signal was sent to notice the main thread. The signal response function which was registered by main thread received the signal, wrote the communication pipe and submitted an event, which was called `qemu_event`, to the virtual layer. A soft interrupt was implemented through the QEMU event mechanism, and the virtual layer started to check the state of communication pipe that written by the signal response function. If the pipe had been written, function `posix_aio_read()` was called to read and empty the pipe. At last, function `posix_aio_process_queue()` was called to complete the remaining work. Above is the process of DMA that simulated by QEMU through asynchronous I/O.

To simulate DMA in kernelized-QEMU, three questions had to be solved: signal, pipe and thread. Because the virtual layer was run in kernel space, through signal and pipe to notice the virtual layer that DMA reading or writing process has been completed was abandoned, function `posix_aio_process_queue()` was called directly without used QEMU event mechanism while the response of DMA request was completed. The kernel thread was used to instead of `pthread` (POSIX thread) to create the thread. And the mutex lock of the Linux kernel was used to protect the critical resource.

In addition, another way to simulate DMA is changing asynchronous I/O to synchronous I/O. In this situation, function `aio_thread()` was called, not created a new thread, when the DMA request was submitted.

## 5 Experiments and Discussion

Due to the limitation of experimental conditions, kernelized-QEMU was deployed on an X86 platform. And a Linux kernel was run on the system-level virtual layer constructed by kernelized-QEMU. The kernel boot time in kernelized-QEMU system was faster than the time in QEMU system (Figure 3).

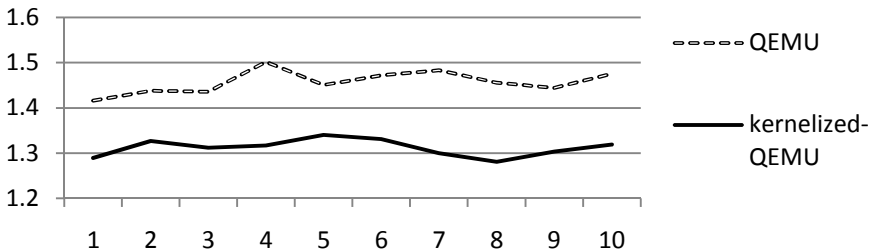
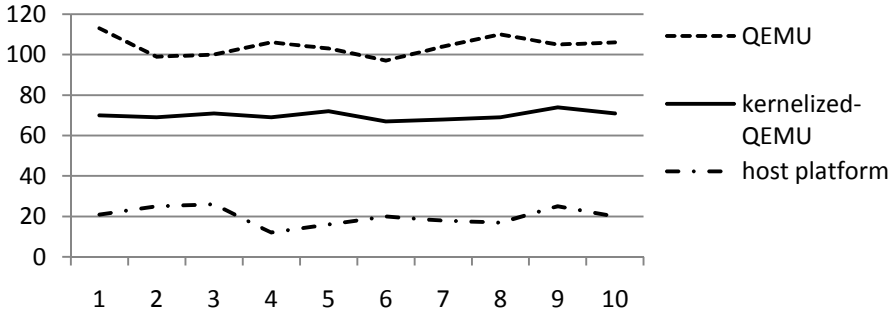


Fig. 3. The kernel boot time (Unit: Second)

The program for calculating the value of  $\pi$  was run to test the performance of the virtual layer. Taylor formula was used to calculate the arctangent value to get the value of  $\pi$  (accurate to 1000 decimal places) in the program. And the virtual layer constructed by kernelized-QEMU used less time than QEMU at all time (Figure 4).





**Fig. 4.** The time to calculate  $\pi$  (Unit: millisecond)

Figure 3 and figure 4 shows that the operation efficiency of kernelized-QEMU system is better than ordinary QEMU system, which consistent with our vision. The virtual layer implemented in the Linux kernel can reduce the multilayer calls and take full advantage of the hardware performance of the host platform, thereby improve the performance of the virtual layer. And the operating system can be started by kernelized-QEMU automatically and immediately, which means the virtual layer is transparent to the users.

Currently, kernelized-QEMU system cannot apply the memory over the kernel address space. In the future, we will solve this problem, deploy the kernelized-QEMU system to the heterogeneous platform and optimize the virtual layer according to the hardware features of the host platform.

## 6 Conclusion

The experiments proved that the kernel based system-level virtual layer can improve its operation efficiency. Thus, it may bring up some new ideas for the development of the full system simulation on the heterogeneous platform.

## References

1. Bellard, F.: QEMU, a fast and portable dynamic translator. USENIX (2005)
2. Kontogiannis, K., Martin, J., Wong, K., et al.: Code migration through transformations: An experience report. CASCON First Decade High Impact Papers. IBM Corp., 201–213 (2010)
3. Rosenblum, M., Garfinkel, T.: Virtual machine monitors: Current technology and future trends. *Computer* 38(5), 39–47 (2005)
4. Butts, T.H., Burris Jr, S.H., Clark, S.J., et al.: Server and web browser terminal emulator for persistent connection to a legacy host system and method of operation: U.S. Patent 5,754,830 (May 19, 1998)
5. Vallee, G., Naughton, T., Engelmann, C., et al.: System-level virtualization for high performance computing. In: 16th Euromicro Conference on Parallel, Distributed and Network-Based Processing, pp. 636–643. IEEE (2008)

6. Kernel Based Virtual Machine, <http://kvm.qumranet.com/kvmwiki>
7. Bovet, D.P., Cesati, M.: Understanding the Linux Kernel, 3rd edn. O'Reilly Media, Inc. (2005)
8. Yabin, H., et al.: An Optimization Approach for QEMU. In: 2009 1st International Conference on Information Science and Engineering (ICISE 2009), pp. 129–132 (2009)

# A Multi-pattern Matching Algorithm Based on Double Array Trie

Miao Hou<sup>\*</sup>, YingHui Song, Dongliang Xu, and Hongli Zhang

School of Computer Science and Technology, Harbin Institute of Technology  
Harbin, China  
houmiao@163.com, xudongliang@pact518.hit.edu.cn,  
{zhanghongli, yhsong}@hit.edu.cn

**Abstract.** With the increasing of network throughput, the size of pattern sets in Network Intrusion Detection System (NIDS) are growing gradually, how to reduce memory space of pattern sets has become one of the research hot spots. This paper presents a multi-pattern matching algorithm which combines the classical Aho-Corasick (AC) algorithm with Double Array Trie (DAT), called DAT-AC. We use two linear arrays to determine the state transition and decrease the memory space by reducing the unnecessary state transition. Experimental results demonstrate that DAT-AC performs better than the classical AC algorithm, when the number of pattern is larger than five thousand.

**Keywords:** multi-pattern matching, AC, DAT, DAT-AC.

## 1 Introduction

Pattern matching is one of the oldest problems in computer science. Due to the high stability of automaton, string matching algorithm which is based on finite automaton is more widely used. For example, Snort is a popular detection system with open resource, and it is implemented by AC algorithm which is based on automaton. With the rapid growth of the network traffic and the continuous improvement of user requirements, the rule number of Snort is increasing continuously. At this time, the memory space of the string matching expands rapidly, which leads to performance degradation of the whole system. For large scale data string matching algorithm, tightening technology of efficient storage has become one of the hot spots of research.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 analysis the theory of AC and DAT. Section 4 shows the design of DAT-AC. In section 5, several experiments are designed to evaluate our approach.

## 2 Related Work

Research of string matching has a long history with so many classical and famous algorithms, such as Knuth-Morris-Pratt algorithm [1], Aho-Corasick algorithm [2],

---

<sup>\*</sup> Corresponding author.

Boyer-Moore algorithm [3], Wu-Manber algorithm [4], SBDM [5], etc. Most of these algorithms are widely used at present. Above all this algorithms, AC which is an exact multi-pattern matching algorithm based on automaton, has high practical value. However, with the increase of the number of patterns, the performance of AC is not well. In order to solve this problem, many tightening technology of memory space have been proposed, such as Dencker et al. [6], Aoe [7], Mizobuchi et al. [8], Aoe [9], Galli et al.[10], etc. Above them, Aoe [7] and Mizobuchi et al. [8] introduced Double Array Trie to solve the shortage of memory effectively.

### 3 Theory of AC and DAT

#### 3.1 Aho-Corasick(AC)

AC automaton can be divided into two stages which include preprocessing phase and search phase. In the preprocessing phase, AC automaton establishes three functions, including transfer function, failure function and output function, and constructs a finite automaton like a tree through these three functions. Transfer function puts a two-tuples which is composed of status and input character mapped to another status. When transfer function is failure, the failure function will be asked. And the output function connects keyword with each state, the output states indicate keywords have been found. In the search phase, when we start scanning the text, the initial state is the current state of the automaton and the first character of the input text is regarded as the current input character, then we scan the text to take out the next character and use both transfer function and failure function to enter into next state which is based on current state and the character we have scanned. These steps are continued to perform, a match is founded until the output function is not empty.

The traditional AC algorithm has two advantages, the one is that it does not need to trace back when scanning the text, the other one is the complexity is  $O(n)$ . Complexity is only related to text length, and not affected by the pattern set.

For example, in Fig.1, we create a finite automaton which uses pattern set  $\{he, she, his, hers\}$ , the solid line shows the transfer function, and the dotted line represents the failure function, while the double circle shows the output node.

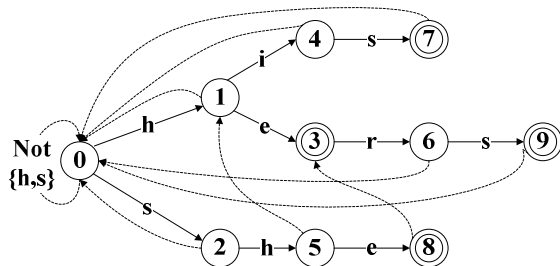


Fig. 1. A state transition diagram of AC algorithm

We put the ASCII set as the set of the text input, and create the state table of Automaton in the form of matrix. The ASCII set have 256 different characters and

each character represents 4 bytes. Each state transition is at least 1KB(256\*4) while we do not contain failure function and output function. In fact, a lot of transition is point to NULL in the state table of automaton. To decrease unnecessary transfer of memory is one of the effective methods to reduce the storage space. So AC algorithm based on Double Array Tire is proposed in this paper.

### 3.2 Double Array Trie(DAT)

For a string of length  $n$ , we can complete the entire search through  $n$  matches by using trie. It is also a common implementation in Chinese word segmentation algorithm. However, the shortcoming of trie matching is the high free rate of the space. In order to reduce the waste of shortage memory and ensure the query efficiency of the trie, DAT algorithm is proposed.

DAT is one of deformations of the trie, it is a data structure that it can improve the space utilization under the premise of guarantee the trie searching speed. It searches the event only related to the length of the data rather than the amount of data storage. DAT is a deterministic finite automaton (DFA) in essence. Each node represents a state of automaton, and conducts state transition according to different variables. The query will be completed when it reaches an end state or cannot transfer. DAT use two linear arrays (*base* and *check*), while array *base* is used to determine the transfer of the status while *check* is used for testing the correctness of transfer.

## 4 Design of DAT-AC

In this paper, we apply DAT to AC in order to reduce the data storage space. The method we used is to introduce double array trie on the design of transfer function which is named *base* and *check*. The *base* value of the current state add the ASCII of input character is the offset of the next state, while table *check* stores the father state information of the current state. Failure function is constructed with transfer function, and the construction method of output function is the same as the classical AC algorithm. Fig.2 shows the relationship among these three tables which include table *base*, table *next* and table *check*.

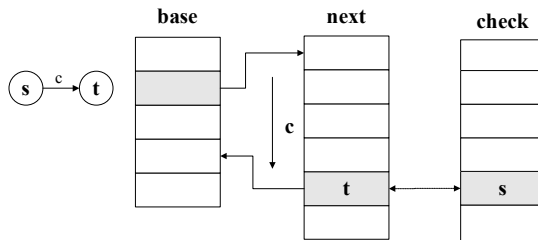


Fig. 2. The relationship of three tables in transfer function

Table *next* represents transfer function and the subscript of *next* is the position offset while the output is the status value. The subscript of the *base* is status value and the output is the *base* value. We suppose that the current status of *next* is *s*, *c* is the

input character, we get state  $t$ . The position of  $t$  in the next table is the location of state  $s$  adds the base value of  $s$  and the value of input ASCII. For the table *check* which subscript is a status value and the output is the father state value of the subscript status. The following pseudo-code is the algorithm of building Double Array Trie.

---

```

Algorithm: buildDAT (s,c)
Input:  $s, c$ 
Output: next state
Procedures:
01:  $t = \text{Next}[\&s + \text{base}[s] + c]$ 
02:   if ( $\text{check}[t] = s$ )
03:     then next state =  $t$ ;
04:   else
05:     fail
    
```

---

Based on the above algorithm, we give an example to illustrate. For example, we want to conduct a finite automaton according to the pattern set  $P = \{he, she, his, hers\}$ , and the input mode of the patterns is in the form of breadth-first. The first layer is  $\{h, s\}$ , the second is  $\{e, h, i\}$ , the third is  $\{e, s, r\}$  and the forth is  $\{s\}$ . To build the table *next*, we apply for 256 storage unit because there are 256 possible inputs of ASCII. The first input character of each pattern is ranked according to the order in the dictionary. The offset of the smallest character in the table *next* is 1 while the other characters offset according to its relative position.

We should make sure the ASCII value of the characters in the pattern set. The value of  $h$  is 104,  $s$  is 115,  $e$  is 101,  $i$  is 105 and  $r$  is 114. We analysis the first layer  $\{h, s\}$ , table *next* creates state 1 and state 2. They are shown in Table 1. In the table *base*,  $\text{base}[0] = -103(1=0+\text{base}[1]+104)$ , and in the table *check*,  $\text{check}[1] = 0, \text{check}[2] = 0$ .

**Table 1.** The first layer

0	1	2	3	4	5	6	7	8	9	10	11	12	...
0	1											2	
	h											s	

For the second layer of the pattern set  $\{e, h, i\}$ , table *next* creates three new status, including state 3, state 4 and state 5. It is showed in table 2. We get status *he* and *hi*, they all regard state 1 as father state and  $\text{base}[1] = -100$ . We can also get status *sh* and it takes state 2 as father state and  $\text{base}[2] = -103$ . In the table *check*,  $\text{check}[3] = 1, \text{check}[4] = 1$  and  $\text{check}[5] = 2$ .

**Table 2.** The Second Layer

0	1	2	3	4	5	6	7	8	9	10	11	12	...
0	1	3	5			4						2	
	h	e	h			i						s	

For the third layer of the pattern set  $\{e, s, r\}$ , table *next* creates three new status, including state 6, state 7 and state 8. As shown in Table 3, states 6 represents *her* and  $\text{base}[3] = -112$ , states 7 represents *his* and  $\text{base}[4] = -116$ , states 8 represents *she* and  $\text{base}[5] = -97$ . And in the table *check*,  $\text{check}[6] = 3, \text{check}[7] = 4$  and  $\text{check}[8] = 5$ .

**Table 3.** The third layer

0	1	2	3	4	5	6	7	8	9	10	11	12	...
0	<b>1</b>	<b>3</b>	<b>5</b>	6	7	<b>4</b>	8					<b>2</b>	
	<b>h</b>	<b>e</b>	<b>h</b>	r	s	<b>i</b>	e					<b>s</b>	

For the forth layer of the pattern set  $\{s\}$ , one new state is created, which is state 9 *her*. As shown in Table 4,  $base[6]=-111$ ,  $check[9]=6$ . At this moment, transfer function is constructed, we can get state transfer diagram which is the same as figure1.

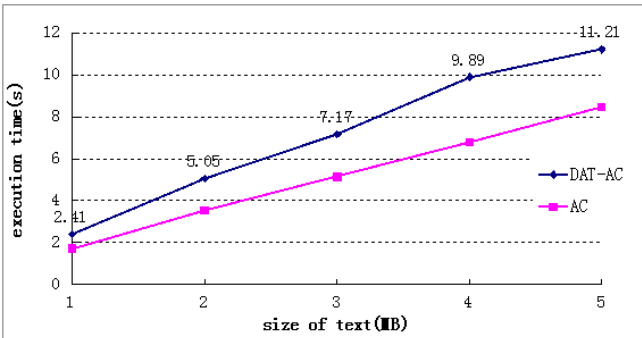
**Table 4.** The forth layer

0	1	2	3	4	5	6	7	8	9	10	11	12	...
0	<b>1</b>	<b>3</b>	<b>5</b>	6	7	<b>4</b>	8	9				<b>2</b>	
	<b>h</b>	<b>e</b>	<b>h</b>	r	s	<b>i</b>	e	s				<b>s</b>	

### 5 Experimental Evaluation

In order to prove the efficiency of DAT-AC in large pattern sets, we have designed the following experiments. Our experimental environment is as follows. CPU is AMD Athlon II X4 620 Processor, memory is 512 and operating system is Fedora 13.

We search different sizes of texts using a same pattern set whose size is 1000, and the text size is from 1MB to 5MB. We perform AC algorithm and DAT-AC algorithm respectively to match different sizes of text. We can draw a conclusion through Fig.3. AC performs better than DAT-AC algorithm when the number of patterns is 1000. And with the increase of the text size, the running time of AC continues to grow regularly, so we can also verify the complexity of AC algorithm is proportional to the size of the text.



**Fig. 3.** The speed of the two algorithms

Next we search the same text using seven different sizes of pattern sets, and the size of the pattern sets is from 1 thousand to 50 thousand while the size of the text is 20KB. Then we observe and record the searching time of AC and DAT-AC respectively. Fig. 4 shows the comparison results on searching time. From Fig. 4, the searching time of DAT-AC is far lower than AC algorithm on large pattern sets.

The searching time of DAT-AC increases slowly while AC grows rapidly when the number of patterns growth. DAT will ease the storage space problem, and for small memory processor, DAT-AC will greatly accelerate the speed of pattern matching.

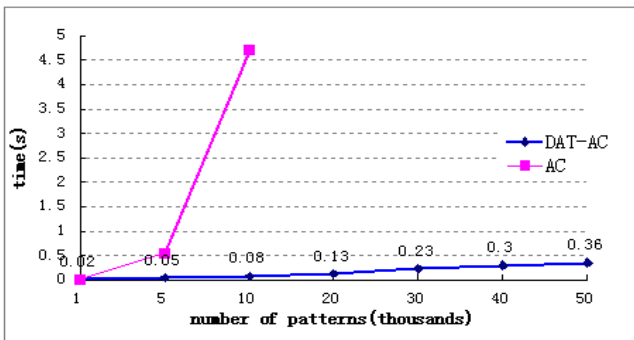


Fig. 4. The searching time of the two algorithms

**Acknowledgment.** This research was partially supported by the National Basic Research Program of China (973 Program) under grant No. 2011CB302605, the National High Technology Research and Development Program of China (863 Program) under grants No. 2011AA010705, No. 2012AA012502 and No. 2012AA012506, the National Key Technology R&D Program of China under grant No. 2012BAH37B01.

## References

1. Knuth, D.E., Moms, J.H., Pratt, V.R.: Fast pattern matching in strings. *SIAMJ. Compt.* 6(2), 323–350 (1977)
2. Aho, A., Corasick, M.: Efficient string matching: An aid to bibliographic search. *CACM* 18(6), 333–340 (1975)
3. Cicero, S., Knuth, D.E., Moms, J.H., Pratt, V.R.: Fast pattern matching in strings. *SIAM J. Compt.* 6(2), 323–350 (1977)
4. Wu, S., Manber, U.: A fast algorithm for multi-pattern searching, Technical Report TR-94-17 (1994)
5. Raffinot, M.: On the multi backward matching algorithm (MultiBDM). In: *Proceedings of the 4th South American Workshop on String Processing*, pp. 149–165 (1997)
6. Dencker, P., Durre, K., Heuft, J.: Optimization of parser tables for portable compilers. *ACM Transactions on Programming Languages and Systems* 6(4), 546–572 (1984)
7. Aoe, J.: An efficient digital search algorithm by using a double-array structure. *IEEE Transactions on Software Engineering* 15(9), 1066–1077 (1989)
8. Mizobuchi, S., Sumitomo, T., Fuketa, M.: An efficient representation for implementing finite state machines based on double-array. *Information Sciences* 129(1), 119–139 (2000)
9. Aoe, J.: An efficient implementation of static string pattern matching machines. *IEEE Transactions on Software Engineering* 15(8), 1010–1016 (1989)
10. Galli, N., Seybold, B., Simon, K.: Tetris-hashing or optimal table compression. *Discrete Applied Mathematics* 110(1), 41–58 (2001)



# Lower Bound Estimation for Required Number of Nodes in the Opportunistic Communication – Based Wireless Sensor Network

Yubo Deng<sup>\*</sup>, Shizhong Wu, Shoupeng Li, Yongping Xiong, and Tao Zhang

China Information Technology Security Evaluation Center  
Beijing, China

dengyb@lzu.edu.cn, {pm,lisp,zhangt}@itsec.gov.cn,  
ypxiong@bupt.edu.cn

**Abstract.** We are developing an opportunistic communication based wireless sensor network system to gather and analyze the data of sediment transport in the desert for desertification research. There are two kinds of nodes in our system, one is base station and the other is sensor node, which can move randomly driven by wind and forward the data to the base station while they are encountering. Since the number of nodes has large impact on performance like success rate of data forward, lifetime and cost of the system, it is significant to estimate the lower bound for the required number of sensor nodes. In this paper, through some assumption and simplification, we give a simple formula to estimate the minimum number of sensor nodes that the system need to ensure opportunistic encounter between nodes will occur to make the data can be forwarded.

**Keywords:** Lower bound, Number of nodes, Opportunistic communication, WSN, Probability.

## 1 Introduction

In China, the land desertification research is very significant since almost one third territory is desertification and more than 95% of them is located at west of China. Meanwhile, the character of WSN (wireless sensor networks) makes it have wide foreground in the environment monitoring scopes. Our work aims to deploy a WSN to monitor the phenomena of sediment transport caused by wind in the desert. However, it is unpractical to deploy a such traditional WSN in the desert for two reasons: (1) the cost is too high, and (2) the desert environment is so complex and extreme that the communication between two nodes would be easily damaged because of sensors' damage, energy consumed, buried by sand, signal interference by strong wind etc.

In this situation, we proposed to deploy a WSN using the framework of opportunistic networks [1], which means the nodes can move independently driven by the wind

---

<sup>\*</sup> Corresponding author.

and communicate only when they encounter each other or base stations. The base stations here are deployed in advance to receive data forwarded from these nodes. In other words, a sensor node will move randomly more like a moving sand but having larger volume in the desert. A GPS chip in the node works intermittently to make it can record the track itself, which actually reveal some regular pattern of sand's movement. At the same time, some other data like temperature, humidity, speed etc. is recorded by relevant sensors as well. Obviously, if we expect the nodes have enough probability to encounter in a desert with vast expanse, the number of nodes is one of the pivotal factors that we cannot ignore. Thus, how to determine the minimum required number of nodes is necessary to discuss.

Generally, for an ordinary WSN, in which all nodes is static, how should the sensors be deployed in the monitored region is an important question and should be taken into consideration firstly. coverage is considered as a measure of the quality of service provided by the sensor network. In order to sufficiently monitor the entire field of interest for the sensor network, every point of the monitored field must be covered by at least one sensor. Therefore, sensor deployment strategies play a significant role in determining the appropriate placement of sensor nodes to meet certain coverage requirements. The quintessence of sensor deployment is that it uses the least number of sensor nodes to satisfy specific coverage requirement, or to maximize the sensing coverage quality within a given economic budget.

Determining the required number of sensors to be deployed is a critical decision for WSN. The art gallery problem is to determine the minimum number of guards required to cover all points in a gallery [2]. Similar works [3] considered the necessary and sufficient conditions for covering a sensor network with nodes deployed in a grid over a square region. Others have mainly focused on the random deployment strategies, e.g., Hall [4] established approximations or asymptotic bounds for area coverage.

However, many related researches assume that the sensors are deployed statically and in an infinite region, rather than boundary region which is more relevant to real application scenarios. In this paper, we don't need to consider the coverage problem due to the movement feature of all nodes. This is an another special circumstance and different from others. Some standalone nodes are placed into a very large boundary region and they can move randomly. We need to estimate fewest number of nodes to make sure they can encounter at least once for the data forward when they are moving.

The remainder of this paper is organized as follows. In Section 2 related works are outlined. Section 3 describes the system models and problems. We give the formula to compute the number of nodes in Section 4. Finally, we draw the conclusion, and point out the future work in Section 5.

## 2 Related Work

Given the sensor and target characteristics, an exposure-based model to find the required number of sensor nodes was presented in [5]. Specifically, a scheme was

developed to determine the density of sensors for complete coverage. The model incorporated a mobile target that moved on a straight line.

If the least number of sensor nodes that cover a region is  $K$ , then the coverage is of  $K$  degrees. In [6], it was shown that for a set of sensor nodes that provide at least one degree of coverage on a convex region, the communication graph was connected if the communication range of sensor nodes was greater than or equal to twice the sensing range. A similar result was provided in [7].

In [8] the author discussed the trade-off between the number of sensors and the breach detection probability considering the effects of sensor parameters. It presented the weakest breach path problem formulation and provide a solution by utilizing the Dijkstra's shortest path algorithm. Then author proposed a method to determine the required number of sensors to be deployed and to gain in sight about the surveillance performance of the network, the maximum detection probability on the weakest breach path was considered as the performance measure. The required number of sensor nodes and their places were determined in [3] to provide a coverage threshold that defines the confidence level of the deployment.

Random deployment refers to the situation in which sensor nodes are uniformly and independently distributed across the monitored field. In [4], Hall studied how many nodes with fixed coverage radius are needed so that every point of a unit square region is covered by randomly placed sensor nodes. The research in [9,10] determined the densities of sensor nodes that achieve a desired area coverage based on Hall's asymptotic analysis. They defined the area coverage as the fraction of the geographical area and determined the minimum number of sensors to be deployed in the infinite plane using homogeneous Poisson point processes.

A general mathematical model was proposed in [11] to determine the number of nodes based on the required working time of WSN in general application. The model treated the CSP algorithms and network protocols for different applications as the parameters of energy consumption in each processing step; then it determined the number of nodes from these parameters.

In addition, the theory of asymptotic analysis also has a great impact on coverage based node scheduling, active node selecting and applications such as intrusion detection. Given the assumption that the nodes are densely deployed, the research in [12] and [13] organized the sensor nodes into disjoint node sets by working alternately to extend network lifetime. The number of nodes in one set was selected according to the coverage requirement. For the reason that the number of sensors required to meet the desired coverage is based on asymptotic analysis, which cannot meet deployment quality due to coverage overestimation in real applications, the search in [14] proposed two deployment strategies. The deployment quality of the two strategies was analyzed mathematically. Under the analysis, a lower bound on the number of deployed sensor nodes was given to satisfy the desired deployment quality.

All these researches focus on the number of nodes in coverage problem since it has large impact on the performance, lifetime and cost of WSN as well as our work. But none of them can be used directly for our work due to the very different scene. In this paper, we intend to study how to analyze lower-bound of the required number of sensor nodes deployed in a boundary region. The goal of our design is to acquire the

minimum number of sensor nodes randomly deployed in a certain area, which ensure data can be forwarded between nodes and base stations.

### 3 Network Model and Problem Description

In this section, we describe the network model and give some definitions to simplify the analytic process in next section. In order to simplify the problem, the following assumptions are made:

- 1) Each node has infinite energy, which can make it works the whole time.
- 2) The deployed nodes and base stations are independent and identically distributed in the monitored region and moved randomly driven by the wind.
- 3) Ignore the encounter duration time between two nodes, which means the data always can be forwarded from one node to the other when they are meeting.
- 4) The monitored area is square.
- 5) Each node will move out the monitored region finally.
- 6) During a node's movement, it never moves back, which means there's no repetitive path.

The definitions to describe movement pattern of nodes are given as follow:

**Definition 1. A monitored region A.** A monitored region is defined as the area monitored by sensor nodes. We consider this area as  $j \times j$  square monitored region, where  $j$  is  $k$  times as much as a node's transmission radius  $r$ .

**Definition 2. A grid.** A grid is defined as a square having the same side length with  $A$  and divided into many equilateral sub-squares, which have the same side length with diameter of a node's coverage ( $2 \times r$ ).

**Definition 3. A encounter.** A encounter is defined as a node and a base station positioned at the same sub-square.

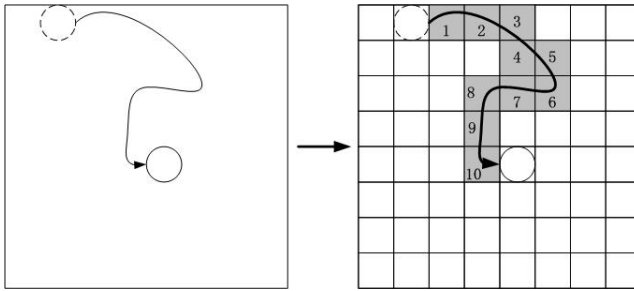
**Definition 4. A start point.** An start point is defined as a sub-square in the grid where a node is deployed originally.

**Definition 5. An end point.** An end point is defined as a sub-square in the grid where a node will move out from.

As showed in Fig. 1, a real movement of a node in monitored area driven by the wind can be mapped in a grid. In this case a whole movement path of the node can be seen as a set of sub-squares.

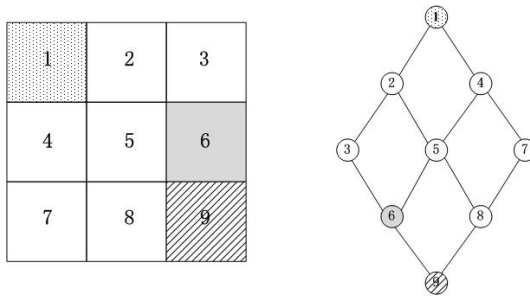
The problem then can be described as follow:

Given a monitored region  $A$ , to estimate the probability that a node encounters base stations in  $A$  means to find out entire paths passing the sub-square located the base station from all of path that a node moves from start point to end point.



**Fig. 1.** Grid representation of a node movement path

For example, the left diagram in the fig. 2 shows three key positions. Sub-square 1 is a start point, and sub-square 6 is where base station located, meanwhile 9 is an end point. The problem now is how to calculate the ratio of the number of paths from 1 to 9 passing 6 and the number of all of paths from 1 to 9, and the repetitive path is out of considering. Then the calculation in the grid can be represented as a path problem in a graph showed as right diagram.



**Fig. 2.** The problem description as graph

### 4 Lowerbound Estimation Formula

Firstly, we consider the situation with one node and one base station in the monitored area. Let a node has  $u$  paths from start point to end point, and  $v$  paths crossing the base station. The probability that the node encounter the base station can be denoted as follow:

$$P(A) = \frac{v}{u} \tag{1}$$

If there are  $N$  nodes, the probability that these nodes encounter the base stations is:

$$P_N(A) = 1 - P(\bar{A})^N \tag{2}$$

Let's discuss a more complex situation that there are  $m$  base stations. Let the probability that a node encounters each base station is:  $P_1(A), P_2(A), P_3(A) \dots P_m(A)$

( $m=1, 2, 3 \dots$ ), then the probability that the node encounters at least one in  $m$  base stations is:

$$P_1^m(A) = 1 - P_1(\bar{A})P_2(\bar{A}) \cdots P_m(\bar{A}) \tag{3}$$

If there are  $N$  nodes, the probability that these nodes encounter  $m$  base stations is:

$$\begin{aligned} P_N^m(A) &= 1 - \left( P_1(\bar{A})P_2(\bar{A}) \cdots P_m(\bar{A}) \right)^N \\ &= 1 - \left( \frac{(u-v_1)}{u} \frac{(u-v_2)}{u} \cdots \frac{(u-v_m)}{u} \right)^N \\ &= 1 - \left( \frac{(u-v_1)(u-v_2) \cdots (u-v_m)}{u^m} \right)^N \end{aligned} \tag{4}$$

where the value of variable  $u$  and  $v$  is related to  $j, k$  and  $r$ ; variable  $m$  is bound up with  $j, k$ , and the coverage area of base station.

Let  $\eta$  be the minimum expectation of encounter probability, which is  $P_N^m(A) \geq \eta, \eta \in [0,1]$ . Then the lower bound of required sensor nodes  $N$  is:

$$N \geq \log \left( \frac{(u-v_1)(u-v_2) \cdots (u-v_m)}{u^m} \right) (1 - \eta) \tag{5}$$

Although the formula (5) is referential, it can help us estimating the number of sensor nodes we need before an actual experiment to insure the nodes will come across a base station.

## 5 Conclusions

In our developing system, there are two kinds of nodes. One is sensor nodes, which is used to collect data about the blown sand environment and the movement law of sand. The other is base stations, which is prepared for receiving data forwarded from sensor nodes. Considering the cost and the efficiency of the system, it is necessary to estimate the required number of nodes. Through some assumption and modeling simplification, we give a formula to compute the lower bound of node number under a simple condition at last. However, our theoretical analysis may not be precise enough due to too many assumption and plain model. Science the real world is very intricate and the node's movement is acted on various factors. The analysis in this paper is just a reference for the implementation of the actual work and need to be ulteriorly validated through practices.

**Acknowledgments.** This work was supported by the Natural Science Foundation of P. R. of China (61073193), the Key Science and Technology Foundation of Gansu Province (1102FKDA010).

## References

1. Pelusi, L., Passarella, A., Conti, M.: Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *Communication Magazine* 44, 134–141 (2006)
2. Hoffmann, F., Kaufmann, M., Kriegel, K.: The Art Gallery theorem for polygons with holes. In: *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, pp. 39–48 (1991)
3. Shakkottai, S., Srikant, R., Shroff, N.: Unreliable sensor grids: coverage, connectivity and diameter. In: *Proceedings of Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM 2003)*, pp. 1073–1083 (2003)
4. Hall, P.: *Introduction to the Theory of Coverage Processes*. John Wiley & Sons Inc., New York (1988)
5. Adlakha, S., Srivastava, M.: Critical density thresholds for coverage in wireless sensor networks. In: *Proceedings of the IEEE Wireless Communications and Networking Conference*, pp. 1615–1620 (2003)
6. Wang, X., Xing, G., Zhang, Y., Lu, C., Pless, R., Gill, C.: Integrated coverage and connectivity configuration in wireless sensor networks. In: *Proceedings of the First International ACM Conference on Embedded Networked Sensor Systems*, pp. 28–39 (2003)
7. Zhang, H., Hou, J.C.: *Maintaining sensing coverage and connectivity in large sensor networks*. In: *Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless and Peer-to-Peer Networks*. CRC Press (2004)
8. Onur, E., Ersoy, C., Delic, H.: How many sensors for an acceptable breach detection probability? *Comput. Commun.* 29, 173–182 (2006)
9. Liu, B., Brass, P., Dousse, O., Nain, P., Towsley, D.: Mobility improves coverage of sensor networks. In: *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 300–308 (2005)
10. Liu, B., Towsley, D.: A study of the coverage of large-scale sensor networks. In: *IEEE International Conference on Mobile Ad-Hoc and Sensor Systems*, Piscataway, pp. 475–483 (2004)
11. Yang, Z., Yong, Y., He, J.: Determining The Number of Nodes for Wireless Sensor Networks. In: *Proceedings of the 6th IEEE Annual Symposium on Emerging Technologies: Mobile and Wireless Comm.*, pp. 501–504 (2004)
12. Liu, C., Wu, K., Xiao, Y., Sun, B.: Random coverage with guaranteed connectivity: joint scheduling for wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* 17, 562–575 (2006)
13. Jiang, J., Li, F., Wen, J., Wu, G., Zhang, H.Y.: Random scheduling for wireless sensor networks. In: *IEEE International Symposium on Parallel and Distributed Processing with Applications*, pp. 324–332 (2009)
14. Fan, G., Wang, R., Huang, H., Sun, L., Sha, C.: Coverage-Guaranteed Sensor Node Deployment Strategies for Wireless Sensor Networks. *Journal of Sensors* 10, 2064–2087 (2010)

# Potential Attacks against k-Anonymity on LBS and Solutions for Defending the Attacks

Pan Juncheng<sup>\*</sup>, Deng Huimin, Song Yinghui, and Li Dong

School of Computer Science and Technology, Harbin Institute of Technology  
Harbin, China

{jchengpan, hmindeng}@126.com,  
{yhsong, lee}@hit.edu.cn

**Abstract.** Widespread using of mobile positioning devices makes location based service (LBS) more and more popular. Since LBS need users' current location and some of users' personal interest as input, it would incur some privacy related issues about the users. One important and comparatively effective method to protect users' privacy in LBS is spatial cloaking based on k-anonymity, however there are some inherent drawbacks of traditional k-anonymity techniques in protecting users' privacy in LBS. In this paper, we analysis some security attacks that utilized these drawbacks to encroach users' privacy in LBS, and then we proposed some novel methods to defend these attacks. At the end, some suggestions for constructing a security scheme to protect the users' privacy in using LBS are given.

**Keywords.** LBS; privacy, k-anonymity, continuous queries, distribution inference, K-disaccord.

## 1 Introduction

Recent years, there have been a tremendous progress in sensor and wireless technology, and due to this progress, many kinds of wireless computing and technology related applications spring up like mushroom[1,2]. One kind of the mobile computing based application called Location Based Service (LBS), which mainly rely on the wireless positioning technology, provides its users useful geospatial information.

Typical applications of LBS are navigation and finding k-nearest Point of Interesting (POI) e.g., restaurant, stores, and entertainment place. However, LBS need the user's current location and some information about user's personal interest as input so that user's location related privacy would much likely be exposed and then be illegally utilized.

Hence, protecting users' privacy when they use LBS is critical. Many existing approaches to protect LBS users' privacy is based on spatial cloaking techniques [3,4,5] to anonymize the user's accurate location.

---

<sup>\*</sup> Corresponding author.



Figure 1 demonstrates how spatial cloaking work. The main idea of spatial cloaking is to enlarges the accurate location of query source to a cloaked region.

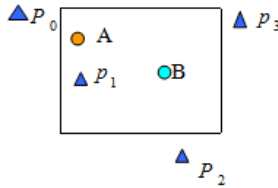


Fig. 1. Spatial cloaking

In this paper we discuss some technologies that protecting the LBS users’ privacy based on spatial k-anonymity, an effective and extensively used spatial cloaking concept.

k-anonymity [9] is first introduced in relational database, where it is used to prevent some records from being related to specific person. Now with new privacy issues arising in spatial query area, we adopt the k-anonymity concept.

Existing architecture of privacy preserving LBS system is illustrated by figure 2. The architecture consists of three main components: User(who issue queries), location anonymizer, LBS server.

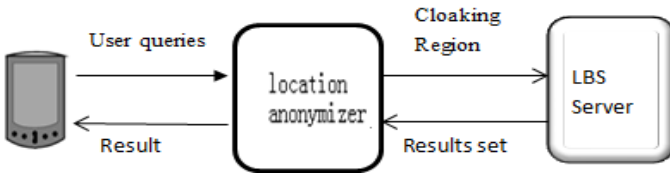


Fig. 2. Architecture of privacy preserving LBS system

The figure 1 show an example of results set (p0,p1,p2,p3) return by LBS, in this example, the K =2, the cloaking region contains two users A,B, and A is the user who forward the query. p0, p1, p2, p3 are candidate POIs for the cloaking region, and p1 is the point that nearest to user A.

Existing technologies [3,4,5,6,7] that realize the K-anonymity in spatial temporal context mainly focus on how to minimize the cloaking area for given K so that minimize the workload of LBS to compute candidate results. Although there are some work [8,11] consider potential attacks on existing k-anonymity technology, they all lack a comprehensive describing and analysis of potential attacks under different context.

In this paper, we will describe and analyze kinds of attacks that utilize the drawbacks of existing k-anonymity technology in different query context. After that, we propose some solutions to remedy these drawbacks.

The rest of the paper is organized as follow: Section 2 reviews some related work. Section 3 introduces and analyzes some privacy attack models on existing k-anonymity technologies. Based on the analysis of the security attacks introduced by section 3, we discuss some solutions to defend the attacks in section 4. Future work and conclusion are presented in section 5.

## 2 Related Work

In this section, we review some cloaking algorithms and some attacks based on these algorithms.

### 2.1 Spatial Cloaking Algorithm

In this subsection we are introducing two cloaking techniques: Interval cloak [3] and Hilbert cloak [5].

Interval cloak is the first cloaking algorithm, which is based on quadtrees [12]. The location anonymizer maintain users set which are indexed by quadtrees. When the anonymizer receive a query and a privacy requirement  $k$  from user A, it traverse the quadree up to down until find a quadrant that contains A and fewer than  $k-1$  other users. Then it chooses the parent of the quadrant as the  $k$ -ASR [5].

Hilbert cloak is a transformed based cloaking technique [5,6] which transform a two dimensional space into a one dimensional encoded space. The main principle is that the algorithm first splits the space into a many sufficient small subspaces, and then present each of the subspace by a node, after that algorithm lists all nodes as a sequence in an encoded order.

### 2.2 Attacks on Snapshot Query

In snapshot query scenario, existing work present two main attacks [5] on existing cloaking algorithms. One is distribution inference [5], the other is outlier.

The *distribution inference* [5]depict such a scenario: in a very short time spell there are two queries issued by different user A,B, they both have the same privacy requirement  $K$ , and the cloaking region generate for A's query include B, however the cloaking area generate for B is not contain A.

The figure 3 illustrates this case. In this case, the attacker would have a probability higher than  $1/k$  (this case  $k$  is 4) to ascertain that the first query is issued by A, Panos and Gabriel [5] give the detailed analysis.

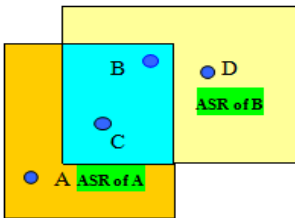


Fig. 3. Distribution inference

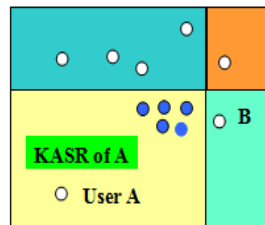


Fig. 4. Outlier distribution

The outlier attack is that, in order to satisfied privacy requirement, the cloaking region that formed by user A's query is so large due to there are few user around her, and at the same time the distribution of users in this area are so nonuniform that the

attacks can ascertain A who issued the query with probability higher than  $1/K$ . In the figure 4, user A is the outlier.

### 2.3 Attacks on Continuous Query

Lin Xin[11] and Aniket[8] give analysis of some privacy attacks against the *clique cloaking*[11] algorithm under continuous query context. The main idea is that queries happen different places by same user may have some preference linkage. This linkage can be explored by historical snapshot, which would exposure the user who issue the query with probability higher than  $1/k$  ( $k$  is the anonymity requirement).

## 3 Privacy Attack Model against K-anonymity

### 3.1 Snap Linkage Attack

**Definition 1.** *snapshot* can be defined as following data structure:

```
Struct
{ int k; //anonymous requirement, usually encoded;
  Quadrant location; //K-ASR that contain user
  Info information;
  int user_id[k]; //encoded user id
}snapshot
```

Consider such a scenario:

user A issue  $N$  queries in one day, the anonymizer generate a set of snapshots for each of the queries. Consider one of the snapshots  $snapshot_i$  contains sensitive information to attacks, and now attacks begin to record all  $k$   $user\_id$  from  $snapshot_i.user\_id[k]$ , and then they begin to process snap linkage attack. The attack algorithm as figure 5:

---

#### Algorithm. Snap\_linkage

---

```
Snap_linkage(int i, snapshot snapshot[N], &int user_id[k])
1. for j=0 —>k-1
2.   user_id[j]=snapshot[i].user_id[j]
3.   float potential[j]=1/k;
4. for j=0 —>N
5.   for K=0—>k-1
6.     if j≠i
7.       m=|snapshot[j] snapshot[K]|
8.       for all users in |snapshot[j] snapshot[K]|
9.         n=user.id
10.        Potential[n]=(1—(1—potential[n])×(1-m/k))*1/m
11. Return potential;
```

---

Fig. 5. Snap\_linkage Algorithm

The ‘potential[j]’ in the figure 5 means the probability of user<sub>j</sub> to be the request issuer.

**Lemma 1:** if user<sub>j</sub> were found in a new snapshot, the potential[j] can be recalculated as:

$$\text{potential}[j] = (1 - (1 - \text{potential}[j]) \times (1 - m/k)) * 1/m \quad (1)$$

The m in the (1) means the number of common users between the new snapshot and the first snapshot that is examined.

*Proof:* the best anonymizer would makes the K users in the cloaking region undistinguishable, so that we think that these K users form a uniform distribution in terms of the probability of being the request issuer. So (1-potential[j]) means in the last queries, the probability of A being not the issuer. This time due to the anonymizer, every user have the same probability of being issuer again, so after this new snapshot, the potential of user<sub>j</sub> can be roughly refreshed and calculated by (1).

From the lemma 1, we can see that with the number of snapshots that contains user<sub>j</sub> accumulated, the probability of user<sub>j</sub> being the request issuer increased.

### 3.2 K-disaccordance Attack

Consider such a scenario, at time t<sub>1</sub>, User Alice issue a query, the anonymizer return a cloaking region Q<sub>1</sub> for the query, the anonymous requirement K for the Q<sub>1</sub> is k<sub>1</sub>, Bob located in Q<sub>1</sub> at time t<sub>1</sub>. Then at time t<sub>2</sub>, Bob issues a same query, his anonymous requirement is K<sub>2</sub>, and this time Alice is included in the Bob’s cloaking region Q<sub>2</sub>. Assuming that K<sub>1</sub> > K<sub>2</sub>, this would give the attacker a message that the probability of first query issued by A is at least 1/k<sub>2</sub> > 1/k<sub>1</sub>.

## 4 Design of Anonymizer

In this section, we will provide a privacy enhanced scheme for anonymizer.

### 4.1 Assumptions

First, we assume that anonymizer is a trustful server while the LBS is not a secure server. There is a secure link between user and anonymizer. However the attacker can solicit information about users from the LBS server. Second, we don’t exclude the possibilities that attacks synthesis the continuous snapshots to analyze the user’s behavior feature. We also don’t exclude the possibilities that the same user may issue same requests in different time and/or in different place. Thirdly, we allow the K value can be known by attacks, Last, we assume that the attacker know the location of every user.

### 4.2 A Privacy Enhanced Scheme for Anonymizer

In this subsection we offer an overall anonymizing scheme based on 4.1.

Before offer the scheme, we give some essential definition.

**Definition 2**

**PKC** : Permanent k-user clique, it is formed by k users, k is the user customized parameter. In the scheme, all members of PKC are behaviorally bound together.

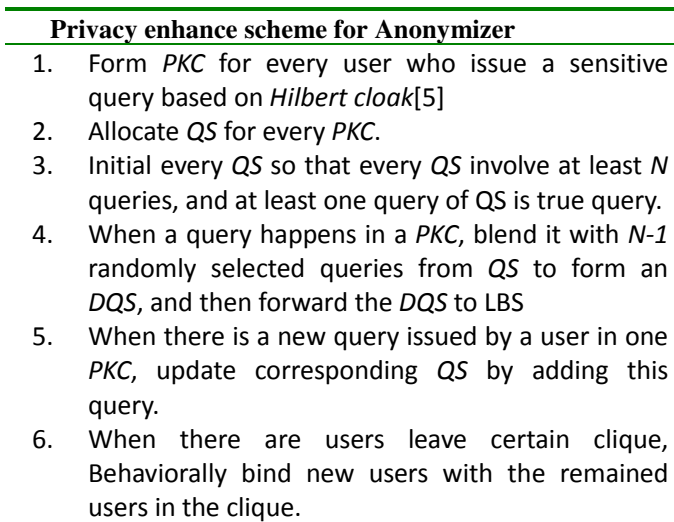
**DQS**: Dummy query set, in the scheme, anonymizer forwards DQS to LBS instead of single query, and only one query of the DQS is the true query.

**Behaviorally Bind**: If there happens same queries in different snapshots and the queries are issued from same clique, the DQS forwarded by anonymizer would be same.

**QS**: queries set, the anonymizer allocate a QS for every PKC. The dummy queries are randomly selected from QS.

**N**: the number of queries in DQS.

The privacy enhance anonymize scheme is described as figure 6



**Fig. 6.** Privacy enhance scheme for Anonymizer

The scheme above, we think can alleviate extent of privacy encroached by kinds of attacks mentions in this paper. Firstly, the PKC is mainly used to attenuate the strength of Distribution inference attacks on user's privacy, beside that it can check some influence of outlier attacks. More important, it offers an solution for defending K-disaccord attacks. Secondly, the DQS is mainly used to defend outlier attack, which can also further lower the probability of attackers pinpoint the issuer. In other word, DQS is a remedy for k-anonymity.

Combine PKC with DQS together can get an enhance privacy scheme, however it may incurs exceptional computational overhead, because they are based on Hilbert cloaking [5] and dummy query [9] technique. But the work in [5,9] do not mention how these two techniques be combined to defend the novel attacks model we mention in this paper.

## 5 Conclusion and Future Work

In this paper, we give a detailed introduction about privacy issues that exist in Location based service system in section 1. Then we discuss some existing techniques that can solve the privacy attacks in LBS system based on some existing privacy attack models in section 2. After that, we propose some novel and potential security attack models in section 3. we then provide a deep analysis of these attacks. Some suggestions for designing of secure anonymizer based on our analysis is given in section 4.

Due to the limited length of the paper, we have to leave details of designing the anonymizer to our future work. Besides, in this paper, we focus on theoretical analysis of privacy attacks in LBS, our future work would shift our focus to experimental work, which is based on the theories and methods we have established in this paper.

**Acknowledgment.** This research was partially supported by the National Basic Research Program of China (973 Program) under grant No. 2011CB302605, the National High Technolgy Research and Development Program of China (863 Program) under grants No. 2011AA010705, the National Key Technology R&D Program of China under grant No. 2012BAH37B01, the National Science Foundation of China (NSF) under grants No. 61173144 and No. 61073194.

## References

1. Culler, D., Estrin, D., Srivastava, M.: Overview of Sensor Networks. *IEEE Computer* (August 2004)
2. Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D., Anderson, J.: Wireless Sensor Networks for Habitat Monitoring. In: *WSNA 2002*, Atlanta, Georgia (September 2002)
3. Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: *MobiSys* (2003)
4. Dewri, R., Ray, I., Ray, I., Whitley, D.: Query m-invariance: Preventing query disclosures in continuous location-based services. In: *MDM* (2010)
5. Kalnis, P., Ghinita, G., Mouratidis, K., Papadias, D.: Preventing location-based identity inference in anonymous spatial queries. *TKDE* (2007)
6. Moon, B., Jagadish, H., Faloutsos, C.: Analysis of the Clustering Properties of the Hilbert Space-Filling Curve. *IEEE TKDE* 12(1), 124–141 (2001)
7. Mokbel, M.F., Chow, C.Y., Aref, W.G.: The New Casper: Query Processing for Location Services without Compromising Privacy. In: *Proc. of VLDB*, pp. 763–774 (2006)
8. Pingley, A., Zhang, N., Fu, X., Choi, H., Subramaniam, S., Zhao, W.: Protection of query privacy for continuous location based services. In: *2011 Proceedings of the INFOCOM*, pp. 1710–1718. *IEEE* (2011)
9. Samarati, P.: Protecting respondent's privacy in microdata release. *IEEE Transactions on Knowledge and Data Engineering* 12(6), 1010–1027 (2001)
10. Samet, H.: *The Design and Analysis of Spatial Data Structures*. Addison-Wesley (1990)
11. Lin, X., Li, S.P., Yang, Z.H.: Attacking algorithms against continuous queries in LBS and anonymity measurement. *Journal of Software* 20(4), 1058–1068 (2009)

# To Practice Management Innovation and Reinforce the Foundation of Intellectual Property Strategy

Wen-Liang Sun<sup>1</sup>, Yong-Fei Ma<sup>2</sup>, Li-Wen Chen<sup>3</sup>, and Shang Yu<sup>4</sup>

<sup>1</sup> School of Economics and Management, Hebei University of Technology,  
Tianjin, 300401, China  
k27277@sina.com

<sup>2</sup> School of Economics and Management,  
Hebei University of Technology, Tianjin, 300401, China  
81254464@qq.com

<sup>3</sup> School of Economics and Management,  
Hebei University of Technology, Tianjin, 300401, China  
lwchen@hebut.edu.cn

<sup>4</sup> Nankai University, Tianjin 300071, China  
810996627@qq.com

**Abstract.** Intellectual property (hereafter referred to as IP) has become increasingly important in comprehensive national strength of a country. Innovative and effective IP management can greatly improve the creation, protection and operation efficiency of IP. Besides expounding the meaning of innovative IP and IP management, this article aims to identify the current problems in the IP management and propose appropriate innovative strategies so as to promote the coordination and smooth movement of the work of IP and accelerate the pace of construction of an innovation-oriented country.

**Keywords:** Innovation, IP, IP management.

## 1 The Original Source and Connotation of Innovation

Among the English words list, the noun form of innovation is “innovation”, while the verb form is “innovate”. “Innovate” is originated from “innovare” in Latin, which has three meanings – to renew, create something new and change. Innovation acts as a means of using the existing natural resources to create new things.

Innovation Theory[1] is first proposed by Joseph Alois Schumpeter, an Austrian-American economist, to explain the theory of capitalist economy development and its periodic process. However, with enterprises as the main object, his theory is to measure the basic interaction mechanisms between technology and economy mainly from the economic point of view. In his own view, innovation is a new combination of production factors made by the entrepreneurs; it belongs to the economic category rather than technology; instead of inventions of science and technology, it refers to introduce the science and technology ever invented into enterprises, forming a new production capacity. This is the definition on innovation made earlier by this

economist. Nevertheless, he did not make a comprehensive explanation and strict definition on the series of innovation theories. As time goes by, Schumpeter's followers has made some more researches on innovation theory, so that this theory has been further enriched and developed.

In the modern times, the research on innovation theory has made a remarkable breakthrough. In his book - Knowledge Management Science, Professor Qiu Junping points out that knowledge innovation is a process of acquiring knowledge on new basic science and science of technology from scientific researches, which is aimed at the pursuit of new discovery, exploration of new rules, creation of neodoxy as well as accumulation of new knowledge. Innovation can contain various ways, such as innovation in thoughts, theory innovation, technological innovation, process innovation, institutional innovation and management innovation, etc.

## 2 IP and IP Management

The primary meaning of IP is "intellectual (property) ownership" or "wisdom (property) ownership", which is also known as "IP right". Before the mid-1990s, most of Chinese scholars in law circle defined IP as the exclusive right legally enjoyed by people who had made creative intellectual achievements.[2] After the mid-1990s, China's academia of IP provided a new summary of the concept of IP. Although there are some differences among the major points, all of them reflect the characteristics of the concept of IP.

**2.1** Different from the special right of traditional ownership, IP is an immaterial property generated in the realm of spirit: the rights generated based on intellectual achievements, operating marks or knowledge information.

**2.2** IP is not equal to intellectual creative achievements right. Not all of the rights - commanded in the name of IP, come from the knowledge domain, nor are generated based on intellectual achievements. Viewed from the sources of rights, IP mainly occurs in the intellectual creation activities and business operation; from the object perspective, it is composed of its creative achievements, operating marks, reputation as well as other knowledge information.

**2.3** The generation of intellectual right - a statutory right, is generally approved by law.[3] IP mainly consists of copyright and its adjacent right, ownership of trademark, geographical indication rights, industrial design right, patents, design right of integrated circuit layout, exclusive right of information which is never published, etc.

The so-called IP management is a comprehensive activity, under certain circumstances, in order to achieve the best goals of the organization, making effective plans, organizing, leading and controlling the IP resources which can be allocated by the organization for the sake of high-efficiency operation. The basic contents of IP management include information query, resource allocation, development of outcomes, archiving of outcomes, the acquisition of rights, asset operations, and the protection of outcomes. The core of IP is to encourage innovation, and respect for labor, knowledge and talents. IP management is a systems engineering of a series of management behaviors, such as the strategy formulation of IP, system design, process monitoring, application and implementation, personnel training, innovative



integration and the like. IP management not only together with the creation, constitutes the main content of Chinese IP system and its operation, protection and utilization of IP, but also runs through all aspects of the creation, protection and utilization of IP. From the perspective of the national macro-management, the system legislation, judicial protection, administrative licensing, administrative law enforcement and policy making of IP can also be included in the content of the macro-management of IP; viewed from the business management, the generation, implementation and right protection of IP of enterprises are inseparable from the effective IP management.

### **3 The Important Value of Innovation of IP Management**

Management is defined as the process of coordinating work activities so that they are completed efficiently and effectively with and through other people.[4] What matters to the innovations of Management is that any areas are in need of creativity consciousness, innovative concepts, strategies and execution. Innovations are required by any discipline and knowledge, including IP management. The innovative management of IP is to abandon the outmoded system in the field of IP management, and actively seek for reasonable, effective model of innovation management. Only if a system is sound can the efficiency of management be improved, and then the protection of IP will be easy to conduct, which is an ongoing process.

IP management is greatly valuable and plays a fundamental and overall important role in the whole IP work. Firstly, effective IP management can not only help to enhance the creativity consciousness of the subject and the consciousness of IP, but also contribute to making innovation, research and development with external forces, so that the innovation ability of the subject is strengthened. Secondly, effective IP management is conducive to the establishment of infringement prevention mechanism of IP and rapid response mechanism while there is infringement, thereby enhancing the protection ability of IP of the creative subject. Thirdly, effective IP management can promote subject's operational ability of IP as well as the abilities coping with IP disputes and related matters. Finally, effective IP management conduces to the collocation and coordination of all functional departments of the subject, so that the organization and coordination ability of creative subject will be improved.[5]

## **4 The Major Problems in China's IP Management**

### **4.1 Superficial Consciousness of IP**

As IP, an intangible asset is not so intuitive as houses, cars and gold, silver and jewelry, people are short of sufficient knowledge of its value. Many companies also embark their money, manpower and materials in industry which has short-term returns, and consider the IP as a distant and even dispensable thing. The insufficient understanding, particularly management layer's lack of awareness, results in that IP is left out in cold which is undeserved. For instance, many local leaders only emphasize the development of industry which can achieve quick results, because these achievements can contribute to their promotion. However, the returns of IP have its

own periodicity. Some cycle is very long. Moreover, as IP is not linked to the performance of the leadership, many local leaders do not attach importance to the protection and management of IP.

## 4.2 The Scarcity of IP Talents

China's scarcity of IP talents has become a bottleneck of the development of IP. Many domestic law teachers (whose research direction is IP, especially patent law), have no background of science and engineering. Most of them graded from doing research of civil and commercial law to that of IP. However, people can not carry out in-depth study without support by science knowledge. Furthermore, IP talents, providing intermediary services, are so sparse. Take patent agents for instance, according to statistics, 76% of our patent applications are accomplished by patent agency and the agents. In 2009, State IP Office accepted more than 970,000 pieces of patent cases, while there were only a few thousand patent agents in the whole country. Such scarcity of patent services talents has greatly hindered the normal development of China's patent industry. Furthermore, most persons are forced to engage in IP. Short of professional and systematic training, they are relatively poor in overall quality. As a result of lacking intelligence support for talents, Chinese enterprises always encounter the barrier of IP in the international competition, which makes us deeply distressed. Sometimes they are even sued wastefully.

## 4.3 Shortage of High and New Technology with Core Competitiveness

IP has been developing greatly in China, while patent applications have sharply increasing each year. Moreover, the realm of copyright also takes on an air of prosperity, while the value of trademark has also been entertained with the attention of enterprises. However, the innovation standard of China's IP is still very low and what is more, there are even fewer technologies with core competitiveness. When Professor Wu Handong, headmaster of Zhongnan University of Economics and Law and president of IP Rights Seminars of China Law Society, gave an academic report in East China University of Political Science and Law, he made a brilliant summary on this fact, "We are less competent in the realm of patent, not amazing in the realm of trademark and not gratifying in the field of copyright." Naturally, part of the reason that create this state of affairs should be attributed to the looser authorization standard, which was due to that innovations were encouraged in the embryonic stage of China's IP system.

# 5 Creative Measures of China's IP Management

According to the strategic goals of The National IP Strategy Outline, by 2020, China will become a country with a comparatively high level in terms of the creation, utilization, protection and administration of IPRs. The legal environment for IPRs is much better, market entities are much better at the creation, utilization, protection and administration of IPRs; the public awareness of IP is increased greatly; the quality and quantity of the self-relied IPRs are able to effectively support the effort to make China an innovative country; the role of the IP system in promoting economic development, the culture prosperity and social progress in China become very apparent. Based on

the Outline, in allusion to the problems existing in the IP management, the following measures can be taken to promote the innovations of IP management after combined with the characteristics of IP.

### **5.1 Carry Out Publicity and Education of Innovations, and Enhance the Awareness of IP and Strengthen Cultural Development of IP**

Rousseau, a French ideologist, referred to customs, traditions and public opinions by the name of “the forth law” after constitution, civil law and the criminal law. In his view, customs, traditions and public opinions were the real constitution, giving people endless strength all the time. This is, in essence, a kind of awareness on mind, the regulation power of which far exceeds any law, because it is a consciousness of maintaining order purposefully. IP and its protection also need this awareness. This consciousness, once formed, will flow like a fountain when needed. Judging from the concept of consciousness, it is an advanced and ordered organization form of materials, the sum of features that live beings can perceive by means of their physical perception system, and related perception processing activities. If there are more processing activities and the harder publicity for the consciousness subject, this consciousness will be more intense, which is conducive to the formation of the inertia of thinking.

It turns out that this consciousness is quite important. Therefore, every conceivable means should be used to carry out publicity and education of innovations. The obsolete way of publicity thinking, working system and forms of publicity and education should be got rid of, so that people will not feel rejected and fatigue. The broad masses begin to appreciate and accept the publicity and education of IP at their options in a relaxed joyful environment, rather than just learning and forced to watch. Publicity and education of innovations can be promoted from the following aspects.

#### **(1) Innovative ways of publicity thinking**

At the work of publicity and education, people should be open-minded, tricky and adept in understanding of IP knowledge. Grasp of the publicity features of IP can make the publicity and education of IP reflect the working characteristics of the new period.

#### **(2) Innovative publicity working system**

People should be active in exploring effective management mechanism of publicity team, and establish rules and regulations. The guiding theory, focal point of propaganda and the examining and incentive methods of the publicity of IP must be specific. The enthusiasm of publicists should be stimulated. The construction of propaganda cultural team at the grassroots level should be emphasized, so that the propaganda work at the grassroots level can be promoted deeply, enduringly and effectively.

#### **(3) Innovative ways of publicity and education**

Publicity and education can be carried out with multi-angle through many channels, at many levels by virtue of various ways, such as documentaries, feature films, art films, reportage, novels, and theatrical performances, etc. For instance, some propaganda team goes deep into the farmer's market in countryside to do propaganda. In order to enhance the effectiveness of publicity and facilitate the smooth progress of IP work, the team carries out publicity and education through various ways, for example, posting the promotion advertising of IP, accepting the consultation of the masses and

distributing propaganda materials. Whether a company and enterprise attached importance to the protection and management of IP of scientific and technology innovations, to a great extent, depends on the concern of the leading cadre over it. Therefore, at the work, the publicity and education of the leading cadre who have the decision-making power should be intensified.

The consciousness and culture construction of IP is the foundation and bedrock. Only when the foundation has been laid can form a good environment of IP career. Meanwhile, other work of IP can proceed smoothly. IP culture must include the substantive characteristics of IP mainstream culture jointly initiated by the international community, as well as the cultural elements with Chinese characteristics that the developed countries do not have any. Thus, IP culture with Chinese characteristics that we are advocating and nurturing is a sub-cultural form of the IP culture of the international mainstream, which absolutely is not self-closed, mutable, exclusive and narrow, but an open, progressive, inclusive and generalized culture. It reflects many aspects of advanced culture with broader connotations.[6]

## **5.2 Conduct Innovative Management and Training of IP Talents, Especially Senior Talent Mode**

The talented person is critical in the implementation of IP strategy. Besides some background in science and engineering, the ideal IP professionals should be equipped with solid legal foundation, and be master of certain knowledge of business administration and economics. The National IP Strategy Outline clearly pointed out that the development of IP human resources should be stimulated. Moreover, an amount of national training base of IP talents should be created, while the construction of senior IP faculty should be accelerated. The mode of innovative management and training of IP talents can be started from the following aspects.

### **(1) Introduction of talents**

Scientific research centers and interest groups should be established in universities, research institutes or in enterprises to attract talents to join in and solve IP problems they confront during communicating and learning.

### **(2) School education**

Firstly, the training approaches of talents should be market-oriented and long-term with reasonable allocation. For instance, the translation talents of IP is now urgently needed, so the addition of related Translation major at the undergraduate or graduate stage can be taken into consideration. Secondly, universal education of IP basic knowledge should be strengthened for primary and secondary school students. Japan, for example, has taken distinguishing measures of IP education for students in different ages. IP teaching materials and reference books are edited and distributed freely to the primary and secondary schools.

### **(3) Training of talents**

Training of IP personnel should be organized in state organs and enterprise and public institution, which must be innovative. For instance, some areas have sent personnel who have some knowledge of IP to the countries where IP have carried out well, like the U.S. and Japan. In this way, on the one hand, they can receive professional training and re-education; on the other hand, their foreign language abilities will be improved.

#### (4) Talent motivation

China's relevant laws have reflected this aspect. Patent law, for instance, ordains that the company should reward inventors who have made a service invention, which can be regarded as a measure encouraging innovations. Nevertheless, companies can expand the mind to find better incentive policies, such as, retaining and motivating talents by virtue of giving them the stock options.

As to the key link of IP talents, the companies should use all wits, open thinking, and strive to create something new and original to follow up the development of talents, when formulating any principles and policies. By means of the above measures, a productive system with orderly operation is formed in attracting, employing, culturing and motivating of talents. Innovation achievements are constantly emerging, and the high and new technologies with core competitiveness can be developed only when the key problem of talents is solved.

### **5.3 Innovate and Improve Management Systems and Coordination Mechanisms of IP**

As is well known, China's IP system is imported from western countries. The basic system and development phase of the western countries are different from that of China. The systems which are borrowed, to a certain degree, will be disagreeing with China's system. Therefore, even though China's current IP legal system is relatively complete, it still needs to be constantly improved and innovated in combination with China's actual conditions.

Comparatively speaking, China's management systems and coordination mechanisms of IP are even more lacking in the vitality of innovation. IP management, in effect, is to allocate and organize all aspects of IP resources to achieve maximum operational efficiency. Hence as a "transmission shaft", management and coordination mechanism plays a crucial role in providing IP information and improving the efficiency of IP coordination. Coordination platforms can be built for the innovations work of IP management and coordination mechanism; IP service platforms are established through the internet and IP service departments to provide information for enterprises and IP talents. For example, Shanghai Intellectual Property Service Center provides an interaction platform of information for enterprises and job seekers. Job seekers can send their resume to the service center. If there is suitable company, the service center will give recommendations, so that the efficiency of recruitment will be improved for the enterprises, while job seekers will also benefit from the information acquisition. From the above it can be seen that IP management is a systematic project, its effective operation depends on the efficient IP management systems and coordination mechanisms. Therefore, the research on effective coordination of IP in different links should be attached more importance to and deserve more time and energy for innovation.

IP management is an art with a blend of Technique, Law, Management, Economics, and can not achieve a very high standard in one single day. Specialization of work, as well as overall consideration, is necessary. There should be talents to do specific work, and management and coordination mechanisms to closely connect every links of the work. Innovative consciousness and thinking mode are needed, while innovation concepts should be learned to use in IP management. However, not everyone has a good sense of innovations. In IP management, people not only should know and understand innovation, but also can take advantage of innovations.

Moreover, they should make flexible use of it, and figure out a way out. In brief, management should be promoted via innovations; the integration of IP and management should be carried forward; IP management education should be strengthened to accelerate the constriction pace of an innovative country.

## References

1. Yong, S., et al.: *The Industrialization of Technical Innovations and Scientific Achievements in the World*. Scientific and Technical Documentation Press, 1 (August 1999)
2. Chengsi, Z.: *Intellectual Property Rights Course*. Law Press China (1993)
3. Qinnan, H.: *A New Intellectual Property Rights Course*. China University of Political Science and Law Press (1995)
4. Wu, H.: *Intellectual Property Rights*. Law Press China, 5 (2007)
5. Robbins, S.P., Coulter, M.: *Management*. Translated by Jianmin, S., Weiwei, H., et al., 7th edn. China Renmin University Press (2004)
6. Zhu, X., Qiao, Y.: *Intellectual Property Management*. Higher Education Press 14 (2010)
7. Haiqun, M., Li, W., Lixia, Z.: *Modern Intellectual Property Management*. Science Press 127 (2009)

# Research of Intranet Security Audit in E-government Management Website Group Based on Multi-agents

Han Bing and Wang Bo

School of management, Tianjin University, China  
z1hzhh@126.com

**Abstract.** E-government website possesses the abilities of government functions so it often receives all kinds of attacks from outward and inward. These attacks can be classified as active and passive aggressions. Active aggressions can modify, fake, destruct and publish viruses which make the network cannot below a security risk. With the appearance of new risk, new loopholes, new attack technologies the safety management is facing growing challenges. In this situation safety management should consider the security configuration, safe operation, and emergency response and security audit problems. Especially intranet security audit is a key technology. In this paper an intranet security audit technology in e-government management website group based on multi-agents is designed and realized.

**Keywords:** E-government management, Multi-agents, Security audit.

## 1 Overview of Security Audit

Security audit is audit, inspection and calculation for system security which is by using technology means to record the event on the internet continuously and then assure the system safety by the way of tracing after the event [1].

In international and domestic norms security audit is seen as an important position. In CC or TCSEC norms audit mechanism should be essential security mechanism for the computer systems above C2 or above C2 levels. In our Chinese computer information system security protection grading criteria security is defined as five levels and after the second level basic audit function is needed. Safety criterion requirement of CC standard defines as much as H safety functions which includes Security audit class [2].

## 2 Security Audit of E-government Management Website Group

### 2.1 Safety Challenges of E-government Management Website Group

E-government management website group will finally build a comprehensive platform and it will possess the functions of public service, government office automation

and oriented decision support. So network safety within the government and for the public service is the main factor need to be taken into account [3].

By investigation with statistics some safety problems lead to data breach and data destruction. From the perspective of hostile attacks 65% is lead from inside the network system. Meanwhile some new safety risks may come from the security design itself. So security management should consider security configuration, safe operation, normal operation, emergency response and security audit problems of network systems. Website security audit of the group of e-government management is a key technology to solve all these problems [4]-[6].

## **2.2 Some Focal Point Auditing Aspects in E-government Management Website Group**

By summarizing the safety requests of e-government management website group we think we should establish information security audit system from aspects of network environment, system platform, data saving and comprehensive analysis. Record and analyze the effects of safety event on e-government management website group from different angles to make sure the security and stability of the system [7].

## **3 Related E-government Management Website Group Security Audit Technology**

### **3.1 Acquisition of the Security Audit Data Source**

Source of data in e-government management website group can be mainly divided into two parts: host data source and network data source. Host data source includes system audit records and system log but it can only detect attacks against the machine. Acquisition of the host data source is usually finished by log of the operating system.

Network data source mainly includes analysis of typical network application protocol, identify, determine and record, online chat and file sharing. Internet data is also the main source in which special data acquisition technology is used.

The capture host data and network data is stored into log database after formatted which is used as audit analysis the data source and User- defined rules to form a security audit rule base.

### **3.2 Data Preprocessing**

Data preprocessing is cleaning, integration, conversion, discrete and reduction for the initial data before the data mining to reach the lowest codes and standards for the data mining algorithms for acquiring knowledge. By data preprocessing it can perfect incomplete data, correct erroneous data, remove the excess data and eliminate redundant data attributes to make data the same, data format consistent, data Information scouring and data storage centralization.



### **3.3 Related Agent Technology**

#### **3.3.1 Concepts of Agent**

Agent is originate from artificial intelligence [8] and we define it as autonomous entities which can sense environment, judge and reason outside information and control of their decisions and actions to completed tasks. Multi-Agent Systems (MAS) is an organized system to finish collective missions which is made up of a set of Agent that logical or physical location on the distribution and connected by internet that can share resources.

#### **3.3.2 Organization Structure of MAS**

Organization structure of MAS offers a framework of the mutual interaction for Agent members and a Multi-Agent solving the problem of high-level perspective and information for each Agent members. Generally speaking organization structure of MAS means communication and control modes between different Agents. From the perspective of operational control MAS can be divided into three organization structure which are centralized, distributed and hybrid mode.

## **4 E-government Management Website Group Security Audit System Design Based on the Agent Technology**

### **4.1 Design and Implementation of the Overall System**

#### **4.1.1 Design Goals**

Based on the features of e-government management website group principles of security audit system design are progressiveness, ease of use, security, multi-functional practicality, openness and scalability.

The design goals of the system are listed as follows.

(1) Multi-point acquisition and centralization of management. If there are the violation of the rules of behavior information management control center will be alarmed and denied access measures can be used if needed.

(2) Host -based security control policy. Some special rules can be defined for specific object of the audit based on administrator, such as make some machines cannot surf on the internet during the special time.

(3) Security log audit. For the logs we can audit real-time or after the event and make responses for the results, such as producing records and alarming.

#### **4.1.2 System Logic Diagram**

According to the above design objectives and related key technologies we design the e-government management the website group security audit system which is shown in Fig.1.

According to the characteristics of the group of e-government management website security audit data and system distributed characteristics we will use each functional module security audit system that is achieved by independent Agent technology. Collaborative Agent to communicate with each other and synchronous interaction are

made to improve the security and stability of the entire system. Surveillance Agent (SA), Cooperating Agent (CA), data collection Agent (DCA), security audit Agent (SAA) and response Agent (RA) are defined in this paper which are distributed in e-government management website group to finish the network security audit function.

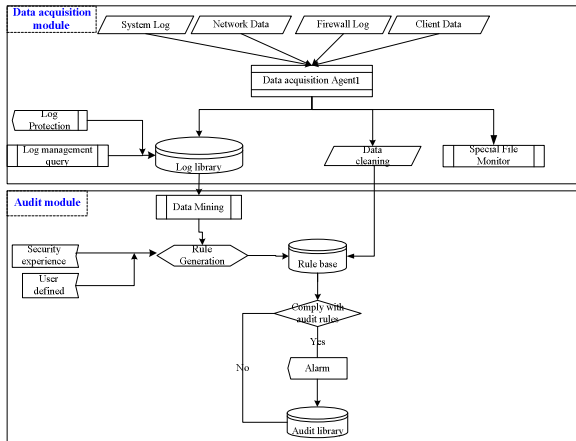


Fig. 1. System framework of e-government management website group security audit

CA keeps in touch with other Agents and makes all kinds of security audit policies which are connection between strategy and control center. SA realizes network monitoring which monitors network activities based on the security policy to detect and track intrusion. DCA can collect all kinds of audit data. SAA analyzes the audit data and produces auditing rules. RA alarm or deal with activities that are compliance with the rules of the invasion line. Logical structure between Agents is shown in Fig.2.

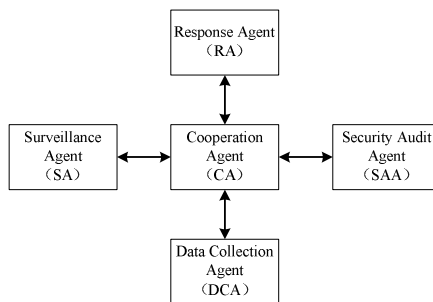


Fig. 2. Multi-Agent logical structure

### 4.1.3 Data Acquisition Module

#### (1) Real-time log collection

An event object EventLog should be created first when getting the host behavior log. Monitor the events by using EntryWrittenEventHandler functions and specified

event object is set to signal state as an event happens. Use MyOnEntryWritten functions to receive signal from events and wait until the signal arrives. Similarly by using rpPcap.PacketArrivalEvent() and device\_PcapOnPacketArrival() to realize collecting network data packets.

(2) History log collection

When security audit system starts running all the history of the day are copied and the past history log information is not recorded which ease the load on the system. Because of the production of a large number of log information at any time a buffer zone as a transit point are need to be built. First system log is stored into buffer and there will be log information inside the log integrity protection module loop detection buffer. If so make it by digital signature and release of the occupied buffer space.

4.1.4 Audit Function Design

(1) The auditing module design ideas

After corresponding log data is obtained by the e-government management website group security audit rule base is generated by using improved association rule mining algorithms and match log data with that in rule base. If log data generated by the operational behavior matches with rule base perfectly response mode will be given as defined in advance. Otherwise ignore this kind of data or judge with more reasonable rules to make sure the criticality.

(2) Security audit module framework design

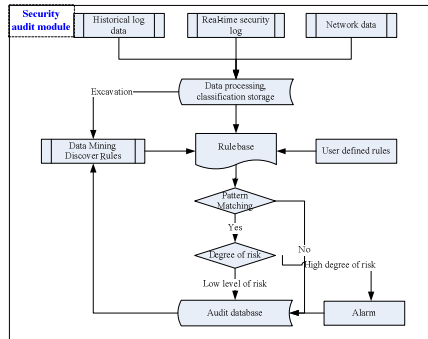


Fig. 3. Flowchart of security audit module

Security audit functions contain the following items.

- ① Complete analysis of the collected data storage
- ② Updates and maintenance of the rule
- ③ Analysis of the results of real-time response
- ④ Audit database to store the contents of the audit and the audit report output

## 4.2 Multi-Agent Based E-government Website Group Safety Audit Management

Multi-Agent Systems as e-government management website group security audit system of data collection and analysis unit has the following advantages.

### (1) Expandability

There is no need to restart to reconfigure because the start and cease is independent.

### (2) Robustness

Because we know the data dependencies between agents in advance, we can predict the failure of the produce. In view of the above analysis, the other Agent will not stop working when single and Agent failure, so that makes the whole system failure risk reduction.

### (3) Flexibility

Network packet data capture and other sources of information access can realize by Agent exploration system. So fusion to improve the accuracy of host-based and based on two methods of network security audit based on Multi-Agent, making the traditional boundaries that exist between them is broken.

## 5 Summary

An e-government management website group security audit system based on multi-Agent and data mining is researched and designed which comes from the safety audit related technology. The main work completed is listed.

(1) Concepts and elements of the security audit are elaborated. The necessity and function of the e-government management website group security audit system are researched.

(2) We focus on safety audit related technology research and solve the key issues in system design.

(3) We do research of the sources of audit information in e-government management website group from two aspects of host and network. By comparing each characteristic we establish the system audit data source.

(4) The design goals of the e-government management the website group security audit system are proposed and design the system from system logic, architecture and overall functional structure.

## References

1. Trusted Computer System Evaluation Criteria. DOD 5200.28-STD, Library No. S225, 711. NCSC (December 1985)
2. The national standard of the People's Republic of China. Computer information system security classification guidelines. GB1559-1999 (1999)
3. Arun Kumar, M., Gopal, M.: Least squares twin support vector machines for pattern classification. *Expert Systems with Applications* 36(4), 235–243 (2009)

4. Baylar, A., Hanbay, D., Batan, M.: Application of least square support vector machines in the prediction of aeration performance of plunging over fall jets from weirs. *Expert Systems with Applications* 36(4), 1068–1101 (2009)
5. Hanbay, D., Baylar, A., Batan, M.: Prediction of aeration efficiency on stepped cascades by using least square support vector machines. *Expert Systems with Applications* 36(3), 4248–4252 (2009)
6. Yu, W., Li, X.: Online fuzzy modeling with structure and parameter learning. *Expert Systems with Applications* 36(4), 111–192 (2009)
7. Xu, S., Fu, X., Hu, J.: Multi-Agent load balancing in the intrusion detection system applications. *Computer Engineering* 34(21), 111 (2008)
8. Guo, H.: Design and Implementation of Agent application process information audit software platform. Nanjing Aerospace University, Nanjing (2003)

# Simulation of Three-Phase Voltage-Source PWM Rectifier with LCL Filter

Liu Hui\*, Xu Chao, Chen Chen, and Wu Yibing

Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, China  
{liuhui, xuchao, chenchen, wuyb}@ipp.ac.cn

**Abstract.** Three-phase voltage source PWM rectifier can attenuate high frequency harmonics effectively with LCL-filter which has the advantages such as smaller inductance and less output current's THD in contrast with traditional L-filter, and suitable for medium and high power applications. In this paper, mathematical model of PWM rectifier with LCL-filter is analyzed and dual closed-loop control structure of inner current loop and outer voltage loop is designed, then the LCL-filter's parameters are analyzed and designed. At last, a simulation model of the rectifier is established; the simulation results show that the grid-side currents have lower harmonic distortions and the rectifier runs at high power factor, which verify the validity of the proposed method.

**Keywords:** PWM rectifier, LCL-filter, unity power factor.

## 1 Introduction

Voltage source Pulse Width Modulation (PWM) rectifier which has the characteristics of small grid side current harmonics, high power factor and bidirectional energy flowing, has become a research hot spot in recent years [1~3]. The traditional three-phase voltage source PWM rectifier adopts L-filter to eliminating the current harmonics, however, in medium and high power occasion where the switching frequency is relatively low(1~2kHz), in order to limit the harmonic current, inductance value of L-filter has to get larger, the problems such as large volume, high cost and slow dynamic response arise[4]. The L-filter is replaced by the LCL-filter, in the same harmonic requirement, the LCL-filter's inductance can be much less than the L-filter's, and suitable for medium and high power applications [5].

In this paper, mathematical model of PWM rectifier with LCL-filter is analyzed, and double closed-loop PI control structure based on the rectifier is designed and the PI parameters of voltage loop and current loop are calculated, and then the LCL-filter is analyzed and designed. At last, a simulation model of three-phase PWM rectifier with LCL-filter is established with Matlab/Simulink; through simulation the superiority of LCL-filter and the correctness of the proposed method are proven.

---

\* Corresponding author.

## 2 PWM Rectifier with LCL-Filter

The topology of three-phase voltage source PWM rectifier with LCL-filter is shown as figure 1.  $L_r$  is the rectifier side reactor, and  $R_r$  is its resistance;  $L_g$  is the grid side reactor, and  $R_g$  is its resistance;  $C_f$  is the filter capacitor and  $R_d$  is the damping resistor;  $s_i$ ( $i=a, b, c$ ) is the switching signal of PWM rectifier switch, when  $s_i=1$ , the switch of the upper bridge arm is on, otherwise it is off;  $C_{dc}$  is the capacitor of the DC side and Resistor  $R_L$  is the load of the rectifier.

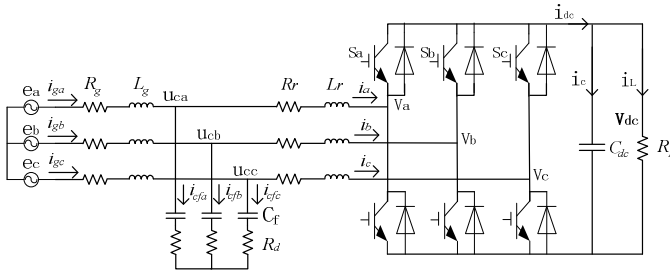


Fig. 1. Topology of three-phase PWM rectifier with LCL-filter

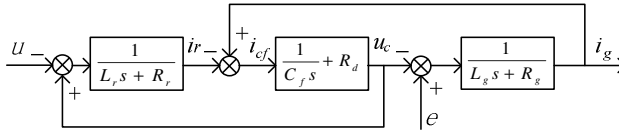
In the LCL-filter, the filter capacitor  $C_f$  is mainly to filter the high order harmonic currents. When analyzing the fundamental wave, the LCL-filter can be modeled as L-filter, by definition  $L=L_g+L_r$ ,  $R=R_g+R_r$ , then the mathematical model of the three-phase PWM rectifier with LCL-filter in two-phase synchronization coordinates( $d$ - $q$  coordinates) can be expressed as follows:

$$\begin{cases} L \frac{di_d}{dt} = e_d - Ri_d - v_d + \omega Li_q \\ L \frac{di_q}{dt} = e_q - Ri_q - v_q + \omega Li_d \\ C_{dc} \frac{dV_{dc}}{dt} = i_{dc} - i_L \end{cases} \quad (1)$$

Where,  $v_d = U_{dc}s_d$ ,  $v_q = U_{dc}s_q$ ,  $i_{dc} = 3/2(i_d s_d + i_q s_q)$ ;  $e_d$ ,  $e_q$ ,  $v_d$ ,  $v_q$ ,  $i_d$ ,  $i_q$ ,  $s_d$  and  $s_q$  are the values of grid voltage( $e_a, e_b, e_c$ ), rectifier bridge side voltage ( $v_a, v_b, v_c$ ), grid current ( $i_a, i_b, i_c$ ) and switching vector ( $s_a, s_b, s_c$ )in the  $d$ - $q$  coordinate respectively.

### 2.1 LCL-Filter

The block diagram of LCL-filter is shown in figure 2.



**Fig. 2.** Single phase LCL-filter model for PWM rectifier

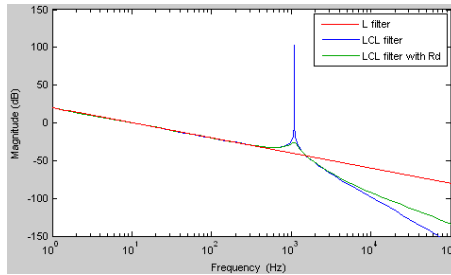
By figure 2, the mathematical models of the LCL-filter can be obtained as:

$$\begin{cases} u_c(t) + R_d \cdot i_c(t) = R_g \cdot i_g(t) + L_g \frac{di_g(t)}{dt} + e(t) \\ u(t) = u_c(t) + R_d \cdot i_c(t) + L_r \frac{di_c(t)}{dt} + R_r \cdot i_r(t) \\ i_r(t) = i_g(t) + i_c(t) \\ i_c(t) = C \frac{du_c(t)}{dt} \end{cases} \quad (2)$$

According to the formula (2), the transfer function of LCL-filter can be written as:

$$G(s) = \frac{i(s)}{u(s)} = \frac{R_d C_f s + 1}{L_r L_g C_f s^3 + (L_r + L_g) C_f R_d s^2 + (L_r + L_g) s} \quad (3)$$

The resonant frequency of the LCL filter is  $f_{res} = \sqrt{(L_g + L_r)/L_g L_r C_f} / 2\pi$ .



**Fig. 3.** Bode diagram of L-filter and LCL-filter

Bode diagram of L-filter and LCL-filter is shown in figure 3. At the low frequency band, LCL-filter can be equivalent to L-filter ( $L=L_g+L_r$ ), and the filter attenuation are all in  $-20\text{dB/dec}$ ; at the high frequency band, the LCL-filter attenuates in  $-60\text{dB/dec}$ , three times than the L-filter, thus the inductance of the LCL-filter can be much less than the L-filter's. At the resonant frequency, if there is no damping, system easily become resonant; through adding damping resistor, the resonance amplitude can be well suppressed; however, the filter attenuation also decreases. The choice of damping impedance is the balance between filter attenuation and system stability.



## 2.2 Design of LCL-Filter

The LCL filter can be designed using the following procedure.

*Rectifier side inductor  $L_r$ .* Select the required current ripples on the rectifier side to design the rectifier side inductor  $L_r$ . If the ripple current is limited in range 15% to 25% of rated current [6], the minimum inductance of  $L_r$  is obtained as:

$$\Delta i_{L_{\text{ripple}}} = \frac{V_{dc}}{8L_f f_s} \leq 25\% \cdot i_{\text{rated}} \quad (4)$$

*Filter capacitor  $C_f$ .* Select the reactive power absorbed at rated conditions to determine the capacitor value. If the maximum reactive power of filter capacitor absorbed is limited to 5% of rated power  $P_{\text{rated}}$ , the parameter's inequality of filter capacitor can be derived as:

$$C_f < 5\% \frac{P_{\text{rated}}}{3 \times 2\pi f_B v_{\text{rated}}^2} \quad (5)$$

*Grid side inductor  $L_g$ .* Select the desired current ripple attenuation rate to design the grid side inductor  $L_g$ . The  $L_g C_g$  link of LCL-filter makes current ripple further attenuated, the ripple attenuation rate is expressed as:

$$r = \frac{i_r}{i_g} = \frac{1}{|L_g C_f \omega^2 - 1|} \quad (6)$$

After selecting  $C_f$ , through substituting the corresponding  $\omega$  and scale coefficient  $r$ , the inductance  $L_g$  can be calculated. The current harmonics mainly concentrates in the vicinity of switching frequency  $f_s$ , so  $\omega = 2\pi f_s$ .

*Damping resistor  $R_d$ .* Adding the damping to avoid oscillation at the resonant frequency, the damping value is set to a similar order of magnitude as the series capacitor impedance at the resonant frequency [7]. the value of the damping impedance  $R$  is set to 1/3 of the capacitor reactance  $1/\omega_r C_f$ , and  $\omega_{\text{res}} = 2\pi f_{\text{res}}$ .

## 3 The Control System of PWM Rectifier

LCL-filter can be assumptive equivalent to L-filter in low frequency current [8], and the  $d$ - $q$  current control loop of the rectifier in the proposed system is shown in figure 4. A double closed-loop control is adopted, the outer voltage loop is used to stabilize the rectifier DC voltage; the inner current loop is used to achieve the control of the power factor and the active current.

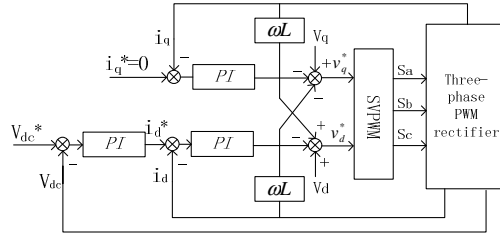


Fig. 4. d-q dual closed-loop control diagram of the PWM rectifier

### 3.1 Design of Current Loop

From (1) that mutual interference exists in the *d-q* current control loops, the voltage decouplers are therefore designed to decouple the current control loops and suitable feed forward control of output DC voltage and the three-phase AC current *d-q* components [9].  $u_q$  and  $u_d$  can be expressed as:

$$u_q = -(K_{ip} + \frac{K_{il}}{s})(i_q^* - i_q) - \omega L i_d + e_q \tag{7}$$

$$u_d = -(K_{ip} + \frac{K_{il}}{s})(i_d^* - i_d) - \omega L i_q + e_d \tag{8}$$

Where,  $K_{ip}$  and  $K_{il}$  are proportional and integral gain of inner current loop respectively;  $i_q^*$  and  $i_d^*$  are reference values of current  $i_q$  and  $i_d$ ;  $s$  is differential operator.

Substitute (7) and (8) into (1), the formula can be obtained as:

$$L \frac{di_q}{dt} = -R i_q + (K_{ip} + \frac{K_{il}}{s})(i_q^* - i_q) \tag{9}$$

$$L \frac{di_d}{dt} = -R i_d + (K_{ip} + \frac{K_{il}}{s})(i_d^* - i_d) \tag{10}$$

According to the formula (9) and (10), the inner current loop of the rectifier achieves a decoupling control and two current loops are symmetrical. For *d*-axis current control loop, the structure can be simplified to figure 5.

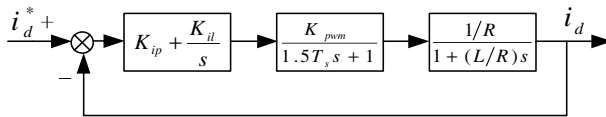


Fig. 5. Current control diagram of the inner *d*-axis current

As  $K_{ip} + K_{il}/s$  can be expressed as  $K_{ip} (\tau_i s + 1)/\tau_i s$ , when design the current PI regulator, only need to take  $\tau_i = L/R$ , for zero-pole cancellation of control transfer function. The open loop transfer function of inner current loop can be expressed as:

$$G_i(s) = \frac{K_{ip}K_{pwm}}{R\tau_i s(1.5T_s s + 1)} \tag{11}$$

According to the parameters relations of the typical I-type system, when the system damping ratio  $\xi = 0.707$ , there is  $1.5T_s K_{ip} K_{pwm} / R\tau_i = 1/2$ . The parameters of the PI regulator can be chosen as:

$$\begin{cases} K_{ip} = \frac{L}{3T_s K_{pwm}} \\ K_{ii} = \frac{K_{ip}}{\tau_i} = \frac{R}{3T_s K_{pwm}} \end{cases} \tag{12}$$

When the switching frequency is high enough, the closed-loop transfer function of the inner current loop can be expressed as:

$$\Phi_i(s) \approx \frac{1}{1 + \frac{R\tau_i}{K_{ip}K_{pwm}}s} = \frac{1}{1 + 3T_s s} \tag{13}$$

### 3.2 Design of Voltage Loop

The control structure of voltage loop of the rectifier is shown in figure 6. The transfer function of voltage regulator is  $K_{vp}(\tau_v s + 1) / \tau_v s$ , the inertia constant of voltage sampling can be combined with that of the current loop by defining  $T_\Sigma = T_v + 3T_s$ .

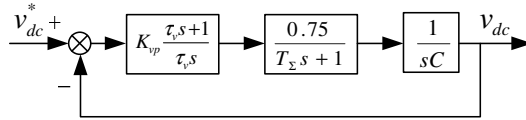


Fig. 6. Control diagram of the outer voltage loop

By figure 6, the open transfer function of the voltage loop can be expressed as:

$$G_v(s) = \frac{0.75K_{vp}(\tau_v s + 1)}{C\tau_v s^2(T_\Sigma s + 1)} \tag{14}$$

Due to the main function of the voltage loop is to stabilize DC side voltage  $V_{dc}$  of voltage source PWM rectifier, so the noise immunity must be taken into account in the course of design voltage loop. The proper choice to the voltage loop is to adopt typical II-type system [10], so the parameters of the regulator can be simulated according to forum  $\tau_v = hT$  and  $0.75K_{vp} / C\tau_v = (h + 1) / 2h^2 T_\Sigma^2$ .

Taking the bandwidth  $h=5$ , the PI regulator parameters of voltage loop are obtained as:

$$\begin{cases} K_{vp} = \frac{C}{1.25T_{\Sigma}} \\ K_{vi} = \frac{K_{vp}}{\tau_v} = \frac{C}{6.25T_{\Sigma}^2} \end{cases} \quad (15)$$

### 3.3 SVPWM Modulation

Three-phase voltage source PWM rectifier can be equivalent represented by six switches, and the upper and lower switch of each phase bridge of the rectifier cannot be turned on simultaneously at any time, they are in reciprocal state. In the actual operation, there are only six effective space voltage vectors  $V_1 \sim V_6$  and two zero vectors  $V_0, V_7$ . Reference voltage vector  $V_{ref}$  is synthesized by the eight basic space voltage vectors [11], and which is synthesized by two adjacent space voltage vectors and one zero vector usually. The vector space is divided into six different sectors I~VI by the six effective vectors, as shown in figure 7.

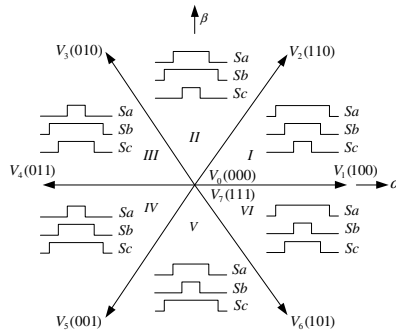


Fig. 7. Voltage space vector, sector number and SVPWM switching signals

## 4 Simulation Results

The simulation model is established using Matlab/Simulink to test the performance of the PWM rectifier with LCL filter as shown in figure 8; in the circuits, the ac source is an ideal balanced three-phase voltage source with frequency of 50Hz, and the phase to phase voltage is 100V. The grid-side inductance of each phase is 6mH, the rectifier-side inductance is 10mH, the filter capacitance is 5.6uF, and the damping resistance is 21mΩ. The output capacitor is 1200uF. In steady state, the dc voltage is set to be 300V. The switching frequency is 2 kHz. The decoupled dual-close-loop control structure is adopted in the model, the inner current loop PI parameter can be calculated according formula (12):  $K_{ip}=40, K_{ji}=0.01$ ; the outer voltage loop PI parameter is calculated using formula (15):  $K_{vp}=0.4, K_{vi}=50$ .

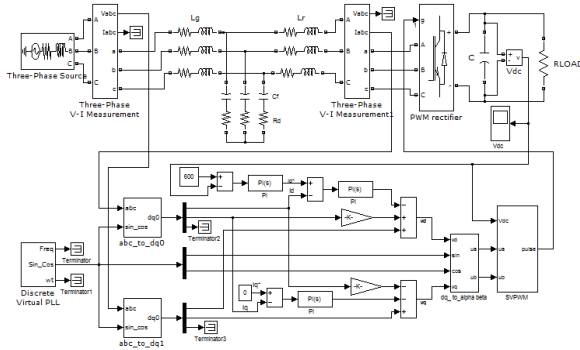


Fig. 8. Simulation model of LCL-filter three-phase PWM rectifier

The simulation results are shown in figure 9~12. In figure 9, active current  $i_d$  and reactive current  $i_q$  get stable quickly,  $i_q$  is close to the given zero to make reactive power to be zero. Figure 10 shows the voltage and current on line side, the current of sinusoidal wave become the same phase with the voltage, and the system run in unit power factor. Figure 11 is  $V_{dc}$  simulation result,  $V_{dc}$  follows the command signal  $V_{dc}^*$ , its value can be controlled to be steady (300V). Figure 12 shows the simulation results of grid-side current  $i_g$  and rectifier-side current  $i_r$ , due to adopting LCL filter, THD of the grid-side current become less. The simulation results show that lesser THD and unity power factor is achieved and DC-side voltage remain steady.

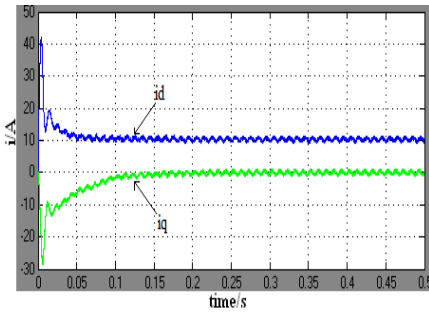


Fig. 9. The curve of  $i_d$  and  $i_q$

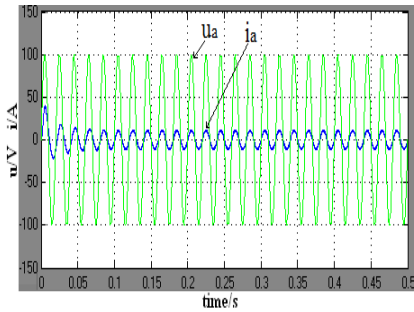


Fig. 10. Current and voltage waveform of A-phase

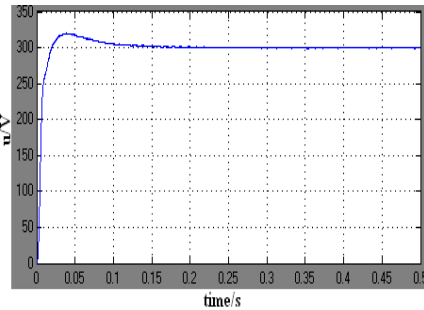


Fig. 11. Waveform of DC voltage  $V_{dc}$

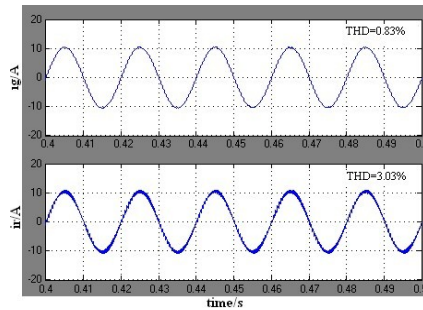


Fig. 12. Grid-side current  $i_g$  and rectifier-side current  $i_r$

## 5 Conclusion

Based on the analysis of three-phase PWM rectifier with LCL filter, a simulation model of the PWM rectifier is established. Simulation results show that the PWM rectifier with double closed-loop control system has fine performance, and with LCL filter, the desired reduction of harmonic distortions can be more easily achieved.

## References

1. Wu, R., Dewan, S.B., Slemmon, G.B.: A PWM ac to dc converter with fixed switching frequency. *IEEE Transactions on Industry Applications*, 880–885 (1990)
2. Ye, Y., Kazerani, M., Quintana, V.H.: A Novel Modeling and Control Method for three-phase PWM converters. In: *Power Electronics Specialists Conference*, pp. 102–107 (2001)
3. Lin, B.R., Yang, T.Y.: Three-phase ac/dc converter with high power factor. *IEE Proc. Electr. Power. Appl.* 152, 757–764 (2005)
4. Lindgren, M., Svensson, J.: Control of a voltage-source converter connected to the grid through an LCL-filter application to active filtering. In: *IEEE PESC*, pp. 229–235 (1998)
5. Liserre, M., Blaabjerg, F., Hansen, S.: Design and control of an LCL-filter based active rectifier. In: *Industry Applications Conference*, pp. 299–307 (2001)
6. Liserre, M., Blaabjerg, F., Hansen, S.: Design and control of an LCL-filter-based three-phase active rectifier. *IEEE Transactions on Industry Applications*, 1281–1290 (2005)
7. Liserre, M., Dell Aquila, A., Blaabjerg, F.: Stability improvements of an LCL-filter based three-phase active rectifier. In: *IEEE 33th Power Electronics Specialists Conference*, pp. 1195–1201 (2002)
8. Zhang, C., Zhang, X.: *PWM Rectifier and Its Control*. China Machine Press, Beijing (2003)
9. Xu, D.M., Yang, C., Kong, J.H.: Quasi soft switching partly decoupled three phase PFC with approximate unity power factor, pp. 953–957. *APEC, USA* (1998)
10. Zhao, Z., Li, H.: PI regulator and parameter design of PWM rectifier. *Journal of North China Electric Power University*, 34–37 (2003)
11. Habetler, T.G.: A space vector-based rectifier regulator for AC–DC–AC converter. *IEEE Transactions on Power Electronics*, 30–36 (1993)

# The Research on the Computer Control Technology of Tobacco Production

Li Zhekun<sup>1,\*</sup>, Song Yanhong<sup>1</sup>, Zou Yusheng<sup>2</sup>, Wu Mingyi<sup>2</sup>, and Liu Geyi<sup>2</sup>

<sup>1</sup> School of Mechanical and Electrical Engineering,  
Kunming University of Science and Technology, Kunming 650500, China

<sup>2</sup> Hongyun Honghe Group Qujing Cigarette Factory, Yunnan Qujing 670000  
zhekunlikust@sina.com, 806550763@qq.com,  
{zeel, liugeyi}@vip.163.com, tripped@163.com

**Abstract.** With the continuous promotion of new technology and new processes, tobacco production tends to be more specialized, refined and intelligent. Bottom layer of control net is not any longer an unaided acnode. It joins automation control system, processing controlling and management system with enterprise resource planning system by the technology of Industrial Ethernet and Fieldbus. As a result, the data acquisition capacity and flexibility are improved, and it achieves from the control indicator to the control parameter and meanwhile the system meets the requirement of the flexible manufacturing process.

**Keywords:** Tobacco, Intellectualization, Control, System Structure.

## 1 Introduction

Intelligent control is necessary to improve the processing and control accuracy, expand the scope of process control, ensure the cigarette products quality to be relatively stable and reach the quality control target.

In order to realize intelligent control, the quality factors to the process and the relevant quality must be researched. We will find out the basic factors of the quality, determine the control contents and the control method of quality elements, and control the standard formation of the key process. By the expanding control range, so that all the factors affecting the sensory quality (running status of equipments, process parameters and so on) are in a controlled state, to realize accurate control by parameter control, realize the change from result control to artificial control, from process control of experience decision to the automatic control of scientific decision.

## 2 System Structure

The control system is mainly combined by three layers, the device control layer, centralized monitoring layer, management layer. The three levels are relatively

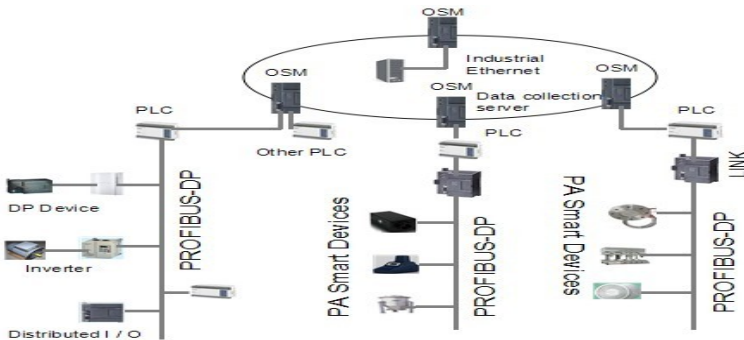
---

\* Corresponding author.

independent functionally, but in the organizational structure they rely on each other, form efficient, reliable, and complete functions. A friendly interface of the control system is integrated organically.

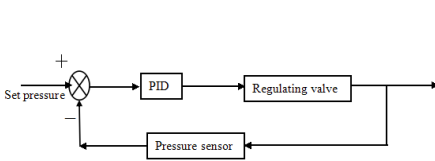
**2.1 The Device Control Layer (PROFINET, PROFIBUS-DP/PA)**

The device control layer adopts Siemens Ltd Industrial Ethernet control mode based on PROFINET bus with PROFINET Industrial Ethernet connection control station (PA interface CPU on its own) and I/O field station. The device does not have the PROFINET interface, but connect through the PROFIBUS port DP/PA bus, to realize the online control function [1]. The device control layer network structure shows in Fig. 1. It is mainly composed of PLC controllers, distributed I/O stations, electronic belt scales, photoelectric switches, proximity switches, field operation terminals, inverters, infrared moisture meters, electromagnetic valves, temperature sensors, flow-meters, pressure gauges, intelligent instruments (through PROFIBUS -DP/PA access) and independent stand-alone equipments etc..

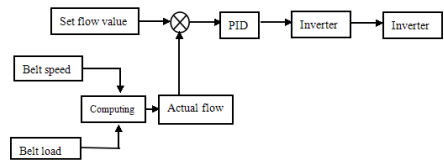


**Fig. 1.** Device control layer network structure diagram

(1) **PLC:** The PLC control system, moisture meters by electronic pressure gauges, processing equipments acquisition of actual water flow, collects the actual values and the actual pressures. As the humidifying tunnel pressure control, single closed loop system control valve and a pressure sensor are composed by the PLC internal PID regulator, gas control, and the control system diagram as shown in Fig. 2.



**Fig. 2.** Pressure control system of humidification tunnel diagram



**Fig. 3.** Principle of an electronic belt



**(2) Electronic Belt Scale:** The electronic belt scale is mainly composed of a machine frame, a conveyor belt, weighing system, transmission system, etc. and is mainly used for the control and measurement of material flow, according to the using functions there are metering type, control type and mixing type. The utility in quantitative feeding in the control system of electron has an independent control system with the PLC control system and has the function of [2] measurement and control of material flow. Its principle as shown in Fig. 3, the weight sensor of an electronic belt scale is on the detection of the materials of instantaneous flow rate and the encoder detection is sent into the PLC control system for the operation. The actual flow value and the set flow value are compared and then through the PLC internal PID regulator the result will control the inverter output frequency regulation of the electronic belt conveying speed, so that the actual flow value tends to equal the set flow value [3].

**(3) Photoelectric Switch:** As pluralities of photoelectric switches according to certain rules are arranged in different measuring points, the measuring point of a predetermined switching can be occluded and each measuring position has certain code. The measurement at each switch of the current object masks namely the position coding and encoding of the current reference position compared with the displacement, representative of the code. So the current position bias can be known. The application should be used to avoid the cyclic code digital change at the gross error [4].

**(4) Proximity Switch:** The proximity switch is also called contact-free overtravel-limit switch. Not only it can complete the stroke control and limit protection, the detection device is a kind of non-contact type and is used for detecting the part size and speeding and so on, can also be used for frequency counter, frequency pulse generator, liquid level control, and processing program of automatic connection.

**(5) Inverter:** A frequency converter speed control motor, a drive belt or a V belt can drive the driven wheel, supporting wheel, and drive the cylinder rotation and control the speed. As a feeding control, by changing the frequency of a metering pump, regulating the speed of a metering pump, the feeding control will be realized.

**(6) Electromagnetic Valve:** The use of the electromagnetic valve is throughout all facilities, such as the door switch electromagnetic valve, water cycle electromagnetic valve, charging electromagnetic valve, metal detector electromagnetic valve, electromagnetic valve for controlling the knife roller clamping, relaxing, and for controlling water, steam temperature, and so on.

## 2.2 Centralized Monitoring Layer (PROFINET)

Centralized monitoring layer mainly consists of monitoring and operational control and field stations, I/O servers, engineer stations, real-time database servers and Ethernet network devices. Its backbone composed of 1000M fiber ring network through the industrial Ethernet switch, multimode optical fiber as the medium, equipped with redundant power supply to ensure stable and reliable communication of the backbone network. This layer is mainly composed of PLC device layer equipments and field distributed control organization, to realize the field data acquisition, alarm, control,

interlocking functions, as well as the data exchange between the device control layer and the management layer. At the same time, the layer control of the user interface of the system can complete the entire line centralized control.

### 2.3 Production management (Ethernet)

Composition of workshop production management layer is mainly composed of database servers, application servers, WEB server management, and management of the computer and network devices. The exchange of information through the Ethernet can realize the entire workshop equipments monitoring and unify equipments monitoring, control and management in one integrated platform. The production management system [5] adopts mature advanced B/S with C/S. The .NET framework based on MICROSOFT has functional connections with supervisory computer network.

Through the analysis of statistical data management system, classifying, sorting and comprehensive analysis of the collected information, forwards, it can get a unified scheduling of production process and monitor, control and manage various production processes, key equipments and technology, in order to achieve the objective of the production management, process management, equipment management, quality management. Backward, as a subsystem of enterprise MES system, it can provide the important basis of making strategic decisions and make the managers understand the real production process immediately.

## 3 System

### 3.1 Centralized Monitoring Layer Function

**(1) Monitoring Function:** Centralized monitoring of real-time simulation shows the operation states of main equipments. According to the production process there are the main equipments, auxiliary equipments and many test points. The running states and main parameters of main equipments will be displayed. The main measurement points will be at motors, conveyor belts, belt scales, pressure gauges, flow meters, water meters, thermometers, frequency converters, limit switches, photoelectric tubes, status of all valve, etc. The number of grades and batches of current production are displayed on the centralized monitor screen. Operation status, equipment parameters, process parameters and other information of the main equipments are reflected through the appearance of equipments and pipeline simulation diagrams. There are the simulation pictures of key process equipments such as damping machine, feeding machine, dryer equipment. The picture with a single device simulation graph and line graph displays the main motor running state, related parameters of the equipment, such as entrance moisture, temperature of water, outlet temperature, motor frequency, motor current, feeding flow accumulated value, steam pressure, etc. and the state of single machine equipment, such as preheating, cooling, work, stopping state, etc. The main valve diagram mainly monitor on-off states of signal valves.

**(2) Control Function:** Control function will complete all controls of whole production processes, so that the control system can monitor different brands and different batches of tobacco, and will make management and control be centralized, and improve the production equipments efficiency, reduce man-made factors, enhance the stability of process control. In order to facilitate the centralized control, convenient operation and strengthen security, all control functions are integrated in a control screen.

**(3) Data Acquisition Function:** The parameters of production line process and equipment operations and other data are unified and stored in information database.

**(4) Display Function:** The current states of motors, electromagnetic valves, photoelectric tubes, limit switches, proximity switches, control and detection devices, alarms, alarm status will be displayed by using graphics and colors. The I/O stations distributed in the field and inverters, electronic belt scales are applied for centralized management. Internet state, distribution status, and manual/automatic states are displayed in real time. The actual feeding states of cabinets, material storage states, material quantity states are displayed and their time information will be recorded. Each process section will be a reasonable layout of the storage cabinet pictures. The cabinet information in various forms is displayed to meet the management of storage cabinet.

**(5) Diagnosis, Alarm Data, Print:** Network (bus, nodes, elements) fault diagnosis and alarms of process quality failure, fault alarms, fault diagnosis of equipments are displayed and could be printed out. The fault alarms include both of real-time alarms and historical alarms.

**(6) Authority Management:** Modifying the local/remote control authority to operate terminal executing local control can be selected either in the control room or remote centralized control according to the actual production needs. Safety management will unify to manage the security permissions of operators, technicians, electrical engineers and other relevant persons, in order to ensure the security of the system, to prevent the illegal user's violation from the production control and the important data.

**(7) Issuance Function:** The authorization supervisory person can view the real-time equipment monitoring pictures of production information through IE browser.

### 3.2 Production Management Function

**(1) System Management:** It includes the functions of the system parameter setting, user management, factory calendar, log management and system maintenance.

- System parameter setting: realizing basic operation in system running environment, running parameters, data maintenance of tobacco production management
- User management: establishing the line of unified user authorization mechanism, realizing the centralized user login authorization, user information maintenance, management of users and user groups
- Log management: providing for centralized monitoring management system and query function of operation log of production management, realizing the function of the process of events that have occurred, making the operation rights and responsibilities be clear.

**(2) Production Modeling:** It is the foundation and core of the whole production management system. According to the requirements of design and production model, in the production management process, production line, material and parameter information are modeled completely. The production line information, process information, brand information, prescription information are transformed as a model of information. Finally the production is configured.

**(3) Production Plan Management:** It is the core module of tobacco strip production management system. Production planning is mainly responsible for what receiving from the MES, to complete the production process organization and executive function in entire production line, including the tobacco strip production plan, maintenance, batch management, task scheduling, scheduling and monitoring process. Considering flow characteristics in tobacco strip production and the time in manufacture implementation, in the production management system, the production plan is generally developed and implemented for only a day.

**(4) Production Data Management:** It will complete the analyses of the production data in the implementation process of tobacco strip production. The main functions are feeding data management, output data management, production operation data management, on-product data management, production reports and other functions.

**(5) Quality Management:** Based on process data, the processing step and process quality in product processing are managed. Process quality management is a process analysis, and will manage and control the factors that are affecting process quality. In the production process, acquisition, calculation, processing and statistical analysis of data are used to evaluate process quality characteristics, changes and rules of control procedures quality, and to make product quality characteristic fluctuating in an allowed scope, which makes different batches of product quality to be homogenized.

## 4 Conclusions

Based on the study of computer control system of tobacco intelligent production, it shows the system structure and the integrated control system, which play an important role in the practical application. Good results are obtained through that technology. It provides important and valuable reference for similar control system.

**Acknowledgement.** This work was sponsored by China Yunnan Cigarette Company Research Foundation Project (No. 2012GY08). We will thanks here!

## References

1. Wu, C., Zhang, J.: Application of field bus technology, modification of tobacco strip of electric control system. *Domestic and Abroad Mechatronics Technology* 2, 14–15 (2005)
2. Zhang, P., Zhang, S., Luo, F.: Electronic belt scale based on PLC in tobacco strip production line application. *Industrial Control Computer* 22(11), 61 (2009)

3. Chong, L.: The design and implementation of electrical control system of electronic belt scale in tobacco strip production line. *Chinese High-Tech Enterprises for 2011* (19), 21 (2011)
4. Yang, H., Xia, J., Li, H., Huang, Z.: The photoelectric switch in displacement monitoring and positioning system applications. *Journal of Scientific Instrument* 22(z2), 165–166 (2001)
5. Songjun, Z.: Based on C/S and B/S the tobacco machinery drawings management system developed. *Modern Manufacturing Engineering* 8, 44–45 (2009)

# A Study on the Application of Financial Engineering in Supply Chain Risk Management

Hao Jianhua<sup>1,\*</sup>, Sun Wenliang<sup>2</sup>, Chen Liwen<sup>2</sup>, and Yang Zhanchang<sup>2</sup>

<sup>1</sup> Tianjin University of Sport, No.51, Weijin Nan Road, Hexi District, China 300381

<sup>2</sup> Hebei University of Technology, Tianjin 300401, China

{45960068, 36847141}@qq.com

k27277@sina.com, lwchen@hebut.edu.cn

**Abstract.** With the constant economic and social development, enterprises are facing more and more diversified and complicated competition. In this paper, the difficult problems in the supply chain risk management are discussed and the internal relations between the financial engineering and risk management analyzed, so as to appropriately apply the financial engineering to the risk aversion. At last, feasible measures and methods are proposed in this paper.

**Keywords:** Risk Management; Financial Engineering; Supply Chain.

## 1 Definition and Application of Financial Engineering

In late 1980s and early 1990s, companies of increasing number participated in the financial services supplied by financial industry and banking industry. The definition of the term of “financial engineering” was put forward by Finnerty, an American financial economist, for the first time. He thought the financial engineering stood for a whole course including the design, development and implementation of the financial instrument and method as well as the effective solution of the financial problems. Later on, the term of “financial engineering” was redefined by the economist John Pfennig, who thought the financial engineering stood for the novel financial instruments and the whole financial process ranging from designing, developing and then using the innovative strategy to solve the enterprise’s whole financial problem[1]. Both Finnerty’s and John’s definitions reveal that the financial engineering contains at least three elements: firstly, it must contain innovative design and development of the financial development process; the second is the creation and design of the new financial tools; and the third is the creative strategy to the overall financial problem of the enterprise.

According to Finnerty’s and John’s definitions on financial engineering, we can roughly classify the research scope of financial engineering as three sections, i.e. the innovative design and development of financial instruments, development and innovation of new financial transaction approaches that reduce the transaction cost, and proposal of sound financial solutions which are used to solve financial problems.

---

\* Corresponding author.

### **1.1 Innovative Design and Development of Financial Instruments**

The content of innovative design and development of financial instruments has a wide range, and, in the meantime, it is now the major issue and objective of the study in the field of financial engineering. Presently, the new financial instruments include the note issuance facility, index futures, security depository receipts, share option, interexchange, forward rate agreement (FRA), synthetic stock and zero coupon bond, etc[2].

### **1.2 Development and Innovation of New Financial Transaction Approaches Reduce the Transaction Cost**

The development and innovation of new financial transaction approaches that reduce the transaction cost includes the internal plan-making optimization, updating and improving of the financial structure; the innovation of the exploration and application of the arbitrage opportunity on the financial market; and the innovation of transaction clearing system, etc. All these means of financial transaction are made to exploit the profit potential of themselves as well as the market fully and thus to control the cost and reduce the cost.

### **1.3 Sound Financial Solutions Designed and Developed for Financial Problems**

This kind of solutions and systematic measures principally include the development and application of the various skills of risk management, creation of the means and tools during the corporate finance, updating the existing managerial strategy, and design and development of the strategy of corporate mergers & acquisitions and securitization of the enterprise's inherent assets.

In recent years, a number of commercial banks have got involved in the financial engineering as many creative financial innovations have been continuously applied to consumption and retail. In their common work of Financial Engineering, Marshall and Bansal use the term of "investment bank" in the broad sense, which covers the emerging financial institution, traditional investment bank and some commercial banks having conducting financial services, which get involved in the financial construction and risk management. However, seen in the angle of financial demand and people's daily life, the financial engineering includes at least four parts, i.e. banking transaction, corporate finance and investment, risk management, and investment and cash management. In these contents of financial engineering, the risk management is deemed as an important part and core content of the financial engineering, and some investors and economists even put risk management in the same place with financial engineering[3].

## 2 Problems in Risk Management of Supply Chain and Their Causes

Supply chain risk means that, due to the uncertain factors which make the enterprise unavoidably subject to the influence of the internal and external environment during its production, the enterprise faces some unavoidable losses leading the enterprise unable to reach the expected income in the supply chain, or leading to the differences between the gained and expected incomes, thus resulting in the enterprise's loss and various possibilities.

The risks in the supply chain are generated due to the following several factors:

**Firstly, Moral Hazards.** The moral hazard means that one party gets the surplus interest from other party in the supply chain, thus leading the agreement to loss efficacy and resulting in the risk and crisis in the supply chain.

**Secondly, Information Risks.** Since every enterprise in the supply chain is an independently operating economic entity, the supply chain, in nature, is a quite loose enterprise assembly. With the increasing expansion of the supply chain, its structure would become more and more complicated and the risks leading to errors of information would increase. The delay and error in information transmission would directly lead to the deviation in the communication between enterprises, thus resulting in the bullwhip effect and overstocks.

**Thirdly, Risks in Selecting the Distributors.** When selecting the distributors, the enterprise many choose the wrong distributors that fail to effectively and sufficiently manage the supply chain in real time, thus leading the core enterprise to failure in management competition and this failure then leads the whole supply chain to become loose and even separated.

**Fourthly, Policy Risks.** When a country's economic policy is readjusted or changed, the industries restricted by the policy need the supply chain to raise a lot of funds the make changes with the country's economic policy.

**Fifthly, Uncertain Risks of the Market Demand.** The market demand guides the direction for the operation of supply chain. With the market competition becoming more intense, the consumers' personal preferences are also changing and becoming more unpredictable. For this reason, the risks in the experience of the whole supply chain are increased.

## 3 Functions of Financial Engineering in Supply Chain Risk Management

Both the risk management and novel techniques are the two critical factors that make the financial enterprises survive the intense market competition in the financial industry. In 2001, China joined in WTO. From then on, China's financial market has become more open to the outside world and the number of foreign trades has increased. In the meantime, a lot of foreign financial institutions swarm into China,



competing fiercely with China's local financial institutions. During the competition, China's financial institutions often lose the game due to their weakness in the risk management and financial innovation. Therefore, some propose that the essence of financial engineering is to adopt the financial innovation to conduct risk management. This does make sense to some degree. Franklin Allen, an American financial economist, holds the opinion that those industries which face high risks, have strong creativity or rely much on the innovation capacity mainly choose the capital market for financing. In this way, they can not only gain many equity capital funds but avoid the capital risks to the maximum, thus transferring part of the risks to the investors via the financial operation on the capital market. This shows that, with the ceaseless deepening and expansion of the financial market, the financial instruments will play a more and more critical part in risk management[4].

### **3.1 Definition of Risks**

The risk is objective, universal, inevitable, recognizable, controllable, uncertain and social, and it may cause losses. Therefore, the risk exists independently from human willpower. Both the natural motion of matter and the rule of social development are determined by the internal factor of objects, i.e. the objective rule beyond the control of people's subjective consciousness.

The risk and uncertainty are different from each other. In the work of *Risk, Uncertainties and Profit*, which was published in 1921 by Frank H. Knight, the difference between the two are described specifically. Mr. Knight thinks if the randomness faced by the economic actor can be measured by using the mathematical probability value—the probabilities in there can not only reflect the individual subjective belief but determine the probability like purchasing the lottery), this condition is regarded as the risk-involved[5]. In addition, if this probability can't be determined, this condition wouldn't belong to the uncertain condition. In a word, we can draw the conclusion on the risks—since people have certain anticipation on a thing or object, but this thing or object is deflected due to some uncertainties, the result is deviated consequently. In here, the uncertainty means the risk.

### **3.2 Solutions of Financial Engineering to Risks**

In the traditional financial activities, the management on risks is almost completed via the financial transaction on the table, which is of no doubt subject to a restriction. As for the traditional approach of risk management, the breakthrough and innovation must be realized in the financial engineering. For example, the off-the-table financial instruments are more and more frequently applied to the risk management. The core theory of the financial engineering lies in the latest results about the theory of risk management; in the meantime, this new theory system also includes the arbitrage theory, assets portfolio theory, value theory, option pricing theory, agency theory and other financial theories as well as the efficient market theory. Apart from the above-mentioned approaches of risk management, the financial engineering also creates and applies the creative non-arbitrage analysis method, which combines a specific

position on the market. This specific position can build a value position composed of profits, which almost needn't stand risks when the market is in balance. The investors estimate the balanced price of the market by using this position. For example, when there are some certain risks on the market, people apply the financial engineering to build a position corresponding to the risk that an on-the-table service stands to realize the offset between the risks on and off the table, thus avoiding risks effectively[6]. To put it simply, the financial engineering allows people to keep the favorable risks while building the position to eliminate or replace the unfavorable risks, i.e. uncertainty being replaced with certainty.

In the financial activities, there are various risks, such as the credit risk, interest rate risk (IRR), operation risk, exchange rate risk, goods price risk and legal risk, etc. In order to handle these risks, a series of useful tools are developed in the field of financial engineering.

**Firstly, Solutions to the Fluctuation Risk of Interest Rate.** The interest rate on the capital market changed a lot in the 1970's, especially in western developed capitalist countries. On the capital market, the bank loan, commercial invoice, bond and other financial products all have fixed interest rates. However, the tremendous fluctuation of the interest rate on the capital market makes the borrowers suffer large risks on various capital markets. In this context, the instrument for floating interest rate becomes the best choice to avoid risks. For this reason, this instrument becomes the most successful result in the financial engineering.

In 1982, Chase Manhattan Corp. invented the preferred stock with adjustable interest rate on the bond and security market. Compared with other preferred stock of that period, the preferred stock with adjustable interest rate is totally different—the ordinary preferred stocks can be adjusted and operated in time in face of the substantial fluctuation. The interest rate change on the market largely affects the changes of the interest rate of the preferred stock with an adjustable interest rate. In here, the adjustment, taking places following a certain criterion, has an outstanding advantage, i.e. guaranteeing maximization of the investors' interests inside the financial assets when the market interest rate changes. The financial assets here include the short-term, medium-term and long-term financial assets[7]. Through the adjustment of the stock with an adjustable interest rate, the issuer affords the risks of changes in the term structure of interest rate. In addition, the market price fluctuation of the above-mentioned preferred stock, due to its adjustable interest rate, is also smaller than other ordinary preferred stocks. Because of this, the risk of interest rate fluctuation is reduced largely. The following are the major tools used to mitigate the risks of interest rate change: interest rate option, interest rate futures and interest rate forward.

**Secondly, Reconfiguration of Risks.** In the transaction of economic activities, some act as the risk avoiders and others as the risk bears, without which the economic market wouldn't be able to exist normally, lest continual development. The theoretical analysis of utility reveals that the risk avoider, risk neutral and risk lover together constitute the mass involved in the economic trading activity. Therefore, on the basis of people's different attitudes to risks, the financial engineers design some professional instruments to meet the need of different investors.

For example, there is a commodity derivative bond which connects the price of goods and yield together. With this bond, the issuer of the bond would afford more risks. The interest-free discount bond, as a kind of bond, was issued by Standard Oil Company in 1986. When this bond is due, 1100 dollars will be returned to the bond holder. If the oil price is above 25 dollars when the oil price is higher than 25 dollars, the bond issuer will get the additional compensation of 170 dollars for each dollar higher than the price, but the compensation wouldn't exceed 2550 dollars. [8] As to the bond issuers, they can issue the bond with a relatively low interest rate, so the bond owners won't worry about the fluctuation in the oil price. In particular, the oil owners will face lower risks. Since the adjustable interest rate makes the issuer assume more risks, the investors would face a lower price risk.

The currency option/changeover, barrier option and the like are also the common financial engineering products which reconfigure the risks. In order to meet the demands of different traders, the financial engineering can effectively combine the forward and option so as to formulate the planning for risky revenue and thus reconfigure the risks.

**Thirdly, Solutions to Information Asymmetry.** Many of the market risks are caused by the information asymmetry. For example, the capital market cannot reflect all the information held by the company management. On the emerging market, in particular, where the laws and regulations are not very sound, the information asymmetry is more noticeable. For example, the packing market is often seen on China's stock issuance. For this reason, the actual condition of the company can't be grasped by the investors accurately and thus lead to the asymmetry of information between the investors and company management. The stock of Hengyang Industry, which went public at the Shanghai Stock Exchange, is a typical case—this stock indicates obvious losses not long after being listed and had the investors suffer great losses. To get rid of this phenomenon, the callable stock was created, which is a new product model designed by the financial economist to make the stock revocable—if the stock price slumps to the set number, the company would buy back the shares at the former price from the investors. In this way, the risk is assumed by the company and the information asymmetry can be eliminated.

**Fourthly, Solutions to Liquidity Risks.** As for the security of assets, liquidity is of great importance. In general, the more liquid the assets are, the fewer risks they bear. For this reason, many investors prefer assets with better liquidity. However, we must confess that many assets on the market are in poor liquidity. It's a big problem to improve their liquidity. To handle this problem, the financial engineer design a series of tools, including the home mortgage which is the most typical. In the US, a lot of home mortgages are held by financial institutions. In order to activate these assets with poor liquidity, the mortgage companies acquire these them from banks and then group them according to their interest rates and deadlines before issuing them as the newly grouped mortgage bonds. In this way, the liquidity of these assets would be improved. The credit card, car loan and securitization of loans are also to some extent designed in this thought, and the securitization is even applied to the assets that are not easily standardized, such as receivable.

The note issuance facility is another quite typical liquid innovative tool. A series of short-term bill portfolios issued on the basis of circulation are created by the engineers who design the note issuance facility. These bill portfolios make up the poor market liquidity. This kind of tools is quite popular with people and is thus maturing at a high speed.

### **3.3 Financial Instruments Designed in Financial Engineering**

In the financial engineering, the common financial instruments include the option, interexchange, forward and future. The financial instrument can be classified as the cash financial instrument and derivative financial instruments. The cash financial instrument is related to the following four categories of market information, i.e. the currency market, foreign exchange market, stock market and bond market; the derivative financial tool is directly related to the financial futures, forward ratio, financial option and interexchange[9].

The fundamental types and characteristics of financial engineering are described as follows: the business information and macro economy belong to the first category of financial information, and the trend of the industrial economic activities and tendency of the macro economy have a great influence on any activity on the financial market, so the first category of financial information includes the information of economic monitor and business cycle, information of macro-economy statistics, information of various macro-economy policies affecting the economic activities and the industrial organizations. The second category of financial information includes various financial market information, which is further classified as the direct financing market information and indirect financing market information. Classified in accordance with its specific information, the financial market can be classified as the capital market, precious metal market, financial futures market, foreign exchange market, currency and bill market and international financial market. The second category of financial information includes the enterprise financial information, which stands for the financial statements and reports of all kinds of enterprises, including public companies.

The financial engineering can't be applied without risk and decision-making analysis, calculating model and financial information pool. The core of financial engineering lies in tool using and technical requirements of the risk management, and its foundation is the financial information held by the enterprises. Therefore, it's necessary that the modern financial theory be applied to the modern financial engineering. In order to control the risks appropriately and effectively, people should sufficiently make use of various useful resources and building the model for analysis, in order to make great breakthroughs.

## **4 Measures to Handle the Problems in Supply Chain Risk Management**

### **4.1 Applicability of Financial Products in Supply Chain Risk Management**

The fluctuation in the market price of the product represents the investment risk borne by every enterprise in the supply chain. When the fluctuation in the product price is

inconsistent with the expected result, the enterprise would face risks, which in turn affect the whole supply chain. Some relevant researches reveal that the tools are upgraded and designed by the financial engineering to avoid and reduce risks. These tools are timely, flexible and accurate in avoiding and mitigating the risks, and are also more convenient and flexible in cost control and management. The financial instruments are useful in financial leverage and hedging transactions, which largely reduce the transaction cost between the enterprises in the supply chain. In addition, the financial instruments can be also used for two-way building of position, which allows many investors to engage the transaction flexibly with better operability. So to speak, with the application of derivative financial instruments, the investors can carry out corresponding operation no matter how the market price goes, so as to avoid or mitigate the risks brought by the fluctuation in the price market. Through this tool, the risks are transferred to the risk lovers or speculators from the hedgers. There are many applicable financial engineering tools, such as the forward, forward interest rate, futures, operation interexchange and the like.

#### **4.2 Analysis on Application of Financial Instruments in Supply Chain Risk Management**

**First of All, Cost Locking and Hedging through Futures Market.** When a large fluctuation emerges on the product price market in the supply Chain, the product marketer, wholesaler and manufacturer would take two measures to handle the fluctuating prices, such as conducting conservative operation on a market with large price fluctuations while engaging profitable operation on another market. In this way, the cost investment is guaranteed and the profit may be even gained, so that the investors won't suffer losses everywhere. With this method, the impact and risk brought by the intense fluctuations in the product market price are mitigated.

**Second, Futures and Options Used to Structure Contract Groups for Effective Risk Aversion.** The so-called future trading is one form of the concentrated trading of standardized forward contract. The process of the transaction is as follows: the two parties of transaction sign a futures contract in the trade; then, they transact the goods for certain quality and quantity according to the contract at a certain location, time and price. However, the real objective of the futures trading is not to transfer the ownership of the goods, but to avoid the spot price risk. The futures trading means that the two parties of the transaction agree to sell or purchase a specific number of goods at appointed prices at a time in the future, but they don't have to do this.

There's one thing to be noticed: since China hasn't become a WTO member for a long time, its financial market advanced at a lower speed than the financial engineering on western capital market; in addition, China's financial market developed a quite low speed due to its being subdivided, so the balanced price on the market is not the financial price[10]. Therefore, in the area of avoiding risks and discovering the market price, the derivative instruments designed and developed by the China's financial engineering haven't been able to give full play to their entire performances. The short position, in particular, hasn't brought into full play or even played a negative role due to the restriction from the system and relevant laws and regulations

of China. Moreover, the reform conducted in many of China's enterprises is backward. In fact, the main projects and products of these enterprises don't need the risk aversion and China's market doesn't have a large demand on the financial engineering. Furthermore, China doesn't have a sound and mature legal system for the derivative instruments of financial engineering, in addition to the insufficient varieties of China's financial instruments and lacked varieties of derivative financial products, so the financial engineering hasn't been able to bring into full play in China's supply chain risk management. Though the introduction of financial engineering has brought tremendous changes to the risk management, the financial engineering is not the approach that can thoroughly solve the problem in the risk management. Therefore, the key point to the enterprise is to continually improve its productivity, enhance its comprehensive strength, conduct management more scientifically, and build an efficient and appropriate operating system. Only in this way will the enterprise handle the risks and challenges more calmly and improve its comprehensive strength increasingly. The financial service doesn't mean all problems related to the financial engineering, whose nature is to emphasize the power of knowledge, cultivate the innovative thoughts ceaselessly, improve the computer technology and engineering technology increasingly, master financial mathematics better, and explore new approaches and technologies to prevent and cope with financial risks, thus having the financial service operation, academic exchange and theoretical study in China ascend to a higher level. In the 21st century, when the opportunities and challenges are everywhere, only the sufficient preparation can make people face various problems in the field of financial engineering and thus run the financial industry better in the new era.

## References

1. Li, M.: Problems about Financial Engineering and Its Application in China (August 2001)
2. Li, D.: Financial Engineering: Technical Assurance of Financial Innovation (August 8, 2001)
3. Marshall, J.F., Ransal, V.K.: Financial Engineering
4. Guanhua, Y.: Rise of Financial Engineering. Financial Times (February 2, 1998)
5. Qiulin, J.: Financial Engineering and Risk Aversion (February 14, 2002)
6. He, W.: Several Problems in Logistic Finance under the Context of Supply Chain. Journal of Chongqing University of Scientific Technology (Social Science Edition) (2009)
7. Feng, J.: Supply Chain Finance: Advantages, Risks and Suggestions. Regional Finance Research (2009)
8. Chen, L., Peng, F.: Problems in Developing Supply Chain Finance and Corresponding Solutions (2009)
9. Finnerty, J.: Financial in Corporate Finance: An Overview. Financial Management (Winter 1998)
10. Yao, F.: Supply Chain Finance: Innovative Financial Services Realizing All-Win. New Finance (2008)

# Logical Symmetry Based K-means Algorithm with Self-adaptive Distance Metric

Wu Zu-Feng, Mu Xiao-Fan<sup>\*</sup>, Liu Qiao, and Qin Zhi-guang

School of Comp. Sci. and Eng.,  
University of Electronic Science and Technology of China,  
Chengdu 610054, China  
{wuzufeng,qliu}@uestc.edu.cn,  
sdmuxiaofan@gmail.com,

**Abstract.** In this paper, we propose a modified version of the K-means clustering algorithm with distance metric. The proposed algorithm adopts a novel weighted Euclidean distance measure based on the idea of logical symmetry of points to its candidate clusters, which challenges the common assumption that the point similarity can only be determined by their physical distance to the centroids of the clusters. This kind of logical symmetry distance can be adaptively applied to many practical data clustering scenarios such as social network analysis and computer vision, in which the logical relationship of the clustering objectives is an important consideration in the design of the clustering algorithm. Several data sets are used to illustrate its effectiveness.

**Keywords:** K-means, logical symmetry, self-adaptive distance metric.

## 1 Introduction

Clustering problem is one of the most important questions in machine learning. Clustering is the process of partitioning or grouping a given data set into groups. This is done that data samples in the same group are similar and data samples from two different groups are dissimilar. With the development of technology and cross-disciplinary cooperation, clustering is widely used in Bioinformatics, image segment, social science. And these many clustering methods can be broadly divided in hierarchical methods and partitioning methods. [1] The K-means algorithm we talked about in this paper is a partitioning method. However, from the practical perspective, the most suitable clustering method depends on the practical problems you met.

The K-means was proposed by Stuart Lloyd in 1957 [3]. There are many advantages of the K-means algorithm, such as it is easy to implement and can be easily applied to processing the large data problem. But there are still some disadvantages like it cannot handle non-spherical clusters well, needs to specify the value of K in advance, different initial centroids may result in totally different final clusters, the clusters' size tend to be similar, and so on. And many research works has been studied to improve

---

<sup>\*</sup> Corresponding author.

the K-means algorithm. Baberjree and Ghost came with the goal function in K-means to get a balanced clustering result [4]; Nazeer and Sebastian improved accuracy of K-means by finding more receivable initial centroids[5][6]. Pelleg and Moore proposed the x-means algorithm based on K-means. By using Bayesian Information Criterion, x-means can efficient estimation the number of clusters [7]. But no matter how many improved algorithms have been proposed, one thing is common recognized that no one variant K-means algorithm can solve all clustering problem. Each has its own special issues. However previous researches have fewer investigations on improving the accuracy of K-means when data set has adjacent and different size clusters. So, this paper we propose a modified K-means based on weighted squared distance. The algorithm we proposed can make up the drawback of tending to partition data set of adjacent, different size clusters into similar size clusters in K-means. Experiments show that when dealing with the above situation, the modified algorithm gets higher accuracy than K-means with a tolerable gain on computation time.

The remaining of this paper is organized as follows. The second section we review the K-means algorithm and show the shortcomings. The details of modified algorithm are provided in the third section. In the forth section are the experiments and results, and in the fifth section is the conclusion. Section 6 is the acknowledgment.

## 2 K-means

K-means algorithm [2] is the simplest method in clustering algorithm. Given a data set  $\{x_1, \dots, x_m\}$  and the number K. K-means divides the data set into K partition. The main idea of K-means is that: minimize the sum of squared distance from each data sample to its cluster centroid. And this is also called goal function in K-means. The formula of the goal function is as follows:

$$J = \sum_{i=1}^m \|x_i - c_{u_i}\|^2$$

$u_i$  represents the cluster  $x_i$  assigned to.  $c_{u_i}$  is the centroid of the cluster  $u_i$ . In order to minimize the goal function, K-means takes following steps.

Input:

1. Data set  $\{x_1, \dots, x_m\}$  which have m data samples.
2. Number of desired clusters, k.

Output: Data samples in the data set are divided into K clusters.

Step 1: Randomly select K data samples  $c_1, c_2, \dots, c_k$  from data set  $\{x_1, \dots, x_m\}$  as initial centroids.

Step2: For each data sample  $x_i, i \in (1, \dots, m)$  in data set, assign it to the nearest centroid.

$$u_i := \arg \min_{j \in (1, \dots, k)} \|x_i - c_j\|^2$$

Euclidean distance is used as distance measurement.



Step3: Recalculate centroid as the center of all data samples in the same cluster

$$c_j := \frac{\sum_{i=1}^m 1\{u_i = j\}x_i}{\sum_{i=1}^m 1\{u_i = j\}}$$

$1\{u_i = j\}$  represents that: when  $u_i = j$  happens, this formula results in 1; otherwise, it results in 0. The left side  $c_j$  is the new centroid for cluster  $j$ .

Step 4: Repeat step2 and step3 until each centroid in the clusters does not change or changes less than a threshold value.

From the Step2, we saw that each time the data sample is assigned to the nearest centroid. This can cause an issue. The data sample in the bound between a larger cluster and a smaller cluster may be assigned to the smaller cluster since the data sample is far away from the larger cluster’s centroid. That means K-means tend to divide the data set into clusters with comparatively similar size, which is the drawback we try to improve.

The problem comes when there are two clusters which are adjacent and have different sizes. So, we add weighted information as a compensation factor to distance measurement in the goal function of K-means. By changing the goal function, we also changing the assignment of data sample in Step2. Thus, we make up the drawback which tends to partition data set into similar size clusters in K-means. We will give a detail explanation of the modified k-means algorithm in next section.

### 3 Modified K-means Algorithm

The modified K-means algorithm takes the hypothesis that if there are more data samples in the cluster which are similar to the assigned data sample, the probability that the data sample be assigned to the cluster should be higher. When we say similarity, we refer to the two distances from two data sample to the same centroid are similar. And in this paper, we still use the Euclidean distance as the distance measure.

Adding weighting factor to the Euclidean distance can rectify the shortage of trending to divide the data set into similar size cluster in the K-means algorithm. And the weight is the reciprocal number of data samples which are similar to the assigned data sample. So, the more similar data samples the cluster has, the less weighted distance between data sample and cluster centroid it will be. Thus the data sample is more likely to be assigned to the cluster which has more similar data samples in it.

We use two parameters  $\lambda_1$  and  $\lambda_2$  to prescribe a limit to the similar between data sample pair. The distance between the assigned data sample and centroid is defined as  $d$ . If the distance between the cluster centroid and the data sample in the cluster is greater than  $\lambda_1 d$  and less than  $\lambda_2 d$ , then, the data sample is similar to the assigned data sample.

Therefore, the modified K-means algorithm modifies the goal function in K-means to the following goal function.

$$J = \sum_{i=1}^m \|x_i - c_{u_i}\|^2 \frac{1}{1 + \sum_{j=1}^m 1\{\lambda_1 \|x_i - c_{u_i}\| \leq \|x_j - c_{u_i}\| \leq \lambda_2 \|x_i - c_{u_i}\|, u_j = u_i\}}$$

This function minimized the weighted sum of squared distances.  $u_j = u_i$  represents that data sample  $x_j$  and  $x_i$  are in the same cluster.

Correspond to the changes in goal function; the new procedure is as follows:  
Input:

1. Data set  $\{x_1, \dots, x_m\}$  which have m data samples.
2. Number of desired clusters, k.

Output: Data samples in the data set are divided into K clusters.

Step 1: Randomly select K data samples  $c_1, c_2, \dots, c_k$  from data set  $\{x_1, \dots, x_m\}$  as initial centroids.

Step 2: For each data sample  $x_i, i \in (1, \dots, m)$  in the data set, assign it to the nearest centroid based on weighted squared distance.

$$u_i := \arg \min_{j \in (1, \dots, k)} \left( \|x_i - c_j\|^2 \frac{1}{1 + \sum_{l=1}^m 1\{\lambda_1 \|x_i - c_j\| \leq \|x_l - c_j\| \leq \lambda_2 \|x_i - c_j\|\}} 1\{u_l = j\} \right)$$

$1\{u_i = j\}$  indicates data sample  $x_i$  has to be within the cluster  $j$ .

$\lambda_1 \|x_i - c_j\| \leq \|x_l - c_j\| \leq \lambda_2 \|x_i - c_j\|$  represents the distance between  $x_i$  and  $c_j$  should greater than the distance between  $x_l$  and  $c_j$  multiply by  $\lambda_1$  and less than that multiply by  $\lambda_2$ .

Step 3: Recalculate centroid as the center of all data samples in the same cluster.

$$c_j := \frac{\sum_{i=1}^m 1\{u_i = j\} x_i}{\sum_{i=1}^m 1\{u_i = j\}}$$

Step 4: Repeat step2 and step3 until each centroid in the clusters does not change or changes less than a threshold value.

## 4 Experimental Results

In this section, firstly, we described the data sets used in this paper and given a brief introduction to the experimental setup. Secondly, we compared the validation performance of the K-means algorithm and the modified K-means algorithms on the data sets.

### 4.1 Data set Description and Experimental Setup

- Synthetic data sets description

The synthetic data sets were generated by the following formula:

$$X = X_0 + r \cdot \sin \theta$$

$$Y = Y_0 + r \cdot \cos \theta$$

With the fixed  $(X_0, Y_0)$  and radius R, each data sample was generated by firstly choosing a random positive number  $r$  which follows the uniform distribution on  $(0, R)$  and a random  $\theta$ ,  $\theta \in (0^\circ, 360^\circ)$ . Then calculated the corresponding  $X$  and  $Y$ . We created three data sets, each of which had two adjacent clusters. The following Table 1 gives a detail description of the synthetic data sets we used.

**Table 1.** Summary of synthetic data sets

Cluster	$(X_0, Y_0)$	Radius R	Number of data sample
Data set 1 cluster No.1	(2.0,2.0)	0.5	100
Data set 1 cluster No.2	(4.0,4.0)	2.3	400
Data set 2 cluster No.1	(2.0,2.0)	1.0	200
Data set 2 cluster No.2	(4.0,4.0)	1.8	300
Data set 3 cluster No.1	(2.0,2.0)	1.4	250
Data set 3 cluster No.2	(4.0,4.0)	1.4	250

The three data sets we finally used in the experiment are show in Figures 1. (a),(b),(c) are synthetic data set1-3, respectively.

- Experimental setup

We implemented both the K-means and the modified K-means algorithms in Python. The experimental operating system was Windows 7 with Intel Core 2Due CPU 2.0GHZ and RAM 2.0GB. We also implemented the method came from K-means++ algorithm [8] to help us find more receivable initial centroids. This cut down the bad initial centroids' effect on the accuracy. Each of the experiments was executed 10 times to make a credible result. We used parameters  $\lambda_1 = 0.85$  and  $\lambda_2 = 1.15$  after we had compared different parameters pairs on the data sets.

### 4.2 Validation Measures and Results

- Validation measures

As a Clustering algorithm, the goal of K-means is to attain high intra-cluster similarity and low inter-cluster similarity. But this is an internal criterion for the quality of a clustering algorithm. For the data sets we used in this paper, we already know the pre-defined classes. So, we use two external criteria, purity<sup>[9]</sup> and Rand index, to evaluate the performance of the algorithm.

The Purity is a simple and straightforward criterion. When the clustering is done, each cluster is assigned to the class which has the maximal number of data samples in the cluster. Then counting the number of correctly assigned data samples in all the clusters and dividing by the total number of data set to get the purity value. Note that some clusters may assign to the same class. And some classes may not share the maximal number of data samples with any cluster. Purity is computed use following formulate:

$$purity(A, C) = \frac{1}{N} \sum_{k=1}^K \max_{j \in (1, J)} |a_k \cap c_j|$$

Where  $A = \{a_1, a_2, \dots, a_K\}$  represents the set of clusters partitioned by the clustering algorithm and  $C = \{c_1, c_2, \dots, c_J\}$  represents the set of pre-defined classes.  $|a_k \cap c_j|$  represents the number of data samples cluster  $a_k$  and classes  $c_j$  shared.  $N$  represents the total number of data samples in the data set.

The greater the purity value, the better performance the algorithm has.

The Rand index is motivated by classification problem. For a data set which contains  $N$  data sample, there are  $C_n^2$  data sample pairs. And the rand index calculates the faction of correctly clustered data sample pairs to all data sample pairs. It uses a contingency table as follows:

**Table 2.** Contingency table

	Same cluster	Different clusters
Same class	TP	FN
Different classes	FP	TN

Where TP is an abbreviation for true positive, which means two data samples in the same cluster actually come from the same class. FN short for false negative, means two data samples in the different clusters indeed come from the same class. FP short for false positive which means two data sample in the same cluster come from different classes. TN short for true negative which means two data sample in the different clusters come from different classes. Formulate of Rand index is as the following:

$$Randindex = \frac{TP + TN}{TP + FP + FN + TN}$$

Like purity, the higher Rand index value indicates a better algorithm performance.

• Results

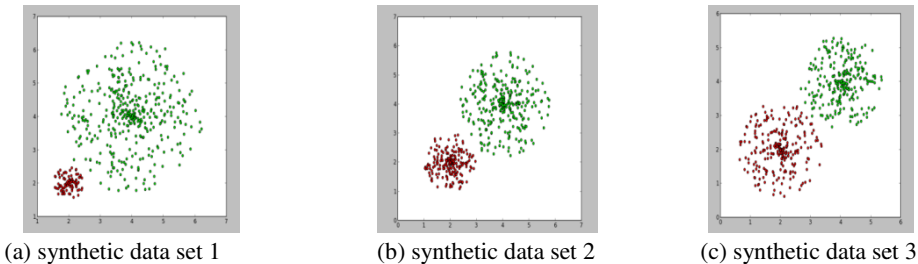
We executed both K-means and modified K-means algorithm 10 times with k=2 on three synthetic data sets. Each time the experiment was run with the same initial centroids for both algorithms. Figure 2 shows the clustering result with initial centroid ((2.415, 2.161), (5.228, 4.640)) on data set 1. On the left side (a) is the clustering

result of K-means algorithm. The data set 1 had been divided to two clusters, represented by red and green colors. We can find that the data samples on the bound of the clusters which also belong to the green cluster had been assigned to the red cluster since they are close to the red cluster centroid. This makes the original small cluster, the red cluster, larger. And the larger cluster smaller. While on the right side (b) is the clustering result of the modified K-means algorithm. The algorithm distinguished the two clusters correctly just as the original distribution on the data set.

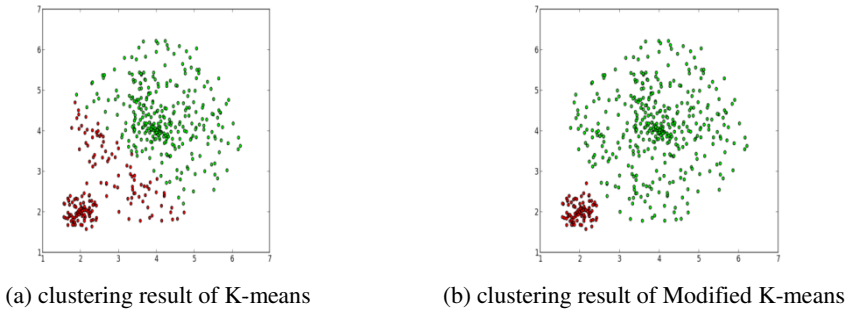
Table 3 is the average performance for both algorithms on three synthetic data sets. From performance of K-means on three data sets, we find that, the K-means get great performance on data set 3 and data set 2 but poor performance on data set 1 with average purity 0.833 and rand index 0.722. This indicated that K-means algorithm

**Table 3.** Purity and Rand index results on synthetic data sets

Data set	Purity		Rand index	
	<i>K-means</i>	<i>Modified K-means</i>	<i>K-means</i>	<i>Modified K-means</i>
Data set 1	0.833	0.958	0.722	0.932
Data set 2	0.975	0.999	0.956	0.999
Data set 3	0.999	0.997	0.999	0.995



**Fig. 1.** Distribution of synthetic data sets. In (a), there are two clusters which are adjacent, and huge difference in size. In (b), the two adjacent clusters are fewer differences in size. And in (c), the two clusters are much similar in size.



**Fig. 2.** Clustering result on synthetic data set 1

couldn't distinguish the clusters with different size. However, the modified K-means algorithm not only performed well on data set 3 and data set 2 as K-means, but also performed great on data set 1 with average purity 0.958 and rand index 0.932. This meant the modified K-means algorithm could make up the weak performance of K-means on data set with adjacent, different size clusters.

## 5 Conclusion

We have proposed a modified K-means algorithm using weighted sum of squared distances as goal function. It is used to improve the clustering result of K-means when the data set has adjacent, different size clusters. In the experimental part, we generated three synthetic data sets and experimented on both algorithms. When using purity and rand index as the validation measure, the experiments result shown that both the algorithms performed well on synthetic data set which had similar size of clusters. However, when the clusters in synthetic data set had adjacent, different size clusters, the modified K-means enhanced purity with 12% and rand index with 20% than K-means algorithm did.

**Acknowledgment.** This work was supported by the National Science Foundation (Grant No. 61133016), the Fundamental Research Funds for the Central Universities (Grant No. ZYGX2012J067), and the National High Technology Joint Research Program of China (Grant No. 2011AA010706).

## References

1. Berkhin, P.: Survey of Clustering Data Mining Techniques (2002)
2. MacQueen, J.B.: Some Methods for classification and Analysis of Multivariate Observation. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistic and Probability, vol. 1, pp. 281–297. University of California Press, Berkeley (1967)
3. Lloyd, S.P.: Least square quantization in PCM. Bell Telephone Laboratories Paper (1957)
4. Banerjee, A., Ghosh, J.: On scaling up balanced clustering algorithms. In: Proceedings of the SIAM International Conference on Data Mining (2002)
5. Abdul Nazeer, K.A., Sebastian, M.P.: Improving the Accuracy and Efficiency of the k-means Clustering Algorithm, vol. I. WCE, London (2009)
6. Chen, Z., Shixiong, X.: K-means Clustering Algorithm with improved Initial Center. In: WKDD, pp. 790–792. IEEE Computer Society, Washington, DC (2009), doi:10.1109/WKDD.2009.210
7. Pelleg, D., Moore, A.: X-means : Extending K-means with Efficient Estimation of the Number of Cluster. In: ICML 2000 Proceedings of the Seventeenth International Conference on Machine Learning, pp. 727–734. Morgan Kaufmann Publishers Inc., San Francisco (2000)
8. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 1027–1035 (2007)
9. Zhao, Y., Karypis, G.: Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering. Machine Learning 55, 311–331 (2004), doi:10.1023/B:MACH.0000027785.44527.d6

# Modeling and Simulation of Network-Based C2 System Based on AOA

Bei Huo<sup>1,\*</sup> and Xin-rong Yang<sup>2</sup>

<sup>1</sup> Xi'an Research Institute of Hi-tech

<sup>2</sup> Shaanxi Xueqian Normal University

Xi'an, P.R. China

lion1109@sohu.com,

renayxr@126.com

**Abstract.** The thesis focuses on Network-Based C2. Firstly, the thesis analyze the workflow of Network-based C2 and key process of building Network-based C2 system. Subsequently, the thesis designed and analyzed the Network-based C2 system based on Agent-Oriented Analysis (AOA) method and Unified Modeling Language (UML). In the paper, the C2 system is regarded as a multi-agent system. The thesis built the model of C2 system and depicted detailedly the elements of system, the static framework and dynamic workflow of C2, the relationship among the elements.

**Keywords:** Network-Based C2, UML, Agent-Oriented Analysis.

Network-Based Command & Control(C2) is a combat command and control system, with which the commander and the command organization can command the information system based on network, aiming to implement the information warfare. By improving the traditional system and workflow, in nature, it maximizes the effectiveness of command network, and enables the commander to master the battlefield situation and deal with battlefield information more quickly and accurately so that the countermeasure will be made. Furthermore, widely dispersed soldiers and diverse weapons can be jointed more flexibly and efficiently to accomplish combat tasks in coordination and synchronization.

Network-based C2 system is a distributed intelligent swarm in dynamic battlefield environment, characterized by the man-machine coordination. The Agent/Multi-Agent System can best solve distributed network-based application problems such as distributed computing, information processing, decision making, and controlling etc. This paper innovatively attempts to apply AOA to the analysis and design of Network-Based C2 System on the basis of UML and the demands in building the system.

---

\* Corresponding author.

# 1 Analysis of Key Points in Realizing Network-Based C2

As shown in Figure 1, the major differences between the Network-Based C2 System and the traditional one are the changes in combat entities' interlink, information exchange and interaction. Combat forces of various armed services in different physical spaces are combined together by intangible data link. Each combat entity, as a node in the system, can be dynamically separated and combined with no affiliation. Military force and organization structure can be tailored for each combat mission based on the information system, possessing higher flexibility in commanding and faster reaction speed, so it will help with the accomplishment of the mission and give full play to each entity's effectiveness. It is this improvement that promotes the combat capability.

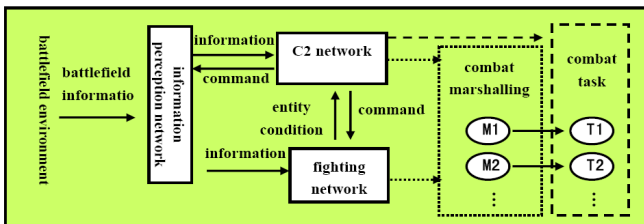


Fig. 1. Network-based command and control

Relying on network information system, building Network-Based C2 System means the close connection between command entity, perception entity, attack entity and other command entities according to mission, and then a network-based intelligent decision set will be formed. With the plug-and-play method, the network is linked by each combat entity so as to establish the seamless link which can shorten the periods of observing, judging, making decision and acting, increase command speed and quicken combat tempo. The core content of building the system is to integrate situational awareness, combat management, task planning and implementation process into the comprehensive and dynamic self-adaption.

# 2 Design of Network-Based C2 System

In order to fully utilize the high-level information-sharing capability of Network-Based C2 System, command levels should be minimized and more authority should be given to the lower-level combat units. At the same time, the multilevel sync coordination decision-making mode should be adopted. In terms of the control way, decentralized control is chosen on the basis of centralization decision due to the changeable situation and frequent emergencies in fighting—the way to accomplish task is relatively independently decided by every combat unit in operation based on the centralization decision.



In the light of analysis above, the design of network-based command and control system is as followings:

The previous tree-like hierarchy structure is changed into the distributed network structure which is wide in horizontal and short in vertical. As for the command and control flow, it is transformed into three-level system which is Command Center—Combat Element Command Entity—Attack Entity.

According to different level and function, the command organization is divided into two levels—command center and element command entity. Command organizations in different levels possess their respective command authority and tasks.

- Command center: It is the highest-level command organization, playing the major role of overall and high-level control and command in battlefield, responsible for acceptance of task order and instruction, battlefield awareness, task planning and distinguishing, marshalling of entities as well as evaluation on the effectiveness of task accomplishment.

- Element command entity: It is a dynamically-set command organization, designed for a particular task, with the main function of controlling and commanding specific operations of combat elements, in charge of synthesizing information from combat elements, controlling and coordinating, commanding etc.

To ensure the sharing of battlefield information and the maximum combat effectiveness, the topological structure of Network-Based C2 System is proposed in this paper, as shown in Figure 2.

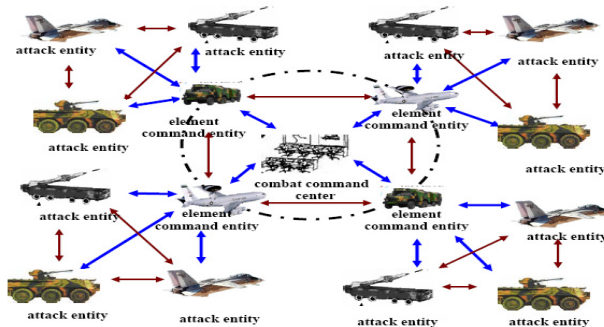


Fig. 2. Topological structure diagram of Network-Based C2 System

### 3 Modeling of Agent-Oriented Network-Based C2 System

#### 3.1 Internal Function Model of Network-Based C2 System

Network-Based C2 System is composed by command center, element command entity and attack entity. The command and control system can be regarded as a MAS system in which the role of Agent = <command center, element command entity, attack entity>. As shown in Figure 3, AUML use case diagram describes the basic function of each Agent role in the system and establishes the internal function model of the system.



Fig. 3. Use case diagram of Network-Based C2 System

3.2 Architecture Model for Network-Based C2 System

Each role in Network-Based C2 System can be regarded as Agent class. Besides, based on the relation between use cases of different roles, the correlation between categories can be described, the class diagram of “entity-responsibility-relation” can be drawn, and the architecture model of command and control system can be built as shown in Figure 4.

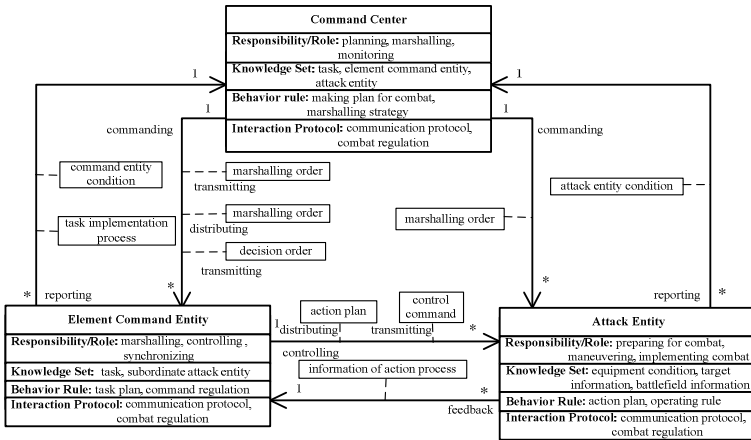


Fig. 4. Agent class diagram of command and control system

3.3 Internal Interaction Model of Network-Based C2 System

Based on the basic workflow of Network-Based C2 System in normal or ideal situation, the interaction model between Agents in the command and control system can be built with AUML interaction diagram as shown in Figure 5.

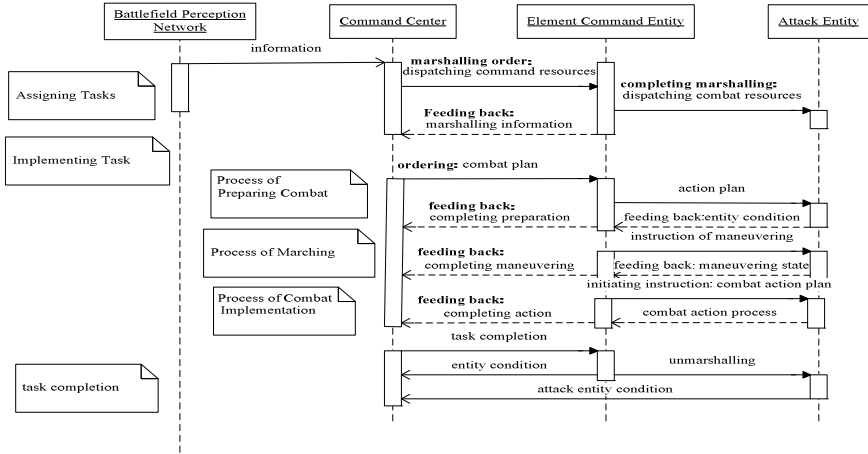


Fig. 5. Agent sequence diagram of Network-Based C2 System

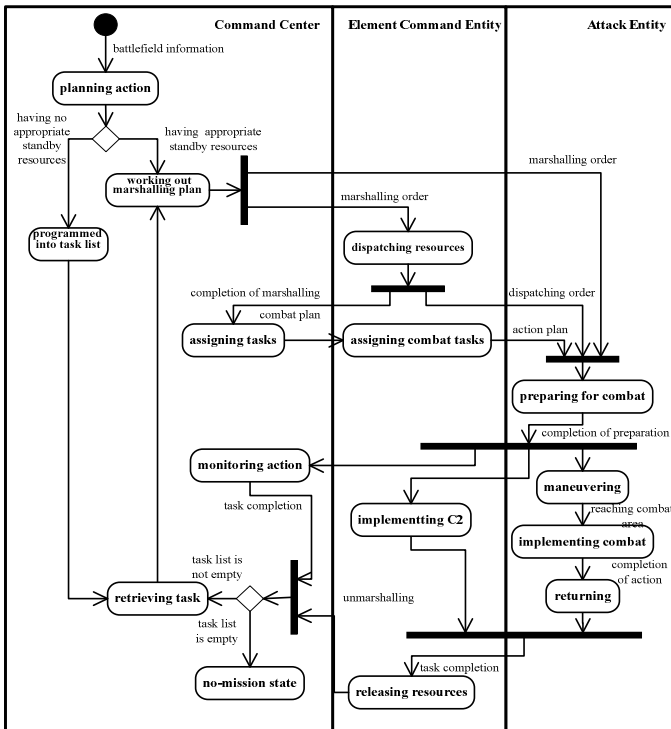


Fig. 6. Activity diagram of Network-Based C2 System

### 3.4 Internal Behavior Model of Network-Based C2 System

Based on the interaction diagram, the activity diagram with swim lane can be used to describe concurrent activity and synchronization requirements of Agents and establish the internal behavior model of Network-Based C2 System, as shown in Figure 6.

## 4 Conclusion

Nowadays, the new military revolution based on information technology has brought unprecedented opportunities and challenges to army. Network-Based C2 System is proving itself the main stream, representing the future development trend of command and control system, based on which a new command and control system can be built. Thus it is necessary to achieve network-based command and control for the construction and development of military information. On the basis of detailed analysis on the concept, features and key points of implementing network-based command and control, this paper proposes the design for Network-Based C2 System. Besides, modeling analysis and design of the system is conducted with Agent-oriented method and AUML language, and also the simulation analysis model is built. Furthermore, this paper elaborates the static structure, control mechanism of dynamic behavior, basic comprising elements as well as the link and interaction between elements of the command and control system. The research results have certain reference and instructive significance for simulation study on network-based command and control and the military information construction.

## References

1. Zhang, O.-Y.: Cruise Missile C3I System Architecture Analysis based on AUML. *Fire Control & Command Control* 31(10), 33–36 (2006)
2. Xiu, B.-X.: The Model and Methodology of C2 Organization Structure Adaptive Optimization. *Military Operations Research and System Engineering* 25(1), 35–40 (2012)
3. Guo, Y.-R.: Modeling about Systematic Operational Capability Based on Network Theories. *Fire Control & Command Control* (4), 26–29 (2012)
4. Deng, X.-Y.: Design and Research on Distributed Support Task Planning and Scheduling Based on Hybrid Multi-Agent System. *Journal of North University of China (Natural Science Edition)* 37(4), 437–442 (2012)
5. Zhang, J.-Y.: A Review on the Methods for Networked C2 Organization Design. *Electronics Optics & Control* 19(5), 63–66 (2012)
6. Garley, K.M., Lin, Z.: Organizational design suited to high performance under stress. *IEEE Transactions on Systems, Man, and Cybernetics* 25(2), 221–231 (1995)

# Application Study of Spatiotemporal Chaotic Sequence in Code Division Multiple Access Communication System

Li Nan\* and Tang Quan-bin

School of Information Engineering,  
Northeast Dianli University,  
JiLin City, JiLin Province, China, 132012  
jllinan@163.com,  
tqb1989@qq.com

**Abstract.** This paper presents a design plan of spatiotemporal chaos system based on the coupled map lattice model, studies the spatiotemporal chaotic sequence of balance, relevance and complexity generated by the model, and applies it into code division multiple access communication system as a spread spectrum sequence. The scheme discussed the effect on the system of multi-user interference and channel interference under different user number, the simulation results show that spatiotemporal chaotic spread spectrum sequence can significantly reduce multi-user interference of communication system, improve the transmission quality of the system and increase the capacity of the system, improving substantially the system performance.

**Keywords:** Spatiotemporal Chaotic Sequence, Coupled Map Lattice Model, Pseudo Random Sequence, Code Division Multiple Access Communication System.

## 1 Introduction

Spread Spectrum Communication has been widely used depend on its advantages of anti-interference, anti-multipath fading, low transmit power density, a low probability of intercept, and realizing the code division multiple access function. Spectrum address code in CDMA communication system still adopted the traditional PN code sequence, but the PN code sequence gradually exposed the disadvantages of few code sequence number, periodic, low complexity, leading to a bad security of communication system, causing a threat to the security of the communication process, thus it cannot satisfy the requirements of the modern communication. Study found that improving the traditional PN sequence cannot effectively solve these problems, looking for a better performance spread spectrum code is the best way to improve system performance [1].

In recent years, the development of chaos theory opens up a new road for the selection of spread spectrum address code. Chaotic systems has sensitivity of the initial value, the tiny differential initial value can produce two different sequences, we only

---

\* Corresponding author.

need to change the initial value or other parameters value to get a huge number of chaotic sequences. Chaos system is a nonlinear system, the generated chaotic sequence has highly randomness and high complexity, which can provided the safeguard for secure communications, and chaotic sequence power spectrum has the character of continuous broadband, making it has better anti noise performance, so it is very suitable for spread spectrum code division multiple access communication [2-4].

At present, the chaotic spread spectrum sequence is generated mainly by Logistic mapping, Kent mapping, Chebyshev mapping, etc. But we find that we can use phase space reconstruction or neural network method to extract the characteristic signal, infer the related properties of low-dimensional chaotic, which greatly reduce the security of low dimensional chaotic communication [5]. Theoretical studies have found that the chaotic sequence generated by high dimensional chaotic system not only has the chaotic behavior in time direction, but also in the space direction during the process of system development, which means it has spatiotemporal chaotic characteristic. Spatiotemporal chaotic sequence has the broadband spectral characteristics similar to the noise and autocorrelation properties of impulse type [6]. We use it as the spread spectrum sequence in spread spectrum communication, which can not only provides conditions for multi-user communication, but also realizes encryption [7]. It has broad application prospects in multiple access communication and safety communication.

## 2 Spatiotemporal Chaotic Sequences

Spatiotemporal chaos system is sensitive in terms of initial value and boundary conditions; mathematical model usually use the partial differential equation form:

$$\frac{\partial u}{\partial t} = f(u) + D\nabla^2 u \quad (1)$$

Where,  $f(u)$  indicates the response term,  $D\nabla^2 u$  indicates the diffusion term.

Theory analysis shows that the solving process of coupling differential equation model is too complex; coupling is a difficult problem to solve. In order to facilitate the theoretical analysis and numerical calculation, the effective method is to use partial differential equation were discretized in time and space domain, but state variables still remain continuous. Coupling image grid model is constructed. The number of space lattice generated by the coupled map lattice model is more, and select any lattice point can constitute an independent chaotic system. Each system has a huge difference, we can get enough number sequence as spread spectrum sequences, thus the cross correlation is better.

### 2.1 Spatiotemporal Chaos Model – A One-Way Coupled Map Lattice Model

One-way Coupled Map Lattice, OCML is defined as:

$$\begin{aligned}
 x_n^1 &= x_n \\
 &\vdots \\
 x_{n+1}^i &= (1 - \mathcal{E})f(x_n^i) + \mathcal{E}f(x_n^{i-1}), \quad i = 1, 2, \dots, L
 \end{aligned}
 \tag{2}$$

Where, n for discrete time variable; i for space coordinates of mapping lattice, L for the system space size;  $f(x)$  indicates the local dynamics of mapping (nonlinear);  $x_n$  for the system drive sequence;  $\mathcal{E}$  is the coupling coefficient.

### 2.2 Spatiotemporal Chaos Sequence Generation

Coupling image grid model based spatiotemporal chaos sequence principle block diagram, as shown in figure 1:

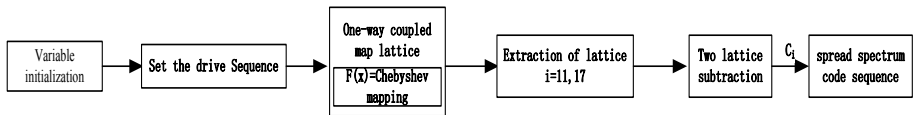


Fig. 1. Generating principle diagram of spatiotemporal chaotic sequence

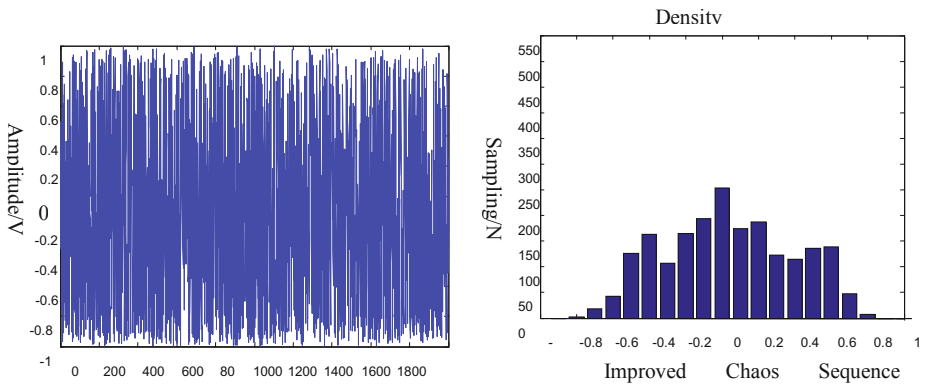


Fig. 2. Time-domain oscillogram

Fig. 3. Density distribution figure

$f(x)$  in chaos system takes Chebyshev chaotic mapping. System space dimension  $L = 30$ , the iteration length for 1000, the iterations number is 5000. In order to avoid the effect of the transition process, the fractal parameter  $r$  is 4 and coupling coefficient  $\mathcal{E}$  is 0.95, the time domain waveform of Spatiotemporal chaos sequence diagram as shown in figure 2, the probability density distribution is shown in figure3.

### 3 Experimental Simulation Analysis

This article applies Spatiotemporal Chaotic sequence to CDMA spread spectrum communication system for single user and multi-user, and carries out simulation analysis. Chart diagram as shown in figure 4:

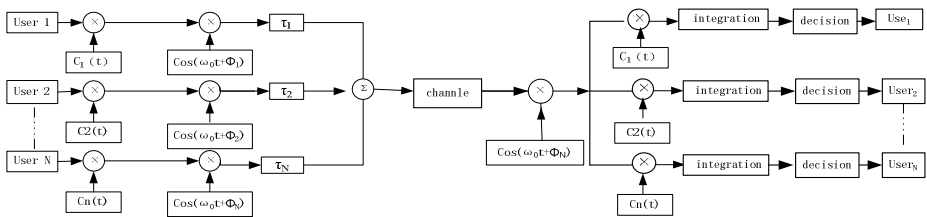


Fig. 4. Block diagram of chaotic spread spectrum CDMA communication system

#### 3.1 Simulated Analysis of System Performance with Different Users

Modeling of CDMA communication system, channel for white Gaussian noise channel, respectively using Logistic sequence and chaotic spatiotemporal Chaotic sequence as spread spectrum code, spread spectrum sequence length  $N$  are all 1000, the system users is set to be 1, 5, and 10, system performance comparison diagram as shown in figure5. We can know from the figure that when using the Logistic chaotic sequence as spread spectrum sequences, as users increases, the system bit error rate will gradually increase, but as for spatiotemporal chaotic sequence, the system error rate keep the same with the increase of users number, which can effectively reduce the multiple access interference (MUI), improve the system capacity.

#### 3.2 Simulated Analysis under Same Users Number with Different Spreading Code

Spread spectrum code length  $N = 1000$ , under the situation of the same users number (10 users), figure 6 shows the performance comparison chart of CDMA communication system used by different spread spectrum code.



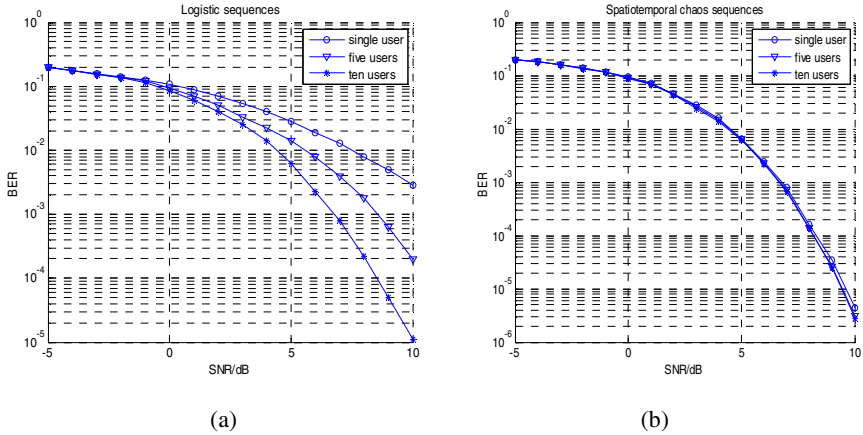


Fig. 5. BER performance chart under different users

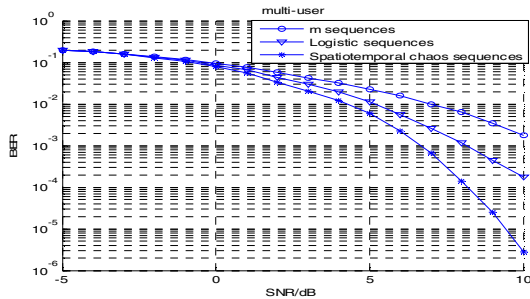


Fig. 6. System performance comparison figure with multi user (users=10)

It can be seen from figure 6 that the BER performance reduces with the increase of signal-to-noise ratio, when the input signal to noise ratio is larger than 0 db, the best performance of system with spatiotemporal chaotic sequence as spread spectrum code.

### 3.3 Simulation Analysis of System Performance under Different Multipath Fading

Considering multipath transmission case, the system simulation model was established under Rayleigh fading environment. The performance of the system is shown in figure7. The system performance is almost the same by using different chaotic code as spread spectrum code under single user. While under multi-user situations, the system performance has relatively improved, the multipath interference resistance ability is enhanced.

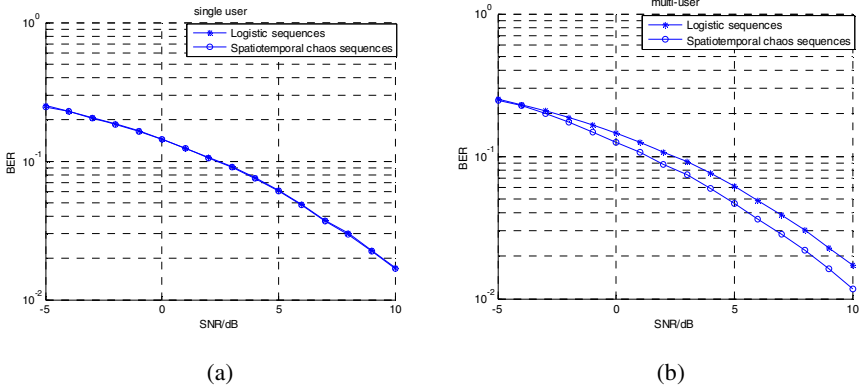


Fig. 7. System performance comparison figure under multipath fading

## 4 Conclusions

The paper established the mathematical model of spatiotemporal chaotic sequence based on one-way coupled map lattice. From the theoretical analysis and simulation prove good pseudorandom properties of spatiotemporal chaos spread, so it is very suitable for spread spectrum communication system. Spatiotemporal chaotic spread spectrum sequence has high complexity. Under the premise of greater assuring of secrecy performance, it still has the ideal autocorrelation and cross-correlation properties, and sequence cycle can be very large. The performance analysis of applying into CDMA shows that it can be used as a optimization object of mobile communication address code, so as to improve the quality and capacity of spread spectrum communication system.

## References

1. Suneel, M.: Chaotic Sequences for Secure CDMA. In: National Conference on Nonlinear Systems & Dynamics, February 6-8, pp. 1-4 (2006)
2. Wang, F.-P., Wang, Z.-J., Guo, J.-B.: Comments on some aspects about chaotic communication. *Journal of China Institute of Communications* 23(10), 71-79 (2002)
3. Kolumban, G., Kennedy, M.P., Chua, L.O.: The role of synchronization in digital communications using chaos. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 44(10), 927-936 (1997)
4. Lau, Y.-S.: Sch. of Electr. & Comput. Eng., RMIT Univ., Melbourne, VIC; Jusak, J.; Hussain, Z.M. Blind Adaptive Multiuser Detection for Chaos CDMA Communication. In: *TENCON 2005*, vol. 10, pp. 1-5. IEEE Region (November 2005)

5. Yu, H., Leung, H.A.: comparative study of different chaos based spread spectrum communication systems. In: The 2001 IEEE International Symposium on, vol. 2, pp. 213–216 (May 2001)
6. Abel, A., Schwarz, W.: Chaos communication-Principles, Schemes, and System Analysis. Proceedings of the IEEE 90(5), 691–710 (2002)
7. Andrievskii, B.R., Fradkov, A.L.: Control of Chaos: Methods and Application. Automation and Remote Control 65(4), 505–533 (2004)

# OPRF Secure Computation for Safety Neighbor Verification Protocols of Wireless Sensor Networks

Mei Meng<sup>\*</sup>, Xu Zhongwei, and Zhu Yujun

School of Electronics & Information Engineering,  
Tongji University, Shanghai 201804, China  
{mei\_meng, xuzhongweish, dyzhuyujun}@163.com

**Abstract.** Due to the resource limitations of sensor nodes, providing security protocols is a particular challenge in sensor networks. A popular proposed method is the neighborhood based key agreement protocol (NEKAP), which is an improvement over the well-known Localized Encryption and Authentication Protocol (LEAP). NEKAP is an efficient and light-weight protocol, but includes loopholes through which adversaries may launch replay attack by successfully masquerading as legitimate nodes and thereby compromise the communications over the network. In this paper, we present a modified security protocol for wireless sensor networks. Secure computation protocol for Naor-Reingold pseudorandom function was studied and a more efficient protocol was proposed based on multiplicative homomorphic encryption in this article. Based on the novel protocol, two private pattern matching protocol was designed. Also the improvement on the performance is discussed and analyzed on several typical attacks found in wireless sensor networks, i.e., replay attack. The performance verification through using a qualitative analysis indicates that our new security protocol can enhance the security resilience of wireless sensor networks better than the conventional methods.

**Keywords:** Wireless sensor networks, Neighbor verification protocols, OPRF, Security protocol.

## 1 Introduction

Wireless sensor networks (WSNs) are distributed systems consisting of a large number of sensor nodes and a base station as a controller which interface the sensor network to the outside network. WSNs may be deployed in unattended and adversarial environments such as battlefields. Compared to conventional networks, they are more vulnerable to physical destruction and man-made threats. Therefore, providing security is a particular challenge in sensor networks due to the resource limitations of sensor nodes, wireless communications and other related concerns. Thus, the key management protocols for sensor networks are based upon symmetric key algorithms, and the design of the security protocols for WSNs should be as light-weight as possible[1-2].

---

<sup>\*</sup> Corresponding author.

Pseudorandom function (PRF)[3] is a kind of efficiently computable function with private key. For randomly chooses private  $k$ , without any information of  $k$ , the value of the PRF is indistinguishable with the value of real random function. Oblivious PRF computation (OPRF) protocol is a two party protocol to securely compute the value of pseudorandom function[4]. In the protocol, one party named server hold the private key  $k$ , and the other party named client hold the input  $x$ . The parties jointly compute the function  $f_k(x)$  through interaction and finally the client gets the result. In the computation progress like in the oblivious transfer protocol[5] and the oblivious polynomial evaluation protocol[6] the server can't get any useful information about  $x$  and the client can't get any information about  $k$ .

## 2 Definition and Basic Tools

### 2.1 Multiplication Homomorphic Encryption Scheme

Homomorphic encryption scheme relative to multiplication is composed of tuple  $(Gen, Enc, Dec)$ , where  $Gen(1^k)$  input security parameter  $1^k$  and output key pair  $(pk, sk)$ ;  $Enc(pk, m)$  input  $pk$ , plaintext  $m \in \{0,1\}^l$ , use random number  $r$ , and output ciphertext  $c$ ;  $Dec(sk, c)$  input  $sk$  and  $c$ , output plaintext  $m \in \{0,1\}^l$ .

The homomorphic feature of encryption relative to multiplication is such that if given two ciphertexts  $c_1 = E(m_1)$  and  $c_2 = E(m_2)$ , a ciphertexts of  $m_1 \cdot m_2$  can be efficiently compute as  $c' = E(m_1) \cdot E(m_2)$  without knowledge of the secret decryption key. As other public key encryption, homomorphic encryption scheme must have semantic security. We show it through a game: a adversary  $A$  choose two plaintext  $m_0, m_1$ ,  $A$  send the message to the user for encryption, user randomly choose  $b \in \{0,1\}$ , encrypt  $m_b$  to  $c = E(m_b)$  and send  $c$  back to  $A$ , then  $A$  give guess of  $b$  as  $b'$ . Then the probability of correct of the guess is negligible.

### 2.2 Naor-Reingold Pseudorandom Function

The Naor-Reingold PRF is defined as  $F_{PRF}(k, x) = g^{a_0 \prod_{i=1}^l a_i^{x_i}} \pmod Q$ . The key of the function is  $k = (Q, g, n, a_0, a_1, \dots, a_l)$ , where  $Q, n$  are primes and satisfy  $n|Q-1$ , and  $g \in Z_Q^*$ , the rank of  $g$  is  $n$ , and  $a_0, a_1, \dots, a_l$  are random element in  $Z_n^*$ . The output distribution of the function is indistinguishable with the uniform random distribution on group  $\langle g \rangle$ .

## 3 OPRF Secure Computation Protocol

### 3.1 Protocol Construction

Protocol  $\Pi$ OPRF that securely compute Naor-Reingold Pseudorandom Function is constructed as below:

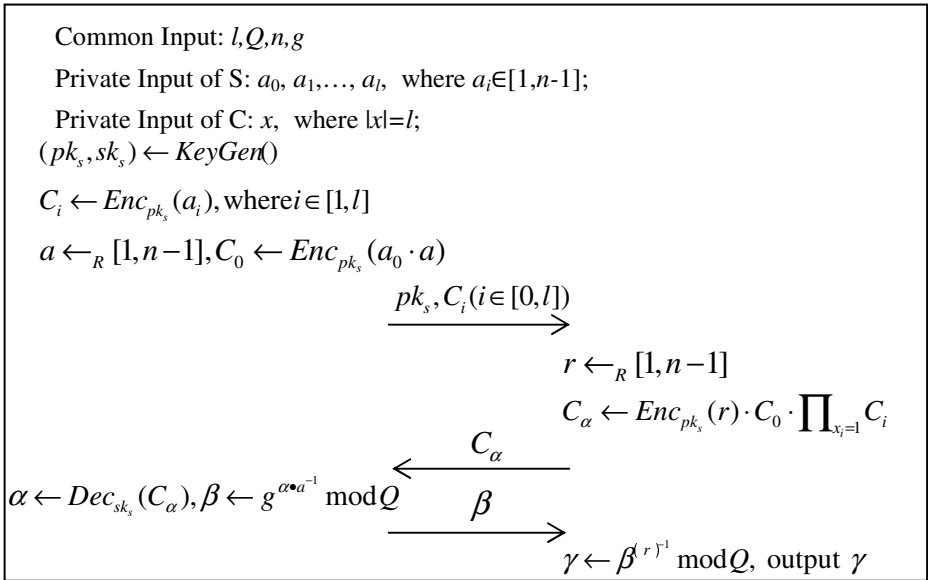


Fig. 1. OPRF protocol construction

### 3.2 Security and Efficiency Analysis

First we give a lemma about the distribution of random variable assembly:

**Lemma 1.** If assembly  $X = \{X_w\}_{w \in S}$  and  $Y = \{Y_w\}_{w \in S}$  are computationally indistinguishable, then assembly  $X' = \{r_w^X, f_w(X_w)\}_{w \in S}$  and  $Y' = \{r_w^Y, f_w(Y_w)\}_{w \in S}$  are computationally indistinguishable, in which  $f_w(w \in S)$  is function sequence that can be computed in polynomial time, and  $r_w^X, r_w^Y$  are random numbers for computing  $X', Y'$  which are uniformly distributed.

We prove the security of the OPRF protocol in the figure 3:

**Theorem 1.** The OPRF protocol  $\Pi$ OPRF in figure 1 securely compute the functionality OPRF under static semihonest adversary model, if a homomorphic semantically secure encryption scheme relative to multiplication exists.

The server is corrupted: In this case, the simulator has  $a_0, a_1, \dots, a_l$  as the input of the server. The view of the adversary includes the input of the server, the random tape, and the single message from the client. It is enough to simulate  $\langle a_0, a_1, \dots, a_l, R^s, C_\alpha \rangle$ , in which  $R^s$  denote the random tape of the Server. The simulator compute  $pk_s', C_i'$  ( $i \in [0, l]$ ) according to the normal step described in the protocol and randomly choose  $x'$  to simulate the input of the client  $x$ , it then compute  $C_\alpha'$  relying on  $x'$  like the client. It is assumed that the corresponding plaintext of  $C_\alpha$  and  $C_\alpha'$  are  $\alpha$  and  $\alpha'$ , respectively. Because  $\alpha$  and  $\alpha'$  are randomized by  $r$  and  $r'$ , their distributions are uniform and indistinguishable. And because  $C_\alpha$  and  $C_\alpha'$  can be computed by  $\alpha$  and

$\alpha'$ , respectively, from lemma 1 we can get that simulated view  $\langle a_0, a_1, \dots, a_t, R^s, C_{\alpha}' \rangle$  and real view  $\langle a_0, a_1, \dots, a_t, R^s, C_{\alpha} \rangle$  are indistinguishable.

Relying on the above discussion it holds that protocol  $\Pi$ OPRF securely realize the functionality OPRF.

### 4 Performance Evaluation in Benign Setting

To evaluate the performance of our protocol in a benign setting, we performed the following simulations. We distributed from 80 to 480 nodes uniformly in a field measuring 400x400m, assigning the nodes' transmission range to be 100m. We repeated this 1000 times. The unit disk graph (UDG) model has been used for determining neighbors of nodes. The values of the maximum distance estimation error  $e$  (as percentage of maximum range  $R$ ) reflect. The value of  $\epsilon_{quad}$  is set to  $2e$ .

**Coverage.** Links have to satisfy the convex quadrilateral test to be verified by our protocol. Yet, even in a benign setting, some links might not belong to any convex quadrilateral, and therefore remain unverifiable. The coverage of the verification protocol is defined as the percentage of verifiable links. Figure 2 shows the coverage of the protocol vs. the average degree of nodes (i.e., average number of nodes that can be ranged by each node). For the networks with average node degree above 7, the coverage is more than 90%. We also note that only a negligible fraction of links cannot be verified due to the distance estimation error.

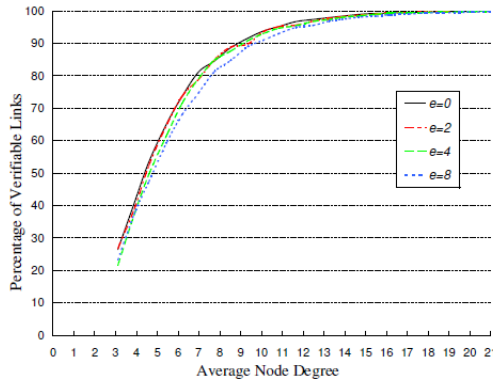


Fig. 2. Coverage of the protocol

**Computational Complexity.** To estimate the computational complexity of the protocol, we count the "number of tested 4-cliques" per node. Figure 3 shows the number of tested 4-cliques vs. the average degree of nodes, for various distance measurement error values  $e$ . As it is shown, the number of tested 4-cliques for each node is almost linear to the number of its neighbors (i.e., constant per link).

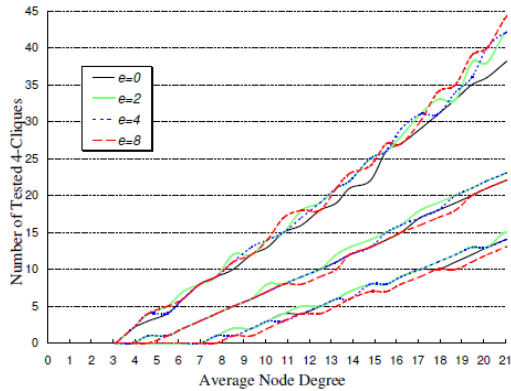


Fig. 3. Computational complexity of the protocol

## 5 Conclusions

We have designed a secure neighbor verification protocol tailored for wireless sensor networks. To demonstrate its applicability to WSN, we have provided a proof-of-concept implementation on existing off-the-shelf hardware (Cricket motes). We have proved that the protocol is secure against the classic 2-end wormhole attack. Yet, our scheme is also effective against more complex relay attacks, as we have demonstrated with simulations in the associated technical report.

**Acknowledgements.** The authors are grateful to Internet forum friends for many interesting discussions about the topic of our article. The authors also thank the anonymous referees for their thoughtful reviews that helped in meaningfully improving the initial version of this paper. Finally, this research is partially supported by the National Natural Science Foundation of China (Grant No.61075002).

## References

1. Zhibin, L.: An Intelligent WPICC Evaluation Model Based on DEA-ANN Hybrid Algorithm. *Journal of Computational Information Systems* 6(13), 4369–4377 (2010)
2. Lu, Z., Mian, C., Chen, L.I.: Software Behavior-Based Trusted Dynamic Measurement. *Journal of Wuhan University (Natural Science Edition)* 56(2), 133–137 (2010)
3. Lu, Z., Mian, C., Shen, C.: Trusted Dynamic Measurement Based on Interactive Markov Chains. *Journal of Computer Research and Development, Chinese* (8), 1464–1472 (2011)
4. Xiaohui, Y., Xuehai, Z., Junfeng, T., et al.: Novel Dynamic Trusted Evaluation Model of Software Behavior. *Journal of Chinese Computer Systems* 31(11), 2113–2120 (2010)
5. Li, K., Gong, L., Kou, J.: Predicting software quality by fuzzy neural network based on rough set. *Journal of Computational Information Systems* 6(5), 1439–1448 (2010)
6. Li, Z., Wu, J., Wu, W.: Power customers load profile clustering using the SOM neural network. *Automation of Electric Power Systems* 32, 66–70 (2008)



# Texture Terrain Based on Land Cover Distribution

Du Jinlian, Xing Lifei, and Wang Song

Software Engineering Department, College of Computer Science,  
Beijing University of Technology  
dujinlian@bjut.edu.cn,  
xinglifei\_xiaoxiao@126.com

**Abstract.** How to reduce the texture data set need to be transmitted is a key technology for network based 3D GIS application. in this paper, We introduced the concept of land cover distribution map(LCD map) which represents the distribution characteristic of land cover displayed in satellite image and design the algorithm for blending a set of small texture samples under the control of LCD map to produce terrain texture close to satellite image in run time. Contrast to the method which directly mapping satellite image to terrain, our method greatly cut down the amount of data needed to be transmitted, reduce the data transmission delay and therefore provides visual continuity, uniform frame rate. Another merit of our method is that it can generate terrain texture with right resolution for different flight height without extra data needed. A potential advantage of our method need to be research is that it can create texture change with time which is valuable for space-time data simulation.

**Keywords:** texture synthesis, satellite image, LCD map, terrain simulation, network based GIS application.

## 1 Introduction

The characteristics of GIS application, such as virtual tour are often base on internet with Browser/Server(B/S) or Client/Server( C/S) architecture. The basic requirement of the application with this architechture is to guarantee continuous of rendering. So if terrain textures have been considered here, research mainly focus on how to organize the single large-scale image data and how to compress it.

In this paper, we propose a novel procedural method which is able to synthesis terrain texture for GIS application. Our contribution focus on two aspects: one is to define the content and storage format of LCD map, the other is to design the method for texturing terrain based on LCD map. LCD map that represents the main rules of land cover is constructed with many regions filled with uniform value, this makes it can be represented by a compact format with much less space requirement. We define the value of each pixel in LCD map with one byte and store the whole LCD map in one channel of an image. In run time, blending algorithm is adopt to generated terrain texture. To smooth the transition of the boundary between two regions, the boundary is filled by searching the nearest pixels of two region's samples along four directions

and then blending them. Texture terrain based on LCD map provides the following advantages over mapping terrain with satellite image directly: Low data transmission delay, Less space requirement and Flexibility for generating higher resolution texture.

## 2 Previous Works

There are two main categories mechanisms for terrain texturing: large single texture, and example-based textures.

Using a large single texture, such as satellite image, to superimpose on the geometry is the simplest method for texturing terrain and often used in GIS application. A key problem of this method is the bottlenecks caused by large single texture in real-time applications that often lead to a loss of interactivity. To solve this problem, two kinds of approaches for improving the performance have been proposed: caching and compression. For example, Cline and Egbert[1] treat one large-scale texture as a bandwidth-limited resource and present a software solution to maintain real-time performance of terrain rendering. Davis and Jiang[2] combine the out-of-core visualization with terrain datasets, use texture paging mechanism for large-scale external texture data to get significantly improved data access and quality of presentation. The clipmap[3] proposed by Tanner and Migdal et al. can efficiently caches textures of extremely large size such as needed by earth-wide visualization. Dollner et. al[4] organize the texture into a texture tree and cache them according to the visual texture error. To reduce the size of data set need to be transmitted, texture compression technology is often used to reduce the size of texture data sets. The popular one is ARB\_texture\_compression [5] which supported by almost all graphics cards to automatically compress the texture image using the appropriate image format, However, the compression rate is too low to really allow for very large textures.

Example-based texture is a powerful mechanism to texture terrain by tiling and blending a set of small texture examples. The common target of this mechanism is to create more realistic terrain texture by blending small samples. Example-based texture has got great development in recent year for its flexibility and low memory usage. One famous method of this mechanism is the splatting technique[6] proposed by Bloom, which can texture terrain by blending a set of small sample textures according to blend weight. Based on splatting technique, many blending methods have been proposed. For example, To solve the feature-agnosticism caused by Bloom's linearly blending, Hardy and Mc Roberts[7] introduce blend maps which are constructed by identifying important features in a texture to obtain more realistic resulting texture in many cases. Zhang et al.[8] try to synthesis terrain texture based on terrain features extracted from multi-resolution mesh model.

Our method is inspired by blending technology to reduce the texture data size used in GIS but generate the real terrain texture, however, there are some key differences. LCD map is not really blend map, the content of LCD map describe the knowledge of actual land cover of the earth reflected by satellite image. Moreover, LCD map express the macro rules for the distribution of land cover in one region.

### 3 Overview

Our method based on the observation of the satellite image: satellite images at whichever resolution all reflect the distribution of land cover and some terrain features by different colors. From the point of image synthesis, the terrain texture represented by satellite image can be regarded as the blending of various texture samples. So if there exist sufficient rules which can guide the behaviour of texture blending, the terrain texture similar to satellite image will be produced. Fortunately, these blending rules are contained in the satellite image itself, as illustrated in figure1, different land cover shows different color region, If we get the rules of land cover distribution from satellite image, we can create the texture in line with practical application.

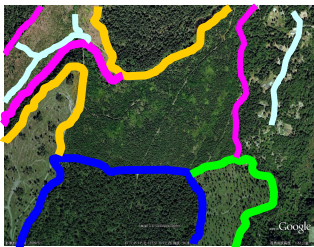


Fig. 1. Land cover distribution expressed by satellite image

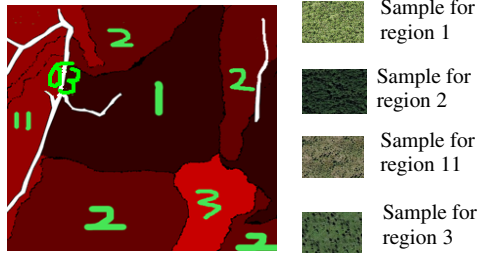


Fig. 2. Land cover distribution map(LCD map) on the left and the corresponding texture samples on the right

In our following work we define and extract the characteristic of land cover distribution from satellite image to build land cover distribution map(LCD map) and use it as the basic rules for texture generating.

#### 3.1 Definition of Land Cover Distribution

As a result of natural forces or role of human beings, there are a variety of land cover on the earth, for example, forest, desert, water, grassland, field, road, snow, etc., reflected on satellite image, different land cover represented by different texture. As show in figure 1, it contains 3 types of forest, two types of grassland and some roads and artificial architectures. To distinguish different type of land cover, we should mark each type of land cover with unique sign which can be number or other kind of identification. To be simple, we use number as the sign, table1 shows an example of the definition of land cover types and their signs. Number 1 to 10 identify 10 types of forests, 11-20 identify 10 types of grass land and so on.

**Table 1.** The signs of different land cover types

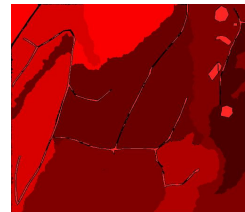
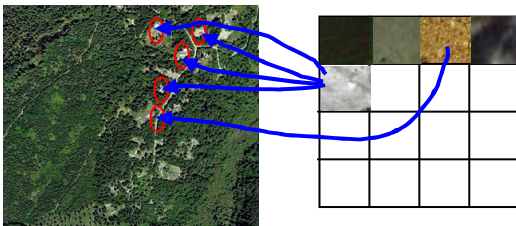
Land cover types	sign	Land cover types	sign
Forest types	1-10	Bare land	32
Grass land types	11-20	Fields	33-40
Artificial architectures	21-25	...	...
Water region	26	Snow	62
Rocks	27-30	Road	63
Desert	31	Boundary between regions	0

Each type of land cover occupies one or more region in the satellite image, which is named distribution characteristics of land cover. The distribution characteristics of land cover can be represented by land cover distribution map(LCD map),as illustrated in figure 2. The LCD map consists of many different regions marked with the land cover type sign defined in table 1, and each region corresponding to a specific land cover. For example, white lines marked with 63 stands for roads, regions marked with 1 to 3 stands for three types of forest cover etc. It should be noted here that there are more content need to be identified if the resolution is high. However, in GIS application, we only pay more attention to the macro distribution rules for land cover, the diversity of the details can be added by other methods. In our method, we use 64 signs for identification without considering city factors.

### 3.2 Samples Organization

In LCD map, each region has one or more texture samples, illustrated as figure 2. The size of sample should be large enough to contain a certain degree structure, for example 64\*64 or 128\*128. Samples are stored in database, the link between samples and regions in LCD map should be recorded.

However, based on the observation of the LCP map, we find that some regions have large area, but others only occupy very small area, especially those lines or points which stand for roads,river or some other features such as buildings. To reduce the data set size of samples, we won't prepare the samples for the line and point shape features in the regular way discribed above. Instead, we regard those small area land



**Fig. 3.** Feature textures which express roads and buildings circled by red line on the left image are packed into one big image as right one

**Fig. 4.** LCP map expressed by red channel of an image

cover as a kind of special features, and represented by feature textures. The feature textures are packed into a small texture sample, as shown in figure 3. The size of feature texture can be  $8 \times 8$ . A sample of  $64 \times 64$  can contain 64 feature textures, which is enough for the application.

### 3.3 Representation of LCD Map

The key to our algorithm is to reduce the data size need to transform. From the discussion above, we know that the LCD map consists of different regions marked with number 0-63. To store the LCD map in a compact format, we use one byte to express the value for one pixel of it. Among the 8 bits in one byte, 6 bits are used to mark the land cover type and 2 bits are used to identify whether the region is special feature, such as road, house, etc. or not. By using this storage format, LCD map can be represented by one channel of an image. The LCD map of figure 2 represented by the red channel of an image looks like figure 4.

This storage format for LCD map greatly reduces the amount of data, and more important the uniform value in one region makes it more conducive to obtain high compression ratio.

### 3.4 Rendering

When rendering, LCD map work as blend map which determine the display of texture samples. In one region, the corresponding samples' blend weight is 1 and other samples' blend weight is 0. At this time, two problems need to be addressed: one is that when a region has more than one sample, how to blend them. The other is how to deal with the border transition.

To increase the variety of the texture, we use more than one sample to generate the result texture for some regions, especial for the grass land regions. Figure 5 show this situation, three regions circled with red line can all be generated by blending two samples on the left.



Fig. 5. Grass land and the samples

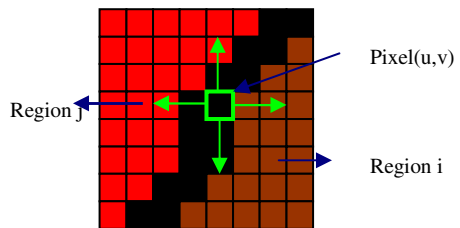


Fig. 6. The search direction along vertical and horizontal for a pixel

We use the following equation to performance the blending of the samples for one region:

$$c(u, v) = \sum_i \alpha_i s_i(u, v) \tag{1}$$

Where  $c$  is the final color of pixel  $(u, v)$ ,  $s_i$  is the  $i$  th sample,  $\alpha_i$  is the weight of  $s_i$ , and its' value is 0 or 1.

Although there is possible that one region has more than two samples, for simplicity we limit the sample number of one region less than 3. You can choose the sample number according to your application. And to get the terrain texture with less structure, we use stochastic process to produce the blending map for these samples.

### 3.5 Border Transition

To eliminate the artefact on the boundary of two regions, blending between the samples of different regions is also executed. In order to make the transition more natural, we use distance-based weight for the blending, the blend equation is:

$$bc(u, v) = \sum_i a_i r_i \tag{2}$$

Where  $bc(u, v)$  is the final color of the boundary pixel  $(u, v)$ .  $r_i$  is the nearest pixel of sample of region  $i$  to pixel  $(u, v)$ .  $\alpha_i$  is blend weight with value between  $[0,1]$  and computed by following equation:

$$\alpha_i = 1 - \frac{near(r_i)}{\sum near(r_i)} \tag{3}$$

$near(r_i)$  is the nearest distance of the vertical and horizontal direction the pixel  $(u, v)$  to the region  $i$ , as illustrated in figure 6.

If a region has more than one sample, the blending equation is:

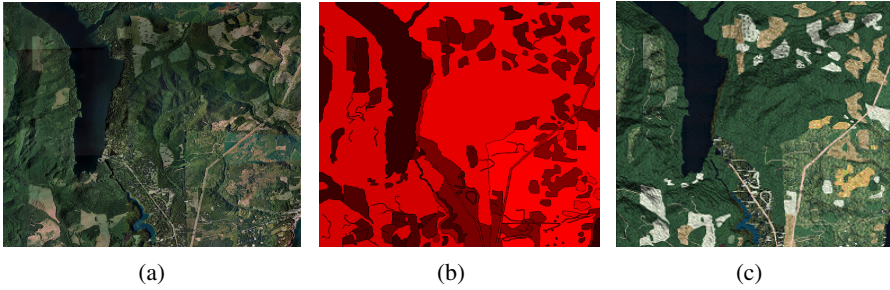
$$bc(u, v) = \sum_i a_i \sum_k \frac{r_{ik}}{n_k} \tag{4}$$

$n_k$  is the number of samples for region  $i$ ,  $r_{ik}$  is the nearest pixel of  $k$  th sample of region  $i$  to pixel  $(u, v)$ .

## 4 Implementation and Result

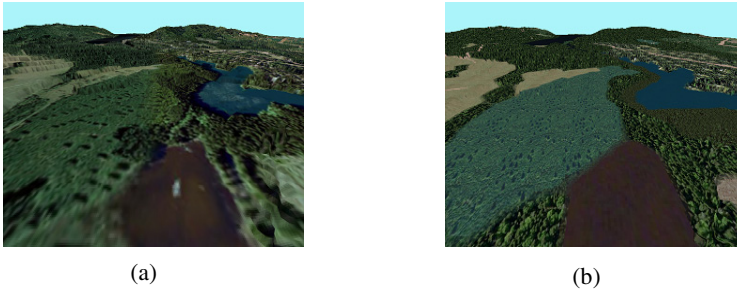
We have experiment our method for the simulation of a part area of The Puget Sound. The LCD map of this area is extracted from the satellite image by PhotoShop under the control of predefined land cover type in table 1. We prepare 10 types texture samples with size of 128×128 and 4 types feature textures with size of 16×16.

Figure7 shows the original satellite image, LCD map, the texture generated with our method. We can see that the result texture generated by our method is more like the original satellite image.



**Fig. 7.** The original satellite image(a) ,LCD map(b), result texture generated by our method(c)

Figure8 shows two snapshots during fly through this terrain at low height. The left side (a) is the result produced by mapped satellite image directly and the right side (b) is the result of our method. With the knowledge expressed by LCD map, we can texture terrain in a similar way to satellite image. It is noteworthy that if the high resolution satellite image is used to texture terrain at low flight height, the data size of texture will become huge. However with our method, only new high resolution texture samples are adopted to add the details, there is no extra data needed. In fact, we can adopt different resolution samples for different flight height.



**Fig. 8.** The snapshots of two methods.(a) is the result of mapping satellite image directly, (b) is the result generated by our method.

Figure9 is the data transmit time, space requirement and frame rate comparison of our method to texture mapping with satellite image directly at the flight height of 2km for one terrain block. In our method, the LCD map is not compressed, the average data transmit time is about 31ms. While if use satellite image as texture, which is represented in JPG format, the average data transmit time is about 116ms. The average fps of the two methods is obtained after all data arrival at client. Our method can obtain 105 fps, a bit slower than mapping satellite image directly. However, long time delay caused by satellite image make computer stop running during data transmission, that further lead to not smooth of rendering. Because our method cuts down the size of dataset need to be transmitted, which greatly deduce the delay of data access, we can obtain smooth frame rate at any flight height.

	Our method	mapping satellite image directly
Average Data transmit time	31ms	116ms
Space requirement	64M	192M
Average frame per second	105 fps	112 fps

**Fig. 9.** The space, frame rate and transmit time comparison of two methods

As for the runtime space requirement, the memory used by geometry data of two methods is the same, however, the memory for textures is very different. In our method, memory size needed for texture data mainly decided by the LCD map, but if use satellite image as texture, memory size will be decided by the resolution of the image. Higher resolution image will need much more memory.

In this paper, we extract the LCD map from the satellite image at height of 2km. thus if we generate the texture at this height, the memory for texture taken by our method is only about 1/3 of that used by directly mapping satellite image. When rendering, there are at least four blocks of geometry data cached in memory, and  $1024^2$  grids in video memory for each frame. If use satellite image directly, there should be 192M image data cached in memory and 48M image data in video memory. But for our method, LCD map only take 64M memory size, and 10 samples cached in memory take about 480K. With an average of 8 samples in video memory, about 16.1M will be taken in video memory for generating texture. This inspired us to cache more LCD map in video memory to improve rendering speed.

## 5 Conclusion and Future Work

Terrain simulation based on network is a wide range of GIS application, in which satellite image is often used as terrain texture. However, the use of satellite image cannot obtain good performance for its' large amount of data. We proposed a real-time method based on LCD map extracted from satellite image to generate terrain texture close to the actual effect of satellite image. Our method uses much less memory space than traditional method which map the satellite image directly to the terrain.

We test our method with a part of Puget Sound data set and get satisfied results which greatly reduced the data transmission delay.

There are still many more problems need to be research in the future: Find the suitable blending algorithm for generating more realistic texture in one region and develop the method of generate complicated land cover, such as city texture under special flight height.



## References

- [1] Cline, D., Egbert, P.: Interactive Display of Very Large Textures. In: Proceedings IEEE Visualization 1998, pp. 343–350 (1998)
- [2] Davis, D., Jiang, T.Y., Ribarsky, W., Faust, N.: Intent, Perception, and Out-of-Core Visualization Applied to Terrain. In: Proceedings IEEE Visualization 1998, pp. 455–458 (1998)
- [3] Tanner, C.C., Migdal, C.J., Jones, M.T.: The Clipmap: A Virtual Mipmap. In: Proceedings of SIGGRAPH 1998, pp. 151–159 (1998)
- [4] Döllner, J., Baumann, K., Hinrichs, K.: Texturing Techniques for Terrain Visualization. In: Proceedings Visualization 2000, pp. 227–234 (2000)
- [5] SGI OpenGL Extension Registry: ARB texture compression & EXT texture compression s3tc, <http://oss.sgi.com/projects/ogl-sample/registry/>
- [6] Bloom, C.: Terrain texture compositing by blending in the frame-buffer (November 2000), <http://www.cbloom.com/3d/techdocs/splatting.txt>
- [7] Hardy, A., Roberts, D.A.K.M.: Blend maps: enhanced terrain texturing. In: Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries, SAICSIT 2006, pp. 61–70. South African Institute for Computer Scientists and Information Technologists, Republic of South Africa (2006) ISBN:1-59593-567-3
- [8] Zhang, H., Ouyang, D., Lin, H., Guan, W.: Texture synthesis based on terrain feature recognition. In: Proceedings of the 2008 International Conference on Computer Science and Software Engineering, vol. 2, pp. 1158–1161. IEEE Computer Society, Washington, DC (2008) ISBN 978-0-7695-3336-0
- [9] Du, J., Wang, J.: Data Transmission and Caching Method for Terrain Rendering on Network. Journal of Beijing University of Technology 35(3), 414–418 (2009)

# Text Mining for Information Screen in Risk Assessment of Environmental Endocrine Disruptive Chemicals

Yong Chen, Yu Wang, Xiaolian Duan, Huatang Zhang, and Yuyue Jia\*

Chongqing Center for Biomedicines and Medical Equipment,  
Chongqing Academy of Science and Technology, Chongqing, P.R. China  
janchenbio@gmail.com, cqjjiawa326@189.cn

**Abstract.** Text mining technique has been widely applied in numerous fields of research, among which, the employment of machine learning in biological text analysis and management has increased its popularity in recent years. In this study, machine learning based automatic classification system has been constructed according to the requirement of the environmental health risk assessment for endocrine disruptive chemicals, as the knowledge explosion has made traditional manual assessment impossible. And the factors that may influence the classifier training and performance were compared and selected. The constructed classifier has been proved to be with high accuracy and efficiency, which has important significance on computer based risk assessment for various potential hazardous chemicals.

**Keywords:** text mining, environmental health risk assessment, machine learning, Naïve Bayes.

## 1 Introduction

In recent years, the popularity of text mining (TM) are increasing in biomedical research, as it assists us in dealing with the tremendous body of biological literature, and the discovery of novel knowledge hidden behind complicated biomedical text [1]. And progress has been made in TM recent years, including increasing volume of annotated corpora, information retrieval (IR) and information extraction (IE) techniques, natural language processing (NLP) techniques, and more challenging tasks, such as, association analysis and the discovery of new information [2]. The emphasis of TM research has been switched from pure theory studies to application of TM techniques to practical tasks, especially in biomedical research.

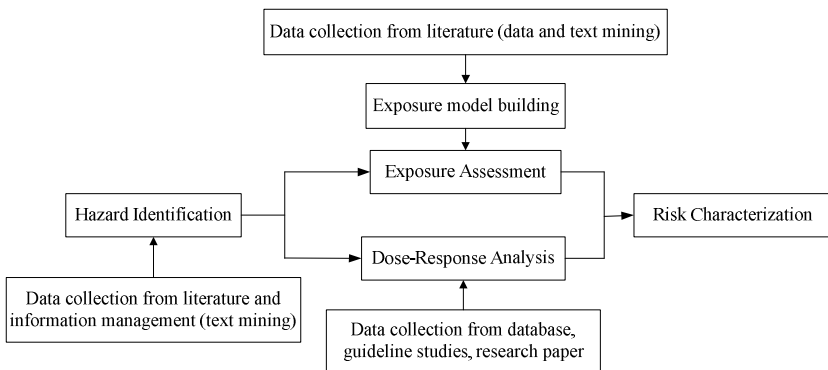
Environmental health risk assessment (EHRA), which involves inspecting published evidence to determine the likelihood of developing adverse effects from potential hazard chemicals, is one of specific task in biomedical area that needs the help of TM [3]. EHRA handles the increasing concern-arousing issue: interaction of pollutant-environment-health [4]. Four steps are included in EHRA process: hazard

---

\* Corresponding author.

identification, exposure assessment, dose-response analysis, and risk characterization, most of which are performed manually by risk assessors with the combination of expert knowledge and extensive literature review [5, 6] (fig.1).

Among EHRA process, the most time-consuming steps are analysis and classification of scientific evidence [7]. Exhaustive literature and data collection can generate consistent, accurate and efficient EHRA results; however, the exponentially growing volume of published literature and data render large burden to risk assessor [8]. Machine learning techniques shift the load from the assessor to computer, which considerably improve the efficiency, save the time required for EHRA, and free assessor for more expert skill needed tasks and more potential hazard chemicals assessment [9]. Moreover, developing standard TM system for EHRA with unified criterion can evade inter-assessor bias and prevent errors.



**Fig. 1.** EPA four-step risk assessment Extensive information collection is needed in first three steps of EHRA task, which influence the accuracy and efficiency of final risk characterization

In this paper, literatures of environmental endocrine disruptive compounds (EDCs) were studied for EHRA purpose, and automatic classification systems were built according to EDCs literatures for information screening and management.

## 2 Literature Screening and Analysis by Machine Learning Based Text Mining

### 2.1 Machine Learning Based Text Screening and Classification

Machine learning which belongs to artificial intelligence has witnessed a booming attention in recent years, and it constructs systems that can learn from training data [10]. In the field of TM, machine learning method has replaced the traditional knowledge engineering (KE) approach, which sets up complicated rules according to expert knowledge about how to implement TM tasks [11]. Regarding to automatic text classification, machine learning process is presented in fig.2, in which, three components are included in machine learning based text classification system: learning, evaluation and classification.

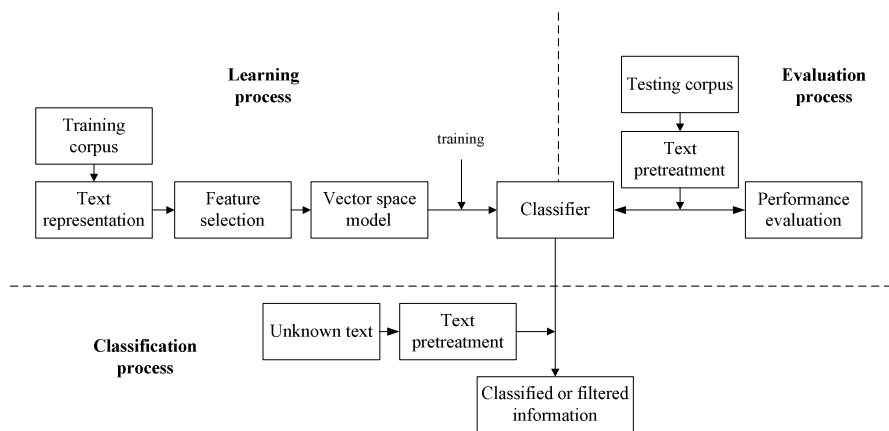


Fig. 2. Machine learning based text classification system

### 3 Manual Data Screening and Pretreatment

#### 3.1 Data and Manual Screening by Expert

Four typical and well-studied EDCs chemicals were selected for classifier training, and 257 documents (including article title and abstract) were downloaded from PubMed in medline format, using the key words “bisphenol A”, “polychlorinated biphenyl”, “cadmium” and “dichlorodiphenyltrichloroethane”.

Documents were manually annotated and screened according to manual classification guidelines and expert knowledge of EHRA. When “relevant” was marked to one document, this document can be used for EHRA purpose in next step; when “irrelevant” was annotated to a document, this document was not suitable for further EHRA analysis. Ultimate decision has been made after discussion: 193 documents were marked as “relevant”, and 64 documents were marked as “irrelevant”.

#### 3.2 Text Representation and Feature Selection

Plain texts were tokenized to single token to form a dictionary of features. And then, stop word filtering and stemming were applied to eliminate features with less information.

Chi-Square test assumes that every feature and class can be described by  $\chi^2$  distribution, and then the contribution of a feature to one class can be presented by  $\chi^2$  statistics. Chi square feature selection was employed in our study for feature dimension reduction, and the  $\chi^2$  value of a feature to a class is described by

$$\chi^2(t, c) = \frac{N(AD-BC)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

Where,  $t$  represents feature,  $c$  is a class;  $A$  is the document frequency that text with  $t$  belongs to  $c$ ;  $B$  is the document frequency that text with  $t$  does not belong to  $c$ ;  $C$  is the document frequency that text without  $t$  belongs to  $c$ ;  $D$  is the document frequency that text without  $t$  does not belong to  $c$ .

## 4 Classifier Building and Evaluation

### 4.1 Classifier Building

The Naïve Bayes (NB) algorithm was applied in the text classification tasks; it deals with questions of text classification by utilizing the joint probabilities of words and categories to filter the category with maximum probability to include a document [12]. The NB function can be written as follows

$$\max_i P(C_i|d) = \max_i P(C_i) \cdot \prod_{w_j \in d} P(w_j|C_i) \quad (2)$$

Where,  $w_j$  represents each word, while,  $C_i$  is the category, and  $w_j$  is conditionally independent of the others in the category of  $C_i$ .  $P(C_i)$  and  $P(w_j|C_i)$  can be estimated from the training corpus. Synthetic minority oversampling technique (SMOTE) was taken to create man-made examples by over-sampling.

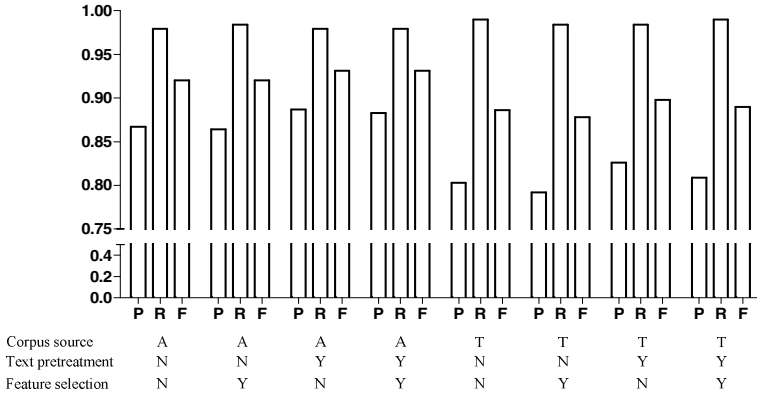
### 4.2 Evaluation

For classifier evaluation, 10-fold cross-validation was employed, and the average precision (P), recall (R), accuracy (A) and F1-measure were calculated. P is the percentage of positive samples that are examined to be positive, while R is the percentage of positive samples that are retrieved. As P and R are not always consistent, the F1-measure which is the harmonic average of P and R, has been used to describe the overall performance of text classifier.

## 5 Results and Perspective

Three factors (corpus source, text pretreatment and feature selection) that may influence the performance of classifier were compared, and each factor contained two levels: for corpus source, corpus was constructed by artical title or abstract; text pretreatment included stop word filtering and stemming, and the two levels consisted of with and without text pretreatment; in feature selection, chi square statistics was employed, and the two levels consisted of with and without feature selection treatment. Results presented in fig. 3 revealed that high recall rate has been achieved in every test group; however, the precision in every group varied, and as a result, the F value which described the overall performance of classifier also varied from group to group, The employment of abstract as corpuse received better results compared the the employment of artical title, as abstract provided more information for classifier

training. Text pretreatment slightly enhanced the classifier performance, which mainly came from stemming treatment (data not shown). Feature selection almost has no influence on classifier performance, but prolonged the time required for model building.



**Fig. 3.** Precision, recall and F1 measure for NB classifiers trained by corpus with various treatments

In our study, the negative samples were far less than positive samples, which came from the data collection principles and annotation guidelines; therefore, imbalanced data set problems may magnify classification errors, especially in small data set. SMOTE was implemented to corpus made of abstract or article title without any pre-treatment or feature selection to make artificial instances, and results was shown in tab.1. SMOTE can largely increase the accuracy and enhance the overall performance.

**Table 1.** SMOTE treatment and classifier evaluation

Corpus source	SMOTE	Category	Classifier evaluation			
			Precision	Recall	F value	Accuracy
Abstract	Y	Relavent	0.931	0.979	0.955	94.39%
		Irrelavent	0.966	0.891	0.927	
	N	Relavent	0.867	0.979	0.920	87.16%
		Irrelavent	0.897	0.547	0.680	
Title	Y	Relavent	0.909	0.824	0.864	84.42%
		Irrelavent	0.767	0.875	0.818	
	N	Relavent	0.803	0.990	0.886	80.93%
		Irrelavent	0.895	0.266	0.410	

To conclude, the selection of corpus source was of great importance to the efficacy of classifier building, and study results proved that abstract corpus was suitable for corpus construction, although it may increase the difficulties in calculation. Stop word

filtering and stemming did not show obvious advantages in classifier training, but they have reduced the dimensionality of feature space vector, and will be considered in future studies. Feature selection slightly influenced the classifier performance negatively; however, it has shown merits in dimensionality reduction and will be studied in following researches. SMOTE has been demonstrated to be an efficient way to enhance the overall performance of NB classifier. Our studies have argued that machine learning based text mining can be applied to EHRA task to help with information classification and screening, and more complicated information filtering system will be developed to meet the requirement of risk assessors.

## References

1. Reibholz-Schuhmann, D., Oellrich, A., Hoehndorf, R.: Text-mining solutions for biomedical research: enabling integrative biology. *Nature Reviews Genetics* 13, 829–839 (2012)
2. Zhu, F., Patumcharoenpol, P., Zhang, C., Yang, Y., et al.: Biomedical text mining and its applications in cancer research. *Journal of Biomedical Informatics* 46, 200–211 (2012)
3. Bascietto, J., Hinckley, D., Plafkin, J., Slimak, M.: Ecotoxicity and ecological risk assessment. Regulatory applications at EPA. Part 1. *Environmental Science & Technology* 24, 10–15 (1990)
4. Backhaus, T., Faust, M.: Predictive environmental risk assessment of chemical mixtures: a conceptual framework. *Environmental Science & Technology* 46, 2564–2573 (2012)
5. Krewski, D., Andersen, M.E., Mantus, E., Zeise, L.: Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Analysis* 29, 474–479 (2009)
6. Klassen, R., Douma, S., Rencz, A.: Environmental and Human Health Risk Assessment for Essential Trace Elements: Considering the Role for Geoscience. *Journal of Toxicology and Environmental Health, Part A* 73, 242–252 (2010)
7. Sun, L., Korhonen, A., Silins, I., Stenius, U.: User-driven development of text mining resources for Cancer Risk Assessment. In: *Proceedings of the BioNLP 2009 Workshop*, Stroudsburg, PA, USA (June 2009)
8. Korhonen, A., Séaghdha, D.Ó., Silins, I., Sun, L., et al.: Text mining for literature review and knowledge discovery in cancer risk assessment and research. *PLoS One* 7, e33427 (2012)
9. Silins, I., Korhonen, A., Högberg, J., Stenius, U.: Data and literature gathering in chemical cancer risk assessment. *Integrated Environmental Assessment and Management* 8, 412–417 (2012)
10. Oquendo, M., Baca-Garcia, E., Artes-Rodriguez, A., Perez-Cruz, F., et al.: Machine learning and data mining: strategies for hypothesis generation. *Molecular Psychiatry* 17, 956–959 (2012)
11. Verhagen, W.J., Bermell-Garcia, P., van Dijk, R.E., Curran, R.: A critical review of Knowledge-Based Engineering: An identification of research challenges. *Advanced Engineering Informatics* 26, 5–15 (2012)
12. Sambo, F., Trifoglio, E., Di Camillo, B., Toffolo, G.M., et al.: Bag of Naïve Bayes: biomarker selection and classification from genome-wide SNP data. *BMC Bioinformatics* 13, S2 (2012)

# The Choices of Issued Inventory Valuation Methods Based on Different Accounting Standards at Home and Abroad

Gu Shusheng

School of Accounting,  
Jiangxi University of Finance and Economics,  
Nanchang, China  
gsswell@sina.com

**Abstract.** This paper, based on the issued inventory valuation methods stipulated in current accounting standards of China, analyzes the different effectiveness resulting from the choices of different inventory valuation method. Taking the actual situation of the international and domestic environment into account, and by adopting the method of case analysis and material analysis, this paper puts forward suggestions and advocates the comprehensive learning of different issued inventory valuation methods under the guidelines of different accounting standards, drawing on their advantages so as to develop new issued inventory valuation methods which are more suitable for the future economic situation.

**Keywords:** inventory, valuation method, current situation, suggestions.

## 1 Introduction of Inventory Valuation Method

The inventory valuation methods are undergoing constant change and development, during which there are several methods as follows: 1) Methods on the condition of non-actual cost: planned cost method, fixed cost method, and selling price method. 2) Methods on the condition of actual cost: first, methods based on time sequence include FIFO (first-in, first-out) method, LIFO (last-in, first-out) method, moving weighted average method, and next-in first-out method. Second, methods based on unit price include specific identification method, high-price, first-out method, low-price, first-out method, and last invoice price method. [1] Third, weighted average at the end of month method, and basic stock method.

Although there are lots of methods of inventory valuation, enterprises are restricted to some degree in the choice of pricing method in terms of inventory ex-factory, no matter in which country's accounting system or the IAS(International Accounting Standards).[2] This paper will analyze the pricing methods in terms of inventory ex-factory based on current accounting standards in our country, combined with United States Generally Accepted Accounting Principles and International Accounting Standards.



## 2 Choices of Issued Inventory Valuation Methods

Rising prices is one of the current hot economic topics. Then what effect can different inventory valuation methods produce in the case of rising prices? In order to make it more intuitive and convenient to judge the results from different inventory valuation methods, what follows will analyze it with an example, together with a table.

Supposing Company A's inventory planning cost is 60, the initial material cost difference is 1000. The issuing and receiving of inventory of the enterprise in January is shown in the following table.

**Table 1.** Company A's Inventory in Jan.1

Inventory Situation	Quantity	Unit price	Balance quantity	Date
Opening Balance		50.00	100	Jan.1
Purchasing	500	60.00	600	Jan.3
Issuing	300		300	Jan.5
Purchasing	200	70.00	500	Jan.14
Purchasing	100	75.00	600	Jan.17
Issuing	400		200	Jan.22
Purchasing	150	80.00	350	Jan.28
Issuing	200		150	Jan.30
Closing Balance			150	

### (1) Specific Identification Method

When the specific identification method is used, the cost flow is completely in consistent with the inventory flow. Thus, this method is the most accurate. The final inventory value and cost of sales are in consistent with the actual situation. Due to the particularity of this method, no comparison will be made here about it, and methods that lead to difference between cost flow and inventory flow will be discussed.

(2) The accounting results from other issued inventory valuation methods are shown in table 2.

**Table 2.** Accounting Results from Different Issued Inventory Valuation Methods

	Current Issued Inventory Cost	Closing Inventory Value
Weighted average at the End of Month Method	58,714.29	9,785.71
Moving Weighted average Method	57,785.71	10,714.29
FIFO Method	56,500.00	12,000.00
Planned Cost Method	58,714.29	9,785.71
LIFO Method	60,500.00	8,000.00

As shown in table 2, in the case of rising prices, different inventory valuation methods can result in different current carry-over inventory cost and inventory balance.

Among these methods, the combining use of planned cost method and weighted average at the end of month method lead to the same accounting method as the weighted average at the end of month method.

As for the other pricing methods, LIFO method results in the greatest current carry-over inventory cost, while the FIFO method leads to the largest current inventory balance. Therefore, among these internationally existing inventory valuation methods, in the case of rising prices, LIFO method results in the least profit, and income tax; FIFO method results in greatest profit and income tax, accompanied by low cash inflows, and high operating cost.

On the contrary, in the case of falling prices, we can draw the opposite conclusion.

### **3 The Problems That the Issued Inventory Valuation System Faces**

Under the circumstance of global convergence of accounting principles, the whole issued inventory valuation system is developing towards being more scientific and united. During the process, however, there are still some problems hindering the inventory calculation and management of enterprises. Firstly, globally, especially domestically, the proportions of actual choices of the inventory valuation methods are, to some degree, seriously unbalanced. Secondly, with the continuously increasing or decreasing commodity prices,[3] the existing inventory valuation methods are insufficient in correctly reflecting accounting information, which has an effect on both enterprises and market.

#### **3.1 Imbalanced Proportions of Choices of Issued Inventory Valuation Methods**

Compared with other methods, the FIFO method is closer to the actual flow of products for most enterprises. [4]Therefore, the FIFO method is more suitable for the actual flow of inventory, which benefits the management of the inventory in the enterprises and makes the inventory balance closer to the market value. But the actual usage of the FIFO method is much less frequent than that of the Weighted Average method.

##### **(1) Choices of Issued Inventory Valuation Methods Globally**

The data in Table 3 are the division of choices of issued inventory valuation methods of the listed enterprises of various countries in 2009. From the table, we can see that in the choices of issued inventory valuation methods, the number of choosing the Weighted Average method is clearly larger than that of choosing the FIFO method. The former is nearly as two times as the latter.

**Table 3.** Choices of Inventory Policies in Different Countries

(Sample Size:179)

Country	FIFO	Weighted average	Combination of Two Methods	Total Number
Australian	6	9	2	17
British	10	11	2	23
China	1	25	1	27
France	5	15	7	27
Germany	4	10	2	16
Italy	1	7	0	8
New Zealand	9	6	1	16
Sweden	7	0	0	7
Switzerland	3	5	1	9
Spain	0	3	1	4
Other European countries	13	10	2	25
Total Number	59	101	19	179

Source: Lao Chuanqi, 2011, *Research on Accounting Cultural Background of Global Convergence of Accounting Principles*

## (2) Choices of Issued Inventory Valuation Methods Domestically

In China, the imbalance of the choices of inventory valuation methods is more obvious. From the Table 3, we can see that 25 out of 27 of the enterprises listed abroad apply the weighted average method. The other companies combine the two methods. Table 4 collects the choices of inventory valuation methods of 897 enterprises listed in China in the end of 2005 and 2009. The Table presents that after the prohibition of the LIFO method in China, the enterprises whose valuation method was LIFO has to choose other ones, among which the FIFO method gains an increased usage of 1.02% while the weighted average method and specific identification method suffer a heavy decline. Besides, the weighted average method has the most extensive usage in China which accounts for about 70% in the two statistics. It is 14 percentage higher than the international average usage of 56.42%.

**Table 4.** Changes of Choices of Inventory Valuation Methods

Methods Time	Weighted Average	FIFO	LIFO	Specific Identification	Others
2005.12.31	69.10%	6.47%	1.31%	7.87%	15.24%
2009.12.31	71.92%	7.49%	1	4.92%	14.65%

Source: Wang Wenqing, 2011, *Research on Inventory Earnings Management and Valuations of Internal Control*.

(3) Through the analysis of Table 3.1 and Table 3.2, we can find out the choices of issued inventory valuation methods as follows: Firstly, the imbalance of choices of

issued inventory valuation methods is common both at home and abroad with the weighted average method as the dominance. Secondly, compared with other countries, China has more serious imbalance of the choices of issued inventory valuation methods. Thirdly, after the cancellation of the LIFO method in China, although a part of the enterprises regard the FIFO method as scientific, changes of accounting system are more inclined to the weighted average method.

### **3.2 Influence of Rising Prices on Issued Inventory Valuation**

Inflation is one of the hot topics of recent years. Now, China, even the whole world, is in the inflation which can't be got rid of in a short period of time. China has been regulating the inflation all the time. In inflation, the increase in price is unavoidable. It is true of the prices of concerning inventory. The LIFO method will accelerate inflation, which makes it impractical in China.

Under the circumstance of prices soaring, the issued inventory valuation methods with the guidelines of the existing accounting standards can deal with the current situation. But as for the calculation of the inventory issuing in the enterprises, there are still some inevitable shortcomings which make it hard to reflect the current accounting information properly.[5] In various inventory valuation methods, it is impossible to keep the differences between valuations of inventory balance and cost of sales of current period to a relatively small range. For instance, using the FIFO method will make enterprises' profits of current period on the high side. Therefore, there is still some room for improvement in the issued inventory valuation methods on the condition of inflation.

## **4 Suggestions for the Management of Inventory Valuation under the Current Circumstance**

Under the current situation of issued inventory valuation methods, the problems in all aspects in China are more serious than that of the whole international community. China has to take measures according to the actual conditions with the purposes of promoting the changes of the current situation of issued inventory valuation and better serve our domestic market.

### **4.1 Improving the Accounting Systems**

Under the circumstance of economic globalization, the accounting principles of different countries are becoming more and more convergent. Our national conditions differ from that of other developed countries as well as other developing countries. The particularities make the accounting standards in China in line with the mainline of IAS. But it does not need to be totally convergent with IAS by keeping its own accounting standards.[6] However, in the future development, great attention still will be paid to the convergence between the accounting standards in China and the IAS.

In the aspect of issued inventory valuation, we have to constantly improve the accounting standards, letting them be more compatible with the current development conditions of China and, at the same time, be closer to the IAS. It will help China gain a more favorable position in the global development and promote the development of our international operation. For example, transnational corporations will confront fewer obstacles in handling the financial data, which will facilitate the operation of enterprises.

While integrating the Accounting Standards for Business Enterprises with IAS, China has to reasonably refine the choosing and operating of inventory valuation methods in order to enable the enterprises to operate in a more standardized way and improve the environment of issued inventory valuation. Combining Accounting Standards for Business Enterprises with our actual national conditions will contribute to solving the embarrassing situation that the inventory valuation faces in China.

#### **4.2 Strengthening the Training of Enterprises Managers**

The professional competence of the accountants and managers of enterprises will, to some degree, influence the quality of final accounting information. Strengthening the training of accountants and managers with constantly upgrading their accounting knowledge can do wonders for them to make concerning accounting decisions better.

#### **4.3 Implementing Pilot Work of New Inventory Valuation Methods**

In recent years, new inventory valuation methods have been put forward constantly, many of which have integrated the strength of several methods such as the widely-used next-high, first-out method, last invoice price method and the latest inventory valuation method. [7]The appearance of these new methods makes the inventory valuation has more choices and lays the foundation for more reasonable inventory valuation methods which adhere to the current economic situation.

With the appearance of new inventory valuation methods, corresponding departments can analyze their reasonability and feasibility. In the case that a method has certain reasonability and feasibility, they can carry out trial investigation on a small scale. The reason is that if the usage and spreading of a new accounting method starts directly, it may influence and impact the existing accounting environment, subsequently hamstringing the stability of the whole accounting environment. Once failure happens, the bad influence can't be removed in a short period of time. [8]Thus in a certain period of time, it is necessary to carry out the pilot work of new inventory valuation methods within some regions, make record carefully, and analyze and summarize the recording processes and the results so as to judge whether the method is suitable for the current situation, thus consequently finding out the more practical inventory valuation methods safely and scientifically.

## References

1. Xun, C., Liang, C.: International Accounting, pp. 10–73. Northeast University of Finance and Economics Press (2012)
2. Guo, Y.: Research on Choices of Enterprise Inventory Valuation Methods and Taxing Planning. *Friends of Accounting*, the First Half of the Second Issue, 86–88 (2012)
3. Lan, H.: Comparison between China and Malaysia in Inventory Accounting Principles. *Guangxi College of Finance and Economics Journals*, the Fourth Issue of the 24th Volume, 11–14 (2011)
4. Hou, J.: Analysis of the Influence of the Choices of Accounting Systems on Enterprise Tax Burden— with the Example of the Choices of Inventory Valuation Methods. *Financial Communications*, the First Half of the 10th Issue, 95–96 (2011)
5. Lao, C.: Research on the Accounting Cultural Background of the Global Convergence of Accounting Principles. working paper (2011)
6. Li, M.: Exploration of the Issued Inventory Valuation Methods of Manufacturing Enterprise. *Economic Forum*, the first issue, 46 (2011)
7. Zou, W.: Analysis of the preferences of Inventory Valuation Methods Choosing of Listed Enterprise. *China's Information Management*, the 18th Issue of the 15th Volume, 5 (2012)
8. Nobes, C., Parker, R.B.: *Comparative International Accounting*, pp. 1–48. Financial Times Prentice Hall (2011)

# Formal Constraints Research of Military Conceptual Model Elements

Ding Xiao-Jian<sup>\*</sup> and Xie Bin

Science and Technology on Information Systems Engineering Laboratory,  
Nanjing, 210007, China  
xbqsb@126.com

**Abstract.** Conceptual model is an essential part of the process of military system modeling, and less detailed description of model elements is available for conceptual model so far. Four core elements of conceptual model are decomposed into twelve elements, and their properties, including attribute, feature, etc., are analyzed. Finally formal constraints of the proposed model elements are given, which can guide develop the military simulation system.

**Keywords:** Conceptual model elements; Military; Object constraint language.

## 1 Introduction

Military system is the main part of military integrated electronic information system. The Conceptual model of military system is a common beginning point of the development C4ISR M&S system. Military conceptual model and related techniques are one of the key answers to reusability, interoperability and VV&A of M&S, and have vital meaning to development of M&S systems. Several papers [1-3] have been discussed the development process of conceptual model, and various formal description and verification methods are given. However, in these literatures, the discussion of simulation elements was not given.

Most models consist of a number of “modes and edges” pictures, lines and some accompanied text. The information obtained by such a model has a tendency to be incomplete, imprecise, and sometimes even inconsistent. A UML diagram, such as a class diagram, is typically not refined enough to provide all the relevant aspects for a specification. There is a need to describe additional constraints about the objects in the model.

Various constraint languages have been used in object oriented modeling methods [4-5], and programming languages [6]. Object constraint language (OCL) is a formal language used to annotate model elements with the constraints [7]. The grammar and structure of an OCL statement is defined in OCL abstract syntax, which is further defined into OCL types and OCL expressions.

---

<sup>\*</sup> Corresponding author.

The rest of this paper is organized into four sections. Firstly, brief introduction to OCL will be given in Section 2. In Section 3, conceptual model elements of military system are shown, and property analysis of elements is described. Then, our approach of OCL constraints of conceptual model elements is elaborated in Section 4. Finally, the conclusion and future work are put forwarded in Section 5.

## 2 Object Constraint Language (OCL)

In practice, there are many restrictions on the elements in a conceptual model that cannot (with great difficulty) be specified in a visual representation. These restrictions can be called constraints. Constrains are always applied to elements of conceptual model, and they always restrict values of these elements. Constrains must be true at a given moment in time for the model to be valid.

OCL is a formal specification language which is used to provide extra precise description and limitation for conceptual model elements. OCL abstract syntax defines the grammar and structure of an OCL statement. The OCL abstract Syntax is further defined into OCL types and OCL expressions. The general format of OCL constraints is as follows: (1) Invariants: The invariants are conditions that have to be TRUE for each instance of the mode; (2) Precondition: A precondition is a constraint that should be TRUE always before the execution of a method starts; (3) Post-condition: A post-condition is a constraint that should be TRUE always after the execution of a method has finished.

The elements constrained by OCL must satisfy all the conditions of invariables, pre-conditions and/or post conditions when situation changes, such as object found or operation executed. Thus, the attributes and/or operations of elements are well constrained by OCL constraints.

## 3 Conceptual Model Elements of Military System

A model element is an abstract description to represent common features of military concepts. A “model element” usually includes the following attributes: id, name, description and constraints. A unique id is used as a distinguished identifier which refers to a model element. Name is meaning to a explanatory name for the model element. Description is a short description of model element, and constraints can be shown in free text form.

### 3.1 Conceptual Model Elements

CMMS describes standards for representing elements of simulation. CMMS defines “EATI (entities, actions, tasks, interactions) representation to define a CSS template”[8] which are independent of the environment used to capture CM. Based on CMMS’s research, this paper gives a four core element set to represent conceptual model description content of military system, as shown in Fig 1. Actor and role elements are derived from entity core element. Mission and input/output elements are



derived from task core element. Measure element is derived from objective core element, and initial state, final state and transition elements are derived from state core element.

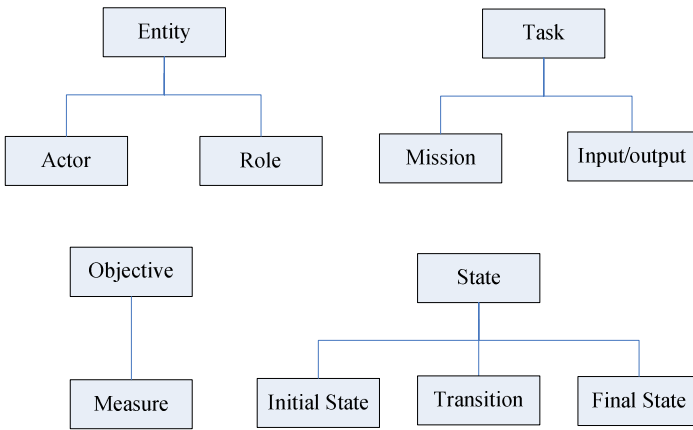


Fig. 1. Model elements

### 3.2 Property Analysis of Elements

#### 3.2.1 Entity Core Elements

Entity is one of the most frequently used model elements. Entity is any being in mission space that has specific properties on its own. An entity may be an actor, organization, facility, network, material, equipment or object. Some types of entities will be differentiated as specific element definitions, like actor. Table 1 shows the properties of entity.

Table 1. Properties of entity

Properties	Description
Id	A unique code for the entity
Name	A specific name for the entity
Type	Shows the entity where entity is derived from
Facility information	Physical establishment, like bridge, airfield etc
kind	Entity that has no function on its own, but yet is a collection of definitions
Force	Has attack capabilities or intelligence on battlefield
Equipment	Is used for and has specific capabilities on battlefield
Material	Has specific capabilities on battlefield
Attribute	Lists all the attributes of the entity
Behavior	Lists all the behaviors and capabilities of the entity
State	The states that the entity may be in
Relation	Lists the relations of the entity between model elements

Actor is an entity whose kind is “actor”. It is also an entity that is responsible from executing tasks. Actor can represent a human being or any other active entity. Although it cannot directly perform tasks, it can define roles to perform these tasks. Actors are derived from entities, so it inherits all properties from an ordinary entity.

A role represents an active actor who is “realize” a task or a mission. Similarly, role has the same properties with an actor.

### 3.2.2 Task Core Elements

A task represents more detailed activities to satisfy a military objective. It is used to detail the mission, and can be reused among different missions. Table 2 shows the properties of task.

**Table 2.** Properties of task

Properties	Description
Id	A unique code for the task
Name	A specific name for task
Explanation	A detailed explanation for the task
Pre-condition	The conditions that are necessary to execute the task
Post-condition	The conditions that are necessary to be executed before the end of the task
Constraint	Informs any constraints about the task

Missions are the high level tasks that the military system is expected to accomplish. The missions of the system can be thought as the objectives for which the system is built. Miss has the same properties with the task, and it involves cooperation of entities and tasks.

The inputs are meaning something to execute the task and the outputs are something resulting from the execution of the task. They are called process product. Process product has the similar properties with the task expect Pre-conditions and Post-conditions.

### 3.2.3 Objective Core Elements

An objective represents the goal of a mission or task expected to fulfill. It is connected to measures to see whether the related task reach its objectives. Table 3 shows the properties of objective.

**Table 3.** Properties of objective

Properties	Description
Name	A specific name to determine the objective
Explanation	The description of the objective
Measure	Lists the measures to evaluate if the objective is achieved
Achievement status	Lists the conditions of measures under which the objective is successful or unsuccessful

The measure element is quantifiable defined to determine the performance of an objective. It is considered to be successful if it satisfies all of the measures connected to it. Table 4 shows the properties of measure.

**Table 4.** Properties of measure

Properties	Description
Name	A name to determine the measure
Unit	The unit of the measure
Value	The value to evaluate the measure

### 3.2.4 State Core Elements

State element is one of the possible conditions in which an entity may exist. An entity may be in various states throughout its lifetime. Table 5 shows the properties of measure.

**Table 5.** Properties of state

Properties	Description
Name	A name to determine the state
Explanation	The description of the state
Single state	Is the state has any sub-states
Sub-State	Is the state the sub-state

Initial state element denotes the entrance of state. It is not a real state but rather used as a formal means to indicate how the control should go. So it must connect an outgoing transition to another state. Final state shows the end of state variety process. It can have any number of incoming transitions but cannot have any outgoing transitions. Likewise, transition is a relationship within a state between source state and target state.

## 4 The Application Domain of OCL Constraints

As described above, model elements properties are not enough to describe the real constraints of the elements. It is essential to propose a formal method to describe the constraints related with military domain. Table 6 shows the OCL constraints of proposed model elements.

As shown in table 6, a mission must have one role and one objective, and a role must be owned by at least an actor. The constraints of task element are similar to mission except for it need one incoming and one outgoing task flow. An objective must has at least one measure, and vice versa. A process product does not have any capabilities. A state must be a single state (does not have any sub-states). An initial state must have only outgoing transition and does not have an incoming transition. A final state may not have an outgoing transition. A transition can have initial state or state as source, and have final state or state as target.

**Table 6.** OCL Constraints of model elements

Element	Constraints
Mission	<b>context</b> <i>Mission</i> <i>self.roleList.size() &gt; 0</i> <b>context</b> <i>Mission</i> <i>self.objectiveList.size() &gt; 0</i> <b>context</b> <i>Task</i> <i>self.roleList.size() &gt; 0</i> <b>context</b> <i>Task</i> <i>self.objectiveList.size() &gt; 0</i>
Task	<b>context</b> <i>Task</i> <i>self.incomingFlows.size()=1 AND self.outgoingFlows.size()=1</i>
Role	<b>context</b> <i>Role</i> <i>self.ownerList.size() &gt; 0</i>
Objective	<b>context</b> <i>Objective</i> <i>self.measureList.size() &gt; 0</i> <b>Notation</b>
Measure	<b>context</b> <i>Measure</i> <i>exists o:Objective in objectives where o.measureList.includes(self)</i>
Process Product	<b>context</b> <i>Process Product</i> <i>self.capabilityList.size() = 0</i> <b>Notation</b>
State	<b>context</b> <i>State</i> <i>isSingle = content.isEmpty()</i>
InitialState	<b>context</b> <i>InitialState</i> <i>(self.oclsKindOf(InitialState))</i> <i>implies ((self.outgoing-&gt;size() = 1)</i> <b>context</b> <i>InitialState</i> <i>(self.oclsKindOf(InitialState))</i> <i>implies ((self.incoming-&gt;size() = 0)</i>
FinalState	<b>context</b> <i>FinalState</i> <i>(self.oclsKindOf(FinalState))</i> <i>implies((self.outgoing-&gt;size() = 0)</i>
Transition	<b>context</b> <i>Transition</i> <i>(self.source.isTypeOf(InitialState) OR self.source.isTypeOf(State)) AND (self.source.isTypeOf(FinalState) OR self.source.isTypeOf(State))</i>

## 5 Conclusions

Conceptual model elements of military system are domain specific elements, which are not dependent on any simulation concept or environment. Conceptual model elements proposed in this study can be used to guide the development of simulation for military system.

The second objective of this study is to present a formal description of model elements by OCL. By this research, elements and given constraints can guide the development of C4ISR simulation system effectively.

## References

1. Wang, Y., Ma, P., Yang, M., et al.: Research on Development Process of Simulation Conceptual Model. *Journal of System Simulation* 18(s2), 17–23 (2006)
2. Huang, J.-L., Tan, D.-F., Zhang, X.-B.: Development Research of Military Simulation Concept Model. *Computer Simulation* 22(2), 15–18 (2005)

3. Hu, J., Chen, X.: Study on the Establishment Methods for Military Simulation Concept Model. *Experiment Science & Technology* 5, 14–17 (2006)
4. Catalysis, <http://www.catalysis.org/>
5. Syntropy, <http://www.syntropy.co.uk/syntropy/>
6. Eiffel, <http://www.eiffel.com/>
7. Gogolla, M., et al.: USE: A UML-Based Specification Environment for Validating UML and OCL. *Science of Computer Programming* 69, 27–34 (2007)
8. Defense Modeling and Simulation Office (DMSO), Conceptual Models of the Mission Space (CMMS) Technical Framework USD/A&T-DMSOCMMS-0002 Revision 0.2.1 (February 13, 1997)

# Leaf Lesion Detection Method Using Artificial Bee Colony Algorithm

Faudziah Ahmad and Ahmad Airuddin

Universiti Utara Malaysia, Sintok, Malaysia  
fudz@uum.edu.my, mrairuddin@gmail.com

**Abstract.** Lesions in images can be detected using several edge detection methods such as Canny, Sobel, Roberts and Prewitt. However, all of these methods are time consuming in detecting lesions. The reason being each of the pixels is serially searched from top to bottom and from left to right in the image. In addition, a lesion can only be detected when all pixels have been completely searched. The methods are inefficient as some of pixels, usually at the edge of a lesion fail to be detected. This paper presents experimental results on an algorithm that was developed using Artificial Bee Colony (ABC). Results showed that ABC produced better percentage of correctness and detection time than Canny, Sobel, Roberts and Prewitt.

**Keywords:** Object detection, Leaf, Lesion, Grayscale pixels, ABC.

## 1 Introduction

Plant diseases increase about 10 to 40 percent per year because of bacteria (Osdaghi, Alizadeh, Shams-Bakhsh, & Lak, 2009). One of the symptom shown causes of those bacteria in leaf is lesion. The leaf of plant will be exposed to damage when lesion exist on the leaf surface (Ahmad, et al., 2010; Chen, 2005). Lesion is a region in an organ or tissues that has suffered damage. Lesion can change the original color of the green leaf into dark brown color. The plant will die if there are many lesions. This is because lesions will cause a reduction in green surface of plant leaves thus, resulting to a reduction in chlorophyll and in turn affect photosynthesis. The total area of lesions can be associated with the severity of a leaf disease. In general, the term "severity" in the study refers to the percentage of seriousness of a leaf disease. The percentage of leaf lesion severity can determine the seriousness of a plant disease (Nutter, Teng, & Shokes, 1991). Leaf disease severity can be divided into five category, grade 0 (apparently infection-free), grade 1 (0 – 25% leaf area infected), grade 2 (26 – 50% leaf area infected), grade 3 (51 – 75% leaf area infected), and grade 4 (>75% leaf area infected) (Horsfall & Heuberger, 1942).

Previous studies used edge detection methods such as Canny, Sobel, Robert and Prewitt to detect lesions. These methods were found to be time consuming and obtained a high percentage of inaccuracies (Patil & Bodhe, 2011). The methods were slow to detect the lesion edge because pixels were searched in sequence to the next

neighbor and both pixel values were compared (Weizheng, Yachun, Zhanliang, & Hongda, 2008). Sobel method can usually detect edge that is thicker than Roberts and Prewitt methods. Canny uses the same edge intensity as that of Sobel method to define edge of the object. Canny can detect much thinner edges than the Sobel. However, Canny has problems in detecting correct edge pixels. Due to the limitation of the current methods, a better edge detection method is thus needed.

In terms of image processing, several studies have found that ABC is able to handle problems related to thresholding, segmentation and object detection. For example, Ye et al. (2011) proposed automatic choosing threshold algorithm using ABC and compared its performance with Otsu algorithm. The results showed that the performance of ABC algorithm was better than Otsu algorithm. In another research by Zhang and Wu (2011), ABC was used to achieve multi-level thresholding image segmentation. They used ABC to overcome time consuming problem. They identified that ABC was faster than GA and PSO. Fast segmentation in Synthetic Aperture Radar (SAR) using ABC was proposed by Ma et al. (2011). The objective of their study was to enhance threshold optimal value of grayscale color between pixels. They found that ABC method produced better quality than Artificial Fish Swarm (AFS) and Genetic Algorithm (GA) in terms of accuracy in segmentation and time. Application of ABC for recognition of an object within certain images was introduced (Chidambaram & Lopes, 2009). The objective of their work was to find a pattern or template of an object anywhere on a target scene. The experimental results, using grayscale and color images showed that the performance of ABC was faster in finding a pattern than a comparable technique such as Evolutionary Algorithm (EA). However, in terms of object detection using ABC, reviews on past literatures showed that no work has been done on lesion detection. The positive reviews produced from past works have resulted in an attempt to use ABC in detecting an object, specifically lesions.

This study will be focusing on detecting lesion in images. An algorithm based on ABC will be developed to detect lesions in grayscale images. The algorithm will be measured in terms of percentage of correctness, percentage of error and detection time.

The organization of the paper is as follows. Section 2 describes the methodology of the study. Section 3 and 4 presents analysis of results and discussion. Conclusions and future research are presented in Section 5.

## 2 Methodology

The research was conducted in 3 phases, data preparation, algorithm construction and analysis.

### PHASE 1: Data preparation

This phase involves three steps: convert images from JPEG or RGB to TIFF; convert TIFF images into grayscale; and perform data cleansing. The steps for the phase are described below.

**Step 1: Convert images from JPEG or RGB to TIFF**

The data used in the study were taken from [www.forestryimages.org](http://www.forestryimages.org). It consisted of four green color images that are in RGB and JPEG format. The images were converted to TIFF format using Microsoft Paint software. Original image is stored in JPEG, a significant change occurred during conversion from an original image to JPEG and this causes some image distortion. However, when an image is stored in TIFF, the change is very minimal. This results to a less distorted image. Therefore, images in TIFF are sharper than images in JPEG.

**Step 2: Convert TIFF Images into Grayscale**

In a color image, each pixel contains three color values, red, green and blue. Converting the color image into grayscale will reduce the number of colors to one. This in turn will reduce the storage size that will result to a decrease in computing time. The conversion was done by using Octave software.

**Step 3: Data Cleansing**

The aim of this step is to filter grayscale images from noises such as dust, blurriness, and unwanted spots. This will produce sharper, smoother and cleaner images (Rajan, 2012). For this process, Gaussian technique was used.

**PHASE 2: Algorithm Construction**

In this phase, an algorithm was constructed based on Sharma et al. (2012). Sharma et al. (2012) had developed an algorithm based on ABC to combine two images into a single image in order to get informative result. This combination is known as image Fusion. Their algorithm was adjusted according to this experiment. The adjustments made were: (i) Eliminate the step for dividing the image into small windows because in this experiment, whole area in an image was analyzed; (ii) Set variable for optima value from 10 to 85 and set amount of bees to 200; (iii) Initialize nectar value to zero; (iv) Change the step of reading two images (sources) into one image; (v) Add time detection step; (vi) Change the step using entropy and spatial frequency to determine highest properties to comparing the optima value to get highest properties; (vii) Produce mark symbol (+) when bees detect the edge. However, other steps are maintained in this experiment. The algorithm was converted from VC++ 6.0 into Octave 3.2.4. The ABC steps are shown in Table 1.

**PHASE 3: Analysis**

The algorithm produced in PHASE 2 was executed on the selected data (grayscale leaf images). The experiment was repeated for four times on each data in order to get a proper output. Every mark symbol (+) was checked to identify whether it is the edge of an object. If it is the edge, then this means that the bees have detected the object. The total number of marks that detected the edge were accumulated and represented as the percentage of correctness. The total number of marks that incorrectly detect an object was also calculated in terms of percentage of incorrectness.



**Table 1.** ABC steps

Step	Procedure	Description
<b>Step 1: Start</b>		Begin the algorithm
<b>Step 2: Initialization Phase</b>	<ul style="list-style-type: none"> <li>-Read image (Source)</li> <li>-Initialize the optimal value of grayscale and total of the bees</li> <li>-Initialize nectar with zero value</li> </ul>	The optimal value of grayscale was set between 10 and 85, the total number of bees was initialized to 200, and the nectar (optimal value) was set to zero. (Initially, there was no nectar value found in the hive (memory).
<b>Step 3: Employed Bee Phase</b>	<ul style="list-style-type: none"> <li>-Select a sources area of size (mxn) in image</li> <li>m=height of the image size, and</li> <li>n=width of the image size</li> </ul>	In this phase, the data from memory had been read in the algorithm to get the size of image.
<b>Step 4: Onlooker Bee Phase</b>	<ul style="list-style-type: none"> <li>-Select the pixels (Nectar) of the sources area (Source) having high property value (Nectar Value)</li> <li>-Mark the selected pixels if edge of object found</li> <li>-Store the mark image in memory</li> <li>- store time</li> </ul>	The onlooker bee searches the nectar according to optima value and detects the edge object if optima value matches.
<b>Step 5: Scout Bee Phase</b>	<ul style="list-style-type: none"> <li>-select next area in the source (New source) and repeat step 3-5 with (p-m) * (q-n) times</li> </ul> <p>p= height and q=width of selected image</p>	The scout bee in this algorithm selects other new area in an image that has potential to explore the nectar.
<b>Step 6: Stop</b>		End of algorithm.

The experiment also captured the time (in millisecond) taken to produce a mark. The time measured was from the start when the algorithm was executed until an edge was detected. In this phase, other detection methods from Gulhane et al. (2011), Weizheng et al. (2008) and Patil et al. (2011) will be compared with ABC.

### 3 Results and Discussion

The performance of the constructed algorithm was measured in terms of percentage of correctness, percentage of error, and detection time (millisecond). Table 2 shows the analysis.

**Table 2.** Comparison of methods

Methods	Image	Lesion	Number of Lesion Found	Percentage of Correctness (%)	Percentage of Error (%)	Detection Time (milli-second)
ABC	P10	7	7	100.00	0.00	0.039065
	P12	14	9	64.30	35.80	0.042973
	P14	8	8	100.00	0.00	0.046878
	P16	2	2	100.00	0.00	0.050785
	<b>Average</b>			<b>91.07</b>	<b>8.95</b>	<b>0.044925</b>
Canny	P10	7	5	71.40	28.60	0.562500
	P12	14	14	100.00	0.00	0.546875
	P14	8	0	0.00	100.00	0.515625
	P16	2	2	100.00	0.00	0.640625
	<b>Average</b>			<b>67.85</b>	<b>32.15</b>	<b>0.566406</b>
Sobel	P10	7	Unknown	Unknown	Unknown	2.890625
	P12	14	Unknown	Unknown	Unknown	2.906250
	P14	8	Unknown	Unknown	Unknown	3.140625
	P16	2	Unknown	Unknown	Unknown	2.906250
Roberts	P10	7	Unknown	Unknown	Unknown	2.390625
	P12	14	Unknown	Unknown	Unknown	1.906250
	P14	8	Unknown	Unknown	Unknown	3.062500
	P16	2	Unknown	Unknown	Unknown	2.609375
Prewitt	P10	7	Unknown	Unknown	Unknown	3.171875
	P12	14	Unknown	Unknown	Unknown	3.031250
	P14	8	Unknown	Unknown	Unknown	3.343750
	P16	2	Unknown	Unknown	Unknown	2.781250

Table 2 shows the comparison results for four detection methods. Column 1 is the list of images that was used in the experiment; column 2 is the total number of actual lesions in the leaf; column 3 is the percentage of correctness in detecting lesions; column 4 is the percentage of error in detecting lesions; and column 5 is the average time taken to detect the lesion.

From the table, it can be seen that ABC could detect lesions correctly for three images, P10, P14, and P16. Canny could only detect two images correctly, while, Sobel, Roberts and Prewitt could not detect any lesions from all images. Column 4 until column 6 were declared as “unknown” because all of these methods could not detect any lesions, thus, unable to calculate percentage of correctness and detection and percentage of error in detection.

From all the methods, it shows that ABC was superior to others in detecting lesions of less than 10. However, if the number of lesions is greater than 10 (P12), ABC fail to identify all lesions correctly. The reason may be due to the number of bees used in the experiment. That is, the probability of finding an object using 200 bees is higher when an image has below than 10 lesions. If the number of bees is to be increased,

then there will be a greater chance of detecting all the lesions. However, the computing time will be increased if the number of bees used is increased.

Canny was able to identify lesions from P12 and P16 correctly but the time taken to detect on average was higher than ABC.

## 4 Conclusions

This study developed an algorithm based on ABC to detect lesions. Four images were tested. ABC and four other edge detection methods were applied to the data. Results showed that ABC could detect lesions faster and more accurately than the other methods. This indicates that ABC could be incorporated in existing object detection algorithms to produce better results.

## 5 Recommendations

Further experiments using various numbers of bees will be conducted to identify an optimal value. Improvements to the ABC will be made so that images with more than 10 lesions can be detected correctly.

## References

- Ahmad, S., Afzal, M., Noorka, I.R., Iqbal, Z., Akhtar, N., Iftkhar, Y., et al.: Prediction of Yield Losses In Wheat (*Triticum Aestivum* L.) Caused By Yellow Rust In Relation To Epidemiological Factors In Faisalabad. *Pak. J. Bot.* 42(1), 401–407 (2010)
- Chen, X.M.: Epidemiology and control of stripe rust (*Puccinia striiformis* f. sp. *tritici*) on wheat. *J. Plant Pathol.* 27, 314–337 (2005)
- Chidambaram, C., Lopes, H.S.: A new approach for template matching in digital images using an artificial bee colony algorithm. Paper presented at the Nature & Biologically Inspired Computing (2009)
- Gulhane, V.A., Gurjar, A.A.: Detection of Diseases on Cotton Leaves and Its Possible Diagnosis. *International Journal of Image Processing (IJIP)* 5(5), 590–598 (2011)
- Horsfall, J.G., Heuberger, J.W.: Measuring magnitude of a defoliation disease of tomatoes. *Phytopathology* 32, 226–232 (1942)
- Ma, M., Liang, J., Guo, M., Fan, Y., Yin, Y.: Sar image segmentation based on artificial bee colony algorithm. *Applied Soft Computing* 11(8), 5205–5214 (2011)
- Nutter, F.W., Teng Jr., P.S., Shokes, F.M.: Disease Assessment Terms and Concepts Plant Disease, pp. 1187–1188 (1991)
- Osdaghi, E., Alizadeh, A., Shams-Bakhsh, M., Lak, M.R.: Evaluation of Common Bean Lines For Their Reaction To The Common Bacterial Blight Pathogen. *Phytopathologia Mediterranea* 48, 461–468 (2009)
- Patil, S.B., Bodhe, D.S.K.: Leaf Disease Severity Measurement Using Image Processing. *International Journal of Engineering and Technology* 3(5), 297–301 (2011)
- Rajan, A.S.: Image Processing Techniques for Diagnosing Paddy Disease. Paper presented at the Proceeding of the World Congress on Engineering, London, UK (2012)

- Sharma, P.K., Bhavya, V.S., Navyashree, K.M., Sunil, K.S., Pavithra, P.: Artificial Bee Colony and Its Application for Image Fusion. *IJ. Information Technology and Computer Science* 11, 42–49 (2012)
- Weizheng, S., Yachun, W., Zhanliang, C., Hongda, W.: Grading method of leaf spot disease based on image processing. In: *Proceedings of the 2008 International Conference on Computer Science and Software Engineering, CSSE, December 12-14*, pp. 491–494. IEEE Computer Society, Washington, DC (2008)
- Ye, Z., Hu, Z., Wang, H., Chen, H.: Automatic threshold selection based on artificial bee colony algorithm. Paper presented at the *Intelligent Systems and Applications (ISA), 3rd International Workshop* (2011)
- Zhang, Y., Wu, L.: Optimal multi-level thresholding based on maximum tsallis entropy via an artificial bee colony approach. *Entropy* 13(4), 841–859 (2011)

# The Accurate Positioning Method of Data Matrix Code Image Marked in Cylinder Glass

Wei-ping He<sup>1</sup>, Gai-fang Guo<sup>2</sup>, Wei Wang<sup>3</sup>, and Xi-zheng Cao<sup>4</sup>

Key Laboratory of Contemporary Design and Integrated Manufacturing Technology,  
Ministry of Education, Northwestern Polytechnical University, Xi'an710072  
3802561002@qq.com

**Abstract.** For the problem of the data matrix (DM) code Positioning marked in cylinder glass, this paper presents an accurate positioning method of DM code image marked in cylinder glasses. Firstly, the method enhances the contrast ratio of the image based on the Retinex theory. Secondly, the paper uses projection and morphology for the early positioning of the image after enhanced. At last, the image after early positioning is positioned precisely based on Hough transform. Experiments show that the way can position the DM data image precisely and that the accuracy reaches to 90 percent. Finally, it is convenient for decoding subsequently.

**Keywords:** Data Matrix Code, Cylinder Glass, Positioning, Retinex, Hough Transform.

## 1 Introductions

The direct parking marking (DPM) is a technology that can directly mark in the surface of the product to form a permanent identification such as words, symbols and pattern. Because data matrix (DM) code has a lot of advantages, people choice DM code more as the identification of the product[1-2].

This paper researches the DM code of the cylinder glass marked by CO<sub>2</sub> laser marking machine, such as liquor and medicine glass. Due to the dimensional structure, the DM code marked in cylinder glass has the features of uneven illumination, highly reflective, low contrast etc. Moreover, according to the principle of the laser marking machine, in the process of marking, the laser beam is irradiated on the glass surface, the glass surface absorbs the laser energy and thermal excitation effect generates in the irradiated regions, so the surface temperature of glass rises, resulting in abnormal, melting, ablation, evaporation phenomena. Because of the different length of the laser beams during marking, the heat of the glass surface is uneven. And because the glass materials of cylinder glasses are composed of not only silica but also some oxides such as sodium oxide, calcium oxide etc, and different materials have different physical and chemical properties. Therefore, the DM code marked in cylinder glass by laser has the features of uneven and discontinuous edges of the DM code. Figure1 is the case diagram of the DM code marked in the cylinder glass.



**Fig. 1.** The DM code marked in the cylinder glass

The DM code is surrounded by detection graphic which is a square making up by two adjacent sides of the solid line and opposites sides of the dotted line, which is used in limiting physical size, location and symbol distortion. Currently, the positioning of most DM codes is positioned based on the geometric features of the detection graphics, the advantage of which is of fast processing speed. But everything has its two sides. The disadvantage of it is that positioning will fail if the boundaries are interfered by contamination or noise [3-5]. For this reason, some scholars have presented some improved methods. Liu has presented positioning of DM code based on wave shape analysis, which position the DM boundaries by wave shape [6]. Jain and Hao have presented positioning of DM code based on the method of texture analysis and machine learning [7-8]. The methods consider the characteristics of DM code rather than limiting to the detection of boundaries, so it has a strong anti-jamming capability. But the disadvantage is more sensitive to changes in perspective.

Therefore, the paper presents a precise positioning way of DM code marked in cylinder glass. First of all, aiming at improving the contrast ratio of the image, the image is enhanced base on Retinex theory. And then the image of the DM code area after enhanced is early positioned using projection combination of morphology. At last, owing to the un-continuous edge of the DM code in cylinder glasses, the picture is positioned precisely based on improved Hough Transform.

## 2 Image Contrast Enhancement Based on Retinex

Due to the process of image acquisition of DM code from devices, the images are inevitably different from the images which are acquired from DM code editors directly. Particularly, it is necessary to process the possible noise, low contrast, blur etc of the images during acquisition. Therefore, for the problem of low contrast, noise from the images of DM code marked by laser, it is necessary to preprocess firstly.

Retinex and Cortex is a model presented by Land etc on how human visual system adjusting perceiving the color and brightness of the object [9]. Compared with other enhancement algorithms, the enhancement algorithm based on Retinex has the advantage of high contrast ratio, less color distortion, large dynamic compression range etc. The paper presents nonlinear enhancement algorithm based on Retinex. Firstly, in order to estimate the luminance image roughly, it transforms the image based on Gaussian Smoothing. Then, the image is enhanced based on Beta transform and wavelet transform. So it can get the reflection image through the ratio of the enhanced image and the image of rough estimation in the log domain. And for the aim at the terminal enhanced image, it merges the reflection image after adjustment and the image of rough estimation. The algorithmic process is shown as figure2.

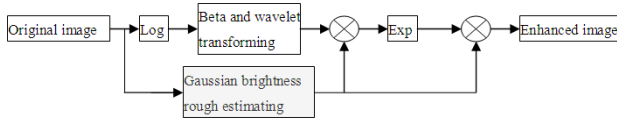


Fig. 2. The nonlinear enhancement algorithm based on Retinex

### 2.1 Gaussian Smoothing on the Original Image

To estimate the luminance image roughly, the image is transformed based on Gaussian Smoothing, denoted as  $I(x, y)$ .

1	2	1
2	4	2
1	2	1

Fig. 3. Gaussian template

### 2.2 The Image Enhancement Based on Beta Transform and Wavelet Transform

The original image is divided into 7 categories based on histogram. Figure4 shows the Classification criteria.

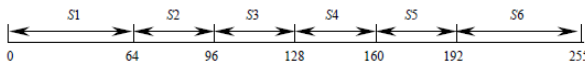


Fig. 4. Image classification schematic diagram based on histogram

It is suppose that the grayscale of input image is divided into 255 gray levels. The entire gray space is divided into 6 corresponding gray levels belonging to 6 child spaces according to figure4, represented respectively by  $S_1, S_2, S_3, S_4, S_5$  and  $S_6$ . It supports the hypothesis that

$$M = \max_{i=1}^6 Si, \quad B_1 = \sum_{i=2}^6 Si, \quad B_2 = \sum_{i=2}^5 Si, \quad B_3 = \sum_{i=1}^5 Si, \\ B_4 = S_1 + S_6, \quad B_5 = S_2 + S_3, \quad B_6 = S_4 + S_5. \quad (1)$$

Then the contrast of the image is classed using the following criteria.

If  $M=S_1$  &  $S_1>B_1$ , image is PD; Else if  $B_2 > B_4$  &  $B_5 > B_6$  &  $B_5 > S_1$  &  $B_5 > S_6$  &  $S_2 > S_3$ , image is MD; Else if  $B_2 > B_4$  &  $B_5 > B_6$  &  $B_5 > S_1$  &  $B_5 > S_6$  &  $S_2 < S_3$ , image is MDS; Else if  $B_2 > B_4$  &  $B_5 < B_6$  &  $S_1 < S_6$  &  $S_6 < S_6$  &  $S_4 > S_5$ , image is MBS; Else if  $S_2 > S_4$  &  $S_5 < S_6$  &  $S_1 < B_6$  &  $S_6 < B_6$  &  $S_4 < S_5$ , image is MB; Else if  $M=S_6$  &  $S_6 > B_3$ , image is PB; Else, image is GGLD; End. Wherein, symbol ‘&’ represents logical AND.

The original image is enhanced based on incomplete Beta transform. The formula of incomplete Beta transform is as follows:

$$F(u) = B^{-1}(\alpha, \beta) \int_0^u t^{\alpha-1} (1-t)^{\beta-1} dt \quad 0 < \alpha, \beta < 10. \quad (2)$$

Table1 is the values of  $\alpha$  and  $\beta$ .

**Table 1.** The range of parameter  $\alpha$  and  $\beta$

Parameters	PD	MD	MDS	MBS	MB	PB
$\alpha$	1.5	1.5	1.5	2.0	2.5	3.5
$\beta$	3.5	2.5	2.0	1.5	1.5	1.5

Among the formula  $u$  represents the image pixel values after normalized. It is necessary for the pixel values to be normalized and inverse-normalized before and after Beta transform. The result of the global enhanced image is represented by  $I_G$ , namely the result after Beta transform.

Then the original image is enhanced using discrete stationary wavelet local domain. Aiming at enhancing the local contrast of the image in the help of nonlinear gain operator, the following transfer function is defined to enhance three high-frequency sub band images in the wavelet domain:

$$g(i, j) = MAG\{f(i, j)\}. \quad (3)$$

Among the formula  $g(i, j)$  represents sub band image after enhanced,  $f(i, j)$  represents original sub band image to be enhanced,  $MAG$  represents a non-linear enhancement operator. We can make  $f_s^r(i, j)$  as the gray level value of pixel in the  $r$  layer  $s$  sub band. Among that,  $s$  represents 1, 2, 3, etc and  $r$  represents 1, 2 or 3.  $\max f_s^r$  represents the maximum values of all pixel gray level among  $f_s^r(i, j)$ . Therefore,  $f_s^r(i, j)$  can be mapped from  $[-\max f_s^r, \max f_s^r(i, j)]$  to  $[-1, 1]$ . Thus the parameter  $a, b$  and  $c$  can be dynamically set separately. Specific enhancement formula is as follows.

$$g_s^r(i, j) = \begin{cases} f_s^r(i, j) & f_s^r(i, j) < T_s^r \\ a * \max f_s^r \{ \text{sigm}[c(y_s^r(i, j) - b)] - \text{sigm}[-c(y_s^r(i, j) + b)] \} & |f_s^r(i, j) \geq T_s^r| \end{cases} \quad (4)$$

$$y_s^r(i, j) = f_s^r(i, j) / \max f_s^r \quad (5)$$

### 2.3 Estimation of the Reflection Image

Assuming the reflected image is  $r(x, y)$ .

$$r(x, y) = \log[I_1(x, y)] - \log[I_0(x, y)] \quad (6)$$



## 2.4 Gamma Correction and Output of the Enhanced Image

On the one hand, gamma correction can compress the reflected image in the dynamic range; on the other hand, gamma correction can adjust the reflected image in the proportion of the enhanced image, which can be expressed as follows.

$$r'(x, y) = [r(x, y)]^{1/\gamma} \quad (7)$$

Among the formula  $\gamma$  is a positive value. After experiments, we can choice  $\gamma$  as 2.2. Then the sum of the reflected image after correction and illumination image is made antilog calculation, we can get the image after enhanced, represented by  $R(x, y)$ .

$$R(x, y) = \exp\{r'(x, y) + \log[I(x, y)]\} \quad (8)$$

Assuming the output image is  $L(x, y)$ .

$$L(x, y) = I_1(x, y) \times R(x, y) \quad (9)$$

Figure5 is the example of the enhanced image based on Retinex.



Fig. 5. Nonlinear image enhancement based on Retinex

## 3 Early Positioning of DM Code Based on Projection

Since there are lots of background and noise in the DM code image, so if the DM code is positioned directly in the process of DM code positioning, it is not only inefficient, but also has a low speed. Therefore, the current positioning of the bar code is divided into early positioning and fine positioning.

Because the DM code area of cylinder glass is uneven that is reflected in the image with a large gradient between pixels, in order to extracting the DM code area, the projection method is used. In this paper, the horizontal and vertical gradient of the image after enhanced based on Retinex are found respectively [10]. Although there are many noises of the background in the image, the projection of background gradient is much smaller than that of DM code. Therefore, it is feasible possible to extract code area and remove background noise in this way. In the algorithm, we can get the horizontal gradient projection, which determines the upper and lower boundaries of code at first. And then we get the vertical gradient projection to determine the left and right boundaries of code in the same way. At last, the bilinear interpolation method is used to cut the image and narrow the DM code image to the desired size. The specific algorithm is as follows.

Step1: Each line of the image is traversed, the horizontal gradient projection of which is calculated, denoted as  $V_i$ . Therefore we can get the average horizontal gradient projection of image, which is denoted as  $V_{avg}$ .

Step2: A cumulative array is established, namely *morpLevel*. The horizontal direction gradient projection  $V_i$  of each line in the image is compared to the average horizontal projection  $V_{avg}$ . If  $V_i > V_{avg}$ , *morpLevel*[*i*] is assigned to 1. On the contrary, *morpLevel*[*i*] is assigned to 0.

Step3: The array *morpLevel* is traversed in descending order. When *morpLevel*[*i*] is 1, stop traverse. At this time the upper boundary *Yup* for the DM code is positioned as 1, namely *Yup*=1. And then the array *morpLevel* is traversed in ascending order. When *morpLevel*[*j*] is *j*, stop traverse. At this time the lower boundary *Ydown* for the DM code is positioned as *j*, namely *Ydown*=*j*.

Step4: The left and right boundaries are positioned using the same way of step1 to step3, namely *Xleft* and *Xright*.

Step5: The four boundaries of the DM code are extended. Every boundary is extended for 20 pixels. Then the width of the image is corrected for 200 pixels. Therefore, the formula for calculating the height *h* is as follows.

$$h = (Yup - Ydown + 1) * (200 / (Xright - Xleft + 1)) \quad (10)$$

Step6: To get the code area, the image within the boundaries is narrowed based on bilinear interpolation method.

The example of early positioning of code is as figure6 shown.

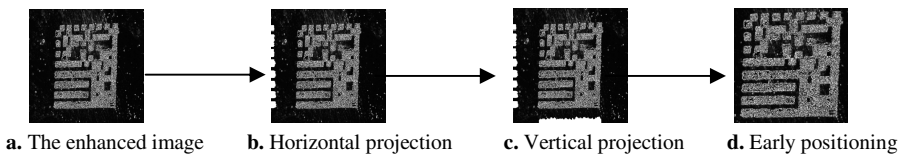


Fig. 6. The process of early positioning

## 4 Precise Positioning Based on Improved Hough Transform

Compared with the printed code, the DM code area of cylinder glasses marked by laser is uniformly filled, and there are lots of noises in the DM code area. Therefore, this paper processes the image after binarized with morphological processing at the beginning. And then the boundaries of the image are extracted by using of morphology. At last, the DM code area is positioned precisely based on Hough transform [11].

During the process of positioning of the code area using Hough transform, the outer contour is extracted from the image after pretreatment at first. Since the DM code of cylinder glasses is complex, the points of the image outer contour extracted are not in a line, but the distribution of these points is in the neighborhood of the boundaries, the points in the outer contour are transformed in Hough domain. To find the boundaries, instead of finding the maximum Hough point directly, the maximum Hough point is found according to the sum of three neighboring Hough width in the same Hough angle. In this way it can reduce positioning mistake and position the DM code area precisely. The specific algorithm is as follows.

Step1: The image after early positioning is binarized based on otsu thresholding. And then binary closing operation is used to process the image after binarized.

Step2: To get the contour of the DM code, the image processed by step1 is transformed based on binary morphological gradient.

Step3: A code contour array is established as buf. The image is traversed form the upper, lower, left and right edges separately. If the image pixel is 255, stop traversing, and add the buf in this pixel.

Step4: A Hough transform array is established as Houghbuf, assuming that the Hough height is Houghheight and the Hough width is Houghwidth. The pixel points which are 1 in the code contour array are transformed with Hough transform, and the Hough transform array is accumulated.

Step5: To get a new Hough array, add the three adjacent arrays, and then we can get the maximum Hough point in the new Hough array. The height of this Hough point is noted as  $\alpha$  and the width of that is noted as d1. The straight-line corresponding to the Hough point is one edge of the 'L' type probe graphics of the DM code, noted as L1.

Step6: Rotate  $\alpha$  form step5 for 90 degrees. The Hough points are traversed from  $\alpha+80$  to  $\alpha+100$  for getting the maximum Hough point in the domain. The height of this Hough point is noted as  $\beta$  and the width of that is noted as d2. The straight-line corresponding to the Hough point is another edge of the 'L' type probe graphics of the DM code, noted as L2.

Step7: The third boundary of the four which is parallel to L1 is to be found, namely L3. Firstly, the sizes of L1 and Houghwidth/2 are judged. If L1 is the bigger size, the maximum Hough point can be found in the domain of Hough height from  $\alpha-5$  to  $\alpha+5$  and Hough width from 0 to Houghwidth/2. On the contrary, the maximum Hough point can be found in the domain of Hough height from  $\alpha-5$  to  $\alpha+5$  and Hough width from Houghwidth/2 to Houghwidth. The straight-line corresponding to the Hough point is noted as L3.

Step8: Using the same way from step7 we can find the fourth boundary which is parallel to L2, noted as L4.

Figure7 is the process of precise positioning of DM code based on Hough transform.

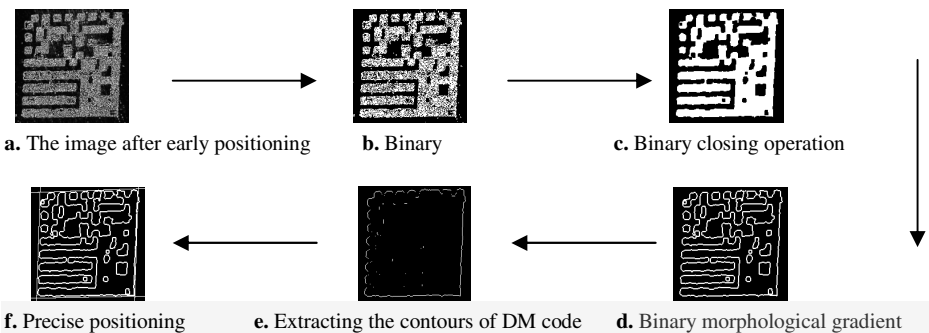
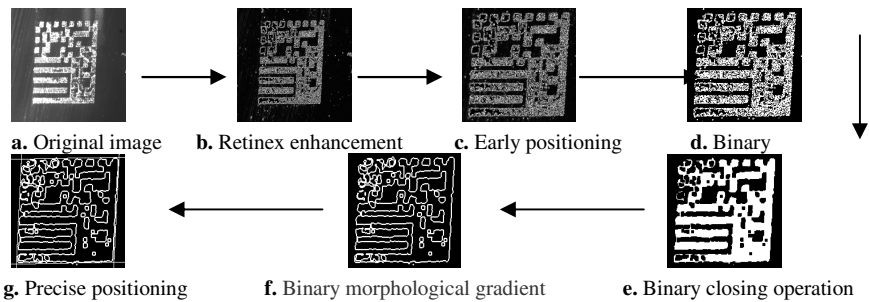


Fig. 7. The process of precise positioning of DM code based on Hough transform

## 5 Experimental Results and Analysis

In order to verify the effectiveness of DM code precise positioning in this paper, the experiments are made in the imaging system of code recognition marked in cylinder glasses. The camera model is selected as MV1300-UM. The image collected by the camera is processed in the Visual studio 2005 platform of the computer. 60 images of DM code marked in the cylinder glasses are collected from the camera. The size of DM code ranges from 4 millimeter to 10 millimeter. Use the way in this paper to position the DM code. Choice  $\gamma$  as 2 in Gamma correction. Under the experimental conditions, the correct rate of DM code positioning is more than 90 percents, and the time is less than 100 milliseconds.

Figure8 is an example of DM code positioning in the way of this paper.



**Fig. 8.** The total process of DM code positioning

The experiment shows that the images of DM code marked in cylinder glass can be positioned precisely in the way of this paper, but some of them can't be positioned well. The reason of failure positioning of DM code is that the code positioning will make mistake from early positioning when there are many noises in the boundaries of the DM code images. Therefore, the method of this paper has certain limitations.

## 6 Conclusions

This paper presents a method of DM code precise positioning marked in cylinder glasses. The experiment shows that it can not only position the DM code area precisely in this way but also remove the background noise in some degrees to convenient the follow up decoding. It applies to practical problems and provides a reference solution for the same problem.

## References

1. Wang, S., He, W., Zhang, W., et al.: Application research on the direct laser marking and 2D barcode technology on tool marking and identification. *China Mechanical Engineering* 18(6), 676–680 (2007)

2. Wang, S.-A., He, W., Wei, Z., et al.: Direct tool marking and identification method. *Computer Integrated Manufacturing Systems* 13(6), 1169–1174 (2007)
3. Liu, N., Yang, J.: Recognition of two-dimensional bar code based on Fourier Transform. *Journal of Image and Graphics* 8(8), 877–882 (2003)
4. Heping, Y., Wang, Z., Guo, S.: A method for 2D bar code recognition by using rectangle features to allocate vertexes. In: Liu, W., Lladós, J. (eds.) *GREC 2005. LNCS*, vol. 3926, pp. 99–107. Springer, Heidelberg (2006)
5. Liang, Y., Wang, Z., Cao, X., et al.: Real time recognition of 2D bar codes in complex image conditions. In: *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, pp. 19–22, 1699–1704 (2007)
6. Liu, N., Yang, J.: Recognition of two- dimension bar code based on waveform analysis. *Journal of Computer Research and Development* 41(3), 463–469 (2004) (in Chinese)
7. Jain, A.K., Yao, C.: Bar code localization using texture analysis. In: *Proceedings of Document Analysis and Recognition, the Second International Conference*, pp. 41–44. IEEE Computer Society Press, Tsukuba Science City (1993)
8. Hao, Y., Qi, F., Jiang, R.: A novel machine learning based algorithm to detect dataMatrix. *Journal of Image and Graphics* 12(10), 1873–1876 (2007) (in Chinese)
9. Land, E., McCann, J.: Lightness and Retinex theory. *Journal of Optical Society of America* 61(1), 1–11 (1971)
10. Liu, N.-Z., Yang, J.-Y.: Recognition of Two-dimension Barcode Based on Projection Algorithm. *Computer Engineering* 9(28), 32–33 (2002)
11. Leavers, V.F.: Which Hough Transform. *Comput. Vis. Graph. Image Proecessing: Image Understanding* 58, 250–264 (1993)

# Analysis of Urban Traffic Based on Taxi GPS Data

Li Meng<sup>1,2</sup>, Li Ru-tong<sup>1,2,\*</sup>, Xia Yong<sup>1</sup>, and Qin Zhi-guang<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

<sup>2</sup> Popular High Performance Computers of Guangdong Province / Key Laboratory of Calculation and Application Service of Shenzhen, Shenzhen, 518000, China  
Lrutong@gmail.com, {mengli,xiayong,qinzg}@uestc.edu.cn

**Abstract.** Recently, the problem of traffic jam in major cities is getting worse. By leveraging the taxi GPS data of Shenzhen, this paper analyzes the urban traffic status and proposes rational suggestions for urban traffic management. In particular, this paper firstly presents the get-on and get-off points on GIS map based on taxi GPS data. Secondly, by using K-Means algorithm to allocate urban traffic cells, the hot areas where passenger flow is huge are pointed out. Finally, based on the taxi speed, we locate the crowded area and find the crowded period, then analyze the reasons that cause the traffic jam and propose rational suggestions for urban traffic management.

**Keywords:** GPS data, clustering algorithm, map matching, traffic analysis.

## 1 Introduction

Currently, many taxis are equipped with GPS, which can record location, speed, direction and other information of taxis. Taxis, as common vehicle, are becoming an important mobile data source because of its large coverage, accurate allocation and high continuity. Referring to the analysis of taxi GPS data, previous work mainly focused on the travelling characteristics of taxis and passengers [9] [10] [11] [12]. Some researchers proposed that 37% time can be saved for taxi drivers by recommending travelling path based on taxi GPS data. Besides, researchers also did abnormal trajectory detection, identification and other regional function work [1] [2] [19].

Taxi GPS data analysis mainly focused on three key points: (1) Taxi trajectory characteristic analysis, which is about how to extract the reliable trajectories from huge original data and then analyze the moving characteristics of taxis or passengers. (2) As taxis are the public carrier for passengers, its GPS data contains a wealth of passengers' information. How to excavate passengers' information from the GPS data (i.e., flow, density) is a hotspot in research community. (3) Social activity pattern excavation, which is mainly research on exploring the social events and their emerging patterns from passengers' information.

This paper is organized as follows. Section 2 states map matching. By using an open taxi GPS dataset, we presented passengers' get-on and get-off points on the GIS

---

\* Corresponding author.

map. Then, we can locate the heavy traffic area on the GIS map. In order to have a better study of the city's traffic flow information, the K-means clustering algorithm is used to divide the city into traffic zones and to locate the hotspots where traffic flow is heavy. In section 3, we divided one day into 12 periods, and defined traffic jam circumstance based on taxis speed. By analyzing the experimental results, we found the congested periods of the corresponding roads and the potential reasons for congestion. Then, based on our results, we can provide reasonable suggestions for city's road traffic management. Major contributions of this paper are as follows: (1) Map-matching taxis data to visualize the passengers' get-on and get-off points; (2) Using K-means algorithm to divide traffic zones and locate hotspots; (3) Locating road congestion for urban road traffic management and providing reasonable suggestions.

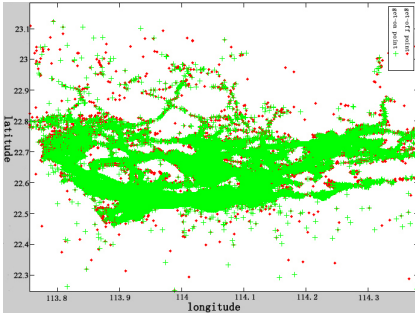
## 2 Road Conditions Analysis Methods

The paper is based on the analysis and processing of Shenzhen taxi GPS data, and mapping the Shenzhen road traffic conditions. In section 2.1, we marked the locations of the taxis at the latitude and longitude coordinates for the horizontal and vertical spatial distribution map, drew a taxis travelling map, and determine taxis driving traces. However, due to the mobility of taxis, taxis driving traces may deviate from the original tracks. In order to properly represent the taxis moving tracks, we matched passengers' get-on and get-off points onto the coordinate map, and located the main areas of social activities on the GIS map. In section 2.2, K-means algorithm was used to conduct a traffic cell division that was further used to analyzed the city's road conditions. In section 2.3, we defined traffic jam circumstance and found the peaks of traffic congestion.

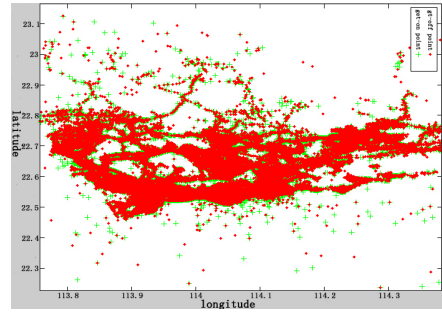
### 2.1 Map Matching

Map matching is mainly used to correct digital maps error, GPS location error and coordinate conversion error. Because of the three kind of errors, there could be a deviation between recorded driving traces and their original traces [15] [16]. Therefore, it is necessary to use map matching to adjust and reposition the taxis traces.

In order to locate passengers' get-on and get-off points and find out the range of latitude and longitude, we matched the GPS data with the coordinate map. Figure 1 shows the map matching result. The green dots are the get-on points and the red dots are the get-off points. The abscissa is longitude and the ordinate is latitude. From the figure, it can be found that passengers' get-on and get-off points cover the entire coordinate map of Shenzhen, and the distribution of such dots is dense in central area and sparse in surrounding, which fits reality circumstance.

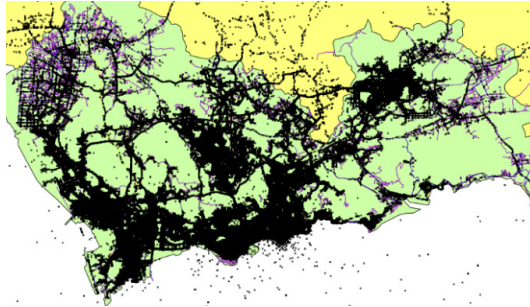


**Fig. 1(a).** Get-on and get-off points



**Fig. 1(b)** Get-on and get-off points

The basic idea of map matching is as follows. We match the taxis trajectory with vector electronic map, find out current travelling road, and project taxis' current anchor point on the road [17]. Its application is based on two premises: one is precision digital map and the other is taxis on the road, which ensures the positioning error will not breaks away from the original track. Additionally, we only keep the component vector that follows current moving path by using projecting. In this way, the positioning accuracy of the taxis is improved [5] [6]. In this paper, we used traditional map-matching algorithm [7]:



**Fig. 2.** Map matching

- Step One: Collecting GPS positioning data;
- Step Two: Determining whether the positioning data is invalid or not; if it is, speculating on the historical location data matching, and then turn to step eight;
- Step Three: Judging whether taxis' current state is in a stopped or taxiing state; and if it is stopped, processing it and then goes to step eight;
- Step Four: Within a threshold, if the number of searchable roads is less than 1, it indicates that the taxi is not on the road. Exiting the matching process and considering the GPS data as the taxis' current position;
- Step Five: Within a threshold, if the number of searchable roads equals to 1, it indicates that the taxi is on the road. Projecting directly and considering the taxi's current position is on this road;



- Step Six: Within a threshold, if the number of searchable roads is larger than 1 and the searchable roads are the same road, it indicates that the taxi is nearby this road;
- Step Seven: With a threshold, if the number of searchable roads is larger than 1 and the searchable roads are different roads, it indicates that the taxi is on one of several similar roads;
- Step Eight: the end of this match.

By using the GIS information provided by Mapinfo, passengers’ get-on and get-off points are matched to the GIS map. The matching result is shown in Figure 2.

### 2.2 Traffic Zone Division

The purpose of traffic zone division is to divide the coordinate map into several traffic zones, calculate the throughput of traffic flows for each zone, and compute the taxis’ dynamic migration from one zone to another. It can be used to locate hotspot areas and to reflect capacity state [8]. We defined hotspot area as the place that passengers’ flow is heavy. Locating hotspot areas can benefit urban traffic management. We defined the accumulated times of passengers’ get-on taxis as discharging amount, and the accumulated times of passengers’ get-off taxis as absorbing amount. Combining the discharging amount and the absorbing amount can measure the impact of one zone to the city traffic [13] [14]. The zone that has high impact suffers more pressure and it is generally crowded.

We use K-Means algorithm [3] to classify the points on the coordinate map and divide the city into zones to locate hotspot areas. K-Means is a classic data processing algorithm using partitioned clustering methods. The purpose is to find out typical point or central point from mass data and use such point for subsequent processing.

### 2.3 Road Congestion Detection

We divide one day 24 hours into 12 periods, with each one 2 hours, to calculate the mean speeds of taxis in different periods. The results are shown in figure 3.

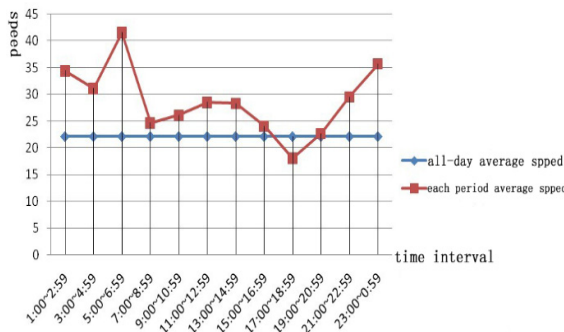


Fig. 3. Speed in each period

It is shown in the figure that two low peaks which indicate the mean speeds of taxis are slow. One of the low peaks is in period 7:00-9:00, and the other in period 17:00-21:00. We can infer that these two periods are associated with high passenger flows. Therefore, we judge these two periods as congested period. Moreover, the mean speed after 19:00 is also low, if we defined the congested path as where the mean speed is below 25 km/h, it is accurately to select 7:00-9:00 (morning peak) and 17:00-19:30 (evening peak) as congested period.

### 3 Experiment

We used the open dataset, which was collected from April 18, 2011 to April 26, 2011, including 13,799 Shenzhen taxi GPS data. It is workday from April 18 to April 22. And it is weekend from April 23 to April 24. The GPS data contains the taxi license plate number, acquisition time, latitude, longitude, taxis passenger status, instantaneous speed, driving directions and other information. By analyzing the taxi GPS trajectory, we dig out potential valuable information, such as hotspot areas, traffic congestion and so on.

#### 3.1 Data Processing

When processing the taxi GPS data, we firstly eliminated duplication, erroneous and incomplete data, then used the K-Means algorithm to cluster points on the coordinate map, finally divided passenger travelling areas into traffic zones to locate hotspot areas. Table 1 shows the database.

**Table 1.** Taxi GPS database

Index	Content	Name	Type	Size	Note
0	GPS_ID	ID	int	20	unique identification code
1	car number	CarNumber	varchar	15	Car identification code
2	longitude	GPS_X	float		coordinate
3	latitude	GPS_Y	float		coordinate
4	date	GPS_Date	varchar	16	0 for vacant 1 for laden
5	capacity state	GPS_State	varchar	3	Crowded identification
6	instantaneous speed	GPS_Speed	float		True

#### 3.2 Traffic Monitoring

Since Shenzhen traffic road situation is complicate, firstly we divided the whole area into seven major areas, then for each major area we subdivide it into 50 small blocks, so it is capable of locating exact congested points. Figure 4 shows the result of zone division and the matching with coordinate map. In figure 4, two different colors

represent different traffic zones. The points, which are divided into the same traffic zone, have certain correlation and similarity. Each zone reflects the temporal and spatial variation characteristics of urban road network traffic. We found that the center point of each zone also matches the center point of seven administrative districts coordinate of Shenzhen.

As shown in figure 4, based on passengers' get-on and get-off points, traffic zones division approximately matches the seven administrative districts of Shenzhen, although there is some difference. In particular, the left black area matches Guangming District, the left green area matches Bao'an District, the left blue area matches Nanshan District, the red area matches Futian, the right blue area matches Luohu District, the right green area matches Yantian District, and the right black area matches Longgang District. Apparently, the traffic zones division is also closely related to urban population, area, economic characteristics and industrial structure. Next, we can analyze the characteristics of residents' travel traces.

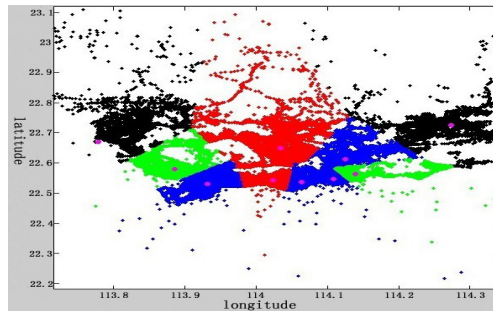


Fig. 4. Passengers' get-on and get-off points divided into traffic zones

In order to verify the accuracy of the zone division, we divide the passengers' get-on and get-off points into 50 blocks with the size of 400\*400 square meters, and calculate the taxi throughput in each block, then illustrate the regional function of each zone. Figure 5 shows the divided 50 blocks.

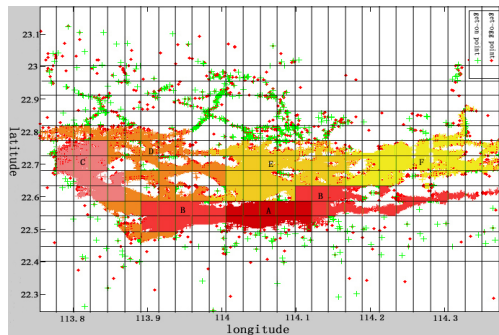


Fig. 5. Flow chart

As shown in figure 5, we use different colors to represent each region's flow. In particular, magenta represents the largest flow area, follows the trend of red, pink, orange, yellow. It shows that the taxi throughput is reducing. Compare with the map of Shenzhen, we infer as follows. Area A is mainly for commercial entertainment. Area B is mainly for resident, government and community, and relatively few for commercial entertainment. Area C is mainly for residential land and industrial land, where the industrial land takes a larger proportion. Area D, E, F are mainly for industrial land, residential land and ecological land, where the ecological land takes a larger proportion.

### 3.3 Road Congestion Detection

Here, we locate the congested roads. As stated in section 2.3, period 7:00-9:00 and period 17:00-19:30 are congested periods. 7:00-9:30 is the morning peak and 17:00-19:30 is the evening peak. If the mean taxis speed is below 25km/h, we consider roads as crowded. Figure 6 shows the corresponding congested roads in the morning peak and evening peak. The abscissa is the longitude and the ordinate is the latitude. The red dots represent the congested roads where the mean taxis speed is below 25km/h, the green dots vice versa.

As shown in figure 6, we can infer that the major congested roads in morning peak are as follows: Nanping Expressway and Qingping speed intersection, Mei Guan interchange at North Central Avenue Road intersection with Nigang, North Central Avenue East, etc. The major congested roads in evening peak are as follows: Fulong Nanping Road and the intersection of Riverside Avenue and the new Island Road intersection, CaiTian Road and Fuhua Road intersection, Honey Lake Road intersection with Riverside Avenue, Riverside Avenue and Hohai Avenue intersection, etc.

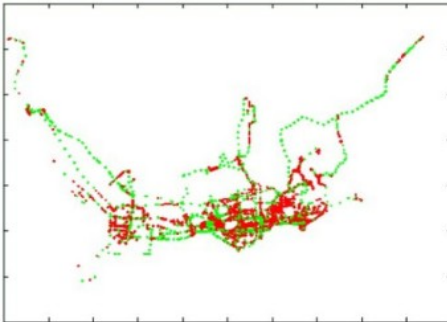


Fig. 6(a) Morning peak road conditions

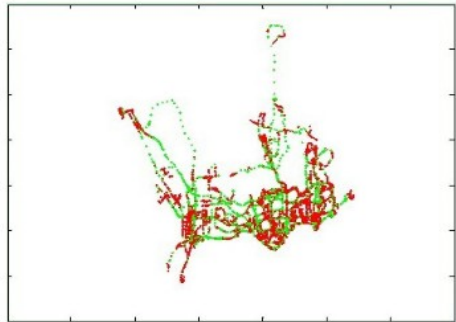


Fig. 6(b) Evening peak road conditions

## 4 Conclusion

In this paper, based on taxis GPS data, we presented passengers' get-on and get-off points on coordinate map, and used the K-means cluster algorithm to divide urban

traffic zones and locate the hotspot areas. By combining taxis' moving speeds and location information, we can infer the congested periods, congested roads, and further provide rational suggestions to improve urban traffic management.

**Acknowledgement.** This work is supported in part by the National Science Foundation of China (No.61133016), Ministry of Education - China Mobile Research Foundation (No.MCM20121041), Shenzhen Key Laboratory of Service Computing and Applications (SZU-GDPHPCL-2012-11), and Program of Sichuan Province Applied and Basic Research (No.2013JY0116).

## References

1. Yu, Z., Xing, X.: Learning travel recommendations from user-generated GPS traces. *ACM Transaction on Intelligent Systems and Technology*, 2–19 (2011)
2. Yu, Z., Liu, Y., Jing, Y., Xie, X.: Urban Computing with Taxicabs. In: *Proceeding of the 13th ACM International Conference on Ubiquitous Computing*, pp. 89–98 (2011)
3. Lee, Y.-J., Vuchic, V.R.: Transit Network Design with Variable Demand. *Journal of Transportation Engineering* 131(1), 1–10 (2005)
4. Tom, V.M., Mohan, S.: Transit route network design using frequency code dgenetic algorithm. *Journal of Transportation Engineering* 129(2), 186–195 (2003)
5. Fan, W., Machemehl, R.B.: A Tabu Search Based Heuristic Method for the Transit Route Network Design Problem. In: *The 9th International Conference on Computer-Aided Scheduling of Public Transport* (2004)
6. Fan, W., Machemehl, R.B.: Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering* 132(2), 122–132 (2006)
7. Fan, L., Mumford, C.: A Metaheuristic Approach to the Urban Transit Routing Problem. *Journal of Heuristic* (2008)
8. Fan, L., Mumford, C., Evans, D.: A simple multi-objective optimization algorithm for the urban transit routing problem. In: *The Eleventh Conference on Congress on Evolutionary Computation*, pp. 1–7 (2009)
9. Qian, Z., Xu, E., Wang, Z., Yafei, D.: DNA Algorithm on Optimal Path Selection for Bus Travel Network, pp. 245–248 (2009)
10. Liu, L.-Q., Zhang, Y.: Research of Urban Bus Stop Planning based on Optimization Theory, pp. 551–554 (2009)
11. Tang, M., Ren, E., Zhao, C.: Route Optimization for Bus Dispatching Based on Genetic Algorithm-Ant Colony Algorithm, pp. 18–21 (2009)
12. Xu, C., Ji, M., Chen, W., Zhang, Z.: Identifying travel mode from GPS trajectory through fuzzy reasoning. In: *Proceeding of the 7th International Conference on Fuzzy Systems and Knowledge Discovery*, Yantai, pp. 889–893 (2010)
13. Chen, W., Ji, M., Shi, B., Xu, C., Zhang, B., Deng, Z.: A prompted recall interview platform for GPS-based household travel surveys: design and development. In: *Proceeding of International Transport GIS Conference*, CDROM, Wuhan (2009)
14. Schaller Consulting. *The New York City Taxicab Fact Book*, Schaller Consulting, Brooklyn, NY (EB/OL)

15. Yang, H., Wong, S.C., Wong, K.I.: Demand supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research: Part B* 36(9), 799–819 (2002)
16. Yang, H., Ye, M., Wilson, H.T., et al.: Regulating taxi services in the presence of congestion externality. *Transportation Research: Part A* 39(1), 17–40 (2005)
17. [19] Jing, Y., Yu, Z., Zhang, L., Xie, X., Sun, G.: Where to Find My Next Passenger? In: *Proceeding of the 13th ACM International Conference on Ubiquitous Computing* (2011)
18. Jing, Y., Yu, Z., Xing, X.: Discovering regions of different functions in a city using human mobility and POIs. In: *Proceeding of the 18th SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 20–29 (2012)
19. Yang, D., Cai, B., Yuan, Y.: An improved map-matching algorithm used in taxis navigation system. In: *Proceedings of Intelligent Transportation Systems, Beijing*, pp. 1246–1250 (2003)

# Research and Application of Embedded System Development Based on Petri Nets

Chen Yu

Department of Computer Science and Engineering, ChongQing University of Technology,  
No.69, HongGuang DaDao, BaNan, ChongQing, China, 400054  
boyyu788@163.com

**Abstract.** Because of the widely application in various industries, the embedded system becomes more and more complicated. The key to the system development is how to divide its functions into multiple tasks and design a detailed corporation process across different tasks. This paper introduces the methods about how to use FSM and Petri nets to design an application system, by an instance.

**Keywords:** Embedded system, Development method, FSM, Petri-nets.

## 1 Introduction

An embedded system is a dedicated computer, which is a combination of hardware and software. It is generally designed to perform a specific function by running its hardware components to execute orders of the software. The key characteristic of embedded system is:

- Interact with the external environment: The embedded system is activated by events, and must responds to external events. In other words, it must make decisions when it received data from external environment, then output control.
- Multi-tasks' synchronous processing: There are generally several events occur at the same time within an embedded system. The function of the whole system is completed by running multiple tasks and each task describes a single event independently.

A good embedded system is minimum system to achieve a specific application of the special function. Especially, the embedded software is the key to realize the system. Because of huge market, advances in technology, and application complexity's increase, the development method's of embedded system is one of the most active direction of computer hardware and software technology development.

## 2 Developing Method of Multitask System Base on Embedded

By multi-task scheduling, the real-time multitask system realize system functions. it ensure that important events could make the right response within a specific time

period, typically without buffering delays. Since multiple tasks are running independently, its CPU provides service to each task by turns.

So the first step to develop an embedded system is to confirm the function of the system according to detailed target for the design. That is the whole application system is divided into several tasks. As a result, each one of them has functions to complete certain requirements of the application. Then priority of the task list should depend on connections among tasks. Finally, the workflow between tasks can be designed as well.

## 2.1 To Divide the Application on Multiple Tasks

In order to corporate tasks and simplify the system, there is often a logical requirement to divide the application as tasks. The cost of switching between tasks a lot could be increased if the application is over-divided. Similarly, if the application is not divided fully, the parallel operation can only be completed serially, the system throughput would decrease. For getting the balance and trade-off between efficiency, real-time and system throughput, The following methods can be used to divide as tasks:

- Draw data flow diagrams of the system, analyze data transformations, and determine which can be parallel, which must be executed in order.
- Divide the application into several tasks with considering I/O Requirement between changes, execution frequency, timing requirement, asynchronous relationship of the function and the priority, and the priority depends on its function or logic should be taken into account as well.
- In order to simplify the design, some connected tasks, such as missions that have same deadline or are activated by same event, can be combined to share resource and execute by same event.

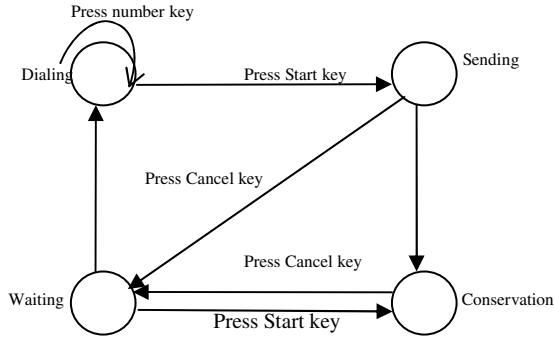
The development of telephone system can be illustrated as an instance. Its application is divided into several tasks including idle (running without other processes), accepting input (press keypad, such as numeric keys and star key), conservation (on a call), and display (shows dates, time and caller ID for an incoming call). By accepting different input event, each task can change its states to complete the appropriate functions.

## 2.2 To Analyze the Workflow within Tasks by FSM

In a embedded system, The running of a task which is activated by related case changes external environment and the state of facilities and instruments to achieve the goal of control. So in order to establish the model of FSM with a clear structure, we can analyze by FSM.

In example of phone system, after segregating and combining its work process by FSM, a has four states related to a conservation event including waiting, dialing, sending and conservation. The phone is switching one of these states to another





**Fig. 1.** A FSM in a conversation event

within the entire calling process. The initial state of the phone is waiting and it can change into a certain state when initiated by a triggering event or condition, such as numbers of its keypad are pressed. the communication task’s FSM is as shown in Figure 1.

**2.3 To Analyze the Relationship among Tasks by Petri Nets**

It is suitable FSM describes the limited state and the transfer between these states. However, when transitions of system parts are a lot, it will be very complicated with an increase of states in the system, so, FSM is not suitable for concurrency, and it can not describe asynchronous system. Petri nets has better modeling capabilities than FSM, it can describe asynchronous system synchronous system.

A Petri nets is a mathematical model for the definition of systems with independent, multiple processes and tasks. And a Petri net have an exact mathematical definition of states to describe events, with a well-developed mathematical theory for process analysis. It is good at describing characteristics of this kind of systems including development, competition, concurrency, and so on.

Petri net basics are including:

- Place, a list of finite states, circle node. Place is marked in the analyzing process and may contain a discrete number of tokens, which are represented by an asterisk. The resource is available if there are sufficient tokens in places at the beginning.
- Transition, rectangle node, and mean events.
- Arc, it is a connection between place and transition. A Arc is from Place to Transition, this Place is called input Place of this Transition, Otherwise, it is called output Place of this Transition. Each transition can has many I/O places.

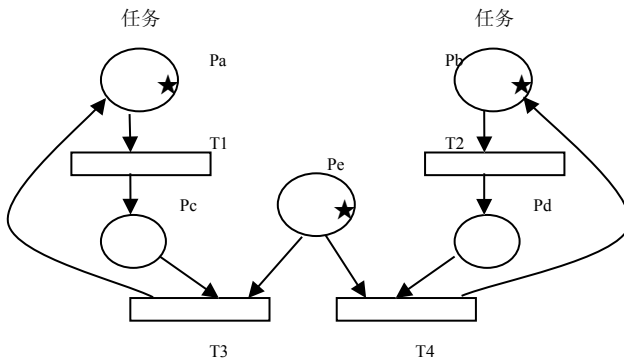
In phone system, the meaning of every Place and Transition is shown as Table 1.

**Table 1.** the definition of Places and Transition

Place	Definition	Transition	Definition
Pa	Initial waiting state	T1	Incoming call
Pb	Initial waiting state	T2	Press number
Pc	Answered State	T3	Answered
Pd	Sending state	T4	Start calling
Pe	Conservation		

Figure 2 illustrate the Petri nets model of conservation task in communication system including two independent tasks. At the beginning of the conservation task, the system is in the state of waiting. The function of the first and second task is picking up when there is an incoming call. Summarily, to start a call is the function of the second one in the system. The initial places that are included Pa, Pb and Pe have Token, which represents that the resource is available.

In Figure 2, The Arc is from Pa to T1, which means Pa is a input Place, and Pa is precondition of T1. The Arc is from T1 to Pc, which means Pc is an output Place, Pc is result of T1. If every input Place of a Transition has at least one Token, this Transition is called enable. For example, T1 and T2 are enable, They would be active. T3 and T4 are not enable.



**Fig. 2.** A Petri's example

In a Petri nets, the definition Fire means the process of consuming one Token from every input place and creating one token in every output places. The Petri nets in Figure 3 is based on Figure 2 after firing T1 and T2. In a Petri nets, transition means events and fire describe their occurrences. As a result, a transition is enabled, if it meets all requirement of appearance of this event, which is shown with an asterisk in places. Two tasks of this system run without interruptions in the start. As we know, T3 is enable, when T1 is fired, and T4 is enable after firing T2. However, the fire of T3 and T4 is competing when both T1 and T4 are fired, and only one of them can get Pe, to run next step (Pe means pressing the key of start or sending). Of course, in order to solve conflicts and run two tasks are running concurrently, two Token would be designed for Pe, if the resource is available.

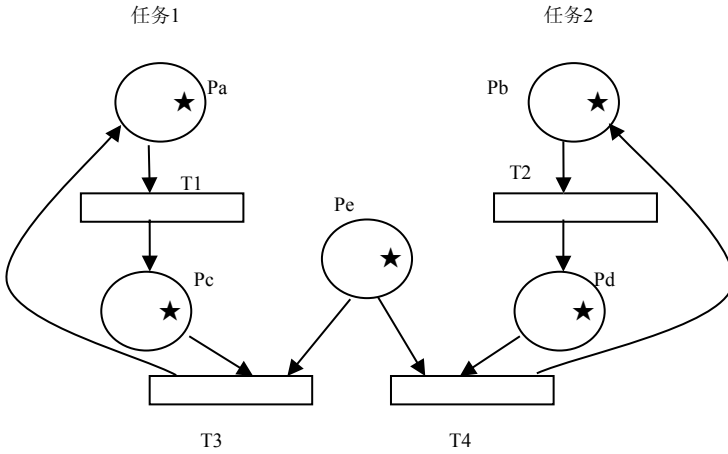


Fig. 3. T1 and T2 fired

### 2.4 The Design of Communication System among Tasks

The structure of tasks depends on the relation of transitions, which is the logic behind tasks. The basic point to transitions and communications among multiple tasks is the real-time core within the system that can change to different tasks and start running new missions according to their states and priorities. Since different systems can use various communication methods, the system can constantly change to several missions and communicate between each other.

## 3 Conclusion

FSM is simple and easy to use, because the system can deal with different states. When the conditions are met, the state-jump function is completed, It is good not only for controlling ranges and structures of the system but also for handling motion characteristics. In sum, are great for building model to describe the concurrency of system, which cover weakness of its synchronized characteristic. Petri net offers great support for modeling, and description of characteristics of system, such as concurrent, synchronize, etc.

## References

1. Shu, Z.B.: The idea and method of embedded software designing, pp. 66–72. Beijing University of Aeronautics and Astronautics Press, Beijing (2009)
2. Ge, J.: The design method of Embedded System base on Petri net (March 2009), <http://www.docin.com>

# CloStor: A Cloud Storage System for Fast Large-Scale Data I/O

Wang Ying<sup>1</sup>, Yang Dongri<sup>1</sup>, and Liu Peng<sup>2</sup>

<sup>1</sup> College of Engineering & Information Technology,  
University of Chinese Academy of Sciences, Beijing, China  
Yingw@ucas.ac.cn, Yang.dr@163.com

<sup>2</sup> PLA University of Science and Technology, Nanjing, China  
gloud@126.com

**Abstract.** Non-structured and semi-structured large-scale data storage has become a hot research topic. This paper has, through analysis and research on the fast I/O mechanism of large-scale block data, designed a fast data access mechanism. The system can achieve efficient storage of large-scale small files, reduce the I/O frequency of metadata control information, improve the storage and access performance of small files, reduce the memory utilization of metadata and maximize the number of files supported by the system; meanwhile, the system can enhance the storage performance of large files by providing such files with high-speed parallel access interfaces.

**Keywords:** Cloud Computing, Cloud Storage, Distributed File System, Fast Access.

## 1 Introduction

In 2006, Google, Amazon, and other companies proposed the concept of cloud computing. This was interpreted in various ways by the industrial and academic communities, but its mainstream idea is basically consistent with the prediction of John McCarthy, the pioneer of AI, that ‘Computing may someday be organized as a public utility’[1]. In January 2011, the National Institute of Standards and Technology(NIST) defined cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction’[2]. The concept of cloud computing and its popularization have resulted in an explosion of data. Google processes over 400PB of data per month, eBay’s analysis platform processes 100PB of data per day(exceeding the daily amount processed by the NASDAQ Stock Market),and Wal-Mart processes 1 million transactions per hour. Some 2.5PB of data, which is 167 times the amount of data held in the Library of Congress, will be stored in the database[3]. The storage and management of such data, including block data that can be packed and scattered into a large amount of small files, in a simple, efficient, and expandable way is a challenge for the industry and the science and technology community.

## 2 Related Work

The core of cloud storage is the distributed file system. Today’s distributed storage technologies are mainly classified by implementation mode: distributed block storage, distributed file system storage, distributed object storage, and distributed table storage (Table 1).The technical architecture can be classified into distributed file systems with or without management nodes (Table 2).

**Table 1.** Comparison of different distributed storage technologies[8-10, 12-16]

Storage Type	Systems
Block storage	IBM XIV storage system
File system storage	IBM SONAS GPFS Google GFS, Hadoop HDFS
Object storage	Amazon Dynamo, openstack swift EMC Atoms
Table storage	Taobao TFS

During the design of a cloud storage system, besides storage technologies and architecture, the fault tolerance of data should also be considered. There are two representative fault-tolerant technologies, replication- and erasure-code-based. Replication-based fault tolerance is easily realized and provisioned. However, multiple duplications of the same size must be created for each data object, taking up a large storage space. Erasure-code technology can integrate information from multiple data blocks, leaving little redundant information and effectively saving storage space. However, separate encoding and decoding resources are needed for the reading and writing of data. In case of data failure, the replication-based technology can easily and rapidly recover the data, where as erasure-code methods require the download of far greater amounts of data than the failure data, thus having a high recovery cost[15].

**Table 2.** Comparison of distributed file systems with and without management nodes [8-10, 12-18]

Architecture	Systems
Distributed file systems with Management nodes	Google GFS, Hadoop HDFS, IBM Blue Cloud
Distributed file systems without management nodes	Amaon Dynamo, Facebook Cassandra, Openstack swift

Today, only a few cloud storage systems are suitable for fast I/O of large-scale block data, and although some are open source, they fail to provide specific product and technical support. Traditional database management systems, e.g., the Oracle Real Application Server (Oracle RAC) parallel cluster, adopt the structure of centralized storage + memory integration. This requires communication between different

nodes via private networks, and has high bandwidth and intra-node clock synchronization requirements. Therefore, in practice, the expansion in the number of nodes is limited, far lower than that of Google's Cloud computing platform, which supports over 1000 nodes[18].

### 3 Architectural Design of the CloStor Cloud Storage System

According to the technical analysis above, and the consideration of a convenient storage function call as well as management and monitoring features, the CloStor cloud storage system is designed to attain excellent I/O performance with files and block data. The system architecture is shown in Figure 1.

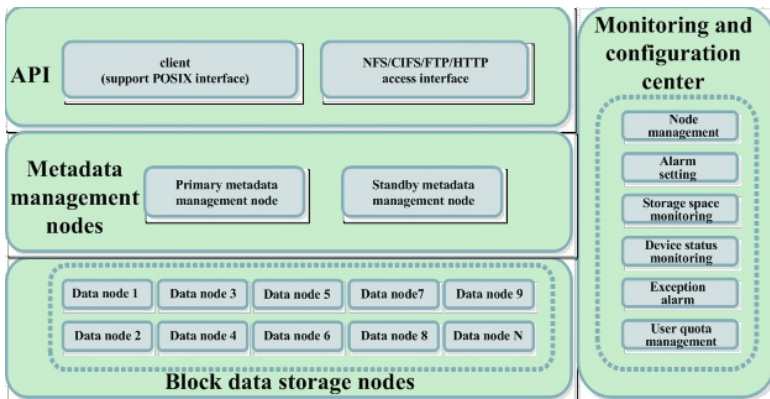


Fig. 1. Architecture of CloStor Distributed File System

CloStor adopts the distributed storage model with management nodes, and consists of an application interface module, metadata management module, block data storage module, and monitoring and configuration center module. The specific functions of each of these are described below.

- Application Interface Module
- Metadata Management Module
- Block Data Storage Module
- Monitoring and Configuration Center Module

The very low coupling degree between modules means that each can be provisioned separately. The system can be scaled up or down by adding or deleting storage nodes, and the system storage service will not be interrupted during this scaling. Meanwhile, because each module is an application running on the operating system, they can be provisioned on different server platforms, thus reducing the differentiation of hardware and facilitating subsequent expansion and maintenance.

## 4 I/O Strategy of the CloStor Distributed File System

### 4.1 Separating Data Flow from Control Flow with the Central Server Model

The specific I/O flow of files is as follows which is shown in Figure 2:

- The client communicates with the central management node network to request file reading and writing;
- The management node distributes storage service nodes under its management according to the client request, and returns storage service node information to the client;
- The client directly conducts data reading and writing with corresponding storage servers;
- The client returns the completion status to the center management node upon completion of data I/O, which indicates the end of the I/O flow.

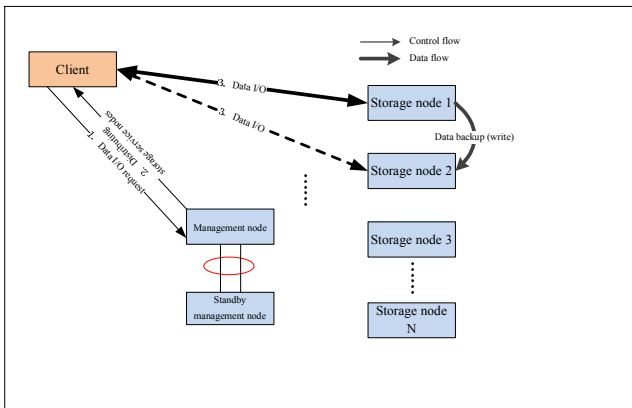


Fig. 2. I/O flow with data flow separated from the control flow

The technology for separating the data flow from the control flow not only transfers the load from the metadata servers to the storage servers, but also enhances the service capacity of the system. This is embodied by the nearly linear rise in system throughput rate. Meanwhile, adopting such a separated structure with low coupling not only maximizes the service capacity of each server, but also greatly enhances the maintainability of the system, providing the system with strong online scalability.

### 4.2 Improving I/O Speed by Reading and Writing of Metadata Memory

The cache mechanism is important in enhancing the performance of the file system. To improve their performance, general file systems usually need to realize complicated cache mechanisms.

Metadata servers need to operate on their metadata frequently. To improve the efficiency of these operations, metadata may adopt the memory cache strategy. This

allows all metadata operations in metadata servers to be directly carried out in the memory. However, in order to enhance the memory utilization, it is compulsory to adopt corresponding compression mechanisms to reduce the space occupied by the metadata.

Memory resources are very valuable in computer systems. For the storage of massive files, the system memory may not be able to meet demand. To solve this conflict, the storage system adopts the idea of hierarchical storage for metadata, and stores metadata on solid state disks to balance the access efficiency and the memory resources. In order to achieve outstanding performance, Clostor works with an in-memory dataset like Redis[20], a open source advanced key-value store. Figure 3 shows the main data structure in in-memory dataset. Through the method, the metadata can be accessed in memory instead of in disk.

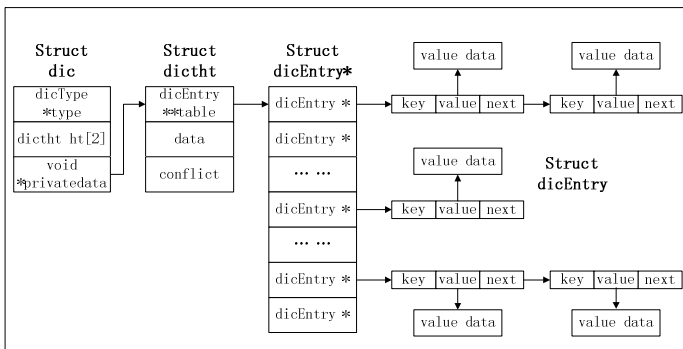


Fig. 3. Data structure in memory

### 4.3 Improving Write Speeds with Write-Combining

When writing data, the client will judge the block data to be read or written at each time step, and then proceed with the next I/O process according to the condition of returned data.

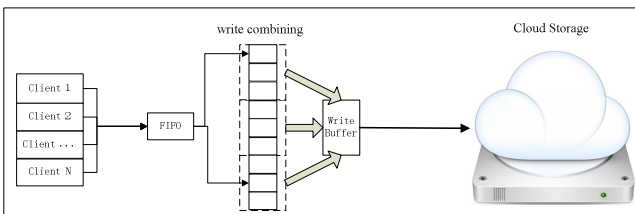


Fig. 4. The diagram of write combining

Figure 4 shows the theory of write combining, which is a computer bus technique for allowing data to be combined and temporarily stored in a buffer — the write combine buffer (WCB) — to be released together later in burst mode instead of writing (immediately) as single bits or small chunks. The write buffer can be treated as a fully



associative cache and added into the memory hierarchy of the device in which it is implemented. Adding complexity slows down the memory hierarchy so this technique is often only used for memory which does not need strong ordering (always correct) like the frame buffers of video cards.

## 5 Testing and Analysis Results

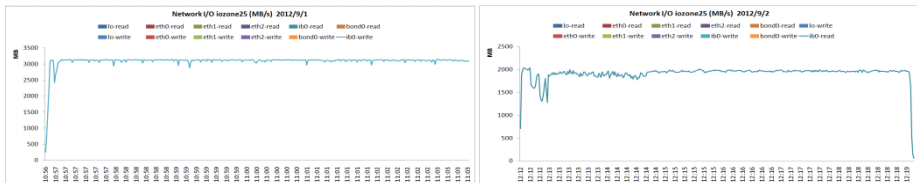
As no complete testing standard for cloud storage exists at present, we refer to the performance testing technologies and approaches of network-based storage systems. During the tests, data is first written to the logical volume, read, and finally checked. The consistency and I/O efficiency of the data can then be checked by comparing the check codes[19]. The test environment is described in Table 3.

**Table 3.** Test environment

No.	Device	Qty.	Model and Configuration
1	Metadata server	2	Dual-channel octa-core CPU, 32GB DDR3 memory, 2TB SATA system disk
2	Storage server	8	Dual-channel octa-core CPU, 16GB DDR3 memory, 500GB system disk, 3TB SATA data disk ×8
3	Client	1	Dual-channel octa-core CPU, 16GB DDR3 memory, 3TB system disk
4	Distributed file system	1	CloStor cloud storage system v1.0

Two test items are selected in the test: 1) single flow I/O speed of large files; 2) real-time I/O performance test of small files. The first item tests the access efficiency of system data flow, and the second item mainly tests the control and deployment capacity of metadata management nodes and the access efficiency of small files.

The test results for only one client and a 250 GB file are shown in Figure 6.



**Fig. 5.** Write performance and Read performance - single client

We can see from Figure6 the CloStor cloud storage system exhibits stable read and write performance with large files, achieving a write speed of 3GB/s, a read speed of about 2GB/s, and a maximum bandwidth utilization of 60%.

The real-time I/O test results are shown in Figure 7.

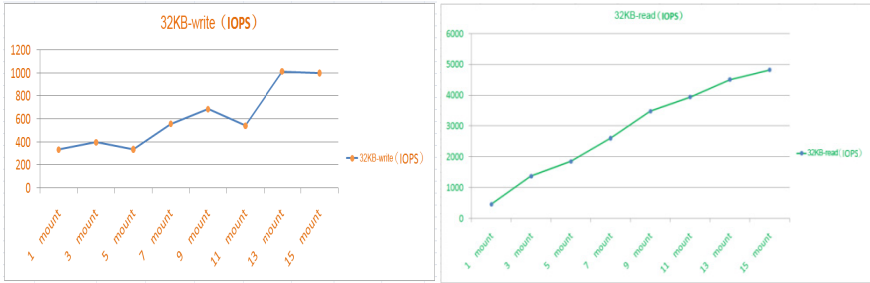


Fig. 6. IOPS for 32KB files

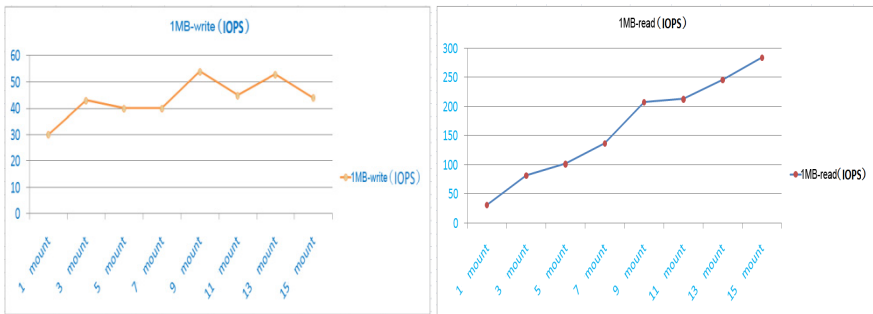


Fig. 7. IOPS for 1MB files

Figure 7 and Figure 8 show that the system has outstanding I/O performance for large files, whereas its I/O performance for small files requires further improvement. The system targets fast I/O applications with large-scale block data, such as video transmission systems. In this respect, the proposed system has reached the expected design goal.

## 6 Conclusion

In this paper, we have summarized the major technical research outcomes of recent years in the field of cloud storage. We have explored the basic architecture of cloud storage, as well as approaches to enhance its efficiency and simplify its application. The I/O strategies for large and small files have been established, and the CloStor system model has been developed according to the application demand for large-scale block data. CloStor may serve as a beneficial reference for the research and development of cloud storage. On the basis of this paper, the focus of future research will be the efficient storage of large numbers of small files. We believe that further development of cloud computing will increase the demand for large-scale block data storage, and this will become a core demand in different fields of the economy and society.

## References

1. Foster, I., Yong, Z., Raicu, I., et al.: Cloud computing and grid computing 360-degree compared, pp. 1–10 (2008)

2. Mell, P., Grance, T.: The NIST definition of cloud computing. National Institute of Standards and Technology (2011)
3. Li, G.J.: Scientific value of the study on big data. *Communications of the CCF* 8 (2012)
4. Kuang, S.H., Li, B.: System architecture and application cases of cloud computing. *Computer & Digital Engineering* 245 (2010)
5. Wu, Y.W., Huang, X.M.: Cloud storage. *Communications of the CCF* 5 (2009)
6. Weil, S.A., Pollack, K.T., Brandt, S.A., Miller, A.E.L.: Dynamic metadata management for petabyte-scale file systems. In: *Proc. of the 2004 ACM/IEEE Conf. on Supercomputing (SC 2004)*. ACM Press, Pittsburgh (2004), doi:10.1109/SC.2004.22
7. Hua, Y., Jiang, H., Zhu, Y.F., Feng, D., Tian, L.: SmartStore: A new metadata organization paradigm with semantic-awareness for next-generation file systems. In: *Proc. of the FAST 2009*, pp. 1–12. ACM Press, Portland (2009), doi:10.1145/1654059.1654070
8. Decandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Voshall, P., Vogels, W.: Dynamo: Amazon’s highly available key-value store. In: *Proc. of the SOSP 2007*, pp. 205–220. ACM Press, Stevenson (2007), doi:10.1145/1294261.1294281
9. Lakshman, A., Malik, P.: Cassandra: A decentralized structured storage system. *ACM SIGOPS Operating Systems Review* 44(2), 35–40 (2010), doi:10.1145/1773912.1773922
10. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. In: *Proc. of the Symp. on Operating Systems Principles (SOSP 2003)*, pp. 29–43. ACM Press, Bolton (2003), doi:10.1145/945445.945450
11. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.E.: Bigtable: A distributed storage system for structured data. In: *Proc. of the OSDI 2006*, pp. 205–218. USENIX Association, Seattle (2006), [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/research.google.com/zh-CN//archive/bigtable-osdi06.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/zh-CN//archive/bigtable-osdi06.pdf)
12. The Apache Software Foundation. HDFS Architecture Guide (2009), [http://hadoop.apache.org/common/docs/current/hdfs\\_design.html](http://hadoop.apache.org/common/docs/current/hdfs_design.html)
13. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The Hadoop distributed file system. In: *Proc. of the IEEE 26th Symp. on Mass Storage Systems and Technologies (MSST)*, pp. 1–10. IEEE, Lake Tahoe (2010), doi:10.1109/MSST.2010.5496972
14. Deng, Q.N., Chen, Q.: Cloud computing and key technologies. *Development & Application of High Performance Computing* 26 (2009)
15. Wang, Y.J., Sun, W.D., Zhou, S., Pei, X.Q., Li, X.Y.: Key technologies for distributed storage in the cloud computing environment. *Journal of Software, Institute of Software Chinese Academy of Sciences* 23(4), 962–986 (2012)
16. Chen, K., Zheng, W.M.: Cloud computing: system cases and research status. *Journal of Software* 20 (2009)
17. Luo, J.Z., Jin, J.H., Song, A.B., Dong, F.: Cloud computing: system architecture and key technologies. *Journal on Communication* 32 (2011)
18. Sun, J., Jia, X.Q.: Technical architecture of Google cloud computing platform and factors impacting its cost. *Telecommunications Science* (2010)
19. Zheng, L., Zhu, L.G., Zhao, Y.T., Yi, Q., Yan, C.P., Hu, H.Z., Yang, F.: Design of evaluation method of storage virtualization. *Computer Engineering and Applications* 46(36) (2010)
20. Zawodny, J.: Redis: Lightweight key/value Store That Goes the Extra Mile. *Linux Magazine* (August 31, 2009)

# Nonnegative Variable Weight Based Auto Parts Combination Demand Forecasting Model Research

Jian Gong<sup>\*</sup>, Chun Lu, and Xin Liu

Shanghai Key Laboratory of Financial Information Technology, Shanghai, 200433, China  
{gongjian, deadeye}@sufe.edu.cn, arrower@263.net

**Abstract.** Demand forecasting of auto parts is important for the management of the auto supply chain. After comprehensively exploring the characteristic of the auto parts and considering the strengths of some individual forecasting methods, this article chose ARIMA, SVR and RBF neural network based regression to develop a nonnegative variable weight combination model to forecast demand of auto parts for the auto aftermarket. This model improved accuracy and stability and had wider applicability.

**Keywords:** Auto-aftermarket, Auto Parts, Combination Model.

## 1 Introduction

With the rapid development of auto industry, the auto aftermarket is playing more and more important role in whole industry. How to increase efficiency and management are emergent problem to be solved. Accurate demand forecasting is important for improving the performance of the supply chain, which can help company to realize reasonable inventory management and improve management. However, there are so many kinds of auto parts and they have miscellaneous characteristics. The demand of auto parts is affected by many factors, such as vehicle population, season, state policy, economic situation and so on. Forecasting demands by experience will cause inadequate response to the market. So it is necessary to develop a widely applicable and accurate demand forecasting method for the auto aftermarket.

## 2 Present Research on Demand Forecasting of Auto Parts

At present, many researches have aimed at the auto parts forecasting and management for individual enterprise, especially for the automobile-manufacturing enterprises. Such research focused on sales order forecasting for auto parts based on inventory control [3]. In the market of auto part, enterprises serving for auto aftermarket also play an important role. But few researches have been conducted in demand forecasting for the enterprises on auto aftermarket. Provided that the auto parts

---

<sup>\*</sup> Corresponding author.

followed erlangian distribution, Viswanathan and Yan Feng adopted weighted averages method, moving average method and probabilistic method for forecasting service arrivals in a repair center [4]. Besides fault law, the demands of auto parts are affected by many exterior factors. Yun Chen and Heng Zhao proposed an ARMA model based on trend and seasonal factors to forecast the demand of auto parts [2]. In the research, some intelligent algorithms are used because of the complexity of demand series. Yun Chen and Ping Lin provided an improved Regression-Bayesian-BBNN (RBBPNN) based model [1] and Li Yu and Yun Chen proposed an artificial neural network based method [5] to realize the demands forecasting.

Most of the above work adopted an individual model. Combined model is seldom been used. But the individual model can't reflect all influence factors and is more vulnerable to influences and impacts, which means that the stability and applicability of the individual model was limited. There are several problems in present forecasting model. Firstly, present models hardly consider outliers which hide the real law of the auto parts demand. Secondly, present models can get good forecasting result when respectively handling data with or not with significant nonlinear characteristics, which means the applicability of models will be limited. Finally, combined models provided have negative weights which don't have practical significance. This article proposes a nonnegative variable weight combined forecasting model considering both significant exterior factors and nonlinear characteristics after handling the outliers. Such method expands the applicability of the forecasting model and is very important for improving the precision and stability of forecasting.

### 3 Characteristics Analysis of Auto Parts

With the development of auto parts market, factors influencing the demand of the auto parts have become more and more complex. Firstly, more and more different types of auto parts have been put into the market, which make the demand for the auto parts more uncertain. Secondly, the exterior factors affecting the demand for the auto parts, such as the disposable income of a family, the GDP, the economic situation, and so on, are very complicated, which cause the demand to fluctuate remarkably. The demand characteristics of auto parts are shown as the following.

- Different influence extent of the exterior factors. Different auto parts are affected by different exterior factors. Even the same factors affect different types of auto parts differently. An individual model wasn't enough for all parts.
- Missing historic data. Auto aftermarket started late in China. The promotion strategies and product strategy are changing frequently. So historic data for the auto parts are not enough and data missing usually happen, which will lead to the problem of small sample and influence the accuracy of forecasting.
- Subject to appear outliers. Many sales data have outliers, which are bigger or smaller obviously than the other data. Such outliers can't reflect the demand law of the auto parts.
- Non-linear characteristics of the historic demand data. The non-linear characteristics include two aspects, the first one is that some data don't have trend and display high fluctuation. The second one is that some historic demand data have non-linear relationship with the influencing factors.

From the above analysis, there is no individual model to forecast accurately for auto parts with different characteristics. So a combination model was necessary to consider the applicability of some individual models.

## 4 Nonnegative Variable Weight Combined Model

The processes of developing nonnegative variable weight combined model of demand forecasting for auto part are described in the following.

### 4.1 Outliers Handling for the Historic Sales Data

Before model construction, we need some methods to find these outliers and normalize them to assure the scientific value of the combined model.

- 1) Handle the missing data as an outlier. In order to detect the missing data, a large number such as  $10^{10}$  need to be selected to replace the missing data
- 2) Use distance-based outlier method [6] to detect the outliers in one auto part's demand time series. Firstly, compute the distance  $p_{ij}$  of each demand data using the following formula in which  $x_i$  represents  $i$  period data in the time series.  $p_{ij}$  denotes the distance between  $i$  period data and  $j$  period data.

$$p_{ij} = |x_i - x_j|, \quad i = 1 \dots n, j = 1 \dots n \quad (1)$$

- 3) Define the distance threshold denoted by  $d$ , and then compare  $p_{ij}$  with  $d$ . For each period, compute the percentage of  $p_{ij}$  which value is more than  $d$ . The percentage is denoted as  $d_i$ .
- 4)  $f$  denotes the detection threshold. If  $d_i > f$ , the  $i$  period data is an outlier.
- 5) Use polynomial fitting method to handle the outliers detected in the above step.

### 4.2 Individual Method Candidates Selection

In order to improve the applicability of combined model, the individual methods selected should be complementary. For each type of demand characteristic, there is a corresponding method which can get good forecasting result.

*Definition 1.*  $T = \{t_1, t_2 \dots t_n\}$  denotes the characteristics set.  $M = \{m_1, m_2 \dots m_s\}$  denotes the selected methods set and  $s \leq n$ . For each  $t_i$ , there is a  $m_i$ , which will forecast well for data with the characteristic  $t_i$ . Then  $M$  is complementary.

The characteristics are analyzed from exterior factors and nonlinear which are shown in table 1. In table 1, the characteristics set have four elements.

**Table 1.** Characteristics set T

Dimension		Influence of exterior factors	
		Significant	Not significant
Nonlinear	Significant	<i>t</i> <sub>1</sub>	<i>t</i> <sub>3</sub>
	Not significant	<i>t</i> <sub>2</sub>	<i>t</i> <sub>4</sub>

For the auto parts with characteristics of *t*<sub>1</sub> and *t*<sub>2</sub>, exterior factors methods are suitable, such as regression and neural network approach. The regression equation is usually known when using regression approach. But the distribution function of the demand for auto parts is commonly unknown. So we use RBF to set up the model. For the auto parts with characteristics of *t*<sub>3</sub> and *t*<sub>4</sub>, time series analysis methods are suitable. In our research, ARIMA and SVR are chosen. ARIMA is a method to analyze random time series and will get good forecast for time series whose nonlinear is not significant. That means ARIMA is suitable for auto parts with characteristic of *t*<sub>4</sub>. SVR is suitable when handling significant nonlinear data, so it can be used for auto parts with characteristic of *t*<sub>3</sub>.

### 4.2.1 Demand Forecast Regression Based on RBF for Auto Parts

The process of demand forecasting for auto parts based on RBF are shown as the following.

- 1) Key exterior factors identification. After collecting ten exterior factors, denoted by *f*<sub>*i*</sub>, *i* = 1, ..., 10, correlation analysis is done firstly to computer the grey correlation between each exterior factor and the demand for the auto parts. Let *g*<sub>*i*</sub>, *i* = 1, ..., 10 represent the grey correlation. Compare each *g*<sub>*i*</sub> with predefined threshold  $\bar{g}$ . If *g*<sub>*i*</sub> >  $\bar{g}$ , then the factor *f*<sub>*i*</sub> is the key factor.
- 2) RBF neural network training. By step 1, there are *k* key factors identified. Then *k* is the number of neurons in input layer. Neuron in output layer is dependent variable which is the forecasted object denoted by *Y*<sub>*t*</sub>. Because the exterior factors has lagged influence on the demand, one period lagged data denoted by *g*<sub>*it*</sub>(-1) and two period lagged data denoted by *g*<sub>*it*</sub>(-2) are selected to train the network.

$$g_{it}(-1), g_{it}(-2), i = 1, \dots, k; t = 1, 2, \dots, N, N + 1, \dots, N + T \tag{2}$$

The first *N* - 1 time series of *k* key factors are used as training input data and the corresponding training output data are *Y*<sub>*t*</sub>, *t* = 3, ..., *N*. The other data are testing data. Let *P* and *Q* respectively represent the input data and output data in training phase.

$$P = \begin{bmatrix} g_{11}(-1) \dots g_{k1}(-1), g_{11}(-2) \dots g_{k1}(-2) \\ g_{12}(-1) \dots g_{k2}(-1), g_{12}(-2) \dots g_{k2}(-2) \\ \dots \dots \dots \dots \dots \dots \dots \dots \\ g_{1j}(-1) \dots g_{kj}(-1), g_{1j}(-2) \dots g_{kj}(-2) \end{bmatrix}, j = N - 1 \quad Q = \begin{bmatrix} Y_3 \\ Y_4 \\ \dots \\ Y_N \end{bmatrix} \tag{3}$$

In our research, ergodic process based on the principle of minimal MAPE is adopted to find the number of hidden layer.

- 3) Demand forecasting by the trained RBF model and evaluation.

**4.2.2 Demand Forecast Model Based on ARIMA**

The process of demand forecasting for auto parts based on ARIMA are shown as the following.

- 1) Test stationary by ADF and make non-stationary data stationary by difference
- 2) Automatically determine parameter p and q. Compute the autocorrelation coefficient (ACF) and the partial autocorrelation coefficient (PACF). Through ACF and PACF, estimate the max p and max q, denoted by  $M_p$  and  $M_q$  respectively. Traverse between 1 and  $M_p$  and traverse between 1 and  $M_q$ . Select the pair of p and q which results in minimal MAPE.
- 3) Estimate the parameters of ARIMA model by the method of least squares.
- 4) Set up ARIMA model, forecast and evaluate.

**4.2.3 Demand Forecast Model Based on SVR**

The process of demand forecasting based on SVR aiming at nonlinear demand data for auto parts are shown as the following.

- 1) Normalize demand data. This article adopts extremum method.
- 2) Model training. In this article,  $\epsilon$ -SVR is select to set up the forecasting model.  $\{Y_t, t=1,2,\dots,N, N+1,\dots,N+T\}$  denotes the time series of one auto part. The relationship between  $Y_t$  and  $(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$  is gotten by SVR model training. The time series  $\{Y_t, t=1,2,\dots,N\}$  is chosen for training. We construct the input matrix  $I$  and the output vector  $O$  as the following.

$$I = \begin{bmatrix} Y_1, Y_2, \dots, Y_p \\ Y_2, Y_3, \dots, Y_{p+1} \\ \dots \\ Y_{N-p}, \dots, Y_{N-1} \end{bmatrix} \quad O = \begin{bmatrix} Y_{p+1} \\ Y_{p+2} \\ \dots \\ Y_N \end{bmatrix} \tag{4}$$

During the training, there are two types of parameters to be given. One is the training dimension P which is searched by traversal between 1 and the max autocorrelation coefficient. The other is the training parameters including penalty factor C, insensitive loss factor  $\epsilon$  and Kernel function parameter  $\sigma$ . This article finds the set of parameters which results in the minimum training error by traversal.

- 3) Forecasting by the trained SVR model and evaluation.



### 4.3 Individual Model Sifting and Nonnegative Weights Computing

#### 4.3.1 Individual Model Sifting

In our research,  $n$  candidate individual methods have been selected through the above step. Before setting up the combined model, individual model will sifted again from the candidate methods by comparing MAPE and standard error. The time series  $\{Y_t, t = 1, 2, \dots, N, N + 1, \dots, N + T\}$  denote the actual demand data of one auto part. Series  $\{Y_t, t = N + 1, \dots, N + T\}$  were chosen as the testing data. The process is shown as the following.

- 1) Predefine the MAPE threshold  $\bar{m}$  and the standard error threshold  $\bar{\epsilon}$ .
- 2) Respectively forecast the demand by every individual model and computer the relative error denoted by  $e_i(t)$ , standard error denoted by  $\epsilon_i(t)$  and MAPE, which are shown as the following.
- 3) If  $MAPE_i(t) < \bar{m}$  and  $\epsilon_i(t) < \bar{\epsilon}$ , then choose method  $m_i$ .

$$MAPE_i(t) = \frac{1}{t} \sum_{j=N+1}^{N+t} e_i(j), (i = 1, 2, \dots, n; t = N + 1, \dots, N + T) \tag{5}$$

#### 4.3.2 Nonnegative Weights Computing

By the above sifting process, we will get  $p (p \leq n)$  individual models. The weight of the combined model at period  $t + 1$  will be computed by the  $p$  individual models' relative error series  $\{e_i(1), e_i(2), \dots, e_i(t), i = 1, \dots, p\}$ .

- 1) Let  $w_i$  denote the weight of method  $m_i$  in the combined model. The relative error of the combined model at each period can be computed by the following formula.

$$e(t) = \sum_{i=1}^p w_i * e_i(t), i = 1, 2, \dots, p \tag{6}$$

- 2) The sum of squares of the relative error of the combined model until period  $t$  can be computed by the following formula.
- 3) Solved the following programming to get the weight of each individual model.

$$\min P = e^2 = w^T * E^T * E * w \tag{7}$$

$$s.t \begin{cases} \sum_{i=1}^p w_i = 1 \\ w_i \geq 0, i = 1, \dots, p \end{cases}$$

$$E = \begin{bmatrix} e_1(1), e_2(1), \dots, e_p(1) \\ e_1(2), e_2(2), \dots, e_p(2) \\ \dots \dots \dots \\ e_1(t), e_2(t), \dots, e_p(t) \end{bmatrix}$$

Where  $w = [w_1, w_2, \dots, w_p]^T$  and

#### 4.4 Set up Combined Model and Forecast

After weights computing, we will set up the final nonnegative variable weight combined model and forecast demand. The model can be expressed as:

$$\hat{Y}(t) = \sum_{i=1}^n w_i(t) * \hat{Y}_i(t) \quad (8)$$

### 5 Case Study

In this section, we applied the above combination model to a 4S shop (denoted by A) and an auto parts supplier (denoted by B) in Shanghai to prove the effectiveness of the proposed method. We selected key auto parts according to sales volume and sales amount and chose those which have few outliers.

- A company: 21 key auto parts, from Jan.2007 to Apr. 2009.
- B company: 12 key auto parts, from Feb.2009 to July 2010.

The demands of these selected auto parts were respectively forecasted by the combination model and three individual models included in the combination model to compare the forecasting precision and forecasting stability. During the phase of outliers handling,  $d = 0.2$  and  $f = 0.2$ . During the phase of individual method sifting,  $\bar{m} = 0.2$  and  $\bar{\varepsilon} = 0.3$  for company A,  $\bar{m} = 0.3$  and  $\bar{\varepsilon} = 0.3$  for company B. the comparison results are shown in table 2.

**Table 2.** Precision comparisons between individual models and the combined model

		RBF	ARIMA	SVR	Combined Model
MAPE of A	Mean	77.55%	30.62%	34.27%	27.29%
	Number (<=20%)	3	10	5	11
MAPE of B	Mean	40.53%	32.00%	47.09%	30.17%
	Number (<=30%)	3	7	2	7
effectiveness of A	Mean	0.41	0.59	0.56	0.62
	Number (>=0.5)	8	16	16	17
effectiveness of B	Mean	0.50	0.60	0.47	0.64
	Number (>=0.45)	8	10	6	11
Standard error of A	Mean	0.44	0.24	0.24	0.23
	Number (<=0.3)	11	16	14	17
Standard error of B	Mean	0.21	0.31	0.28	0.19
	Number (<=0.3)	11	10	9	12

- Precision Evaluation. Both MAPE and effectiveness were selected to reflect the precision of different forecasting model comprehensively. In the case of small actual value, the MAPE will also be big even if  $Y_i - \hat{Y}_i$  is small. But the effectiveness will avoid such result. The effectiveness  $M(Y, \hat{Y})$  was calculated as:

$$M(Y_i, \hat{Y}_i) = \frac{1}{t} \sum_{i=1}^t a_i \left[ 1 - \sqrt{\frac{1}{t} \sum_{i=1}^t \left( a_i - \frac{1}{t} \sum_{i=1}^t a_i \right)^2} \right], \text{ where } a_i = \begin{cases} 1 - e_i, & |e_i| < 1 \\ 0, & |e_i| \geq 1 \end{cases}, \quad e_i = |Y_i - \hat{Y}_i| / Y_i \quad (9)$$

The less MAPE, the more accuracy of the model. The more effectiveness, the more accuracy of the model. In table 2, the combined model gets the lowest MAPE and the highest effectiveness at the same time, which means the combined model can produce more accurate forecasting result and has wider applicability.

- **Stability Evaluation.** Standard error is selected to evaluate the stability. From the results shown on table 3, the combined model has the lowest mean and the highest number, which means that the combined model is more stable than the individual models.

## 6 Conclusions

This article set up a nonnegative variable weight combination model for auto parts demand forecasting which could reflect the complex characteristics of various auto parts during different periods and reduce manual intervention. The case study indicated that such combined model has more precision and stability than some common individual models. But there is still some problem worth of being studying in the future. One problem is about the parameter optimization which is set by experience or by traversal. The other problem is about the optimization rule used in the combined model. In the proposed model, relative error is selected as the optimization rule. There are some others to be researched to improve the combined model.

## References

1. Chen, Y., Liu, P., Yu, L.: Aftermarket Demands Forecasting with a Regression-Bayesian-BPNN Model. In: IEEE 2010 International Conference on Intelligent System and Knowledge Engineering, pp. 52–55. IEEE Press, Hangzhou (2010)
2. Chen, Y., Liu, P., Yu, L.: Demand Forecasting in Automotive Aftermarket Based on ARMA Model. In: 2010 International Conference on Management and Service Science, pp. 1–4. IEEE Press, Wuhan (2010)
3. Partovi, F.Y., Anandarajan, M.: Classifying inventory using an artificial neural network approach. *J. Comput. Ind. Eng.* 41(4), 389–404 (2002)
4. Viswanathan, S., Yan, F.: Forecasting of Service Arrivals in a Repair Center. In: IEEE 2005 International Conference on Automation Science and Engineering, pp. 285–289. IEEE Press, Edmonton (2005)
5. Yu, L., Chen, Y.: A neural network based method for part demands prediction in auto aftermarket. In: IEEE 2010 International Conference on Software Engineering and Service Science, pp. 648–651. IEEE Press, Beijing (2010)
6. Edwin, M.K., Raymond, T.N.: Algorithms for Mining Distance-Based Outliers in Large Datasets. In: 24th International Conference on Very Large Data Bases, New York, pp. 392–403 (1998)

# A Data-Hiding Method Based on TCP/IP Checksum

Zhen Liu<sup>\*</sup>, Yuyu Jiang, and Ping Qian

College of Computer Sci. Jiangsu Univ. of Sci. & Tech. Zhenjiang, China  
{lz, jyy}@just.edu.cn, just\_qian@qq.com

**Abstract.** Data-hiding has become one of the focuses in the field of information security. Steganographic methods based on TCP/IP are widely used in network environment. In this paper, a new steganographic method based on TCP/IP is proposed, which hides encrypted secret information randomly in the checksum domain of TCP segment. Because this kind of domain is never used for data-hiding media and secret data has been encrypted before being embedded and is embedded at random, the approach enjoys better invisibility and anti-statistic attacking performance than other current steganographies based on TCP/IP.

**Keywords:** data-hiding, steganograph, TCP/IP, TCP segment, checksum.

## 1 Introduction

Steganographic is an important aspect of communication and information security technology. Information hiding is a significant method of steganographic communication. In other words, the important information to be transmitted should be embedded in the carrier, so that it cannot be easily detected. There are a lot of carriers, such as text, image, audio, video, TCP/IP header and so on. The existing information hiding methods based on TCP/IP protocol realize the steganographic depending on the following domain:

- (1) TCP or IP head options domain and the domain rarely used by data transmission [1][2];
- (2) Mandatory field for hiding information in data transmission, such as source address field, destination address field, identification domain in IP packer header and source port domain, destination port field domain, serial number domain in TCP packer header.

The technology using these domains for hidden is unreliable because these hidden fields have been known for attackers. Aimed at these techniques, some attack methods such as packet filter, chi square analysis has occurred. Because the TCP or IP head hidden information can be detected or extracted by using these attack methods, it is easily to be attacked with low security level.

In allusion to the defects which is easily hacked and unsafe based on the existing TCP/IP protocol hiding technology, this paper presents a method of sending/receiving

---

<sup>\*</sup> Corresponding author.

steganographic communication in computer network environment, which makes the secret information have better secrecy and anti-aggressive. The following will introduce the sending/receiving method of the TCP protocol in transport layer based on TCP/IP.

## 2 Sending Method for Steganographic Communication

The TCP protocol is a significant protocol in transport layer of TCP/IP. The head of TCP includes the source port, destination port, serial number, confirmation number, checksum and so on(Fig.1).

Source port		Destination port	
Sequence number			
Acknowledgement number			
TCP header length	Retain	Tag	Window size
Checksum		Urgent pointer	
Data(optional)			

Fig. 1. TCP segment header format

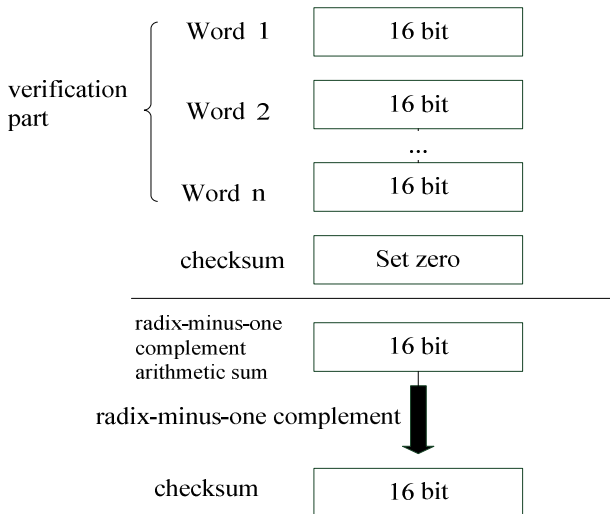


Fig. 2. Flow chart of TCP protocol checksum generation

The generation process of checksum is that it calculates the checksum of TCP segment head, data and pseudo head. Firstly, we divide the check part into 16-bit sequences. Secondly, we set the check zero. Finally, we add all 16-bit data using digital radix-minus-one complement technology, and the radix-minus-one complement of the result is as the generated checksum. (Fig.2)

TCP protocol checkout process is that it calculates the verification part (including pseudo head, head and data fields) and checksum by radix-minus-one complement. If the result is 1 by adding all 16-bit, the checksum is correct, otherwise failed (Fig 3).

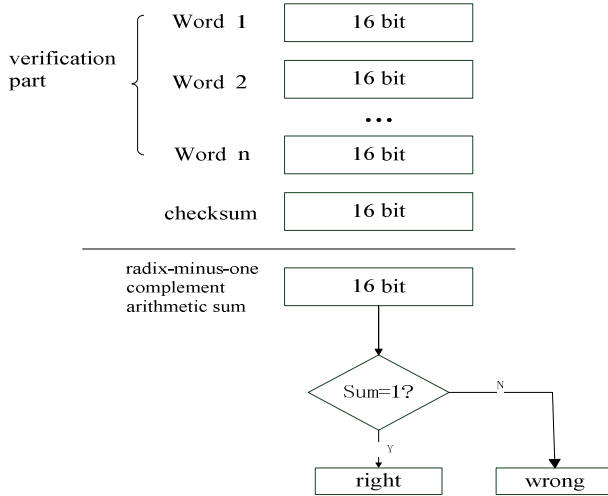


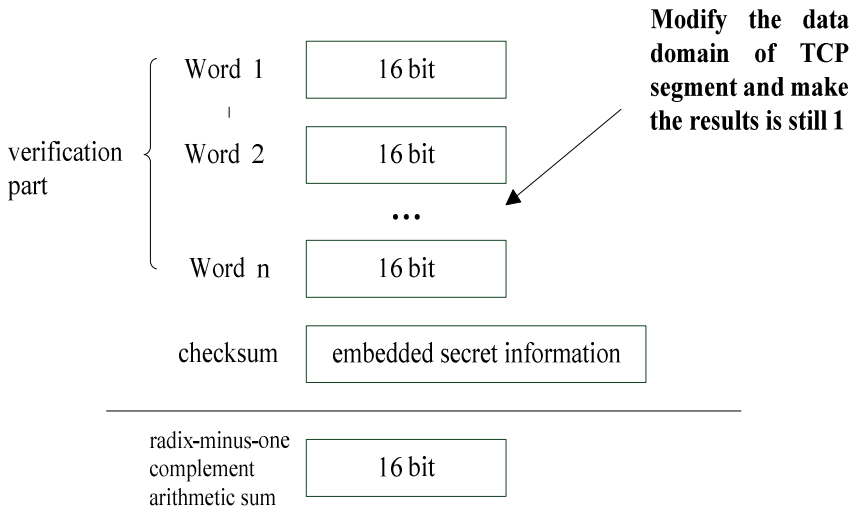
Fig. 3. TCP protocol verification flow chart

The core idea of our steganographic method is that when sender sends data to receiver through network, information is embedded randomly in the TCP head check field. Then sender uses the changed check filed to modify the data field of message segment, in order to ensure checksum verification is correct. Detail steps are as follows:

- (1) Encrypting secret messages. The sender uses RC4 cryptographic algorithm to encrypt the secret information with the key shared with the receiver. The encrypted information is divided into 8 bit every group, and m groups are formed.
- (2) Constructing m TCP segments. We input the correct value to source port, destination port and the serial number of the head sections in each TCP segment, and then generate a 16-bit checksum.
- (3) Extracting 8 bits from step (2) obtained 16 bits in each TCP segment header checksum. Which 8 random bits in 16 bits are chosen is agreed with both sides in advance.
- (4) Replacing the extracted 8 bits data in each group from step (3) with encrypted 8 bits data in each group obtained from step (1) to in turn.

(5) Putting the 8 bits data each group obtained from step (4) to check field according to the same random bit position in step (3).

(6) Using the changed check domain in step (5) to modify the data domain of message segment, in order to ensure the checksum correct. The amendment process of checksum is similar with the verification process. We put the check parts (including pseudo header, header and data domain) and checksum that is embedded secret information together with radix-minus-one complement arithmetic operations. Modify the data domain of TCP segment after adding 16 bits to make the results is still 1, in order to ensure the verification right. The amendment process of checksum is shown in Fig4.



**Fig. 4.** check and modify data domain

(7) The sender sends all TCP segments to receivers in sequence.

### 3 Receiving Method for Steganographic Communication

The receiver extracts the secret information from the check domain of the TCP segment after he gets data from network. The specific steps are as follows:

- (1) The receiver receives all  $m$  TCP segments.
- (2) Extracting 16-bit checksum from the header of each TCP segment.
- (3) Extracting 8 bits from step (9) obtained 16 bits checksum according to 8 random bits' positions agreed upon in advance.
- (4) Extract all bits sequentially from step (10) to form information.
- (5) Decipher information using the shared key and RC4 encryption algorithm, so as to obtain the secret information to be received.

## 4 Conclusion

This paper presents a method of sending/receiving steganographic communication in network environment. When the sender transmits data through network, information is embedded randomly in the TCP head check field, then we use the changed check field to modify the data field of message segment in order to ensure correction of checksum verification. The receiver extracts the secret information from the check domain of the TCP segment after getting the data from network.

This paper only discusses the situation of TCP as the steganographic carrier, because TCP segment is widely used in network. The method is also applicable to other network protocol data unit, such as UDP datagram. The generation and verification of the UDP datagram is as same as the TCP segment, so information hiding and extraction method are same. But UDP datagram header is less than TCP segments, especially there is no sequence number. When the sender hides the information to UDP datagram after dividing secret information into groups, it also must choose some bits in UDP data domain to write serial number in it. So the receiver can restore the information sequentially.

From the preceding analysis we can see the method presented in this paper has the following advantages:

(1) Since the most important technique in steganographic is the imperceptibility of carrier, the method relies on the hidden field is the verification domain of protocol data unit header. This field has not been used by other hiding system, so it has better invisibility;

(2) In order to ensure the secret of hidden information, the method firstly encrypted the information by RC4 before the secret information to be hidden. It has better security because even if the attacker finds the hidden information, it is also very difficult to decrypt the information.

(3) Against statistical characteristics is a significant factor to evaluate the merits of the hidden technology, this method divide the embedded information into groups, each group is 8 bits. It just uses the 8 bits of 16 bits in check domain. The statistical characteristics of check domain become less, thereby it can effectively resist attack based on statistical analysis. And this method uses the random method in embedding the information, which is embedded in 8 random positions of 16 bits in each protocol data unit. The random positions in every protocol data unit are not same, so that the attacker is very difficult to find the rules to attack.

## References

1. Ahsan, K.: Covert channel analysis and data hiding in TCP/IP. Masters thesis., University of Toronto (2002)
2. Ahsan, K., Kundur, D.: Practical data hiding in TCP/IP. In: Proc. Workshop on Multimedia Security at ACM (2002)



3. Xu, B., Wang, J.-Z., Peng, D.-Y.: Practical Protocol Steganography: Hiding Data in IP Header. Paper presented at the First Asia International Conference on Modelling & Simulation (AMS 2007), 584–588 (2007)
4. Trabelsi, Z., El-Sayed, H., Frikha, L., et al.: A novel covert channel based on the IP header record route option. *International Journal of Advanced Media* 1(4), 328–350 (2007)
5. Dhobale, D.D., Ghorpade, V.R., Patil, B.S., et al.: Steganography by hiding data in TCP/IP headers. In: Paper presented at the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE) (August 2010)
6. Ahn, G.-J.: Efficient construction of provably secure steganography under ordinary covert channels. *Science China (Information Sciences)* (July 2012)
7. Yang, C.H., Weng, C.-Y., Wang, S.J., et al.: Varied PVD+LS. Bevading detection programs to spatial domain in data embedding systems. *The Journal of Systems and Software* (2010)
8. Mare, S.F., Vladutiu, M., Prodan, L., Opritoiu, F.: Advanced Steganographic Algorithm Using Payload Adaptation and Graceful Degradation. In: *Lecture Notes in Information Technology—Proceedings of 2012 International Conference on Information Engineering (ICIE 2012)* (June 2012)

# A Parallel Full-System Emulator for Risc Architecture Host

Xiao-Wu Jiang, Xiang-Lan Chen, Huang Wang, and Hua-Ping Chen

Department of Computer Science, University of Science and Technology of China  
No. 443, Huangshan Road, Baohe District, Hefei City, Anhui Province, PRC  
{wesker,ustc}@mail.ustc.edu.cn,  
{xlanchen, hpchen}@ustc.edu.cn

**Abstract.** In this paper, we port a parallel full-system emulator to RISC host to achieve higher performance by utilize all the multi-core resources from physical CPU, in contrast the traditional full-system emulator is sequentially in SMP emulation and can only use one core of host machine. We mainly deal with the atomic instruction translation to RISC ll/sc pairs, and apply lightweight lock-free FIFO queue algorithms using both interleaving and non-interleaving ll/sc pairs. The tests show that the performance of parallel full-system emulator have high efficiency.

**Keywords:** Parallel emulation, Atomical, lock-free queue.

## 1 Introduction

RISC is a type of microprocessor architecture that utilizes a small, highly-optimized set of instructions. It is in the late 1970s and early 1980s that the first RISC projects came out, they are IBM 801, Stanford MIPS, and Berkeley RISC 1 and 2. In 1980s and the early 1990s, a wide variety of similar RISC processors were used in Unix workstation market as well as in printers, routers, etc. It is the beginning of 21st century when RISC architecture dominate the majority of low end and mobile systems. This situation happened mainly because the low power and low cost compared to X86. Now the typical RISC architectures are arm and MIPS.

As the performance of a single processor had nearly reached its rooftop. Other technologies are used to ensure the Moore's Law. Symmetric multiprocessing is the most efficient one. Manufacturers typically integrate multiple cores into a single integrated circuit diet, which is known as a chip multiprocessor or CMP. After 2006, intel and AMD first introduce x86 CMP cpu Core Duo and Athlon64 X2 to desktop user. Four years later, armv7-a based multi-core cortex-A9 and MIPS64-compatible quad-core loongson 3A appear to the public. Now desktop CPU has reached deca-core (e.g. Intel Xeon E7-2850) and mobile CPU has reached octa-core (e.g. samsung Exynos 5 Octa).

Now RISC architecture are developing towards desktop. AMD has announced to produce arm cpu for server, google had already run ChromeOS on arm, and Microsoft also establish windows RT to support arm. Besides some desktop and laptop are

inspired by loongson 3 family, a multicore MIPS64-compatible cpu developed by Chinese Academy of Sciences.

Even through RISC Architecture, especially arm and MIPS, are rapidly developing, In desktop area these Architecture still lack of applications compared to X86. Full-system emulation has been proved to be an effective way to imigrate existen applications to other architectures. There is already some full-system emulators that enable X86 OS running on arm or MIPS. A typical tool is QEMU, which is sequential in SMP emulation and can emulate multiple architecture on multiple architecture, including X86 on arm/MIPS. Thanks to the high efficiency of X86, Sequential QEMU is fast enough to use On x86 machine. Even though MIPS and arm have no less cores than X86, for a single core, it's far less fast compared to X86. For example, MIPS64-compatible loongson 3A has 4 cores at 900Mhz while Intel i5-2400 has 4 cores at 3.1Ghz. when running a 7z in one thread, Intel is 5.8x faster than loongson. So when emulating a X86 machine on arm/MIPS, it is very important to use all of the cores rather than only one to make the guest machine fluently.

X86 have some parallel full-system emulator which can use all of the resources of host machine. But on arm or MIPS there is no parallel full-system emulator now.

In order to make full-system emulation faster, a parallel full-system emulator COREMU is ported to RISC architecture. In the remainder of this paper we presents how we port this parallel full-system emulator to RISC. Section 2 introduce full-system emulator and the parallelizing strategy of full-system emulator on X86. Section 3 focus on solving the atomic instruction translation from CISC x86 to RISC MIPS/arm. Section 4 present lock-free FIFO queue algorithms using ll/sc pairs in interruption simulation. Section 5 report experimental results comparing our Parallel full-system emulator to ordinary full-system emulator and host OS. This paper mainly talk about MIPS, arm is the same except lock-free queue algorithm in section 4.3.

## 2 QEMU and Parallel Full-System Emulator

### 2.1 QEMU and QEMU's Multiprocessor Emulation

QEMU[1] is a hosted virtual machine monitor: It emulates CPU through dynamic binary translation(DBT) and provides a set of device models, enabling it to run a variety of unmodified guest operating systems. QEMU has user-mode emulation and full-system emulation mode, in which QEMU emulates a full computer system, including one or more processors and peripherals.

For a single emulated processor, QEMU translates the emulated code to TCG(Tiny Code Generator) and then translates the TCG to host instructions. After a block of emulated code translated to instructions on host machine, QEMU will execute it. In QEMU's multiprocessor emulation, QEMU emulates a SMP machine with multiple processors and a certain device to support inter-core communications (such as APIC in x86). QEMU emulates these processor sequentially in a round-robin strategy: each emulated processor has a time slice to execute. After that, physical CPU turn to the next emulated processor to execute. And between each time slice, physical CPU turn to execute some peripherals simulation and inter-core communications.

## 2.2 Parallel Full-System Emulator

While QEMU being as a sequentially full system emulator, there exist a few kinds of parallel full-system emulator: Parallel SimOS, COREMU[2], PQEMU[5] and HQEM[6]. Parallel SimOS is designed for alpha architecture and the other three are specially designed for X86 machine, both of them are not able to run on MIPS architecture now. Compared to other parallel emulator, COREMU have high scalability and high performance. Our job is majorly based on it.

COREMU is hosted on X86 and targeted on multiple architecture especially x86 and arm. It wraps the translation-execution logic to a single thread, and then bind these threads to different physical CPU cores. Besides, it wraps all peripherals emulation to an individual thread called IO-thread. COREMU majorly use multithread to achieve parallel. It offered an efficiently emulate synchronization primitive to coordinate concurrent access to the emulated shared memory from each emulated processor. To deal with inter-core communications COREMU use lock-free FIFO queue.

When building a parallel full system emulator like COREMU on MIPS, we mainly deal with the atomic instruction translation strategy for lightweight memory transactions and lock-free FIFO queue for inter-core communications.

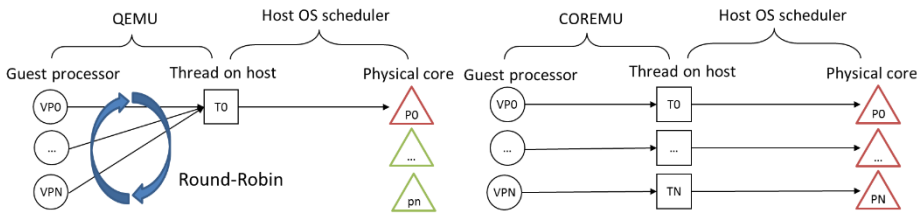


Fig. 1. Sequential and parallel full-system emulation

## 3 Atomic Instruction on MIPS Host

### 3.1 Atomic Instruction on X86

CAS(compare-and-swap) is an atomic instruction which is widely used in multithreading to achieve synchronization. The C function of CAS in Figure 1 shows the basic behavior of CAS, which provide the guarantee of atomicity.

```
int CAS(int *mem,int oldval,int newval){
    int old_reg_val=*reg;
    if(old_reg_val==oldval)
        *reg=newval;
    return old_reg_val;
}
```

Fig. 2. CAS in C

```
void inc(int *reg){
    do{
        int old=*reg;
        int new=old+1;
    }while(CAS(reg,old,new)!=old)
}
```

Fig. 3. translation of atomic inc use CAS

COREME use CASN(Multiword CAS) algorithm in atomic instruction translation, which execute multiple CAS to simulate the atomic instruction on guest machine.

Figure 2 shows the translation of atomic inc use CAS in C.COREMU use CASN majorly because it targeting at X86 host, and X86 has cmpxchg as its CAS instruction.

### 3.2 Atomic Instruction on MIPS

Different from X86, MIPS is a RISC architecture and has no CAS instructions. MIPS provides ll(Load Linked) and sc(Store Conditional Word) (on arm the instructions named llrex, screx) to achieve atomic read-modify-write (RMW) operation. ll reg,mem load a word from memory to reg, and remember this operation. sc reg,mem store a word to the same location in memory. When a sc instruction fetch memory, it will check whether the location is modified after the last ll instruction. If it didn't modified, reg will set 1 for the success of execution, while if it has been modified the reg will set 0 for the failure of execution. LL/SC has two advantages over CAS: reads and writes are separate instructions, and both instructions can be performed using only two registers.

### 3.3 Aligned Instruction

X86 atomic instruction contains inc, dec, add, xchg, and, or, xadd, bit\_testandset, bit\_testandreset, etc. But OS and applications won't use them all. From experiment we find that linux kernel and applications on it only use inc, dec, xchg, cmpxchg and xadd. This paper use ll/sc pair and inline assembly to achieve lightweight memory transaction. Figure 3-6 show the translated inc, xchg, xadd, cmpxchg in MIPS.

```
1: ll t,*mem
   addi t,t,1
   sc t,*mem
   beqz t,1b
```

**Fig. 4.** inc mem

```
1: ll temp1,*mem
   move temp,reg
   move reg,temp1
   sc temp,*mem
   beqz temp,1b
```

**Fig. 5.** xchg reg,mem

```
1: ll temp,*dst
   add temp1,reg,temp
   move reg,temp
   sc temp1,*dst
   beqz temp1,1b
```

**Fig. 6.** xadd reg,mem

```
1: ll temp,mem
   bne temp,old,2f
   move temp,new
2: sc temp,mem
   beqz temp,1b
```

**Fig. 7.** cmpxchg mem,old,new

### 3.4 Unaligned Instruction

The above research shows the solution to all 32bit aligned memory access, but as a CISC architecture, X86 has non-32bit aligned memory access while MIPS required 32bit aligned memory access. The experiment result shows that all these unaligned memory access exist in 8bit or 16bit bit xchg and cmpxchg, and the memory will not across two 32bit memory address.

With this feature, we deal with this unaligned instruction as below: when QEMU got a unaligned instruction, then just expand the address to 32bit aligned (`new_addr=addr & ~0x3`) and operate the whole 32bit atomically, then we can ensure the atomicity of the original operation.

## 4 Lock-Free Queue in Interruption Simulation

### 4.1 Interruption Simulation

As emulation in QEMU is sequential, the asynchronous communication between core to core/device emulated in a synchronous way. All of the processor running logic are schedule by round-robin fashion. When a core is schedule out, QEMU will do those synchronous events including device interruption and inter-processor interruption. However in parallel emulation more than one emulated core are running at the same time, interrupt vector may be modified parallel by each running core. COREMU use a lock-free FIFO queue to achieve asynchronous communication.

### 4.2 Lock-Free Queue in X86

Unlike ll/sc pair in MIPS, CAS in X86 can't not detect ABA problem[8]. A typical ABA problem like below:

- Process1 reads value A from shared memory
- Process1 then preempted allowing process2 to run.
- Process2 modifies the shared memory value A to value B and back to A before preemption.
- Process1 begins execution again, sees that the shared memory value has not changed and continues.

ABA problem is a major problem when designing a Lock-free queue algorithm because the node type of queue are always pointer, and a same pointer may result from an enqueue with the same malloc. COREMU add a counter to each queue node and use CAS2 to avoid ABA problem in lock-free queue. CAS2 check a queue node which contains a pointer and a counter. The counter never be the same after each en/dequeue operation. Besides X86 has native CAS2 instruction: `cmpxchg8b/16b`.

### 4.3 Lock-Free Queue in MIPS

There is a way to use interleaving ll/sc pairs directly to form a lock-free FIFO queue as Claude Evequoz talk about in his paper[3]. We apply this algorithm on arm because arm support interleaving of ll/sc pairs.

```

Q: array[0..Q_LENGTH-1] of *NODE;
unsigned int Head, Tail;
bool enqueue(node *p){
    unsigned int t,tail;
    node *slot;
    while(true){
        t = Tail;
        if(t == Head + Q_LENGTH)
            return FULL_QUEUE;
        tail = t % Q_LENGTH;
        slot=LL(&Q[tail]);
        if(t == Tail)
            if(slot != null){
                if(LL(&Tail))
                    SC(&Tail,t+1);
            }
            else if(SC(Q[tail],node)){
                if(LL(&Tail)==t)
                    SC(&Tail,t+1)
                return OK;
            }
    }
}

// Circular list initialized with null
// Extraction and insertion indices
node *Dequeue(void){
    unsigned int h,head;
    node *slot;
    while(true){
        h = Head;
        if(h == Tail)
            return null;
        head = h % Q_LENGTH;
        slot = LL(&Q[head]);
        if(h == Head)
            if(slot == null){
                if(LL(&Head) == h)
                    SC(&Head,h+1);
            }
            else if(SC(&Q[head],null){
                if(LL(&Head) == h)
                    SC(&Head,h+1);
                return slot;
            }
    }
}

```

**Fig. 8.** Lock free FIFO queue using ll/sc pair

As algorithm using ll/sc always need either nesting or interleaving of ll/sc pairs, we can't use the algorithm based on it because MIPS do not support it. Actually a single en/dequeue contains two operations: modify Head/Tail pointer and en/queue node. Two operation must be execute at one atomic time while MIPS only support one.

Paper[7] offered us a way to build CASN which can atomically run multiword CAS. We first use ll/sc pair to build a software version of CAS, and then generate the CAS2. In this way we can use lock-free queue in COREMU, but CAS2 in MIPS is much heavier than CAS, not to speak of there is no native support of CAS. So it is very important to reduce the use of CAS2 in lock-free queue algorithm.

We use lock-free algorithm found by John D.valois[4], which especially reduce CASN instruction. Both enqueue and dequeue has only one CAS2 and one xadd. This algorithm is based on a standard circular array. There are three special values, HEAD TAIL and EMPTY, and node value. Initially, two adjacent locations are set to HEAD and TAIL while others are set to EMPTY. To enqueue the value x, a process find the unique location containing the special TAIL value.CAS2 is then used to change two adjacent location from <TAIL, EMPTY> to <x, TAIL>.The dequeue operation is similar, using the CAS2 operation to change <HEAD, x> to <EMPTY, HEAD> and the return the x.

Besides, we keep two counters: the number of enqueue and the number of dequeue. Both of them are increase by FAA whenever an en/dequeue process complete. These two counter helps to quickly find the HEAD and TAIL.FAA can be simulated by xadd 1,mem in Figure 4, CAS2 can be generate by multiple CAS.

When reaching the beginning and ending of the array, this algorithm still work, because software CAS2 do not require two memory adjacent.

## 5 Experiments and Discussion

In order to test the performance between origin QEMU and our modified QEMU, and test the performance between multithread program in native machine and in our modified QEMU, two benchmark are chosen. These benchmark are performed on a 4 core (900Mhz) loongson 3A , a quad core MIPS cpu, running Debian 6 with kernel version 2.6.36.3. The guest OS is Debian 6 with version 2.6.32-5.

Firstly, we write a simple multithread pi which is designed to calculate pi in totally N step in T threads concurrently. Each thread calculate  $\pi^t$ , and finally calculate  $\pi$ .

$$\pi^t = \sum_{i=1}^{N/T} \frac{8}{16(iT + t)^2 - 16(iT + t) + 3}$$

$$\pi = \sum_{i=1}^T \pi^i$$

All these result shows below. OriQemu short for origin QEMU ModQemu short for our modified QEMU, n(1,2,4) means QEMU run with `-smp n` option(emulating an n core machine). The result shows the efficiency of modified QEMU, It is 3x faster than original QEMU when the number of core on emulated machine is set to 4 and the number of thread is set to 4. The speedup rate reached 3 and efficiency reached nearly 3/4 compared to the number of physical core.

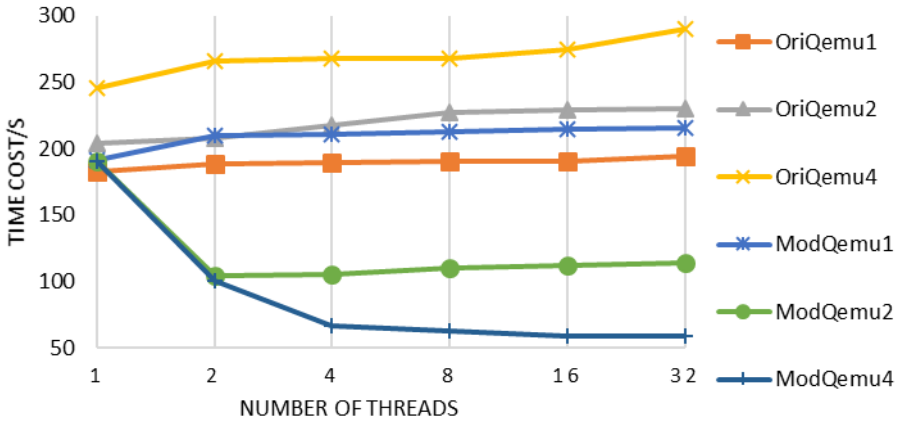


Fig. 9. The time of multithread pi

Secondly, we test the performance between modified QEMU and native machine though 7z, a widely used multithread compress application which contains a building benchmark. The dictionary size is set to 256KB. The result shows the compress and decompress speed on both native and modified QEMU, A higher threadnum makes a higher compress/decompress rate, and both in native and modified QEMU the application hold almost the same speedup rate: compress 2.79 to 2.68 and depress 3.68 to 3.65.



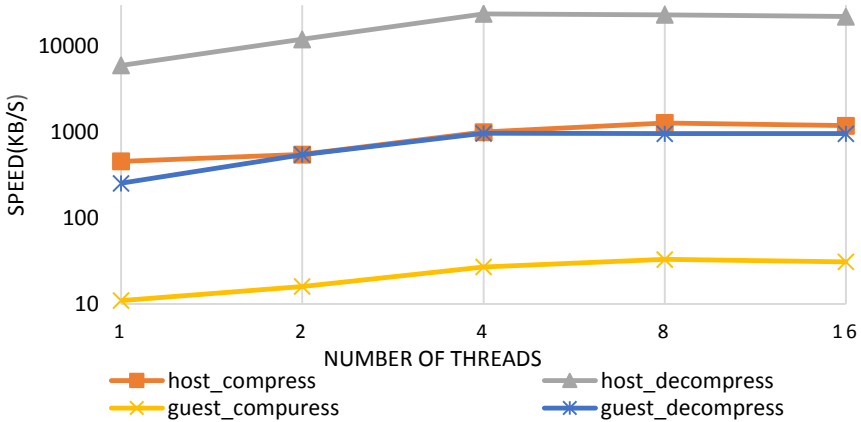


Fig. 10. The speed of 7z compress/decompress

## 6 Conclusion

We find an atomic instruction translating strategy for ll/sc pairs on RISC and use a more light-weight lock-free FIFO queue on asynchronous communication emulation. Finally we successfully emulate X86 in parallel on MIPS target. The experiments proved its efficiency compared to original QEMU and host machine.

## References

1. Bellard, F.: QEMU, a fast and portable dynamic translator. USENIX (2005)
2. Wang, Z., Liu, R., Chen, Y., Wu, X., Chen, H., Zhang, W., Zang, B.: COREMU: a scalable and portable parallel full-system emulator. In: Cascaval, C., Yew, P.-C. (eds.) PPOPP, pp. 213–222. ACM (2011)
3. Evéquo, C.: Non-Blocking Concurrent FIFO Queues with Single Word Synchronization Primitives. In: ICPP, pp. 397–405. IEEE Computer Society (2008)
4. Valois, J.D.: Implementing Lock-Free Queues. In: Proceedings of the Seventh International Conference on Parallel and Distributed Computing Systems, Las Vegas, NV (1994)
5. Ding, J.-H., Chang, P.-C., Hsu, W.-C., Chung, Y.-C.: PQEMU: A Parallel System Emulator Based on QEMU. In: ICPADS, pp. 276–283. IEEE (2011)
6. Hong, D.-Y., Hsu, C.-C., Yew, P.-C., Wu, J.-J., Hsu, W.-C., Liu, P., Wang, C.-M., Chung, Y.-C.: HQEMU: a multi-threaded and retargetable dynamic binary translator on multicores. Paper presented at the Meeting of the CGO (2012)
7. Harris, T.L., Fraser, K., Pratt, I.: A Practical Multi-word Compare-and-Swap Operation. In: Malkhi, D. (ed.) DISC 2002. LNCS, vol. 2508, Springer, Heidelberg (2002)
8. Dechev, D., Pirkelbauer, P., Stroustrup, B.: Understanding and Effectively Preventing the ABA Problem in Descriptor-Based Lock-Free Designs. Paper presented at the Meeting of the ISORC (2010)

# Density Based Active Self-training for Cross-Lingual Sentiment Classification

Mohammad Sadegh Hajmohammadi<sup>\*</sup>, Roliana Ibrahim, and Ali Selamat

Software Engineering Research Group, Faculty of Computing, Universiti Teknologi Malaysia,  
81300 UTM Skudai, Johor, Malaysia  
shmohammad2@live.utm.my, {roliana, aselamat}@utm.my

**Abstract.** Cross-lingual sentiment classification aims to utilize annotated sentiment resources in one language (typically English) for sentiment classification in another language. Most existing research works rely on automatic machine translation services to directly project information from one language to another. However, since machine translation quality is still far from satisfactory and also term distribution across languages may be dissimilar, these techniques cannot reach the performance of monolingual approaches. To overcome these limitations, we propose a novel learning model based on active learning and self-training to incorporate unlabeled data from the target language into the learning process. Further, in this model, we consider the density of unlabeled data to avoid outlier selection in active learning. The proposed model was applied to book review datasets in two different languages. Experiments showed that the proposed model could effectively reduce labeling efforts in comparison with some baseline methods.

**Keywords:** Sentiment Classification, Self-training, Active Learning, Density.

## 1 Introduction

Text sentiment classification is the process of automatically predicting the sentiment polarity of a given text document[1]. Although traditional classification algorithms can be used to train sentiment classifiers from labeled text data, construction of manually labeled data is a very expensive and time-consuming task. However, since most labeled sentiment resources are in English, there are not enough labeled sentiment data in other languages [2]. Therefore, the challenge is how to utilize labeled sentiment resources in one language (source language) for sentiment classification in another language (target language) and leads to an exciting research area called cross-lingual sentiment classification (CLSC).

Most existing works employed machine translation to directly project the data from the target language into the source language [3] and then treated the problem as mono-lingual sentiment classification in the source language. However, since machine translation quality is still far from satisfactory and also term distribution across

---

<sup>\*</sup> Corresponding author.

languages may be dissimilar due to the difference in cultures and writing styles, these methods cannot reach the performance of monolingual methods. To solve this problem, making use of unlabeled data from the target language can be helpful because they are always easy to obtain and have the same term distribution and writing style with the target language. Active learning (AL) and semi-supervised learning (SSL) are two well-known techniques that make use of unlabeled data to improve classification performance. In this paper, we propose a new model based on a combination of Active learning and self-training in order to incorporate unlabeled data from the target language into the learning process.

The rest of this paper is organized as follows. The next section presents related work on CLSC. The proposed model is described in Section 3 while evaluation and experimental results are given in Section 4. Finally, Section 5 concludes this paper.

## 2 Related Works

Cross-lingual sentiment analysis has been extensively studied in recent years. These research studies are based on the use of annotated data in the source language (always English) to compensate for the lack of labeled data in the target language. Most approaches focus on resource adaptation from one language to another language with few sentiment resources. For example, Mihalcea, Banea [4] generate subjectivity analysis resources into a new language from English sentiment resources by using a bilingual dictionary. In other works [5, 6], automatic machine translation engines were used to translate the English resources for subjectivity analysis. In [6], the authors showed that automatic machine translation is a viable alternative for the construction of resources for subjectivity analysis in a new language. Pan et al. [7] designed a bi-view non-negative matrix tri-factorization (BNMTF) model to solve the problem of cross-lingual sentiment classification. Another approach is that of cross-lingual classification, that is translating the features extracted from labeled documents [8]. It can, however, suffer from the inaccuracies of dictionary translation, in that words may have different meanings in different contexts. In another work, Wan [3] used the co-training method to overcome the problem of cross-lingual sentiment classification. The author exploited a bilingual co-training approach to leverage annotated English resources to sentiment classification in Chinese reviews.

## 3 The Proposed Model

As mentioned before, because translated data in cross-lingual sentiment classification cannot cover all vocabularies used in test data, the performance of sentiment classifier in this case is limited. To increase the performance, making use of unlabeled data from the target language can be helpful since these data are always easy to obtain and have the same term distribution as test documents. However, manually labeling unlabeled data is a hard and time-consuming task. To reduce the labeling effort, we propose a new model based on the combination of active learning and self-training.

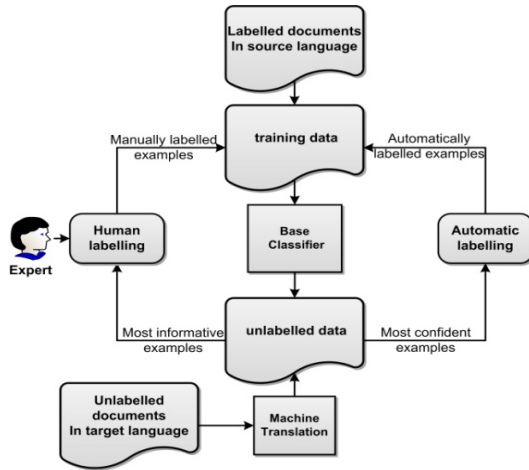


Fig. 1. Framework of the proposed approach

This model attempts to enrich initial training data through manually (AL) and automatically (self-training) labeling of some unlabeled data from the target language in an iterative process. The framework of the proposed model is illustrated in figure 1.

The query function is essential in the active learning process. The simplest query function is uncertainty sampling [9] in which unlabeled examples with the maximum uncertainty are selected for manual labeling in each learning cycle. Entropy is a popular uncertainty measurement widely used in recent researches [10]. Formula (1) shows the uncertainty function calculated based on the entropy estimation.  $P(.)$  is the posterior probability of the classifier and  $H(.)$  is the uncertainty function.

$$H(x) = \sum_{y \in Y} P(y|x) \log P(y|x) \tag{1}$$

As reported in [11, 12], many unlabeled examples selected by the uncertainty sampling cannot help the learner since they are outliers. It means that a good selected example for manual labeling should not only be the most informative, but also the most representative one. Jingbo, Huizhen [12] proposed a density based technique to select the most informative and representative example to solve this problem. To determine the density degree of an unlabeled example, they used a novel method called  $k$ -nearest neighbor based density ( $k$ NN density). In this measure, the density degree of an example is computed by average similarity between this example and  $k$  most similar unlabeled examples in the unlabeled pool. Suppose  $S(x) = \{s_1, s_2, s_3, \dots, s_k\}$  is a set of  $k$  most similar unlabeled examples to the  $x$ . Therefore, average similarity for  $x$  ( $A(x)$ ) can be computed based on the following formula:

$$A(x) = \frac{\sum_{s_i \in S(x)} \text{Similarity}(x, s_i)}{k} \tag{2}$$

We employ this density degree to avoid selecting outlier example in active learning. We use cosine measure as the similarity function to compute the pair-wise similarity value between two examples. In this model, an unlabeled example with the maximum uncertainty and density is selected based on the following formula for manually labeling.

$$u = \arg \max_{x \in U} (H(x) \times A(x)) \quad (3)$$

On the other hand, the self-training algorithm is used to label the most confident examples and generate new training examples along with active learning. These most confident classified documents are selected and added to training data with corresponding predicted labels in each step (automatic labeling). Confidence in each newly classified example is computed based on the distance of each example from the current decision boundary.  $p$  positive and  $n$  negative the most confident examples are selected as auto labeled examples for the next iteration. These two groups of selected examples are then added to the training data and removed from the unlabeled data. We called this model density based active self-training (DBAST).

## 4 Evaluation

In this section, we evaluate the proposed approach in CLSC on two different languages in the book review domains and compare it with some baseline methods.

### 4.1 Datasets

Two different evaluation datasets have been used in this paper.

1. English-French dataset (En-Fr): This dataset contains Amazon book review documents in English and French languages. This dataset was used by Prettenhofer and Stein [13].
2. English-Chinese dataset (En-Ch): This dataset was selected from Pan reviews dataset [7]. It contains book review documents in English and Chinese languages.

All review documents in target languages are translated into the source language (English) using the Google translate engine<sup>1</sup>. In the pre-processing step, all English reviews are converted into lowercase. Special symbols, words with one character length and other unnecessary characters are eliminated from each document. Unigram and bi-gram patterns were extracted as sentimental patterns. To reduce computational complexity, we performed feature selection using the information gain (IG) technique. We selected 5000 high score unigrams and bi-grams as final features. Term presence was used as feature weights because this method has been confirmed as the most efficient feature weighting method in sentiment classification [14].

---

<sup>1</sup> <http://translate.google.com/>

## 4.2 Based Lines Methods

The following baseline methods are implemented in order to evaluate the effectiveness of proposed models.

- Active Self-Training model (AST): this model is similar to DBAST but without considering the density measure of uncertain examples.
- Active learning (AL): this model is based on the simple uncertainty sampling.
- Random Sampling (RS): In random sampling approach, in each cycle, one example is randomly selected from unlabeled data for manually labeling.

## 4.3 Experimental Setup

In all experiments,  $SVM^{light}$  (<http://svmlight.joachims.org/>) is used as the base classifier with all parameters set to their default values. However, SVM does not directly output the posterior probabilities of predicted labels. Therefore, we use a strategy that introduced in [15] to compute the probabilities. In the experiments, we used the 5-fold cross validation to obtain the results. In this setting, translated documents are split into five groups. In each cycle of cross validation, the text documents from 4 groups are considered as unlabeled data and the remaining group being used as test data.

In order to compare the proposed active learning methods, we used the deficiency metric [16] that has been employed in recent papers [12]. The deficiency metric between two methods  $BASE$  and  $ALG$  is defined by:

$$Def_n(ALG, BASE) = \frac{\sum_{t=1}^n (Acc_n(BASE) - Acc_t(ALG))}{\sum_{t=1}^n (Acc_n(BASE) - Acc_t(BASE))} \quad (4)$$

Where  $BASE$  is the baseline method (in our experiment, uncertainty sampling) and  $ALG$  is the proposed methods such as DBAST and AST.  $Acc_t(\cdot)$  refers the accuracy of active learning method in  $t$ th learning cycle and  $Acc_n(\cdot)$  denotes the accuracy of active learning at the end of the learning process. This metric is always non-negative measure, and smaller values (i.e.,  $< 1.0$ ) indicate that  $ALG$  is better than  $BASE$  method.

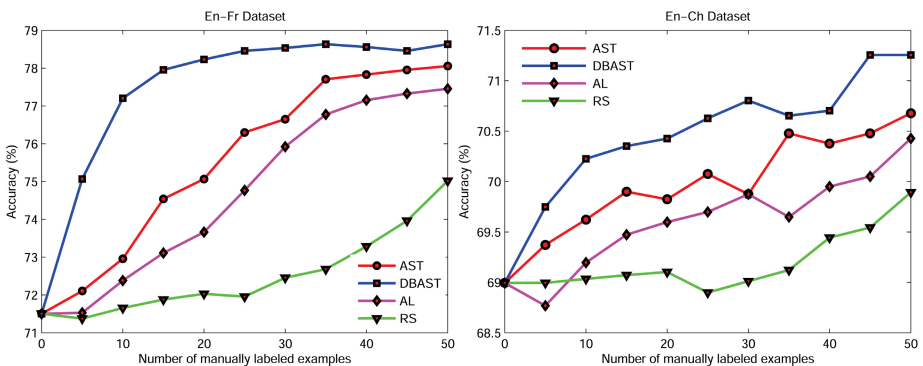


Fig. 2. The classification accuracy over the number of manually labeled examples

#### 4.4 Results and Discussions

In this section, the proposed method is compared with three baseline methods. We set  $k=20$  in the  $k$ NN density measure. We also used  $p=n=5$  for the self-training algorithm. The total number of iterations is set to 50 iterations for all algorithms. After full learning process, test data is presented into learned classifier for evaluation.

Fig. 2 shows the classification accuracy of various methods on two evaluation datasets. As shown in this figure, by comparing the proposed method (DBAST) with the AST model, the classification accuracy of the proposed model improves very quickly in the first few cycles (specially in French language). This is due to the examples, selected based on density and uncertainty, are more representative than examples, selected only based on uncertainty in active learning. This figure also shows that combining active learning with self-training helps to obtain better accuracy. This is most likely due to the augmentation of most confident automatic classified examples, along with manually labeled examples, into training data during the learning process.

Table 1 shows the deficiency metric of DBAST and AST method in compare with uncertainty sampling active learning (AL). DBAST achieves smallest deficiency in all datasets, which indicates better performance than AST and AL method.

**Table 1.** Deficiency metric - compared with uncertainty sampling (AL)

Dataset	Methods	
	DBAST	AST
En-Fr	<b>0.0248</b>	0.7010
En-Ch	<b>0.0384</b>	0.5571

## 5 Conclusion

In this paper, we have proposed a new model by combining active learning and self-training in order to reduce the human labeling effort in CLSC. We also considered a density measure to avoid selecting outlier examples from unlabeled data to increase the representativeness of selected examples for manual labeling in the active learning algorithm. We applied this method to cross-lingual sentiment classification datasets in two different languages and compared the performance on the proposed model with some baseline methods. The experimental results show that the proposed model outperforms the baseline methods in all datasets.

**Acknowledgement.** This work is supported by the Ministry of Higher Education (MOHE) and Research Management Centre (RMC) at the Universiti Teknologi Malaysia (UTM) under Research University Grant Scheme (Vote No. Q.J130000.2628.07J52).

## References

1. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, vol. 5, p. 167. Morgan & Claypool Publishers (2012)
2. Montoyo, A., Martínez-Barco, P., Balahur, A.: Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems* 53(4), 675–679 (2012)
3. Wan, X.: Bilingual co-training for sentiment classification of chinese product reviews. *Comput. Linguist.* 37(3), 587–616 (2011)
4. Mihalcea, R., Banea, C., Wiebe, J.: Learning multilingual subjective language via cross-lingual projections. In: *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (2007)
5. Banea, C., Mihalcea, R., Wiebe, J.: Multilingual subjectivity: are more languages better? In: *Proceedings of the 23rd International Conference on Computational Linguistics 2010*, pp. 28–36. Association for Computational Linguistics, Beijing (2010)
6. Banea, C., et al.: Multilingual subjectivity analysis using machine translation. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing 2008*, pp. 127–135. Association for Computational Linguistics, Honolulu (2008)
7. Pan, J., Xue, G.-R., Yu, Y., Wang, Y.: Cross-Lingual Sentiment Classification via Bi-view Non-negative Matrix Tri-Factorization. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) *PAKDD 2011, Part I. LNCS*, vol. 6634, pp. 289–300. Springer, Heidelberg (2011)
8. Moh, T.-S., Zhang, Z.: Cross-lingual text classification with model translation and document translation. In: *Proceedings of the 50th Annual Southeast Regional Conference 2012*, pp. 71–76. ACM, Tuscaloosa (2012)
9. Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In: *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval 1994*, pp. 3–12. Springer-Verlag New York, Inc., Dublin (1994)
10. Zhu, J., Ma, M.: Uncertainty-based active learning with instability estimation for text classification. *ACM Trans. Speech Lang.* 8(4), 1–21 (2012)
11. Tang, M., Luo, X., Roukos, S.: Active learning for statistical natural language parsing. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics 2002*, pp. 120–127. Association for Computational Linguistics, Philadelphia (2002)
12. Jingbo, Z., et al.: Active Learning With Sampling by Uncertainty and Density for Data Annotations. *IEEE Transactions on Audio, Speech, and Language Processing* 18(6), 1323–1331 (2010)
13. Prettenhofer, P., Stein, B.: Cross-Lingual Adaptation Using Structural Correspondence Learning. *ACM Trans. Intell. Syst. Technol.* 3(1), 1–22 (2011)
14. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: *Proceedings of the ACL 2002 Conference on Empirical Methods in Natural Language Processing*, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
15. Brefeld, U., Scheffer, T.: Co-EM support vector learning. In: *Proceedings of the Twenty-First International Conference on Machine Learning 2004*, p. 16. ACM, Canada (2004)
16. Baram, Y., El-Yaniv, R., Luz, K.: Online Choice of Active Learning Algorithms. *J. Mach. Learn. Res.* 5, 255–291 (2004)



# Constructing 3D Model Based on Panoramic Images

Jia Li, Yehua Sheng, Ping Duan, Siyang Zhang, and Haiyang Lv

Key Laboratory of Virtual Geographic Environment of Ministry of Education, Nanjing Normal University, Nanjing, 210023  
{lijia8425, duanpingshai}@163.com, shengyehua@njnu.edu.cn,  
dyeone.corey@gmail.com, 592411972@qq.com

**Abstract.** Constructing a 3D model usually needs professional graphic technology and scanning equipment. This often in vast time consumption and high costing budgets. To solve the problem of high cost and complicated process, a simple, fast and low cost method is proposed. Firstly take images with 360 degree panorama of the object that is going to be modeled. Then update the image set to the database in the cloud. With the powerful cloud technology and computing power, the image set will be transformed to realistic 3D model that can be downloaded by clients. Professional modeling technique of users is no need for this method. To get a realistic 3D model all we need to do is just taking panoramic images of the object and finish few steps of simple settings.

**Keywords:** 3D model, panoramic images, cloud, virtual reality.

## 1 Introduction

As we live in a 3D world which is full of 3D objects, to reconstruct them accurately in computer, we must figure out how to describe these objects properly in 3D space [1]. Currently, the chief methods of constructing a 3D model are follows [2]: (1) Constructing 3D model manually (2)3D laser scanning (3) Image-based 3D model constructing.

3D reconstruction is a technology with higher threshold and requirement of much manpower and time. To solve the problem of high cost and complicated process, this paper proposes a method with cloud-based process technology for quick and easy 3D model reconstruction, and can spread among the majority of users to become easy to use. This paper take Autodesk 123D Catch [3] as an example to explore the method of operation, which is the image set being converted a realistic 3D model fast, and simplify the model to reduce the cost and time requirement of reconstructing the object.

## 2 Automatic Modeling Method and Technique Principle

Autodesk 123D Catch is the software which is based on image modeling and rendering, making use of powerful cloud computing and analytical ability of converting panoramic images to realistic 3D model of the object or portrait and can be downloaded by

the client. Cloud computing is a technology that can provide users with computing, software, data access and storage services without the consideration of the configuration of user's computers or physical locations.

## 2.1 Autodesk 123D Catch Workflow

Autodesk 123D Catch is the software which is based on multiple real images analysis and calculation, can automatically generate photo-realistic 3D models and scenes, and is client - server architecture. Autodesk 123D Catch allows any user capture a series of photos from different angles on the subject through a digital camera to generate 3D model of the subject, which demands the adequate overlap and coverage between different images. If the picture is uploaded to the client through the desktop server, Autodesk cloud-based applications using the photographic measurement principle to reconstruct the relative position between the images and produce a 3Dpoint of the target object grid model, and finally the image will be projected onto the grid model to generate realistic rendering [4, 6-8].

## 2.2 The Technology Principle of 123D Catch

The core technology Photo generating grid model is based on a set of computer graphics, computer vision and other related disciplines algorithm composed by order of evaluation include [5]:

Photo identification → Camera calibration and position outside camera computing → Panorama Stitching → Extracting pixels from the image → Automatic Point Cloud → Reconstructing the 3D Mesh → Mapping the extracted pixel as materials and textures on the corresponding parts of the mesh as binding operation → Generating the photo-realistic 3D scene model.

## 2.3 Operation Process

### Shooting the Right Images

In accordance with Autodesk 123D Catch shooting tips, should shoot two laps of panoramic images around the object, shown in Fig.4, the first lap angle of the lens should be horizontal, the second lap the angle should be set at 45 degree. Every lap about 16-24 pieces will be shot, about 4-6 sheets every 90 degrees, increases the number of shots depending on object size [5].

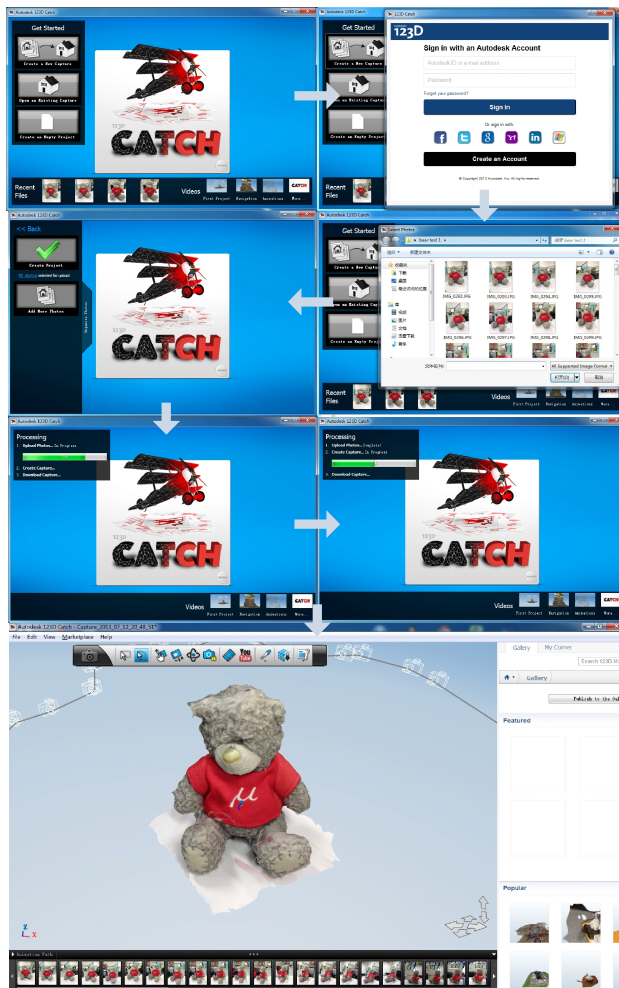
- Accessibility
- Occlusions and number of photographs to shoot
- Photographs need features
- No transparent, reflective or glossy subjects
- Subjects cannot move
- Consistent lighting

### Model Building

Before uploading the image Autodesk 123D Catch on the site <http://www.123dapp.com/catch> must be downloaded and installed. The application is not a 3D modeling software, but an image up loader, image processing are all carried out in the cloud, so ensure smooth network.

In Autodesk 123D Catch client, select "Create a new Photo Scene" You can create a new model, log Autodesk account, will shoot a panoramic image objects submitted to the backend server to be calculated.

Panoramic image sets were submitted to the background server, Autodesk 123D Catch use high-performance cloud computing server group to do a series of analyses and calculations, to generate a photo-realistic 3D scene model (in Fig.1 and Fig.2).



**Fig. 1.** Showing loading of images onto Autodesk software with subsequent computing of the images over the cloud to create a 3D model

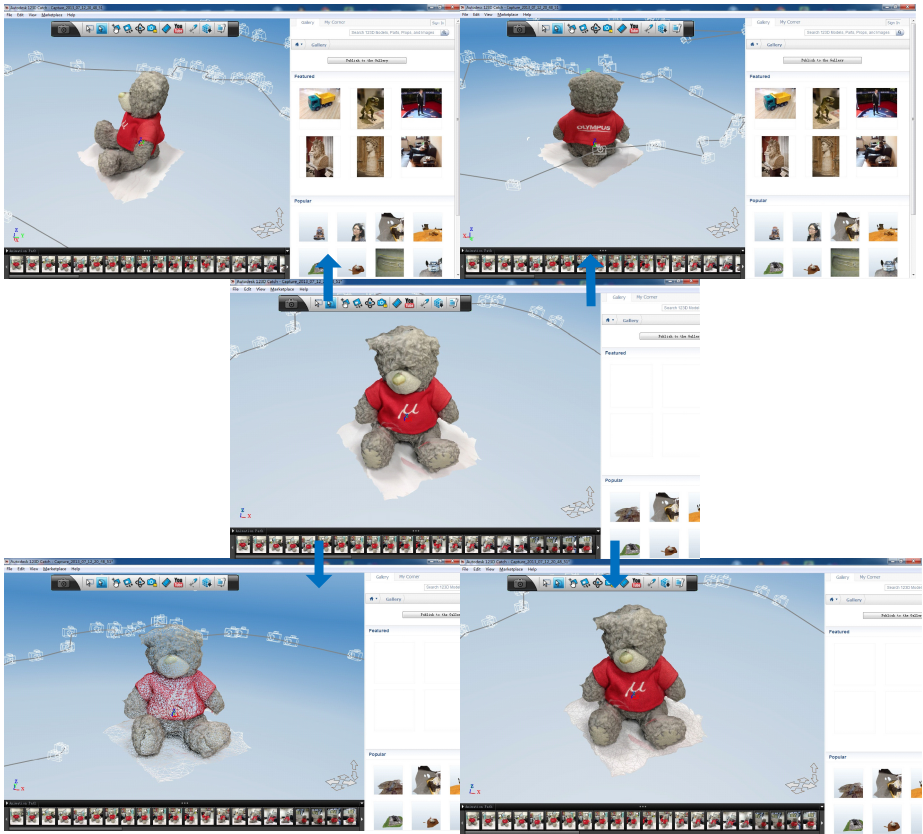


Fig. 2. The display of the 3D model

### 3 Conclusion

This paper take Autodesk 123D Catch as an example, from the point of view which be focused on application discussing the its basic method, the corresponding processes, operating points, and exploring the principle of realization. Take advantages of cloud calculation, which are fast, exactness and efficiency to create the 3D mesh of complex objects, to access to photograph-level the real effect, to simplify the reconstruction and to shorten the time. But at present the cloud software requires higher quality of image material, in the modeling process due to technical limitations the failures always come out. Nonetheless, Autodesk 123D Catch with the thinking of convenience uses the image to carry out 3D model will pioneered a new era in the future.

**Acknowledgments.** This work has been supported by the foundation of Innovative research program of Jiangsu Province of China (No.CXZZ13\_0403).

## References

1. Venkatesh, S., Ganeshkar, S.V., Ajmera, S.: Image-based 3D Modeling: A Simple and Economical Technique to Create 3D Models of the Face. *International Journal of Health Sciences and Research* 2, 93–99 (2012)
2. Wen-Kuo, H., Pi-Ling, P.: 3D Modeling and Application from Panoramic Photography. In: *Geographic Information Society Annual Conference*, Taiwan (2012)
3. Autodesk 123D Catch, <http://www.123dapp.com/catch>, 2012
4. Seitz, S.M., Dyer, C.R.: Photorealistic scene reconstruction by voxel coloring. *J. International Journal Computer Vision* 35(2), 151–173 (1999)
5. Autodesk Inc. Autodesk Project Photofly, Getting Started with the Photo Scene Editor 2.0[EB/OL] (2011), <http://labs.autodesk.com/>
6. Simões, F., Almeida, M., Pinheiro, M., et al.: Challenges in 3D Reconstruction from Images for Difficult Large Scale Objects: A Study on the Modeling of Electrical Substations. In: *IEEE 14th Symposium on Virtual and Augmented Reality*, Rio Janiero, pp. 74–83 (2012)
7. Brutto, M.L., Meli, P.: Computer Vision Tools for 3D Modeling in Archaeology. *International Journal of Heritage in the Digital Era*, 1–6 (2012)
8. Kersten, T.P., Lindstaedt, M.: Image-Based Low-Cost Systems for Automatic 3D Recording and Modeling of Archaeological Finds and Objects. In: Ioannides, M., Fritsch, D., Leissner, J., Davies, R., Remondino, F., Caffo, R. (eds.) *EuroMed 2012*. LNCS, vol. 7616, pp. 1–10. Springer, Heidelberg (2012)

# QoE Evaluation Model of Anti-Spam Allergy Problem Based on Fuzzy Entropy Method and Hidden Markov Model

Haizhuo Lin\* and Jilong Wang

Academy of Network Science and Cyber Space, Tsinghua University  
Beijing, China  
lin.haizhuo@163.com

**Abstract.** The most important innovation in this study is to introduce the concepts and methods of QoE of traditional video and audio experience quality into the allergy problem in the anti-spam field for the first time. We put forward a new multi-user QoE comprehensive evaluation model based on fuzzy entropy method as well as a new single user QoE evaluation model based on hidden markov model to make the quantitative analysis and build the complete computing framework for allergy index in the anti-spam field. We realize the complete principal derivation process for allergy index, all the Observation states and quantitative indicators of fuzzy entropy method and Hidden Markov model. With experiments and simulation based on data of Tsinghua University anti-spam gateway, we find the efficiency and characteristics of these two models and also summarize the relation and difference between these two methods. The experimental results confirm that these two models can efficiently solve the quantitative evaluation problem of the allergy phenomenon in the anti-spam field and get the reasonable allergy index.

**Keywords:** QoE, Anti-Spam, Allergy Problem, Fuzzy Entropy Method, Hidden Markov Model.

## 1 Introduction

Compared to vector space model[1], AHP[2] and other QoE evaluation model[3][4][5] based on artificial intelligence[6], a major advantage of fuzzy entropy method is using fuzzy matrix standardized and entropy conversion methods to reduce user subjective one-sidedness[7], also quantify the qualitative problem - allergy index. Compared to the other QoE evaluation models based on stochastic process[8], a major advantage of Hidden Markov model is considering the historical experience of the users[9], while other existing evaluation models have neglected the QoE impact of the past on that of the future.

---

\* Corresponding author.

## 2 Multi-user QoE Comprehensive Evaluation Model Based on Fuzzy Entropy Method

The experiments are based on data statistics of 90 mailbox users in Tsinghua University anti-spam gateway, including recalling rate, user positive feedback rate, and the suspected spam rate in June 2012 to August 2012. With these three rates as evaluation indicators, build the fuzzy entropy framework and the evaluation objective – allergy problem.

Complete multi-user QoE comprehensive evaluation model based on fuzzy entropy method is described as follows.

- (1) Determine the fuzzy matrix of second level indicators to first level indicators with M users' recalling rate, user positive feedback rate, and the suspected spam rate during the N months as the second level indicators for the upper indicators.

$$F=(f_{ij})_{m \times n} .$$

$$F=(f_{ij})_{m \times n} = \begin{bmatrix} 0.041 & 0.000 & 0.073 \\ 0.107 & 0.039 & 0.082 \\ 0.000 & 0.000 & 0.000 \\ & \dots & \\ & \dots & \\ 0.184 & 0.021 & 0.000 \\ 0.008 & 0.000 & 0.015 \\ 0.033 & 0.031 & 0.000 \end{bmatrix} \tag{1}$$

- (2) Make standard transformation for fuzzy matrix, eradicate overall error generated by extreme values to the QoE comprehensive evaluation.

$$f_{ij}'=(f_{ij}-\bar{f}_j)/\sigma_j+1 \tag{2}$$

- (3) Calculate proportion of each indicator to the sum of all the indicators for each user.

$$P = \frac{f_{ij}'}{\sum_{i=1}^M f_{ij}'} = \begin{bmatrix} 0.0087 & 0.0004 & 0.0166 \\ 0.2286 & 0.0076 & 0.0186 \\ 0.0011 & 0.0004 & 0.0017 \\ & \dots & \\ & \dots & \\ 0.0393 & 0.0041 & 0.0017 \\ 0.0017 & 0.0004 & 0.0034 \\ 0.0071 & 0.0060 & 0.0017 \end{bmatrix} \tag{3}$$

- (4) Calculate the entropy for each indicator.  $e_j(f)$ .

$$e_j(f) = -k \sum_{n=1}^n p_{ij} \ln p_{ij} \tag{4}$$

$$e_1 = 0.6434, e_2 = 0.4909, e_3 = 0.7824$$

(5) Calculate Difference coefficient  $V_j$  based on the entropy

$$v_j = 1 - e_j \tag{5}$$

$$v_1 = 0.3566, v_2 = 0.5091, v_3 = 0.2176$$

(6) Calculate the weights for each second level indicator to the whole second level indicators.  $w_j$ .

$$w_j = \frac{v_j}{\sum_{j=1}^N v_j} \tag{6}$$

$$w_1 = 0.3291, w_2 = 0.4699, w_3 = 0.2010$$

$$w = (0.1097, 0.1566, 0.0670, 0.1039, 0.1125, 0.1169, 0.1864, 0.1087, 0.0383)$$

(7) Calculating an index relative to the ultimate goal of allergy index weight ratio, and according to analytic hierarchy process jointing second level indicators, first level indicators and the ultimate QoE goal. Calculate average statistical values of 9 second level indicators by 90 mailbox and the allergy index as follows.

$$AI = w * C = 0.5769 \tag{7}$$

### 3 Single User QoE Evaluation Model Based on Hidden Markov Model

Complete single-user QoE evaluation model based on Hidden Markov model is described as follows.

(1) Divide the QoE of spam misjudgment into five hidden states based on QoE-MOS standard classification made by International Telecommunication Union[10].

$$Q_i \in (M_1, M_2, M_3, M_4, M_5), i \in (1, 2, \dots, x)$$

(2) Calculate the hidden state initial probability matrix as follow.

$$\pi = [\pi_1 \ \pi_2 \ \pi_3 \ \pi_4 \ \pi_5] = [0.10 \ 0.20 \ 0.20 \ 0.15 \ 0.35] \tag{8}$$

$$\pi_i = P(q_1 = M_i), i \in (1, 2, 3, 4, 5)$$

(3) Calculate the hidden state transition probability matrix  $A = (mos_{ij})_{M \times N}$  as follow.



$$A = \begin{bmatrix} mos_{11} & mos_{12} & mos_{13} & mos_{14} & mos_{15} \\ mos_{21} & mos_{22} & mos_{23} & mos_{24} & mos_{25} \\ mos_{31} & mos_{32} & mos_{33} & mos_{34} & mos_{35} \\ mos_{41} & mos_{42} & mos_{43} & mos_{44} & mos_{45} \\ mos_{51} & mos_{52} & mos_{53} & mos_{54} & mos_{55} \end{bmatrix} = \begin{bmatrix} 0.500 & 0.250 & 0.125 & 0.075 & 0.050 \\ 0.373 & 0.204 & 0.187 & 0.098 & 0.138 \\ 0.299 & 0.342 & 0.225 & 0.102 & 0.032 \\ 0.100 & 0.100 & 0.180 & 0.500 & 0.120 \\ 0.013 & 0.012 & 0.025 & 0.200 & 0.750 \end{bmatrix} \quad (9)$$

$$mos_{ij} = P(q_{t+1} = M_j | q_t = M_i), 1 \leq i, j \leq 5$$

Transition relation of QoE hidden states is as follows.

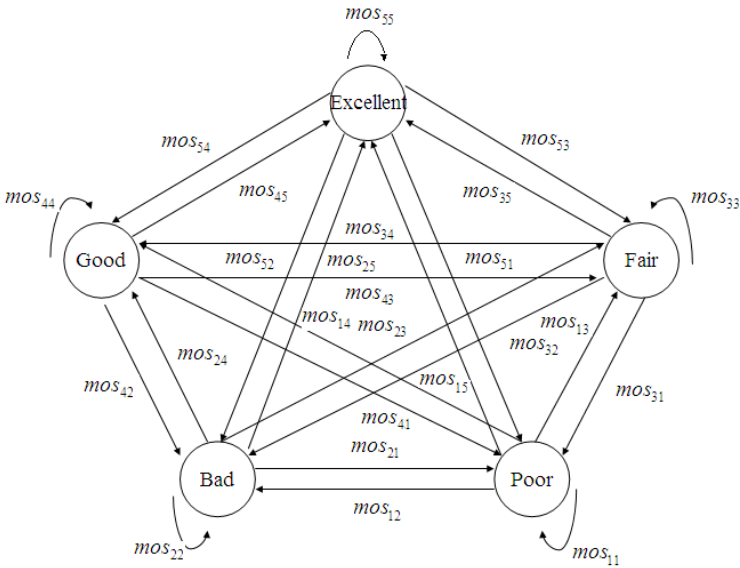


Fig. 1. Transition relation of QoE hidden states

(4) Status - observation two-state probability matrix  $B=(obs_{jk})_{N \times M}$  as follow.

$$B = \begin{bmatrix} obs_{11} & obs_{12} & \dots & obs_{1Y} \\ obs_{21} & obs_{22} & \dots & obs_{2Y} \\ obs_{31} & obs_{32} & \dots & obs_{3Y} \\ obs_{41} & obs_{42} & \dots & obs_{4Y} \\ obs_{51} & obs_{52} & \dots & obs_{5Y} \end{bmatrix} = \begin{bmatrix} 0.600 & 0.350 & 0.025 & 0.025 \\ 0.550 & 0.250 & 0.150 & 0.050 \\ 0.250 & 0.400 & 0.200 & 0.150 \\ 0.200 & 0.150 & 0.350 & 0.300 \\ 0.050 & 0.050 & 0.100 & 0.800 \end{bmatrix} \quad (10)$$

$$obs_{jk} = P(l_t = O_k | q_t = M_j), 1 \leq j \leq 5, 1 \leq k \leq M$$

(5) Calculate QoE hidden state sequence at the maximum probability significance according to the given observation sequence by viterbi algorithm.

i) First observation sequence is  $L_2$ , calculate  $\delta_1(i)$

$$\begin{cases} \delta_1(1) = \pi_1 obs_1(L_2) = 0.1 \times 0.350 = 0.0350 \text{ MOS}_1 \\ \delta_1(2) = \pi_2 obs_2(L_2) = 0.2 \times 0.250 = 0.0500 \text{ MOS}_2 \\ \delta_1(3) = \pi_3 obs_3(L_2) = 0.2 \times 0.400 = 0.0800 \text{ MOS}_3 \\ \delta_1(4) = \pi_4 obs_4(L_2) = 0.15 \times 0.150 = 0.0225 \text{ MOS}_4 \\ \delta_1(5) = \pi_5 obs_5(L_2) = 0.35 \times 0.005 = 0.00175 \text{ MOS}_5 \end{cases}$$

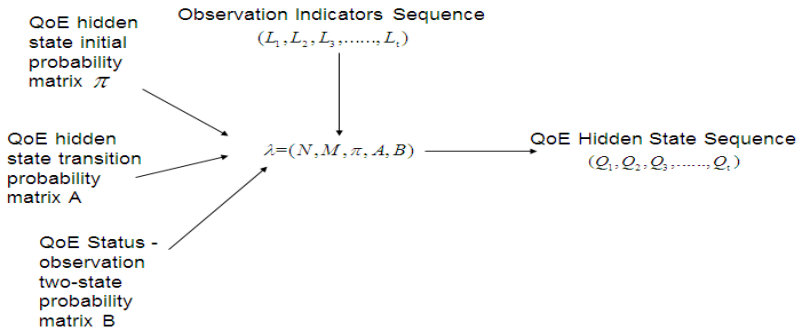
ii) Second observation sequence is  $L_1$ , calculate  $\delta_2(i)$

$$\begin{cases} \delta_2(1) = \max_{1 \leq i \leq 5} [\delta_1(2) mos_{i1}] obs_1(L_1) = \max \{ (0.035 \times 0.5 \times 0.6), \\ (0.05 \times 0.373 \times 0.6), (0.08 \times 0.299 \times 0.6), (0.0225 \times 0.1 \times 0.6), (0.00175 \times 0.013 \times 0.6) \} \\ = 0.02392 \text{ MOS}_1 \\ \delta_2(2) = \max_{1 \leq i \leq 5} [\delta_1(2) mos_{i2}] obs_2(L_1) = \max \{ (0.035 \times 0.250 \times 0.55), \\ (0.05 \times 0.204 \times 0.55), (0.08 \times 0.342 \times 0.55), (0.0225 \times 0.1 \times 0.55), (0.00175 \times 0.012 \times 0.55) \} \\ = 0.01848 \text{ MOS}_2 \\ \delta_2(3) = \max_{1 \leq i \leq 5} [\delta_1(2) mos_{i3}] obs_3(L_1) = \max \{ (0.035 \times 0.125 \times 0.25), \\ (0.05 \times 0.187 \times 0.25), (0.08 \times 0.225 \times 0.25), (0.0225 \times 0.18 \times 0.25), (0.00175 \times 0.025 \times 0.25) \} \\ = 0.01800 \text{ MOS}_3 \\ \delta_2(4) = \max_{1 \leq i \leq 5} [\delta_1(2) mos_{i4}] obs_4(L_1) = \max \{ (0.035 \times 0.075 \times 0.2), \\ (0.05 \times 0.098 \times 0.2), (0.08 \times 0.102 \times 0.2), (0.0225 \times 0.5 \times 0.2), (0.00175 \times 0.2 \times 0.2) \} \\ = 0.00225 \text{ MOS}_4 \\ \delta_2(5) = \max_{1 \leq i \leq 5} [\delta_1(2) mos_{i5}] obs_5(L_1) = \max \{ (0.035 \times 0.05 \times 0.05), \\ (0.05 \times 0.138 \times 0.05), (0.08 \times 0.032 \times 0.05), (0.0225 \times 0.12 \times 0.05), (0.00175 \times 0.75 \times 0.05) \} \\ = 0.000345 \text{ MOS}_5 \end{cases}$$

iii) Third observation sequence is  $L_4$ , calculate  $\delta_3(i)$

$$\begin{cases} \delta_3(1) = \max_{1 \leq i \leq 5} [\delta_2(1) mos_{i1}] obs_1(L_4) = \max \{ (0.02392 \times 0.5 \times 0.025), \\ (0.01848 \times 0.373 \times 0.025), (0.018 \times 0.299 \times 0.025), (0.00225 \times 0.1 \times 0.025), \\ (0.0069 \times 0.013 \times 0.025) \} = 0.000299 \text{ MOS}_1 \\ \delta_3(2) = \max_{1 \leq i \leq 5} [\delta_2(1) mos_{i2}] obs_2(L_4) = \max \{ (0.02392 \times 0.25 \times 0.05), \\ (0.01848 \times 0.204 \times 0.05), (0.018 \times 0.342 \times 0.05), (0.00225 \times 0.1 \times 0.05), \\ (0.0069 \times 0.012 \times 0.05) \} = 0.0003078 \text{ MOS}_2 \\ \delta_3(3) = \max_{1 \leq i \leq 5} [\delta_2(1) mos_{i3}] obs_3(L_4) = \max \{ (0.02392 \times 0.125 \times 0.15), \\ (0.01848 \times 0.187 \times 0.15), (0.018 \times 0.225 \times 0.15), (0.00225 \times 0.18 \times 0.15), \\ (0.0069 \times 0.025 \times 0.15) \} = 0.0006075 \text{ MOS}_3 \\ \delta_3(4) = \max_{1 \leq i \leq 5} [\delta_2(1) mos_{i4}] obs_4(L_4) = \max \{ (0.02392 \times 0.075 \times 0.3), \\ (0.01848 \times 0.098 \times 0.3), (0.018 \times 0.102 \times 0.3), (0.00225 \times 0.5 \times 0.3), \\ (0.0069 \times 0.2 \times 0.3) \} = 0.0005508 \text{ MOS}_4 \\ \delta_3(5) = \max_{1 \leq i \leq 5} [\delta_2(1) mos_{i5}] obs_5(L_4) = \max \{ (0.02392 \times 0.05 \times 0.8), \\ (0.01848 \times 0.138 \times 0.8), (0.018 \times 0.032 \times 0.8), (0.00225 \times 0.12 \times 0.8), \\ (0.000345 \times 0.75 \times 0.8) \} = 0.00204 \text{ MOS}_5 \end{cases}$$

Based on maximum probability significance, the maximum probability hidden state sequence is  $(MOS_3, MOS_2, MOS_3)$  according to the observation sequence  $(L_2, L_1, L_4)$ .



**Fig. 2.** Process of single user QoE evaluation model based on Hidden Markov model

## 4 Conclusion

In this paper, we promote and improve traditional two types of QoE evaluation model in the video and audio field -- artificial intelligence methods and stochastic modeling methods. Aiming at the two types of allergy problem in the anti-spam field -- spam filters and blacklists for IP filtering, we put forward two new QoE evaluation models to quantify the experience degrees for different users by defining and calculating the allergy index. Finally, we realize the whole quantitative analysis through the experiments and simulations based on the Tsinghua University anti-spam gateway.

## References

1. Menkovski, V., Sanchez, A.C.: Predicting quality of experience in multimedia streaming. In: Proceedings of the MoMM, pp. 52–59 (2009)
2. Du, Y., Zhou, W.-A., Chen, B.-F., et al.: A QoE based evaluation of service quality in wireless communication network. In: Proceedings of the New Trends in Information and Service Science, pp. 552–557 (2009)
3. Huang, R.L., Guan, Y.H.: Data Statistics Analysis. Higher Education Press, Beijing (2010)
4. Urban, F.M.: The weber-fechner law and mental measurement. *Journal of Experimental Psychology* 16(2), 221–238 (1933)
5. Reichl, P., Egger, S., Schatz, R.: The logarithmic nature of QoE and the role of the weber-fechner law in QoE assessment. In: Proceedings of the IEEE ICC 2010 (2010)
6. Fiedler, M., Hossfeld, T.: A generic quantitative relationship between quality of experience and quality of service. *IEEE of Network* 24(2), 36–41 (2010)
7. Mikhailov, L., Tsetinov, P.: Evaluation of services using a fuzzy analytichierarchy process. *Appl. Soft Comput.* 5(1), 23–33 (2004)
8. Schuller, B., Rigol, G.L., Lang, M.: Hidden Markov model-based speech emotion recognition. In: Proceedings of ICASSP, pp. 1–4 (2003)
9. End-2-end Multimedia services performance metrics. Technical Report: 3GPP TR 26.944 (2007)
10. International Telecommunication Union, Geneva. Methods for subjective determination of transmission quality. Report:ITU-T-P. 800 (1996)

# Topic Identification Strategy for English Academic Resources

Huang Li<sup>1,2</sup>, Hu Qing<sup>1,2</sup>, Xiong Xin<sup>2</sup>, and Hua Lijun<sup>1,2</sup>

<sup>1</sup> College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430080, China

<sup>2</sup> Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan 430080, China  
huangli82@wust.edu.cn

**Abstract.** The network resource grows up exponentially with the rapid development of Internet. It is very important to identify whether the resources belong to the professional field or not. However, resources in professional field are sorted up mostly manually at present, and some drawbacks exist in the process. To tackle with these problems, the Topic Identification Strategy of Academic Resources proposes topic identification method based on the main text of resources. The strategy first transfers different forms of academic resources into a unified format. And then break up the whole text into parts to calculate the topic degree of relevance. Samples are used to decide the weight of the thresholds. The proposed method could also add labels for resources automatically. Experiments show that, this strategy could effectively deal with English academic resources topic identification.

**Keywords:** Topic identification, Information extraction, Topic degree of relevance, Automatic labels.

## 1 Introduction

The network resource grows up exponentially with the rapid development of Internet. Since the 1970s, it is estimated that each year there are more than 100,000 kinds of journals, about 900,000 scientific reports, over 10 million conference articles and nearly 5 million scientific papers published in the world [1], and the total number is increasing year by year. Although there have been a large number of efficient and effective information retrieval system [2] to help users, how to find the academic resources of professional field becomes a tricky problem.

Currently, studying topic identification of academic resources from the text classification perspective is a basic method that used in topics identification system[3]. Whereas there are some shortcomings in the traditional text classification methods, like Naïve Bayes[4], KNN[5], Neural Network, SVM[6], maximum entropy model, regression model, genetic algorithm and so on. The above methods perform good in topic identification, but the operation process is quite complex. In order to simplify the

process flow, a new identification strategy for English academic resources has been proposed in this paper.

## 2 Preprocessing and Identification Principle

The proposed strategy will transfer heterogeneous academic resources into text format. There are some tools can assist that, so the proposed strategy takes advantage of these methods to associate with the file conversion. There are some auxiliary tools for converting files into text format, Lius[7] is one of them.

### 2.1 Building of Field Feature Set on the Basis of Wikipedia

We extract fields feature words from Wikipedia, using breadth-first traversal algorithm. Feature word according to the gradation weights allocated dynamically different. Show as an allocation formula 1. Where  $x_i$  represents entry where the number of layers,  $n_i$  represents entry layers where the number of feature words extracted.

$$w_i = \frac{1}{\sum \chi_k n_k} \chi_i n_i \quad (1)$$

### 2.2 Principles for Topic Identification of Academic Resources

Calculating the relevance between the filed subject and the academic resources is basic for the identification strategy, which means the relevance between academic resources and the field feature words have to be figured out at the first stage, and then the relevance value can be used to judge whether the academic resource belongs to the field.

## 3 Topic Identification Strategy for Academic Resources

### 3.1 Calculation of the Relevance

For the academic resources  $P_i$ , we extract all the words in it and put them in set  $R_{P_i} = \{l_k\}$ . And we name the Field Feature Words set as  $Q = \{L_i\}$ . Entries matching is the intersection part between  $R_{P_i}$  and  $Q$ . Naturally, the entries will vary their forms with the change of the context. Therefore, when do the entry matching, small range floating of the relevance are allowed. The matching formula as shown in formula 2.

$$|1 - Simword(l_k, L_i)| \leq \delta \quad (2)$$

The existing tools are not able to merge word frequency of the same term in different forms. To tackle this problem, we use formula 2 to computer frequency of the feature words.

With the above processing, we can get a set of all the relevant features words in academic resource  $P_i$ , i.e.  $R_{P_i} \cap Q = \{L_u\}$ ; the word frequency of each entry  $f_u$ ; and the value of each entry in the field  $w_u$ . Then the relevance can be figured out with referring to the identification principles. See as equation 3.

$$Sim(P_i, Q) = \frac{\sum_{L_u \in R_{P_i} \cap Q} f_u * w_u}{\sum_{L_t \in Q} w_t} \tag{3}$$

However, many academic resources contain large vocabulary. For example, an ordinary academic paper has vocabulary of more than 5000 words. Increasing in vocabulary will definitely led to exponential growth in calculating. Taking this factor into consideration, the efficiency of formula 3 is relatively low in such kind of condition. Therefore, we proposed a dismembered approach to improve the processing efficiency.

In order to improve the efficiency of the computation, the key idea of the method is that the academic resource should be divided into several small pieces for distributed processing and calculation but not processed as a whole one. Each segmentation is based on the number of entries. The principle of the division is the use of the number of entries. Set the number of entries for each one of the *seg*, reached that number was cut into a document. In this paper, we use *Seg* to represents the total number of entries in each segmentation. *Simseg* represents relevance computation of a certain term  $Seg_{P_{iv}}$  in the segmentation. Then equation 3 can be changed into equation 4.

$$Simseg(Seg_{P_{iv}}, Q) = \frac{\sum_{L_v \in Seg_{P_{iv}}} f_v * w_v}{\sum_{L_t \in Q} w_t} \tag{4}$$

The relevance value of the entire academic resource is the average relevance value of all the segmentation, as shown in formula 5.

$$Sim(P_i, Q) = \sum_{seg\_num} Simseg(Seg_{P_{iv}}, Q) \tag{5}$$

By formula 5, we can calculate the relevance value between the academic resource and the field, which can provide data support for subsequent judgment.

### 3.2 Strategies for Sample-Study-Based Threshold Value Setting

In the previous section, the relevance between the academic resources and the field has been calculated out through the dismembered method. The following step will be topic identification. The identification principle is the same as the general judge principle. First, set a threshold value  $\delta$ , and then compare it to the relevance value. If the relevance value is greater than  $\delta$ , we can draw the conclusion that this academic resource belongs to the field.

We choose  $sam\_num$  papers from a certain field called  $Q$  as samples for analysis and then calculate the relevance value with the methods mentioned above and finally draw the relevance curve. We choose the average relevant as the relevant threshold of  $Q$ , and use the symbol  $\delta_Q$  to represent it. The theoretical calculation formula is shown in equation 6.

$$\delta_Q = \frac{\sum_{sam\_num} Sim(P_{sam_t}, Q)}{sam\_num} \quad (6)$$

In equation 6,  $P_{sam_t}$  represents the  $t^{\text{th}}$  academic resource sample and  $Sim(P_{sam_t}, Q)$  represents the relevant value between  $t^{\text{th}}$  sample and the corresponding field.

The threshold value calculated by equation 6 has strong dependence on the field and reflects the average relevant value of the resource legitimately. It is an average level and is greater than the critical value, i.e. the actual threshold value. As a result, the threshold value calculated by equation 6 should be declined by a little bit to work as the actual one. Then equation 6 can be improved to equation 7.

$$\delta_Q = \frac{\sum_{sam\_num} Sim(P_{sam_t}, Q)}{sam\_num - \varepsilon} \quad (7)$$

In equation 7,  $\varepsilon$  represents the decline value, which is obtained from sample analysis curve. The specific solution method will be described in section 4.

### 3.3 Auto-labeling

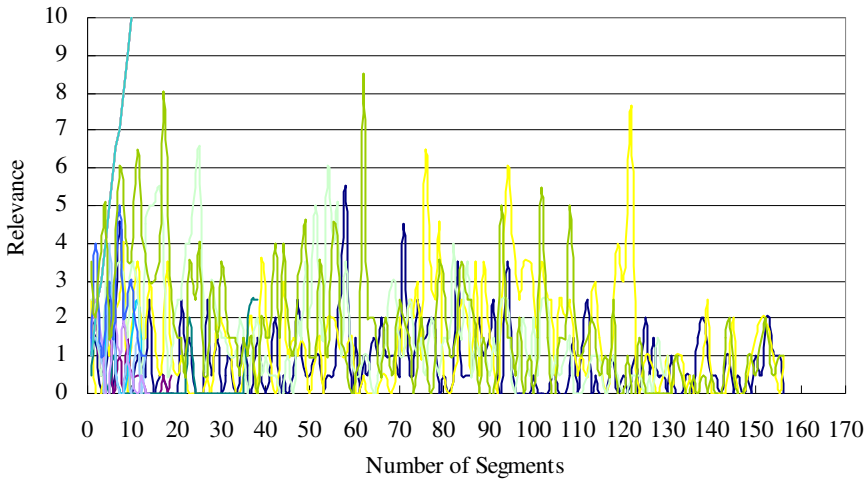
Both word frequency and weight should be considered for the automatic label selection of academic resources. When merging these two factors, the smaller one might be omitted because of different value range of these two factors. To address this problem, the geometric mean of these two values are used to solve the pivotal degree as shown in equation 8.

$$degree_i = \sqrt{fre_i * w_i} \quad (8)$$

There is no need to do pivotal degree calculate for all the entries, only the first  $m$  entries with higher word frequency need this calculating and resorting. The key is need to calculate and sort, then for researchers' reference, choose  $x$  ( $m > x$ ) items with higher value as labels of the academic resource.

## 4 Experiment

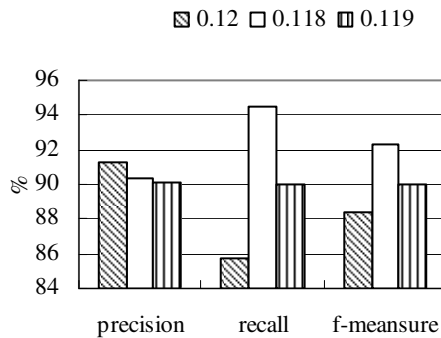
Select 100 computer-related scientific and technical literatures from ACM and Citeseer for sample analysis. Each article will be divided to 200 words of a similar calculation for the value of the block. The result is shown in Figure 1. The abscissa represents resource segmentations, and the ordinate represents relevance value. The wave lines in different colors represent different academic resources.



**Fig. 1.** Analysis of the tentative threshold value

Calculated by formula 7, the average threshold value of computer field is  $\delta=0.12$ .

This paper chooses 200 papers of academic resources including technical literatures, personal pages and laboratory pages as examples. The standard result is obtained by means of manual processing. The final identification result is shown in figure 2.



**Fig. 2.** Analysis of the identification result

Different threshold values are analyzed as seen in figure 2, it clear that when the threshold value is 0.118, the result is the most excellent one. Therefore, the weight of  $\epsilon$  in formula 7 should be 0.002.

Similarly also can be seen in figure 2, the topic identification strategy in quite helpful in topic identifying for the value of the Precision and Recall are relatively high.



## 5 Conclusion

The proposed strategy mainly focuses on a variety of English academic resources. Various academic resources are converted to a unified format, and then the field identification is carried out by calculating the relevance between feature words and extracted words in academic resources. The field feature words are selected from the hierarchical dictionary established on Wikipedia entries. In order to reduce the workload, the calculation of the relevance uses a statistical dismembered method. And finally, according to the calculated relevant degree, this strategy can add labels to academic literature automatically for researchers' reference. Tests on the English academic resources in computer science show that this method can effectively identify the topic of the English academic resources.

**Acknowledgment.** This work was supported by the National Natural Science Foundation of China (Grant Nos. 61100055), Natural Science Foundation of Hubei Province (Grant No.500104), the Hubei Province Key Laboratory of Intelligent information processing and real-time industrial systems (Wuhan University of Science and Technology, Grant Nos. znss2013B013).

## References

1. Cao, Y.: Scientific literature Resources Analysis and Retrieval. *Information Development & Economy* (33), 120–123 (2011)
2. Ricardo, B.-Y., Berthier, R.-N.: *Modern Information Retrieval*, pp. 1–20. ACM Press, New York (1999)
3. Yuan, P., Chen, Y., Jin, H., et al.: MSVM-kNN: Combining SVM and k-NN for Multi-Class Text Classification. In: *IEEE International Workshop on Semantic Computing and Systems (WSCS 2008)*, pp. 133–140 (2008)
4. Kolcz, A., Yih, W.-T.: Raising the Baseline for High-Precision Text Classifiers. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2007)*, San Jose, pp. 400–409 (2007)
5. Hao, X., Tao, X., Zhang, C., et al.: An Effective Method to Improve KNN Text Classifier. In: *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, pp. 1379–1384 (2007)
6. Li, X., Cervantes, J., Yu, W.: Two-Stage SVM Classification for Large Data Sets via Randomly Reducing and Recovering Training Data. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC 2007)*, Montreal, pp. 3633–3638 (2007)
7. Zhu, X.: Publishing models of merged based on Lucene text search engine for Applied Research. *ITS Applications* (22), 3–5 (2010)
8. Yan, X.: Semantic Wiki-based Knowledge Retrieval. *Research Library Studies* (13), 75–80 (2010)
9. Ju, Y., Liu, C.: Abroad typical semantic annotation platform comparative study. *Modern Information* 29(1), 215–217 (2009)

# Performance Comparison between MLP Neural Network and Exponential Curve Fitting on Airwaves Data

Muhammad Abdulkarim<sup>1,\*</sup>, Afza Shafie<sup>2</sup>,  
Wan Fatimah Wan Ahmad<sup>1</sup>, and Radzuan Razali<sup>2</sup>

<sup>1</sup>Dept. of Computer & Information Sciences,  
Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

<sup>2</sup>Dept. of Fundamental & Applied Sciences,  
Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia  
mmmhammad@gmail.com,  
{Afza, fatimhd, radzuan\_razali}@petronas.com.my

**Abstract.** This study aims at comparing the performance of a Multi-Layer Feed-Forward Neural Network and exponential curve fitting Models for the estimation of airwaves associated with shallow water Controlled Source Electro-Magnetic (CSEM) data. The performance measure is based on Mean Square Error (MSE), Sum of Squares Error (SSE) and coefficient of determination ( $R^2$ ). The MLP-NN network produced better and superior results with low MSE of  $1.13e-7$ , SSE of 0.00017 and higher  $R^2$  of 99.35%.

**Keywords:** Airwaves, Coefficient of Determination, Exponential Curve Fitting, CSEM, Multi-Layer Perceptron, MSE, SSE.

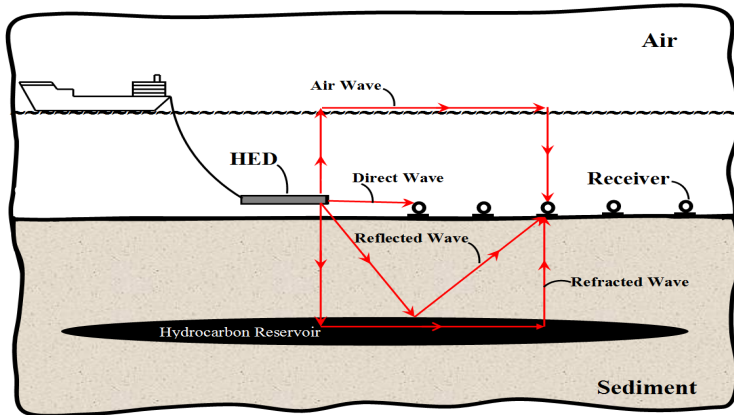
## 1 Introduction

Controlled Source Electro-Magnetic (CSEM) is an electromagnetic technique that provides a means to remotely detect the presence of high resistive sub-sea floor structures of the earth's interior such as the hydrocarbon reservoir. The technique can be describe thus; an electrode refers to a Horizontal Electric Dipole (HED) which serves as electromagnetic source is disposed approximately 30 – 50m above the seabed and connected to the survey recording vessel. The electrode is being charged by the power source on the survey vessel at selected magnitude of alternating current and transmission frequency or frequencies. At a selected source receiver distance (offset), an Ultra-low ( $\sim 0.1 - 5$ Hz) transmission frequency is commonly passed through the seabed into the subsurface of the earth formation. Electric and magnetic sensors/receiver designed with a voltage measuring circuit are placed strategically either on the seabed or on a different survey vessel (see Figure 1). The imparted voltages recorded by the receivers on the seabed are then analyzed to presume the structural formations beneath the earth surface through their electrical properties [1]. The physics of the CSEM is basically the knowledge that when the electromagnetic (EM)

---

\* Corresponding author.

field is propagated through a conductive subsurface, the induced signal is mainly affected by spatial distribution of resistivity. Sediments filled with saltwater in a marine environments typically represent good conductors, whereas the example of resistive bodies that scatter the EM field include carbonates, hydrocarbon filled sediments, salt and volcanic rocks.



**Fig. 1.** Schematic Illustration of CSEM Survey Environment.

Part of the electromagnetic field signal that were scattered by subsurface inhomogeneities propagates back to the seafloor where the signal is recorded by receivers equipped with electric and magnetic field sensors.

The data recorded in shallow water CSEM survey are known to be affected by a noise called “airwaves”. The noise component (airwaves) are generated predominantly by the vertically up going diffuse electromagnetic signal component that propagates in form of wave at the air/sea interface with speed of light and without attenuation before it diffuse back through the water layer vertically down where it is recorded by the electromagnetic receivers [2] as illustrated in Figure 1.

In order to estimate and remove the noise associated with oil exploration data, petroleum industry has found Artificial Neural Networks (ANN) to be useful in processing seismic and potential-field data for hydrocarbon explorations [3]. The neural networks has been used in works such as interpreting well logs, processing of EM sounding data, recognizing seismic waveforms, function approximation, electromagnetic, magneto telluric and seismic inversion purposes and for many other problems [4, 5] .

The ANN technique has the advantage of firstly, ability to be used as an arbitrary function approximation mechanism which ‘learns’ from observed data. Secondly, it has the advantage of not being constraint to satisfy any parametric assumptions. These are two main reasons that motivate the authors of this study to apply ANN technique for the estimation of airwaves in marine CSEM survey data.

Simulations using Computer Simulation Technology (CST) software were carried out in this study to obtain the shallow water airwaves data. Multi-Layer Perceptron (MLP) neural approach and exponential curve fitting method were used to determine

a model that can best fit the airwaves data. Mean Square Error (MSE), Sum of Square Error (SSE) and Coefficient of determination ( $R^2$ ) are used to measure the performance of the two approaches. The paper is organized as follows; Section 2 discusses MLP neural networks and exponential curve fitting algorithms followed by methodology in section 3. The study results are presented in section 4 and the conclusions in section 5.

## 2 The Proposed Artificial Neural Networks

Artificial Neural Networks (ANNs) simply called neural networks are systems for processing massively parallel distributed information. Through the process of learning, a neural network system is capable of gaining and storing experiential knowledge for future use [6]. The artificial neurons are considered to be motivated by the functional units of the human brain that are responsible for computation and processing information. Mathematically, neural networks are defined as universal approximators capable of solving large-scale complex problems like classification, control, estimation, nonlinear modeling, pattern recognition and time series analysis/forecasting.

The ANN system achieved those attributes through identifying relationships that exist between the patterns in a given problem. There are many different types of neural networks that exist in the literatures, but in this work we consider the Feed-Forward Neural Networks (FF-NN) which is the most popular and commonly used neural network architecture. More specifically the MLP is considered and employed for comparison in this study as non-linear process to model the airwaves data.

### 2.1 The Multi-layer Perceptron Model

The MLP neural network model used in this work consists of five (5) input neurons, three (3) hidden neurons, and one (1) output neuron as shown in Figure 2. The architecture of neural network can be represented symbolically by ANN ( $i, j, k$ ).

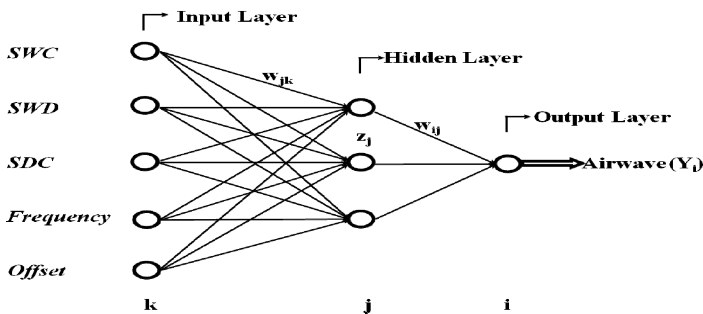


Fig. 2. A Multi-Layer Perceptron Feed-Forward Neural Network Model

The multi layer FF-NN considered in this study is fitted with log sigmoidal (logsig) and linear (purelin) activation functions for the neurons in hidden and output layers of the network respectively. Displayed in Figure 2,  $W_{jk}$  indicates the weight for the connection between the  $j^{\text{th}}$  hidden neuron and  $k^{\text{th}}$  input neuron. Likewise, the weight of the connection between the output neuron and the  $j^{\text{th}}$  hidden neuron is denoted by  $W_{ij}$ . Frequency, Sediment Conductivity (SDC), Sea Water Conductivity (SWC), Sea Water Depth (SWD) and offset signifies the network input variables while  $Y_i$  is the network output variable (i.e. the magnitude of airwaves) for the case of this study. Data normalization is important due to the nature of log sigmoidal activation function [7]. The range of the data after normalization is [0 1].

Bayesian-regularization back propagation is used as the training algorithm for the FF-NNs because the algorithm is known for improving the networks generalization property. Moreover, neural networks model developed by Bayesian-regularization back propagation algorithm generally have smaller biases and weights, and thus producing smoother response with less likelihood to result in over-fitting [8].

## 2.2 The Exponential Function

An exponential function is a widely used function in fundamental and applied sciences to model a relationship in which a constant change in the independent variable gives the same proportional change (i.e. percentage increase or decrease) in the dependent variable. All exponential functions are relatives of the primitive, two parameter families [9] presented in equation (1). Variations within the extended family are merely shifts, stretches, and transformations of this common stock as:

$$Y = f(x) = \alpha\beta^x \quad (1)$$

Where, the parameter  $\alpha$  is called the function's Y-intercept, the parameter  $\beta$  is called the base and  $x$  is the random variable. Together, they completely determine an exponential function's input-output behaviour. The base  $\beta$  in an exponential function must be positive in order to make real and algebraic sense. Also because we only work with positive bases,  $\beta^x$  is always positive. The values of  $f(x)$  therefore, are either always positive or always negative, depending on the sign of  $\alpha$ .

Indeed, in calculus and in many applications,  $\beta = e$  is the base of choice for exponential functions. This can always be arranged, since for any  $\beta > 0$ , we can find a constant  $\lambda$  so that  $e^\lambda = \beta$ . We can then re-write equation (1) as:

$$f(x) = \alpha\beta^x = \alpha e^{\lambda x} \quad (2)$$

However, sometimes the term exponential function is used more generally for functions of the form  $\alpha\beta^x$ , where the base  $\beta$  is any positive real number, not necessarily  $e$ .

### 2.3 Algorithm of Exponential Curve Fitting

Curve fitting technique is a process of determining mathematical function that has the best fit to a series of data points, possibly subject to constraints [10]. Given the points  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , where  $x_i$ 's and  $y_i$ 's ( $i = 1, 2, \dots, n$ ) represent the offset values and the corresponding magnitude of the airwaves respectively. Looking at Equation (2), the nonlinear least-squares procedure requires finding a minimum of:

$$E(\alpha, \lambda) = \sum_{i=1}^n (\alpha e^{\lambda x_i} - y_i)^2 \quad (3)$$

The partial derivatives of  $E(\alpha, \lambda)$  with respect to  $\alpha$  and  $\lambda$  are:

$$\frac{\partial E}{\partial \alpha} = 2 \sum_{i=1}^n (\alpha e^{\lambda x_i} - y_i) (\alpha x_i e^{\lambda x_i}) \quad (4)$$

And

$$\frac{\partial E}{\partial \lambda} = 2 \sum_{i=1}^n (\alpha e^{\lambda x_i} - y_i) (\alpha x_i e^{\lambda x_i}) \quad (5)$$

When the partial derivatives in equations (4) and (5) are set equal to zero and then simplified, the resulting normal equations are:

$$\lambda \sum_{i=1}^n x_i e^{2\alpha x_i} - \sum_{i=1}^n x_i y_i e^{\alpha x_i} = 0 \quad (6)$$

And

$$\lambda \sum_{i=1}^n e^{\alpha x_i} - \sum_{i=1}^n y_i e^{\alpha x_i} = 0 \quad (7)$$

Equations (6) and (7) are nonlinear in the unknowns  $\alpha$  and  $\lambda$ . To avoid a time-consuming computation and the iteration involved that requires good starting values for  $\alpha$  and  $\lambda$  for equations of this nature. We therefore utilize MATLAB software package which has a built-in minimization subroutine for functions of several variables to minimize  $E(\alpha, \lambda)$  directly.

## 3 Methodology

Computer Simulation Technology (CST) software was used generate the study data by simulating different CSEM environment. The area simulated is 25Km. The transmitter is modeled as a short 1250A AC line current segment of length 270m

located 30m above the sea bed. The transmitting frequency was varied as: 1Hz, 0.5Hz, 0.25Hz, 0.125 Hz and 0.0625Hz. The Maxwell's electromagnetic field wave equation in vacuum in the absence of electric or magnetic sources is solved for the electric field vector  $\mathbf{E}$  inside the computational domains. Figure 3 is a 1D geologic model depicting "No Air Model" and "With Air Model" configurations that were simulated to obtain the air wave data. Note that the only difference between "No Air Model" and "With Air Model" is the changing of the sea water depth and replacing the space with air layer. We changed the sea water depth at interval of 100m from 1000m down to 100m. The contribution of the airwaves to the CSEM data were computed by the method for removing the air wave effect as patented by [11] through the following steps:

- Constructing a CSEM geologic model of the region having a top air layer, a middle sea water layer, and a bottom earth layer, with the model reflecting known bathymetry of the region and known conductivities of the air, seawater and earth;
- Using the model to compute the electromagnetic field at all receiver locations for each source location;
- Replacing the air layer in the model with sea water to create a no-air model;
- Computing the fields for the same source-receiver geometries for the no-air model; and
- Computing the airwave effect by subtracting the No-Air Model fields from the corresponding fields of the With-Air Model.

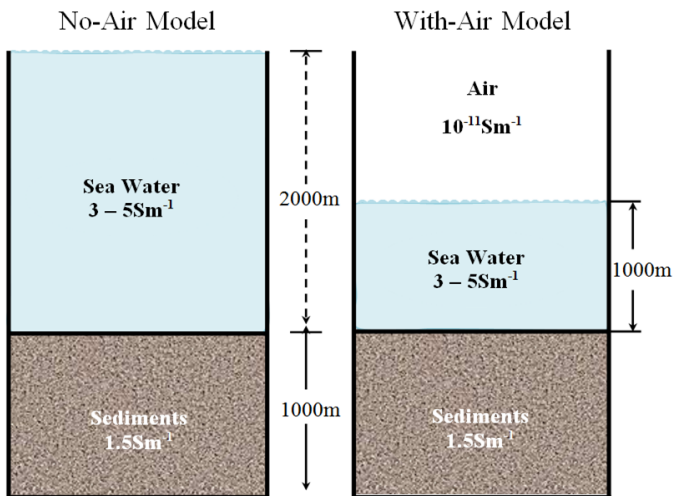


Fig. 3. 1D Simulation Set-up of the No Air and With Air Layer Geologic Models.

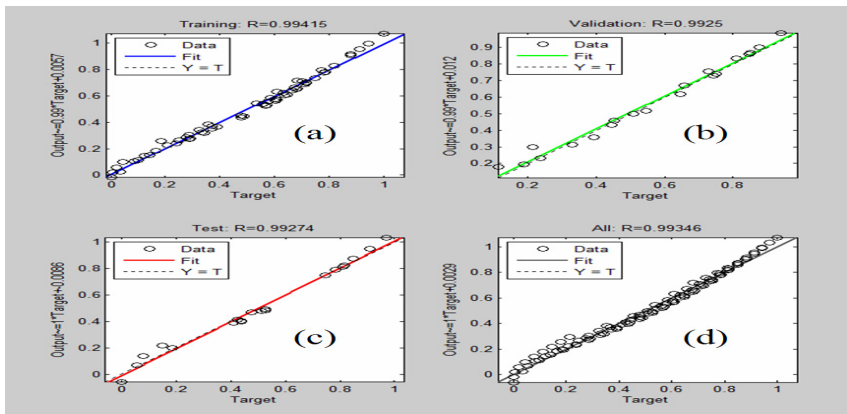
### 3.1 Data Analysis

MATLAB software package which has a built in neural network kit and minimization subroutine for functions of several variables to minimize  $E(\alpha, \lambda)$  was

utilized to run the MLP and exponential curve fitting respectively. The study data were first divided into training set (70%), validation (15%) and testing (15%) for the ANN model. Training data were used to train the application; validation data were used to monitor the neural network performance during training and the test data were used to measure the performance of the trained application. The performance of two models in this study are based on Mean Square Error (MSE), Sum of Squares Error (SSE) and coefficient of determination ( $R^2$ ) obtained from the result of the analysis.

## 4 Results

Simulated airwaves data from five different seawater depth 100m, 200m, 300m, 400m and 500m was used to train the MLP neural model. Figures 4(a) – (d) displays the networks training, validation, testing and the overall networks performance having 0.99415, 0.9925, 0.99274 and 0.99346 as the coefficient of determination respectively.



**Fig. 4.** (a) Training Regression Plot for the MLP. (b) Validation Regression Plot for the MLP. (c) Testing Regression Plot for the MLP. (d) Overall Regression Plot for the MLP.

Even though the result indicated that  $R^2$  for the training is slightly higher than for the testing and validation, this is expected and it might be as the effect of the sample size of the training i.e. 70% compared to 15% each for the testing and validation. The overall performance suggests a remarkable performance by the MLP having near 1.0 for the coefficient of determination. Table 1 present the parameter estimates of  $\alpha$  and  $\lambda$  from the curve fitting model. It can be seen from the tables that the shallower seawater depth of 100m has the highest value of the intercept  $\alpha$ .



**Table 1.** Curve fitting parameter estimates

Sea Water Depth	A	$\Lambda$
100M	1.142e-5	-3.05e-4
200M	8.656e-6	-3.07e-4
300M	5.839e-6	-2.98e-4
400M	3.033e-6	-2.79e-4
500M	1.967e-6	-2.73e-4

This might be as the result of early onset of the airwaves due to the shallowness of the seawater depth. Furthermore, it can also be observed that while value of  $\lambda$  remains fairly within the same range, the value of  $\alpha$  appears to be decreasing as the seawater is getting deeper.

**Table 2.** Average Performance comparison between MLP and Curve Fitting Method

Measure	MSE	SSE	R-Square
MLP Model	1.13e-7	0.00017	0.9935
Exponential Model	1.19e-5	0.0013	0.9759

The results in Table 2 clearly show the superiority of the MLP over the non-linear exponential curve fitting method. The MLP neural network indicate a lower both the mean square error and the sum of squares error of 1.13e-7 and 0.00017 compared to curve fitting having 1.19e-5 and 0.0013 respectively. This can be explain as the advantage neural network has in terms of learning input/output pattern and ability to adjust its weight during training without being restricted to any parametric assumption.

## 5 Conclusion

The MLP neural network approach and non-linear curve fitting methods has been successfully applied to the airwaves data from shallow water CSEM. The result has shown that MLP model can achieve  $R^2$  of 0.9935, MSE of 1.13e-7 and SSE of 0.00017 compared to the curve fitting having  $R^2$  of 0.9759, MSE of 1.19e-5 and SSE of 0.0013. Since the coefficient of determination is very close to 1.0 with small values for the MSE and SSE, it indicates that the MLP fits the airwaves data better than the curve fitting.

**Acknowledgments.** This research is carried out under the PETRONAS Research Fund (PRF). The authors would like to acknowledge Universiti Teknologi PETRONAS for giving the opportunity to carry out this research work.

## References

1. Constable, S., Srnka, L.J.: An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration. In: *Geophysics*, vol. 72. WA3 (2007)
2. Eidesmo, E., et al.: Sea Bed Logging (SBL): A new method for remote and direct identification of hydrocarbon filled layers in deepwater areas. *First Break* 20(3) (2002)
3. Brown, M., Poulton, M.: Locating buried objects for environmental site investigations using neural networks. *Journal of Environmental and Engineering Geophysics* 1, 179–188 (1996)
4. Neyamadpour, A., Taib, S., Abdullah, W.A.T.: Inversion of 2D DC resistivity data for high resistivity contrast regions using artificial neural network. In: *WSEAS International Conference on Engineering Mechanics, Structures, Engineering Geology (EMESEG 2008)*, Heraklion, Crete Island, Greece, July 22-24 (2008)
5. Van der Baan, M., Jutten, C.: Neural Networks in Geophysical Applications. *Geophysics* 65(4), 1032–1047 (2000)
6. Haykin, S.: *Neural networks: A comprehensive foundation*, 2nd edn. MacMillan, NY (1999)
7. Muhammad, A., et al.: Multi-Layer Perceptron Neural Network for Airwave Estimation in Marine Control Source Electromagnetic Data. *GSTF Int'l Journal on Computing* (2012)
8. Demuth, H., and Beale, M. *Neural network toolbox learning. For use with MATLAB: The Math Works Inc., Natick* (2001)
9. Lorentzen, L., Waadeland, H.: *The exponential function, Continued Fractions. Atlantis Studies in Mathematics*, 268
10. Coope, I.D.: Circle fitting by linear and nonlinear least squares. *Journal of Optimization Theory and Applications* 76(2), 381 (1993)
11. Xinyou, L., Srnka, L.J., Carazzone, J.J.: Method for Removing Air Wave Effect from Off-shore Frequency Domain Controlled-Source Electromagnetic Data. U.S. Patent 7 277 806 (October 2, 2007)

# A Private Cloud Storage System Based on Multithread Optimization

Zhang Wei<sup>1,2,\*</sup>, Xu Tao<sup>3</sup>, and Li Kun<sup>1</sup>

<sup>1</sup> School of Computer Science,  
Beijing Information Science and Technology University, Beijing 100101, China

<sup>2</sup> Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,  
Beijing Information Science and Technology University, Beijing 100101, China

<sup>3</sup> Microprocessor and SoC Technology R&D Center,  
Tsinghua University, Beijing 100084, China  
{zhwei02,xutao1979,41k87927}@163.com

**Abstract.** A private cloud storage model was established in this paper by combining local-storage and public cloud storage. In this model, the stored data show the following three stages: data preprocessing, data concentration storage and distributed management, data transmission. Parallel multithread was adopted in the client and the cloud storage server to deal with concurrent requests. At the same time, the lock mechanism was utilized to ensure data consistency. In this research, Python language was used in the Linux system to realize the establishment of the prototype system. According to this system together with the practical public cloud storage platform (Amazon S3), the test was conducted. The results indicate that the multithread optimization technique can contribute to the approximately linear improvement of the system performance.

**Keywords:** Cloud Computing, Cloud Storage, Multithread Optimization, Distributed System.

## 1 Introduction

In 2006, Google, Amazon, and other companies proposed the concept of cloud computing [1]. As an important part of cloud computing, cloud storage integrates different network devices, storage devices, servers, and client programs in the network with applications. This is achieved through features such as cluster applications, grid technologies, and distributed file systems. This enables collaboration and the joint provision of data storage and business access. In brief, cloud storage is a high-performance on-line storage service (Data as a Service) based on information networks provided for users. It has some basic features of cloud computing: 1) on-demand service, planning physical storage capacity with a logical storage view and ease of use; 2) based on information networks (LAN, MAN, or the Internet), it is online in real time; 3) ease of

---

\* Corresponding author.

expansion and release of storage capacity; 4) easily measured service quality and content; 5) ease of management.

Today’s cloud storage systems face two major problems: 1) scattered small files have a low storage efficiency for block data[2]; 2) cloud storage systems without central management nodes, namely those adopting the P2P model, have low data synchronization efficiency[3, 4], e.g., Amazon’s Dynamo[5] and Facebook’s Cassandra[6], whereas cloud storage systems with central management node share likely to encounter bottlenecks in overall system performance at the node servers[2, 7, 8], e.g., Google’s GFS[7] and the open-source HDFS[9].

## 2 The Overall Structure

Cloudkey is mostly a client system resulting from the combination of private cloud and public cloud. The workflow is presented as follows. First, the client transmits requests to the server and then client authentication is confirmed by the server. Afterwards, metadata (a relevant attribute of data) is sent to the server that creates or updates the meta-information in the database. Then, the client can upload (download) the data that are received by the server and saved in the local cluster (The local data are obtained and transmitted to the client). This is the end of the interaction between the client and the server. But the server is also required to monitor the updated local data. To fulfill this task, the server encrypts the data first and then calls the interface supplied by the public cloud storage platform to save the encrypted data in the cloud. The structure is demonstrated in Figure 1. As obviously shown in the figure, this system includes the client and the server. The client is mainly in charge of data preprocessing and the realization of repeating data de-duplication with the aid of the server. There are three modules in this system: data preprocessing, data management and storage, data encryption and backup.

The prototype system mainly shows four advantages: (1) Private cloud is adopted in the system to save data and public cloud is utilized to realize remote data backup. In this way, the system can ensure data security efficiently. (2) This system locally encrypts the data whose backup copy is uploaded in the public cloud storage platform.

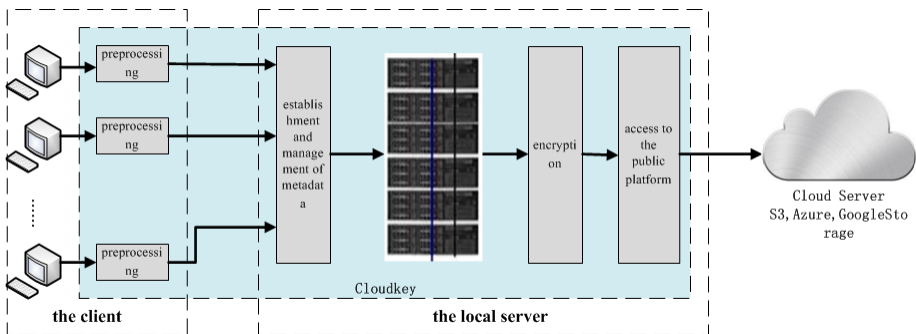


Fig. 1. The overall structure

So, data privacy in cloud is guaranteed effectively. (3) By separating data and metadata, data concentration management and distributed storage are achieved in private cloud. (4) The data preprocessing link is designed in this system and data de-duplication is used to optimize data storage. As a result, the storage space and backup time are saved.

### 3 Multithread Optimization

Figure 2 shows the flow of the multithread optimization technology. Efficiently it deal with the concurrent requests from multiple users, this system adopted the mode of the parallel multithread, as illustrated in the above figure. Its principle is displayed as follows: a server acting as the main thread is started first and used to monitor client requests circularly. In each loop, if there are client requests, they are added into the task queue. Then, tasks in the task queue are extracted in a circular pattern. A thread

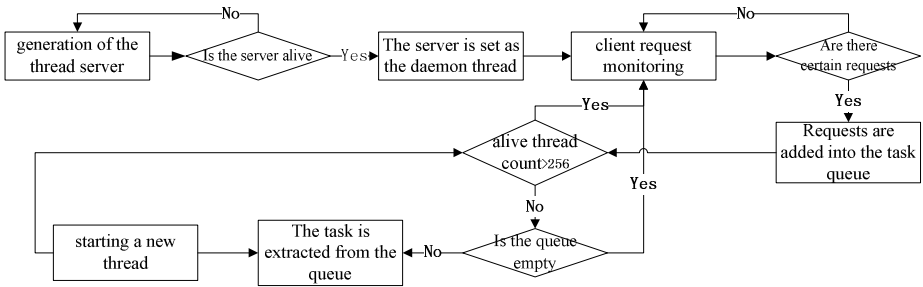


Fig. 2. The schematic diagram of parallel multithread

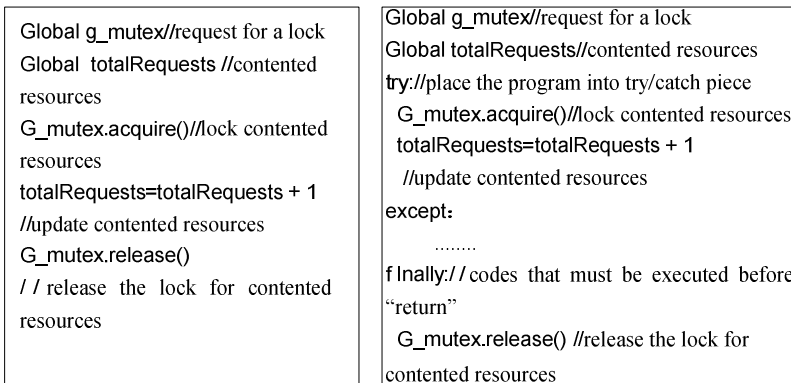
```

server = YunServer()//generation of the main thread
Server.setDaemon(True)//The server is set as the daemon thread
Server.start()//starting the thread
while True:
    if not server.isAlive()://The main server is not alive
        server = YunServer()//regeneration of the main thread
        server.setDaemon(True)
        server.start()
    time.sleep(120)//If there is no task, the main thread sleeps
Class YunServer(threading.Thread):
    socket.listen(200)//client request monitoring
    While(1):
        sock,clientaddr = socket.accept()//recieve socket
        Taskqueue.append(sock)//The accepted socket is added into the task queue
        while threading.activeCount() < 256://determination of the alive thread count
            if len(self.taskqueue) != 0://There are tasks in the task queue
                St = ServerThread(taskqueue.pop(0))//One task is extracted and a new thread is
                generated to execute it
                st.setDaemon(True)
                st.start()
            else://If there is no task, the server continues to accept socket
                break
    
```

Fig. 3. Realization of parallel multithread pseudo codes

is started to execute each task until the alive thread count reaches to a certain threshold (This threshold is usually set according to the bandwidth and computer hardware equipments: 256 in this paper) or there is no task in the task queue. Afterwards, the server continues to monitor client requests. Figure 3 shows the realization of parallel multithread pseudo codes.

As the lock mode is frequently used to prevent contention, it is adopted in this research to solve the contention problem in shared memory. When the invariant is broken, the lock mode can prevent other threads from accessing to the memory related to this invariant. In the lock mechanism, there are two methods: Enter and Exit. When a thread calls Enter, Enter calling performed by other threads is blocked (in the waiting state) until this thread changes to call Exit. The thread calling Enter is the lock holder. If other threads call Exit, generally, the program is abnormal. The lock mode expressed using pseudo codes is exhibited in Figure 4 (left). But a problem is presented in this method: if abnormal cases occur in the lock process, Exit cannot be called, which can induce other threads attempting to execute this code to be blocked permanently. To recover the program from abnormal situations, try/finally is utilized to improve program robustness. Its corresponding pseudo codes are demonstrated in Figure 4 (right).



**Fig. 4.** Left: pseudo codes in lock mode; Right: pseudo codes for solving abnormal lock

## 4 The Test

The test environment parameters are list in Table 1.

**Table 1.** Test Environment

	client	server
hardware	Inter Core(TM) i3 M3802.53GHz RAM: 2.00GB	Inter(R) Core(TM) i5 650 3.20GHz RAM: 4.00GB
OS	windows 7	ubuntu10.10
software	VS2010, mysql5.0	python2.6.5, mysql5.0

First, Cloudkey functions are tested, as illustrated in Figure 4 (a denotes the client interface; b1 represents data in “storage”; b2 refers to metadata records; c denotes data in Amazon S3). The client is presented in the form of the graphical interface, showing many functions including session authentication, user management, path management, content management, task, sharing, etc. Users can select and upload local files freely and the metadata information of selected files can be displayed on the interface of the client software, as exhibited in Figure 5 (a). The local server creates and saves metadata for the files uploaded by the client (Figure 5 (b2)). Of the data contents saved in “storage”, only one content ID is of 182 attribute (Figure 5 (b1)). Thus, the separation of data and metadata is realized and the expected effect of data storage is achieved. In Amazon S3, there is a ciphertext whose file name consists of the file name of “storage” plus the suffix .enc (Figure 5 (c)). In this way, the data backup and privacy are ensured. Therefore, functionally, the Cloudkey system has achieved the expected goals and this provides the experimental platform for further researches in the future.

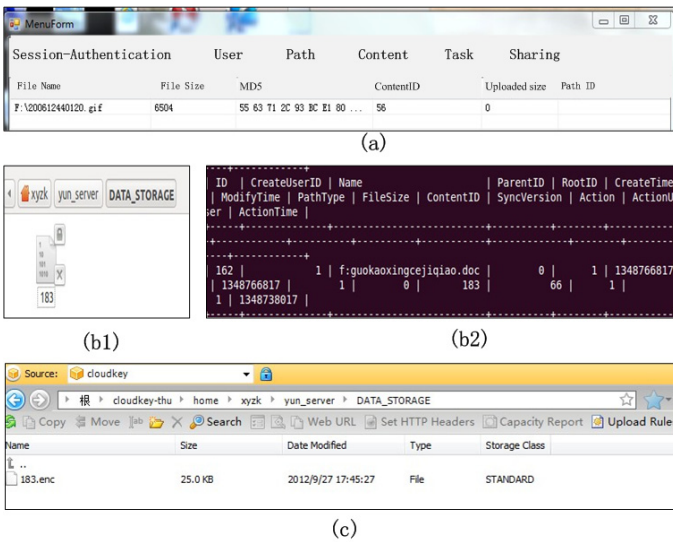


Fig. 5. The functional test for Cloudkey system

In the paper, the C language with linux pthread technology is used for implementing the multi-thread private cloud storage system, which is now generally considered to provide higher speed than the widely used single-thread algorithm. As can be seen from Figure 6, the multi-thread technology has an average line performance enhancement as compared to the traditional single-thread algorithm. The all performance data’s unit is second. About 1.7 time performance improvement can be achieved through 2-thread, and about 2.8 time through 4-thread, and about 4.4 time through 8-thread.

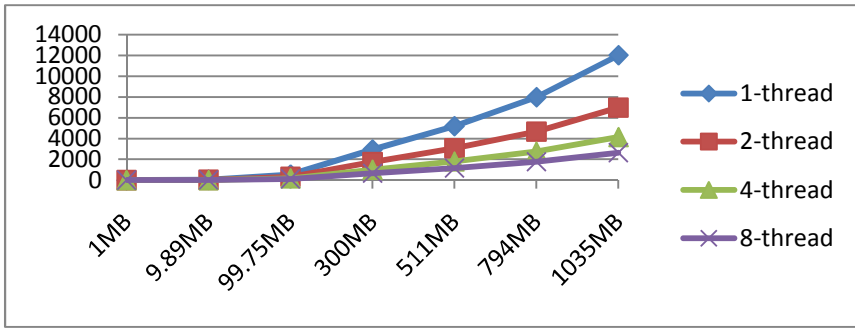


Fig. 6. Multithread Performance compared to single-thread

## 5 Conclusion

For cloud storage system the maximum concurrent request number processed by the system has a great influence on the overall system performance. In this research, the parallel multithread mechanism is used to deal with concurrent requests in the client and the server. In this way, the system performance is optimized. At the same time, the lock mode is also utilized to efficiently prevent the contention that frequently occurs in the parallel multithread. Then, an overall and detailed test is conducted to verify the feasibility of the private cloud model proposed in this paper. Meanwhile, the multithread technique is adopted to facilitate the approximately linear improvement of the system performance.

**Acknowledgment.** Funding for this research was provided in part by the Beijing Excellent Talent Training Project(2012D005007000009), General Program of Science and Technology Development Project of Beijing Municipal Education Commission(KM201110772014), and The Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges Under Beijing Municipality(IDHT20130519). We like to thank anonymous reviewers for their valuable comments.

## References

1. Foster, I., Yong, Z., Raicu, I., et al.: Cloud computing and grid computing 360-degree compared, 1–10 (2008)
2. Wu, Y.W., Huang, X.M.: Cloud storage. Communications of the CCF 5 (2009)
3. Weil, S.A., Pollack, K.T., Brandt, S.A., Miller, A.E.L.: Dynamic metadata management for petabyte-scale file systems. In: Proc. of the 2004 ACM/IEEE Conf. on Supercomputing (SC 2004). ACM Press, Pittsburgh (2004)
4. Hua, Y., Jiang, H., Zhu, Y.F., Feng, D., Tian, L.: SmartStore: A new metadata organization paradigm with semantic-awareness for next-generation file systems. In: Proc. of the FAST 2009, pp. 1–12. ACM Press, Portland (2009)



5. Decandia, G., Hastorun, D., Jampani, M., Kakulapati, G., Lakshman, A., Pilchin, A., Sivasubramanian, S., Vosshall, P., Vogels, W.: Dynamo: Amazon's highly available key-value store. In: Proc. of the SOSP 2007, pp. 205–220. ACM Press, Stevenson (2007)
6. Lakshman, A., Malik, P.: Cassandra: A decentralized structured storage system. *ACM SIGOPS Operating Systems Review* 44(2), 35–40 (2010)
7. Ghemawat, S., Gobioff, H., Leung, S.T.: The Google file system. In: Proc. of the Symp. on Operating Systems Principles (SOSP 2003), pp. 29–43. ACM Press, Bolton (2003)
8. Chang, F., Dean, J., Ghemawat, S., Hsieh: Bigtable: A distributed storage system for structured data. In: Proc. of the OSDI 2006 (2006)
9. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The Hadoop distributed file system. In: Proc. of the IEEE 26th Symp. on Mass Storage Systems and Technologies (MSST), pp. 1–10. IEEE, Lake Tahoe (2010)

# Cloud Storage-Based Medical Data Integration Technology

HongJun Zhan<sup>1</sup> and Wei Zhang<sup>2,3,\*</sup>

<sup>1</sup> Transportation Management College, Dalian Maritime University,  
Dalian City, Liaoning Province, 116026, China

<sup>2</sup> School of Computer Science, Beijing Information Science and Technology University,  
Beijing 100101, China

<sup>3</sup> Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,  
Beijing Information Science and Technology University, Beijing 100101, China  
zhanhongjun0612@163.com, zhwei@bistu.edu.cn

**Abstract.** Recent years has witnessed a spurt of medical information development and the formation of isolated hospital information islands followed. The case data and image data are desired to be integrated on a medical cloud storage information platform to balance the medical resources, improve the diagnosis and treatment level of primary hospital, reduce patient's medical expenditure, and solve the hard and expensive healthcare problem. On the basis of the whole framework of medical cloud storage and data integration technology, this study builds a unified medical cloud storage information platform by fast, efficiently, and compatibly integrating the heterogeneous information sources of each hospital, aiming at realizing an unified, convenient, integrated external service interface and the real sharing of medical information.

**Keywords:** Cloud storage, Heterogeneous data integration, Medical information platform, Medical cloud storage.

## 1 Introduction

In recent years, the rapid development of medical informatization triggers increasing new technologies and equipments, including spiral 320 slice CT, super high field magnetic resonance angiography, molecular imaging, functional imaging, multimodal fusion imaging etc.. These technologies greatly enrich doctors' diagnosis approaches and improve the disease diagnosis effects meanwhile, while also bring some problems:

- The medical cost stays constantly high due to the expensive high-end imaging equipments costing upward of millions to tens of millions Yuan;
- Hundreds to thousands of images are generated in a single scanning of medical imaging equipment, while only a few images are given to patients. This situation

---

\* Corresponding author.

hinders the parameter adjustment and 3D dynamic display and thus greatly discounts the diagnosis value. Patients are required to be reexamined after transferred. Unnecessary reexamination further increases patient's medical burden;

- Small hospitals are widely provided with imaging equipments such as X-ray and ultrasound imaging equipment etc., but lack of professional imaging diagnostic doctors.
- Owing to severe deficiency of funds, equipment, technology, and talents in primary medical institutions, patients over-crow in large hospitals while rarely visit primary hospitals. Such imbalance contributes greatly to the difficult and expensive healthcare.
- Image diagnosis calls for doctor's solid basic knowledge and abundant image-reading experiences considering its difficulty. Meanwhile, the constantly emerged new equipments and technologies further the requirements in imaging diagnosis teaching. In addition, traditional teaching means and equipments in medical colleges fail to meet the need of expanding enrollment scale.

The information and communication technology, with computer technology at the core, has been applied to each field of medical and healthcare industry. The medical and health industry[1] costing of tens of billions Yuan, consisting of the hospital information system, public health information system, remote medical treatment, home nursing, and regional collaborative medical treatment has arisen and drawn the widespread attention from academic and industrial fields. At present, the difficult and expensive healthcare should be mainly solved by increasing government's investment and supervision, balancing regional medical resources, and comprehensively carrying out regional medical sharing and collaboration.

## 2 Related Works

In recent years, European Union, America, and Australia etc. invested heavily in the regional health informatization centered on data sharing of electronic medical record (EMR) and electronic health record (EHR) on state and local level, like Regional Health Information organization, RHIO[2-3], Strategic Health Informatics Network in Europe, SHINE[4], National Switching Point[5], AMIA[6-7], and so on. Mainly driving by the demands in ethics, namely, maximizing the assurance in the medical quality and safety of residents, these initiatives attempts to promote the whole medical service quality and accessibility of medical services and reduce medical cost and risks.

Stanford translational research integrated database environment (STRIDE)[8], researched and developed by Stanford University for translational medical research project support, is a standard information integration model combining various medical information, such as geographic distribution, management autonomy, modal heterogeneous clinical data, molecular data, image data, and spatial data etc.. The 3 components in STRIDE are the HL7 RIM-based clinical data base (including the clinical data of 1,300,000 pediatric and adult patients), application development

framework for clinical data management, and the management system of biological specimen data. Wang et al. [9] proposed a database-based individualized biomedical information integration method in a research concerning translational medical data using the entity relationship model integrated by domain ontology.

As a major trend in information technology in recent years, the integration of heterogeneous information, the rapid development of cloud computing in particular, effectively address the problem mentioned above. Great progress has been made in heterogeneous information integration methods with the efforts from industrial and academic fields. The IT giants, such as Google[10-11], Facebook[12], Amazon [13-14], Yahoo[15], Microsoft[16] etc. have launched their own cloud computing platform for the integration of heterogeneous information.

### 3 The Whole Framework of Medical Cloud Storage

Referring to the framework of Google’s distributed storage system, the whole framework of medical cloud storage is established to realize the interconnection of the isolated information islands in medical information platform, as shown in Figure 1. However, the medical cloud storage framework is disparate with the framework of Google’s distributed storage system. The main difference is indicated as follows. While in the medical cloud storage information platform proposed in this study, case data are shared intensively through an unified entrance served for doctors and patients in a main server of medical information platform, while the image data sharing requires the authority of related hospital in case of further need after obtaining case data. This is because of that, image data, as the private data of related hospital, are dispersedly distributed relying on the isolated information islands.

Case data and image data are valuable for hospital and are unlikely to be shared completely freely. Thus the access authentication and flow charge of each isolated hospital information island in medical cloud computing need to be addressed. In this

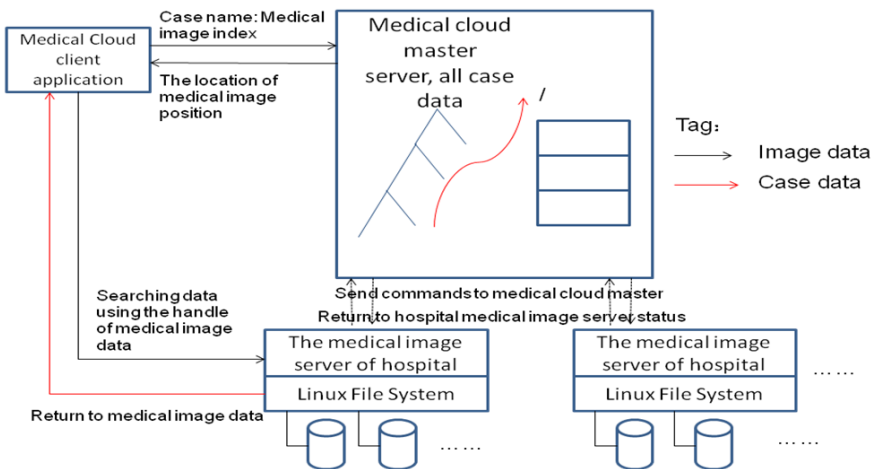
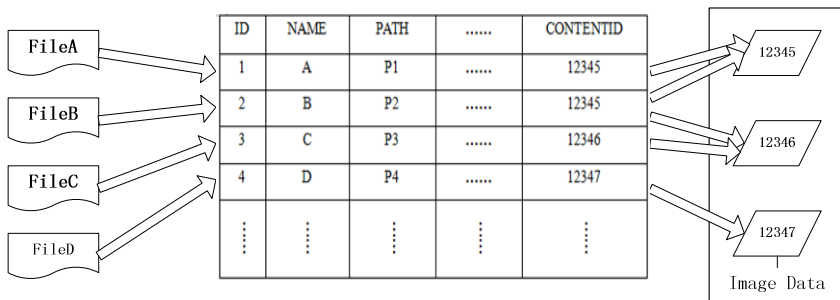


Fig. 1. The whole structure of medical cloud storage

study, they are developed from Google's distributed storage model in purpose of sharing case data and charging imaging data. This encourages hospitals' data sharing and ensuring the privacy of the private data of hospital.

#### 4 The Intensive Storage of Case Data and the Distributed Storage of Image Data

In this study, medical cloud storage employs case data and image data as metadata and data separately. Case data and image data are separately stored in specific warehouse-similar storage spaces in local server terminal in flattening state. The separating principle of case data and image data is depicted as follows: for the data to be stored in client terminal, the basic case data are segregated from inspection results, B ultrasound, CT imaging data. Case data are all stored in a specific space in server terminal, which is regarded as a "warehouse" containing piles of case data. Except a unique ID number, they have no other attributes. Image data are mapped with case data in the data base and searched from the "warehouse" via the case data in database in case of need.



**Fig. 2.** Storage Structure of Case Data and Image Data

The mechanism shows the following advantages: firstly, it reduces the data quantity in the "warehouse" of server terminal and saves the storage space in server terminal. For instance, for two cases with the same image description, the two copies of case data recorded in the database are mapped to one copy of image data in "warehouse". Thus inter-hospital data sharing is realized and repeat inspections are avoided; secondly, it decreases the time consumed in uploading the data from local service terminal to cloud platform and the cost of using cloud platform. This attributes to the reduction in the data amount that needed to be uploaded to cloud platform with the decreasing of the data amount in "warehouse".

As exhibited in Figure 3 (a), the client is presented in the form of the graphical interface, showing many functions including session authentication, user management, path management, content management, task, sharing, etc. Users can select and upload local files freely and the medical case data information of selected files can be

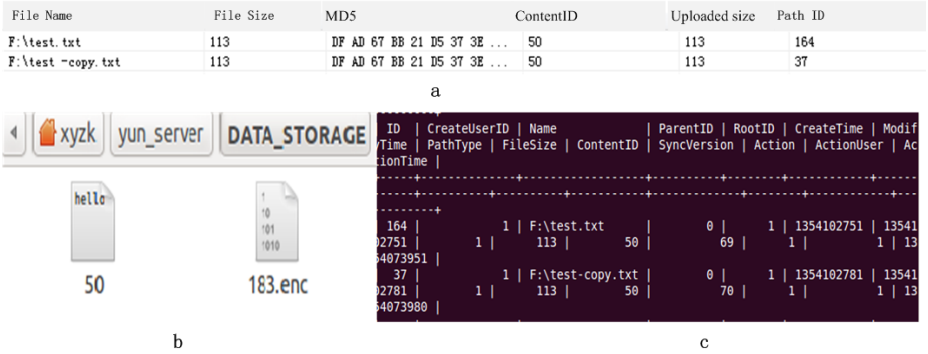


Fig. 3. The functional test for medical Data integration system

displayed on the interface of the client software. Of the medical image data contents saved in “storage”, only one content ID is of 182 attribute (Figure 3 (b)). The local server creates and saves medical case data for the files uploaded by the client (Figure 3 (c)). Thus, the separation of case data and image data is realized and the expected effect of data storage is achieved. Therefore, functionally, the system has achieved the expected goals and this provides the experimental platform for further researches in the future.

### 5 Conclusion

Hospitals are considered as isolated information islands. Medical cloud storage information platform construction, with the integration of heterogeneous medical data information at the core, really functions basing on fast, efficiently, and compatibly integrating the heterogeneous information sources of each hospital. The medical cloud storage information platform constructed finally is capable of realize the realizing a unified, convenient, integrated external service interface and real sharing of medical information.

This study proposed an improved GFS distributed data storage model to better the interconnection of the isolated information islands in medical cloud computing. The master medical cloud server applied allows intensive sharing of case data. In case of need after case data visit, dispersedly distributed images can be accessed through a unified entrance considering that the isolated hospital data island.

**Acknowledgment.** Funding for this research was provided in part by the Beijing Excellent Talent Training Project(2012D005007000009), General Program of Science and Technology Development Project of Beijing Municipal Education Commission(KM201110772014), and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research. We like to thank anonymous reviewers for their valuable comments.

## References

1. Haux, R.: Medical informatics: Past, present, future. *International Journal of Medical Informatics* 79(9), 599–610 (2010)
2. HIMSS RHIO/HIE[EB/OL]  
[http://www.himss.org/ASP/topics\\_rhio.asp](http://www.himss.org/ASP/topics_rhio.asp), 2011-03-15
3. White House Website. Transforming Health Care: the president's health information technology Plan [EB/OL],  
[http://www.starcareonline.com/Transform\\_healthcare\\_WhiteHousePaper.doe](http://www.starcareonline.com/Transform_healthcare_WhiteHousePaper.doe), 2011-03-10
4. KruitD, C.P.: SHINE: Strategic Health Informatics Networks for EuroPe. *ComPut Methods Programs Biomed.* 45(1), 155–158 (1994)
5. AHA. Fast Facts on US Hospitals [EB/OL] (November 7, 2008),  
[http://www.aha.org/aha/content/2008/pdf/fast\\_facts\\_2008.Pdf](http://www.aha.org/aha/content/2008/pdf/fast_facts_2008.Pdf)
6. Health Level Seven International. Health Level Seven Standards [EB/OL] (March 10, 2011), <http://www.hl7.org>
7. Kaplan, B.: Health IT success and failure: Recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association* 16(3), 291–299 (2009)
8. Lowe, H.J., Ferris, T.A., Hernandez, P.M.: STRIDE-An integrated standards-based translational research informatics platform. In: *Proceedings of the American Medical Informatics Association Annual Symposium, Washington USA*, pp. 391–395 (2009)
9. Wang, X., Liu, L., Fackenthal, J., Cummings, S.: Translational integrity and continuity: Personalized biomedical data integration. *Journal of Biomedical Informatics* 42(1), 100–112 (2009)
10. Ghemawat, S., Gobiuff, H., Leung, S.-T.: The Google File System. In: *The ACM Symposium on Operating Systems Principles, SOSP* (2003)
11. Melnik, S., Gubarev, A., Long, J.J.: Dremel: interactive analysis of web-scale datasets. *Communications of the ACM* (2011)
12. Finding a needle in Haystack: Facebook's photo storage. In: *OSDI 2010* (2010)
13. DeCandia, G., Hastorun, D., Jampani, M.: Dynamo: Amazon's highly available key-value store. In: *2007 SOSP* (2007)
14. Lakshman, A., Malik, P.: Cassandra: A Decentralized Structured Storage System. *SIGOPS Oper. Syst. Rev.* 44(2), 35–40 (2010)
15. Isard, M., Budiu, M., Yu, Y.: Dryad: Distributed Data-Parallel Programs from Sequential Building Blocks. In: *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007*, pp. 59–72 (2007)
16. Fetterly, D., Haridasan, M., Isard, M., Sundararaman, S.: TidyFS: A Simple and Small Distributed File System. In: *FAST* (2011)

# The Operation and Maintenance Management System of the Cloud Computing Data Center Based on ITIL

HongJun Zhan<sup>1</sup> and Wei Zhang<sup>2,3,\*</sup>

<sup>1</sup> Transportation Management College, Dalian Maritime University,  
Dalian City, Liaoning Province, 116026, China

<sup>2</sup> School of Computer Science,  
Beijing Information Science and Technology University,  
Beijing 100101, China

<sup>3</sup> Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,  
Beijing Information Science and Technology University, Beijing 100101, China  
zhanhongjun0612@163.com, zhwei@bistu.edu.cn

**Abstract.** The cloud computing data center mainly consists of numerous computers, network equipments, storage devices, power supply equipments, and many business systems serving for various departments in diverse purposes. The challenges encountered in the operation and maintenance management of IT (Information Technology) greatly restrain the reliability of cloud computing services. So, the idea of this paper was to realize the integrated operation and maintenance services characterized by the overall, comprehensive, standard and automated monitoring. That is, this research aimed to provide an integrated solution to improve the IT management level comprehensively and standardize IT environmental management. Also, the value and profit resulting from the IT resource input can be maximized.

**Keywords:** ITIL, Cloud computing, OAM management, Event management.

## 1 Introduction

In the Seventeenth Congress of the CPC, a new strategic thought was proposed: we should accelerate the development of industrialization, informatization, urbanization, internationalization, and marketization simultaneously; besides, IT application should be integrated with industrialization. As a result, informatization is dramatically stressed and obtains a new historical development opportunity. Currently, it has become an important issue in the informatization development process in China to achieve scientific and standardized management in the operation and maintenance (OAM) monitoring of informatization.

At present, there are a lot of components in the cloud computing data center, mainly including numerous computers, network equipments, storage devices, power supply equipments, and many business systems providing services for various

---

\* Corresponding author.



departments in diverse purposes. Based on the basic construction of informatization, the post-construction stage mainly focuses on large-scale applications, information security, and daily OAM. Consequently, the OAM challenges in informatization become increasingly severe and prominent. Hence, to ensure informatization to play its effective role in the development, we need to standardize the processes, work, management, and performance evaluation related to OAM. Thus, we are required to establish the service management system of the OAM processes and monitor OAM in an overall and comprehensive way. Moreover, various statistical forms must be provided; all the system monitoring results and statistical contents should be displayed fully and clearly in graphs. Finally, the comprehensive supervisory system of OAM is formed. Thus, automated management generalized monitoring, intelligent analysis and scientific decision-making can be realized. In this way, the business management level is radically improved.

## 2 Related Work

Numerous literatures allow us to have a better understanding of foreign famous products from IT outsourcing service providers and corresponding methodologies based on ITIL (Information Technology Infrastructure Library)[1]: such as the ITPM (Intelligent Thyristor Power Module)[2] model in IBM (International Business Machine), the ITSM (Information Technology Service Management)[2] model in HP (Hewlett Packard), the BSM (Business Service Management)[3] model in BMC (British Motor Corporation), the MOF (Microsoft Operations Framework)[4] model in Microsoft, the ESM (European Stability Mechanism)[5] methodology in CA, etc. Then, this paper introduced representative IT service management models and methodologies at home and abroad.

Various models for the IT service management at home and abroad are analyzed and compared. Generally, they present these main characteristics:

- Different organizations have a diverse understanding of the IT service management. As a consequence, the management realization mode and the corresponding degree show certain difference among various organizations. No uniform standard and evaluation system are formed.
- Theoretical models for foreign IT service providers are more advanced. The products are designed according to the standard ITIL processes. So, they are not fully consistent with the localized demand in China. Furthermore, foreign products are more expensive.
- Domestic IT service providers conduct few researches on theoretical models for IT service management. Instead, they pay more attention to implementation of technologies.
- There are few studies on IT service management models or methodologies for specific industries.

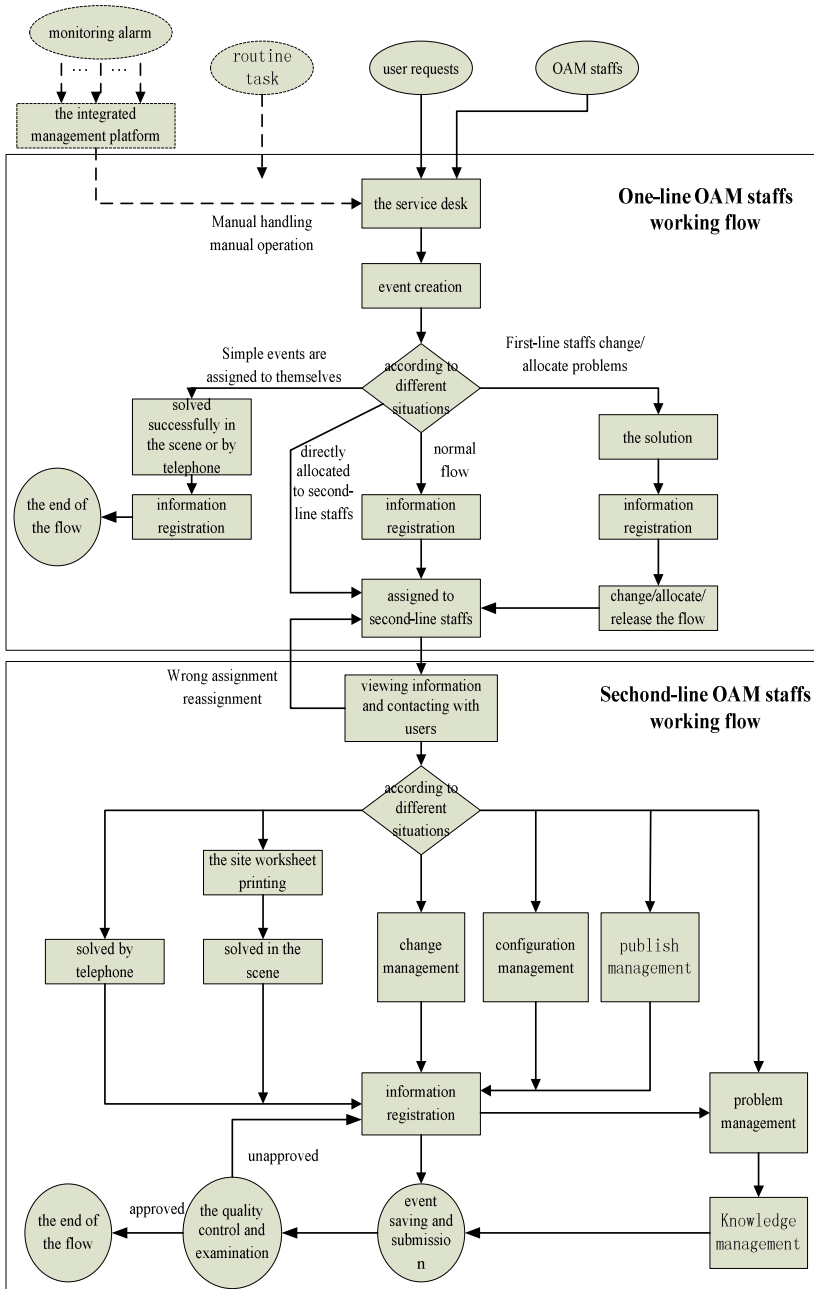


Fig. 1. The workflow of first-line and second-line OAM staffs

### **3 The OAM Management Model Based on ITIL**

The functional description: the events are started; event requires, such as user requires, OAM staffs' requests, alarm requests, should be created; the event list must be available. The functional description: all the created events are displayed and many management operations like searching and changing can be undertaken. The system should show the functions of event starting and event handling, specifically including the event addition, deletion, modification, assignment, enquiry, event processing, attachment download, and processing tracking.

Shown in figure 1, the system should show the function of event handling: specifically, unapproved event handling, event modification, event enquiry, event handling, the attachment download, and processing tracking, problem upgrade, new knowledge addition. Functional description: all the events that are assigned to second-line staffs should be displayed; management operations like handling, searching, and changing can be conducted.

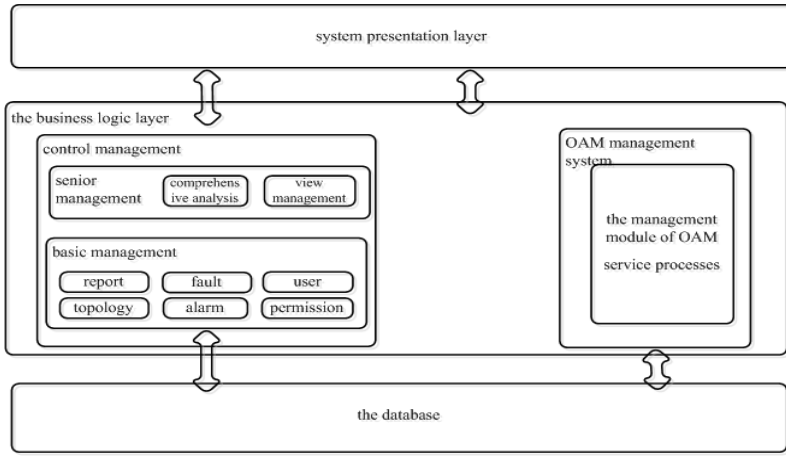
The system adopts industry-standard, uses the ITIL theory and best practices map out an overall plan from techniques, process and staff and construct the IT operation and maintenance systems. First of all, the paper introduces the needs of the business and analyses the functional requirements of the key aspects for operation and maintenance process management system. And from the operation and management, process and statements the point of view of business processing function demand and operation and maintenance of the business demand function are analyzed, and briefly introduces the system of the performance, the information security and system related non-functional requirements. Also, this system takes an event-oriented service process management as a main line, provides several roles such as front-line staff, second staff, problem administrators and knowledge administrators, and designs several business processes of different roles. Furthermore, this system designs the relevant database based on business process requirements and demonstrates the main interface after the realization of the system. The paper the detailed design events processing flow, the problem treatment process, the knowledge process, and so on. According to business process required event table, the event log table, watch, watch problems table, knowledge table are designed, and the paper shows the event processing, one-line-event-processing, second-line-event-processing, problem management processing, knowledge management processing, statistical analysis and the main interface.

### **4 Design and Realization of the System**

According to the standard in this industry together with the ITIL theory and the best practice, the complete OAM system for IT system monitoring is established, by taking technologies, processes and staffs into account comprehensively.

Considering complexity and diversity of the integrated business data platform, this system utilizes the mature and widely-used multi-tier architecture (J2EE) in the industry. As an enterprise-level middleware platform, J2EE connects together many

resources and applications that are scattered on the network. The framework has been designed as a 3-tiered architecture shown in Figure 2, including data collecting layer, business logic layer and data exhibition layer. And, the data collecting layer includes topology, configuration, performance, alarms, business acquisition and preprocessing; the business logic layer includes the management system and service management; the data exhibition layer realize the information dissemination, system monitoring view, the business monitoring view and reports through the operation and management portal.



**Fig. 2.** The overall technology architecture of the system

All the business functions that are required are included in the system and fully realized. The system shows stable functions and high operating efficiency. The interface test indicates that the system interface is satisfactory. Interaction with users can be undertaken successfully and a complete set of system operation and information prompts are also provided. In the test for various business processes, the system works well during the whole procedures including data input, saving, uploading, examination, and data statistics. In the process, the system shows high efficiency and stability. At the same time, the requirements for system reliability and security are satisfied. Figure 3 shows the interface image of the function creating event, one-line/second-line OAM staffs working flow step.

This paper shows the testing process and results. The system applies manual tests by integrating tests and applications. And this system tests the system functions by functional testing, interface testing and threshold testing, analyze the faults of the original system and revise and improve it. Testing and application results show that the system can meet the business needs. The implementation of the system realizes the whole operation and maintenance services to the "comprehensive, integrated, standard and automation control" for bank. It is to provide a set of overall solutions to improve the management level, regulate IT environment use, the maximum of the IT resources for the investment of the value and benefits generated.



Fig. 3. The user interface of the system

## 5 Conclusion

The aim of this paper is to manage OAM service processes of the overall IT equipments and business systems. The corresponding specific managements are presented as follows. The management system of asset and OAM process services should be established to realize the comprehensive OAM monitoring. Various statistical forms ought to be supplied and monitoring results and statistical contents of the integrated system are presented fully and clearly in graphs. Finally, the comprehensive OAM monitoring system is formed to improve IT management level radically and standardize IT environmental management. Moreover, the value and profit resulting from the IT resource input can also be maximized.

**Acknowledgment.** Funding for this research was provided in part by the Beijing Excellent Talent Training Project(2012D005007000009), General Program of Science and Technology Development Project of Beijing Municipal Education Commission(KM201110772014), and the Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research. We like to thank anonymous reviewers for their valuable comments.

## References

1. ITIL, <http://zh.wikipedia.org/zh-cn/ITIL>
2. IBM-tivoli, <http://www-01.ibm.com/software/tivoli/>
3. HP ITSM, <http://www8.hp.com/cn/zh/software-solutions>
4. BMC BSM, <http://www.bmc.com/solutions/bsm/business-service-management.html>
5. Microsoft MOF, <http://technet.microsoft.com/en-us/library/cc506049.aspx>
6. CA ESM, <http://www.ca.com/us/~media/files/brochures>
7. Silvius, A.J.G.: Business & IT Alignment in Theory and Practice. In: Proceedings of the IMB 2005 Conference, pp. 23–25 (February 2005)
8. Van Bon, J., et al.: Foundation of IT Service Management:base on ITIL, pp. 10–24. Van Haren Pulishing (2005)
9. Moura, A., Sauve, J., Jomada, J., et al.: A Quantitative Approaeh to IT Investment Allocation to Improve Business Results. In: Proceedings of the Seventh IEEE International Workshop Policies for Distributed Systems and Networks, pp. 87–95. IEEE ComPuter Society, Washington DC (2006)

# Underwater Object Detection by Combining the Spectral Residual and Three-Frame Algorithm

Zhe Chen, Huibin Wang\*, Jie Shen, and Xin Dong

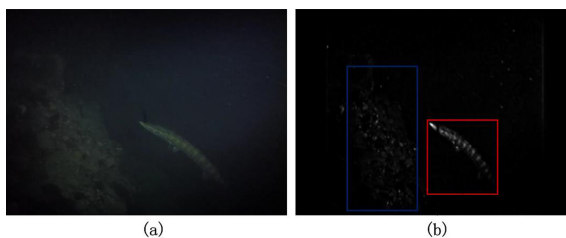
College of Computer and Information Engineering,  
Hohai University, Nanjing, 210098, China  
chenzhe@hhu.edu.cn

**Abstract.** In this paper, a hierarchical system, in which each level is composed by a specialized processor, is proposed to detect objects in underwater videos. The system is designed to assist the underwater monitor system survey operations, specialized to the task of object detection. The input image is firstly transformed into the spatial frequency domain to detect the saliency regions including the object region and parts of the background. Then, the saliency maps are further analyzed by the frame-difference algorithm which contributes to the background noise reduction. Experimental results, which have been performed on a set of real underwater images acquired in different environments, demonstrate the robustness and the accuracy of the proposed system.

**Keywords:** Underwater image, Object detection, Spectral residual, Three-frame difference.

## 1 Introduction

In the last years, several systems have been loaded to monitor the underwater world, such as the “eye-in-the-sea” project[1] and the system in the Hongkong ocean park[2].



**Fig. 1.** The saliency map extraction (a) Original underwater image, (b) Saliency map

In 2004, Walther[3] used the Itti model to generate the saliency maps for the underwater images. In 2006, Barat[4] employed the visual attention system to detect manufactured objects in the seabed. The main drawback of these methods is that they cannot best handle the image with inhomogeneous intensity. Fig.1 shows the result of

---

\* Corresponding author.

the saliency map extraction, we can find that parts of background regions (the blue rectangle) are mistaken as the interesting object (the red rectangle). To this problem, our solution is to perform the spectral residual method to enhance the correctness of the saliency map, and the frame-difference algorithm to reduce the background noise.

## 2 Underwater Object Detection

### 2.1 Spectral Residual Based Saliency Map Extraction

For the saliency map extraction, we use the spectral residual based method that has been proved to work well for a variety of applications[5, 6]. The discipline of the spectral residual method is that frequently occurring feature is the redundancy, while at the same time feature that deviate from the norm input frames are the valuable information. The image information  $I(image)$  can be decomposed into two parts:

$$I(Image) = I(Innovation) + I(Prior Knowledge) \tag{1}$$

Where,  $I(image)$  denotes the novelty part, and  $I(Prior Knowledge)$  is the redundancy. In the spectral residual method, the log spectrum representation is adopted to represent the image:  $L(f) = \log(A(f))$ . It is interesting that the log spectra of different images share similar trends. This indicates a local linearity in the averaged log spectrum. The prior knowledge  $I(Prior Knowledge)$  can be described as:

$$AV(f) = h_n(f) * L(f) \tag{2}$$

Where,  $h_n(f)$  is a mean filter. Therefore the spectral residual  $R(f)$  can be obtained by:  $R(f) = L(f) - AV(f)$ , and the  $I(Innovation)$  is given by:

$$I(Innovation) = g(x) * F^{-1}(\exp(R(f) + P(f)))^2 \tag{3}$$

Where,  $g(x)$  is the Gaussian function and  $P(f)$  is the phase spectrum of the image. While in the underwater environment, we find that the statistics on the underwater images also has invariant factors. Fig.2 shows the examples of the log spectra of the underwater images. We find that the log spectra of different underwater images share a similar trend. Hence, we have full confidence to generalize the spectral residual method into the underwater images.

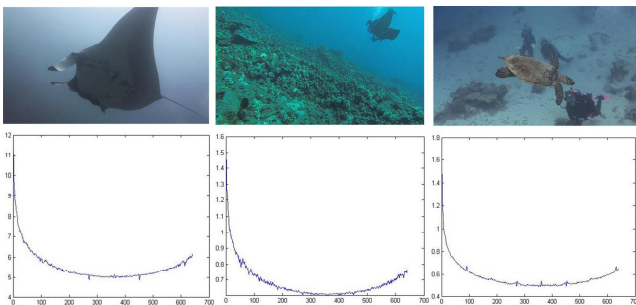


Fig. 2. Log spectra for underwater images

## 2.2 Three-Frame-Difference

Due to the inhomogeneous lumination effect in the underwater environments, the intensity distribution of the underwater images is commonly distorted. Hence, many background regions which are relatively brighter are often mistaken as the object saliency region. From the Fig.1 (b), we can find that besides the interesting fish, parts of the background region irradiated by the artificial source are marked in the saliency map as well. In order to solve this problem, the salient background region should be reduced in the saliency map. Here, this project is achieved by the three-frame-difference method[7]. Generally, the three-frame-difference method falls into the strategy of the optical flow field which detects the moving object by the spatial-temporal gradient of the image sequence, and extracts the moving target from background image by analyzing the change of the moving field. The advantage of the frame-difference method lies in its simplicity and feasibility in practice, while the poor anti-noise performance also blocks its applications. In order to enhance the robustness of the frame-difference method against the background noise, the three-frame-difference method is proposed. In the three-frame-difference algorithm successive pair of them is processed by operating the frame-difference calculation.

$$D_1(x, y) = \begin{cases} 1 & f(k) - f(k-1) > T \\ 0 & f(k) - f(k-1) < T \end{cases} \quad D_2(x, y) = \begin{cases} 1 & f(k-1) - f(k-2) > T \\ 0 & f(k-1) - f(k-2) < T \end{cases} \quad (4)$$

The final step is to make an and operations on the parameters  $D_1(x, y)$ ,  $D_2(x, y)$ , as

$$D(x, y) = \begin{cases} 1, & D_1(i, j) \cap D_2(i, j) = 1 \\ 0, & D_1(i, j) \cap D_2(i, j) = 0 \end{cases} \quad (5)$$

We notice that since the background region commonly keeps quasi-static, the three-frame-difference method may contribute to the background reduction for the saliency map.

## 2.3 Morphological Processing

In this paper, the mean filter is firstly used to smooth the saliency map. Then, the simple erosion and dilation methods are employed. Fig.3 shows the result given by the morphological processing.



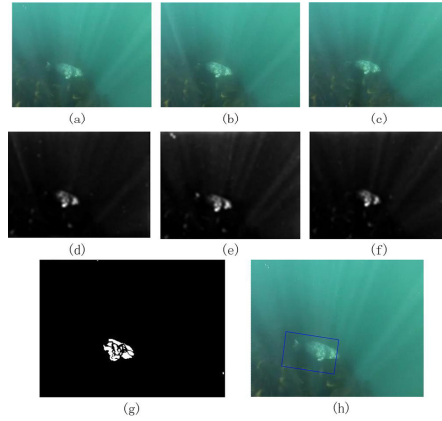
Fig. 3. Saliency map after the morphological processing



### 3 Experimental Results

#### 3.1 Underwater Object Detection Results

In the images (Fig.1(a)), the saliency mechanism identifies particles of marine snow, parts of the background and the animals as being the most salient ones (Fig.1(b)). Following with the three-frame-difference calculation, the false saliency regions belonging to the background and the optical marine snow are both reduced (Fig.3(a)).



**Fig. 4.** Object detection process (a)-(c) Frame 34, 35, 36, (d)-(f) Saliency map extracted by the spectral residual method, (g) Saliency map after three-frame-difference processing, (h) Final result

The saliency map given by the three-frame-difference method is further corrected and the object region is marked with the minimum enclosing rectangle (Fig.3(b)). Another experiment is shown in Fig.4 and similar results are given.

#### 3.2 Performance Evaluation

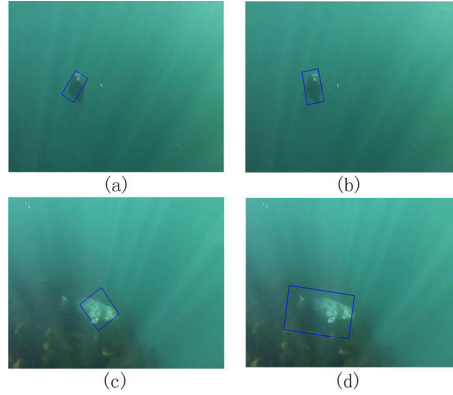
We compare the proposed method with the three-frame-difference method. The results given by the three-frame-difference method are shown in Fig.5. The criterion for expressing the segmentation quality is used to evaluate the performance:

$$C_{good} = \frac{card\{\Omega_{in} \cap \Omega_o\}}{card\{\Omega_o\}} \quad C_{false} = \frac{card\{\Omega_{in} \cap \Omega_b\}}{card\{\Omega_b\}} \quad (6)$$

Where,  $\Omega_{in}$  is the internal region of the extracted area,  $\Omega_o$  is the region of the object to be detected, and  $\Omega_b$  is the background region. Compared to the three-frame-difference method in the same experiments, the superior capability of the proposed method is demonstrated by the increased value of  $C_{good}$ .

**Table 1.** Performance evaluation

	Fig.10 (a) (b)		Fig.10 (c) (d)	
	$C_{good}$	$C_{false}$	$C_{good}$	$C_{false}$
Three-frame-difference	0.5253	0.2133	0.6786	0.1603
Proposed method	0.8021	0.1112	0.8477	0.0924



**Fig. 5.** Performance comparison, (a),(c) the results given by the three-frame-difference, (b),(d) the results given by our method

## 4 Conclusion

Combining the spectral residual and three-frame-difference methods, we have presented a novel hierarchical underwater object detection method. Compared to the traditional strategy, this approach's novelty and advantage lie in the adaptability to the challenging underwater inhomogeneous environments. The experimental results demonstrate the enhanced precision in underwater object detection of the proposed method over traditional ones.

## References

1. Kocak, D., Caimi, F.: The current art of underwater imaging with a glimpse of the past and vision of the future. *Marine Technology Society Journal* 39, 5–26 (2005)
2. Lam, K., et al.: Application of a real-time underwater surveillance camera in monitoring of fish assemblages on a shallow coral communities in a marine park. In: *OCEANS 2007*, pp. 1–7 (2007)
3. Walther, D., et al.: Detection and tracking of objects in underwater video. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004*, vol. 1, p–544 (2004)

4. Barat, C., Rendas, M.J.: A robust visual attention system for detecting manufactured objects in underwater video. In: OCEANS 2006, pp. 1–6 (2006)
5. Guo, C., et al.: Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8 (2008)
6. Cheng, M.-M., et al.: Global contrast based salient region detection. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 409–416 (2011)
7. Zhang, Y., et al.: Three-Frame Difference Algorithm Research Based on Mathematical Morphology. *Procedia Engineering* 29, 2705–2709 (2012)

# The Data Mining of Breast Cancer Based-on K-Means

Zhang Ruilan<sup>1,\*</sup> and Feng Zhixin<sup>2</sup>

<sup>1</sup> College of Basic Medicine Beihua University Jilin, Jilin, China

<sup>2</sup> College of Computer Science and Technology Jilin University Jilin, Jilin, China  
bhdxzr1@163.com, 1063475008@qq.com

**Abstract.** This paper presents a K-means method for the data mining of breast cancer. With the development of science and technology, Data mining has become a new method which is more simple and directly. This article gives a brief introduction about the approach and results of it. We conjecture that the K-means will make a big help to the Clinical diagnose.

**Keywords:** Data mining, Clinical diagnose, Clustering algorithm.

## 1 Introduction

Early detection and timely treatment are the primary problem of disease cure. Currently, because of the limitations of medical technology, earlier diagnosis still relies clinical experience, which bound to bring some misdiagnosis rate. With the development of computer technology, different forms of databases appear in different occasions. It draws many useful basis for clinical treatment through the continuous mining of these data. This paper set experimental results divided into two categories, namely benign cluster and malignant cluster by setting Wisconsin breast cancer data for analysis and using K-means to do some clustering operation. Compare the clustering results with the known experimental data, to arrive at the results.

This topic has important significance. Increasing the accuracy of early diagnosis of major diseases on the one hand can help to improve the chances of cure, on the other hand it can also avoid unnecessary waste of medical resources.

## 2 Breast Cancer and Medical Data Mining

Breast cancer's incidence was in the first place in the world medical records of female cancer. Especially in the past few decades, the incidence of breast cancer in developing countries is high. China, Africa is a relatively low incidence of breast cancer area. But in recent years, some reports show that the incidence had a slow upward trend, and an annual about 3-4% of the speed increase.

In some complex clinical disease (such as breast cancer), the timely diagnosis would help patients recover soon. Medical data mining can help doctors a lot in

---

\* Corresponding author.

diagnosing diseases. According to the association rules can do some data mining to the diseases which has some connection with such as occupation, age, gender, working conditions and living areas. Using the series model can find out the condition of patients and make predictions. Based on Rough theory of data mining technology for cancer and some traditional Chinese medicine such as rheumatoid diseases, greatly improved the accuracy of diagnosis. Rough set theory can also predict new diseases according to past cases of diagnostic rules. In this paper, it put K-means algorithm into Wisconsin breast cancer dataset, experimental methods and results are given to support data mining in disease diagnosis.

### 3 K-means Clustering Algorithm

#### 3.1 Algorithm Pseudo Code

Algorithm. K-means. For dividing on the K-means algorithm, the center of each cluster is used to represent the mean of all objects.

**Input:**

K: the number of clusters

D: contains n objects datasets.

**Output:**

A set of K clusters

**Method:**

- (1) Arbitrarily selected k object from D as initial cluster centers.
- (2) Repeat
- (3) Each object is assigned to the most similar cluster according to the initial value.
- (4) Update the cluster means. (recalculate the mean of each cluster object)
- (5) Until no change.

#### 3.2 K-means Algorithm Analysis

**Advantages.** K-means is more simple to prepare and execute faster than other algorithms. The complexity of the algorithm is  $O(nkt)$  (t is the number of iterations). The complexity of the PAM is  $O(k(n-k)^2)$ . And that of CLARA is  $O(ks^2+k(n-k))$ . When there is significant differences among the clusters and the results are dense, the K-means will give full play to its advantages.

**Shortcomings**

- (1) K-means only can be used when the average of the cluster is defined. It may not apply to some aspects such as those involving a categorical attribute data.
- (2) It need to pre-refer the number of cluster named K.
- (3) Can not deal with noise data and outliers.
- (4) Not suitable for discovering clusters with non-convex shape.
- (5) For different initial values may lead to different results. K-means has poor stability.

## 4 K-means on Wisconsin Breast Cancer

### 4.1 Purpose

Get the K-means applied to the breast cancer data to determine whether the method can diagnose disease.

### 4.2 Data Selection

Experimental data selection must first make sure the source of the data, to ensure data is true and reliable. This experimental data is Wisconsin breast cancer data set, it is real clinical data. Fig1 is a part of the data set. The data set 569 instances, each has 32 property. The first column A is the patient ID, the second column B is the diagnosis result(“M”indicates malignancy, “B”indicates benign).The remaining thirty attributes are divided into three groups, each group of ten, maximum, mean and variance. These 30attributes are about breast lumps cells on medical digital images. The ten properties are nuclear radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, fractal dimension.

#	A	B	C	D	E	F	G	H	I	J	K	L
1	842302	M	17.99	10.39	122.8	1001	0.1184	0.2776	0.3001	0.1471	0.2419	0.07871
2	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017	0.1812	0.05667
3	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	0.05999
4	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052	0.2597	0.09744
5	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043	0.1809	0.05883
6	843786	M	12.46	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089	0.2087	0.07613
7	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074	0.1794	0.05742
8	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985	0.2196	0.07451
9	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353	0.235	0.07389
10	84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543	0.203	0.08243
11	845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323	0.1528	0.05697
12	84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606	0.1842	0.06082
13	846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118	0.2397	0.078
14	846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364	0.1847	0.05338
15	84667401	M	13.73	22.61	93.6	578.3	0.1131	0.2298	0.2128	0.08025	0.2069	0.07682
16	84799002	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364	0.2303	0.07077
17	848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259	0.1586	0.05922
18	84862001	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028	0.2164	0.07356
19	849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498	0.1582	0.05395
20	8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781	0.1885	0.05766
21	8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311	0.1967	0.06811
22	8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076	0.1815	0.06905
23	8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756	0.2521	0.07032

Fig. 1. part of Wisconsin breast cancer data set

### 4.3 Data Preprocessing

First put the data set into MYSQL database. Compare them with the experimental requirements, eliminate the undesirable data. And get each property normalized or standardized to increase the accuracy of K-means.

## (1) Common normalization process

Change the number into a 0-1 of decimal, which means the ratio of each number of the sum of all of them.

Such as:  $a[4]=\{3,7,6,4\}$

Normalization process:  $3+7+6+4=20$ ,  $3 / 20 = 0.15$ ,  $7 / 20 = 0.35$ ,  $6 / 20 = 0.3$ ,  $4/20 = 0.2$ .

Result:  $b[4]=\{0.1,0.2,0.3,0.4\}$ (as method 1)

There is another common normalization formula:

The conversion formula of linear function(as method 2)

$$y = (x - \text{MinValue}) / (\text{MaxValue} - \text{MinValue})$$

Logarithmic conversion equation:  $y = \log_{10}(x)$

Anti-cotangent conversion equation:  $y = \text{atan}(x) * 2/\pi$ .

## (2) Standardization process

As the different units between different variables, the results obtained will be no useful significance after performing operations. For example: 1cm+3kg. It does not make any sense. In order to eliminate the influence of different units, we get data normalized.

Commonly standardized methods:

Standardization of dispersion:  $x'_{ik} = [x_{ik} - \text{Min}(x_k)] / R_k$  (Min( $x_k$ ) is the minimum,  $R_k$  is the range)

Standard deviation:  $x'_{ik} = [x_{ik} - \bar{x}] / S_k$  (as method 3) ( $S_k$  is deviation,  $\bar{x}$  is mean.)

This paper will use the above methods to deal with the data.

#### 4.4 Proper Processing

This paper uses K-means to classify the processed data. K is equal to 2 here, which means divided them into two categories, one is benign, the other is malignant. It will get the percentages compare the experimental results with the known B. We can determine whether the K-means can be used in the diagnosis of breast cancer. Specific process:

Step 1. Get the Wisconsin breast cancer data into databases. 32 properties represented by the letters A-Z, and AA-AF. A is patients' ID. B is whether the patient is sick. C-L is the maximum of the ten specific features of nucleus. M-V is the mean of the ten specific features. W-AF is the variance of the ten specific features.

Step 2. Get the data you need from the database.

Step 3. The C-L (or M-V, or W-AF) normalization or standardization.

Step 4.

Begin

Do K-means algorithm, in which  $k = 2$ , data is from the ten-dimensional array sets composed of C-L (or M-V, or W-AF).

End

Step 5. The results are divided into two clusters, one cluster is benign, the other is malignant. Get a percentage by comparing the result with the known B. Ends

## 5 The Experiment Results

Data mining is just a method of discovering some nature regular. It still needs validation the obtained experimental results by the clinical and experts in related fields. Data mining is hoped used to do some predict to clinical diagnosis.

**Table 1.** The processing results of original data

category \ times	First time	Second time	Third tiem
maximum	0.8348	0.8383	0.8348
mean	0.7557	0.7557	0.7557
variance	0.8541	0.8541	0.8541

**Table 2.** The processing results of percentage data

category \ method	Method 1	Method 2
maximum	0.8717	0.8981
mean	0.7645	0.7865
variance	0.9121/0.9156	0.9420

## 6 Conclusion

This paper applied data mining method to medical diagnostics, which has practical significance. It also provide much reference for doctors' clinical diagnosis. This article uses K-means to Wisconsin breast cancer data and gets satisfactory results. It is helpful for clinical diagnosis to get some information by using data mining.

## References

1. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn., pp. 225–243. Morgan Kaufmann Publishers, San Francisco (2006)
2. Biswas, S.K.D.R., Roy, A.R.: An Application of Intuitionistics Fuzzy Sets in Medical Diagnosis. Fuzzy Sets and Systems 117(2), 209–213 (2001)



3. Cabena, P., Hadjinian, P., et al.: *Discovering data mining: From concepts to implementation*, p. 195. Prentice Hall, New Jersey (1998)
4. Cios, K.J., Moore, G.W.: Uniqueness of medical data mining. *Artificial Intelligence in Medicine* 26(1-2), 1–24 (2002)
5. Kusiak, A., Kernstine, K.H.: Data Mining: Medical and Engineering Case Studies. In: *Industrial Engineering Research Conference*, Cleveland, pp. 1–7 (2000)
6. Parpinelli, R.S., Lopes, H.S., Freitas, A.A.: An Ant Colony Based System for Data Mining: Applications to Medical Data. In: *Proceedings of Genetic and Evolutionary Computation Conference*, pp. 791–797. Morgan Kaufmann, San Francisco (2001)
7. Siromoney, A., Raghuram, L., Siromoney, A., et al.: Inductive logic programming for knowledge discovery from MRI data. *IEEE Engineering in Medicine Biology* 19(4), 72277 (2000)

# Compositional Modeling for Underwater Warfare Simulation Based on HLA and Flames

Hongtao Liang<sup>\*</sup>, Fengju Kang, and Yanyang Zeng

College of Marine Engineering of Northwestern Polytechnical University  
127 Youyixi Road, Xi'an, 710072, Shaanxi, China  
{lianghongtao.789, zyyhost}@163.com, 834428927@qq.com

**Abstract.** To solve the problem of compositional simulation for Underwater Warfare Simulation System (UWSS), a novel approach based on High Level Architecture (HLA) and Flexible analysis modeling and exercise system (Flames) was presented, which was compositional and reusable. Orthogonal relations between the level of UWSS and hierarchy of composability was established; then on the basis of HLA and Flames, the compositional simulation system framework of UWSS was designed, which can build and adjust compositional models, meanwhile carry out evaluation by a variety of tactical indicators for process performance. The practical application results show that the system meets the demands of various underwater conflicts and has good performance.

**Keywords:** Compositional modeling, Underwater Warfare, HLA, Flames.

## 1 Introduction

Modeling and Simulation (M&S) technology has been used for complex military system design, development and integration, which plays an extremely important role in the whole life cycle of the weapon system. In the research of combat system analysis and weaponry argument, because of the behavioral logic complex and dynamic evolution characteristics, the combination of different resolution models has become a very acute problem. A new method is urgently needed to solve the problem in order to cope with the continuous evolution of complex military system.

Compositional simulation was presented under the background of the above demand as a kind of new thought and method for the development of simulation technology, some studies has been carried out on compositional Modeling. Domain ontology driven technology was introduced in the compositional simulation framework in [1]; [2] showed the need for better conceptual alignment in compositional simulation, which included to make sure that components were conceptually composable as well; [3] mainly took into account compositional underwater counterwork visual systems, which was customizable and reusable; Compositional modeling of electronic warfare conceptual framework was studied in [4], in which both the Flames flexible simulation

---

<sup>\*</sup> Corresponding author.

framework and Association Assumes was analyzed and introduced to establish simulation system.

The above works in compositional Modeling were only focus on the research of the analysis with compositional framework and simple compositional modeling etc. which can not meet the need of increasingly complex underwater warfare. In this paper, a new design method based on HLA and Flames is put forward, in which both the underwater war and compositional modeling are studied.

## **2 Compositional Modeling for Underwater Warfare**

### **2.1 The Concept of Compositional Simulation**

Compositional M&S emphasizes the maximum reuse of existing simulation models, in order to quickly build and assemble target simulation system. The composability is the core issue of compositional simulation. Professor Petty gave the influential definition of compositional simulation which referred to the ability "the simulation components are chosen by compositional mechanisms to meet specific application requirements of simulation system"[5]. This definition highlighted the several aspects:

- Simulation component can be stored in the component library, appropriate components is selected to build a simulation system under certain conditions.
- Simulation component can be assembled into a simulation system, and this assembly capacity is focus of the study, namely composability.
- Component composability's goal is to meet the needs of users, so the selection of components and assembly depended on demand.

### **2.2 The Level of UWSS and the Hierarchy of Composability**

Underwater warfare [6] generally includes anti-submarine counterwork, submarine warfare, mine warfare, anti-submarine counterwork, aquatic confrontation and underwater information transmission, etc. In this paper, the UWSS is abstracted into three levels [7]: the Simulation Component Level (SCL), the Simulation Entity Level (SEL) and the Simulation System Level (SSL). SCL constitutes the basis of the simulation model, which has reusable nature; SEL has certain autonomy in the simulation system that has been identified as the logical behavior or physical behavior; SSL refers to collection of entities that can perform a simulation task. Therefore, multi-resolution and compositional system is created in vertical direction, as shown in Fig.1.

By reference to the concept of interoperability hierarchical model [8], composability is divided into the Technical Compositional Hierarchy (TECH), Syntactic Compositional Hierarchy (SYCH), Semantic Compositional Hierarchy (SECH). TECH can only require components to interconnect, which belongs to the lowest level in the hierarchy (such as HLA); SYCH is the basis of composability. Namely, components can communicate with each other through certain data protocol specification, correctness of SYCH should meet the relationship of the input and output interfaces between components; SECH is an upgrade based on SYCH, consistency of the data between components is needed to understand by the components.

The levels of UWSS and the hierarchies of composability are orthogonal [9], as shown in Fig.2. For each of the vertical and horizontal level of UWSS, compositional correctness and validity should be solved. Namely, the UWSS needs to meet TECH, SYCH and SECH so that the final system can be effective.

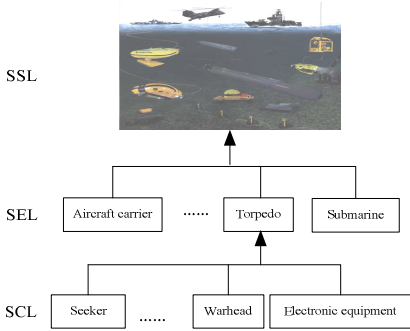


Fig. 1. Level of UWSS

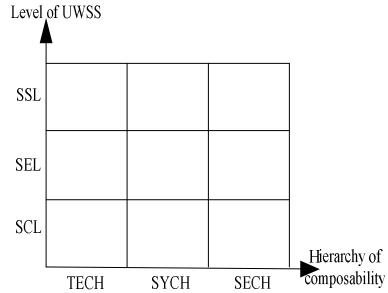


Fig. 2. Orthogonal relation

### 3 System Framework of UWSS

The compositional design based on HLA and FLAMES for UWSS is shown in Fig.3. The Federate was designed according demand, the pRTI1.3 was adopted as network support environment, Scenario Generation and Federal Management (SG&FM) was responsible for the entire underwater counterwork process and the entire federal process. Evaluation system primarily used Monte-Carlo analysis method to implement probability statistics including Miss Distance, Detection Probability and Hit Probability, etc, which can provide the credibility of simulation results.

The Flames [10] was an open architecture and cross-platform simulation framework software. The Flames mainly included standard application components and Flames kernel, which were used to develop environment for M&S and provide scenario model, more detailed content in [11]. In view of the characteristics of the Flames, it was used as a simulation framework combination with pRTI1.3 to provide M&S resources. A scenario process was divided into platforms, sensors, information fusion, weapons, platforms and ammunition interaction by Flames.

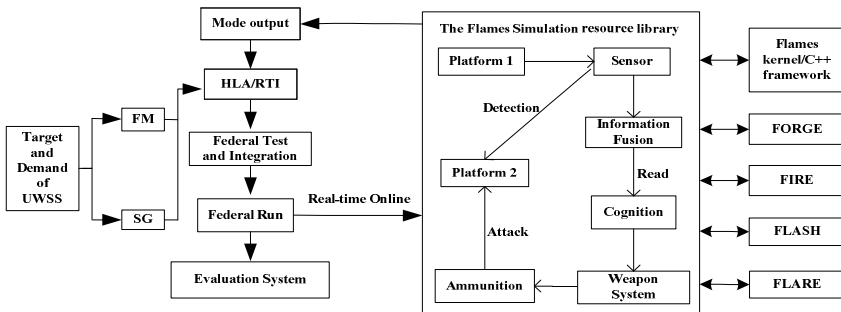


Fig. 3. Design of compositional UWSS

## 4 Compositional Method Based on Flames

The main elements of compositional modeling which is based on Flames are introduced as follows[11]: Model component development is mainly to develop model source code by Flames kernel and to form model component fragments in C++ environment, which can develop Model component library; then based on assumptions associated method, selection of model and the initialization of the parameters of the model component are completed; Thirdly, in the editing interface of component, according to compositional principle of low-SYCH, to initialize the component performance parameters, a certain functional component unit can be formed; the last in the script editing interface, via scripting language to form a entity model, which has all the functions and features. The model link module can complete the model configuration and the test of combinations as well as error information that is adjusted by visual and script editing. Therefore, this compositional modeling method can be divided into component development model, the model component editing, model combination, script analysis and model testing phases, which are shown in Fig.4.

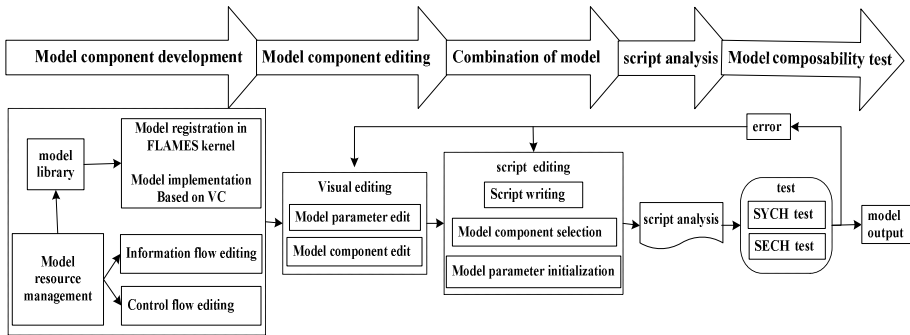


Fig. 4. Compositional modeling process of Flames-based

Qualifying simulation components are searched and corrected online by the above compositional method.

## 5 Typical Application

According to typical scenario of UWSS, the system has a surface ship, an anti-submarine warfare helicopter and two submarines. Through researching this particular application, the important experience will be provided for how to build other compositional UWSS.

### 5.1 The Scenarios of Typical Underwater Warfare

Assume that the involved entities in UWSS are as follows. Red: submarine, decoy. Blue: warship, anti-submarine helicopter, torpedo, airdrop torpedo, sonar. Initial

battlefield situation: The red side and the blue side are set in a sea battle. At the beginning of the simulation, two submarines are cruising at low speed in the same direction; the Surface ship are patrolling and searching submarine with anti-submarine helicopters, and torpedo and airdrop torpedo will be prepared to attack submarine.

According to framework of the HLA and Flames working mechanism, all federal members are added to the HLA/pRTI1.3. Federation members send real-time online data to the Flames for the display of the simulation process.

### 5.2 Situation Display and Analysis

According to scenario, the simulation system performs the initialization to complete the assembly different granularity models. Two sides use different ways at different stages to attack or defense each other. Flames simulation trend is shown in Fig.5-8.

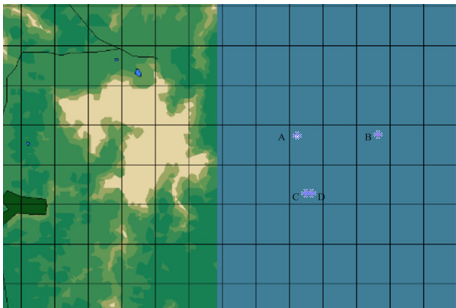


Fig. 5. System initialization

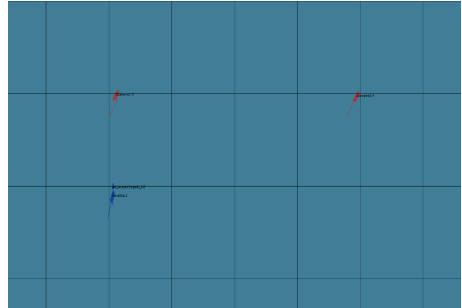


Fig. 6. Beginning stage

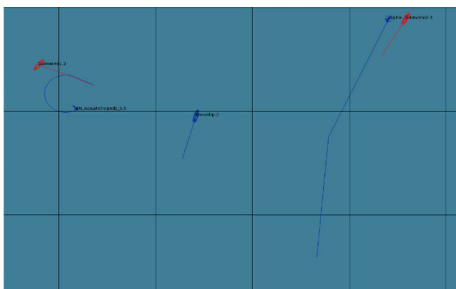


Fig. 7. Detection stage

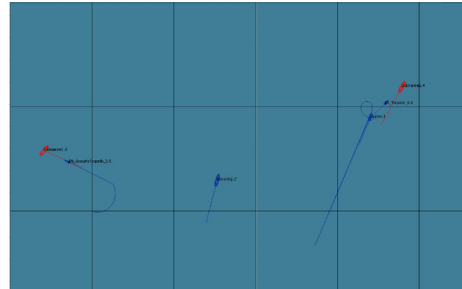


Fig. 8. Attack stage

In Fig.5, A and B are the submarines, C and D are warship and anti-submarine helicopter. In Fig.6, Shipboard sonar is detecting enemy submarine, dipping sonar of anti-submarine helicopters is detecting the other. In Fig.7, according to the sonar information fusion results, the surface is launching torpedo to attack submarine A, however submarine A is launching a decoy jamming to confront torpedo. Anti-submarine helicopter is hanging dipping sonar to continually detect submarine B. In Fig.8,

torpedo is attacking submarine A, anti-submarine helicopters airdrop torpedo is attacking submarine B. there are 200 trials by Monte-Carlo with 95% confidence level, statistical results of different counterwork ways are shown in Table 1.

**Table 1.** Statistical Result

counterwork way	Miss Distance	Detection Probability	Hit Probability
ship-to-submarine	4.89m	0.78	0.67
helicopter-to-submarine	3.25m	0.92	0.75

The compositional UWSS has been successfully applied to a virtual underwater combat with good performance.

## 6 Conclusion

The large and complex underwater warfare simulation research is very important. This paper aims at the issues including poor composability and reusability as well as insufficient strain capacity of combat style in UWSS. A novel approach based on HLA and Flames is presented to solve the problems, which can be used for establishing different granularity models and assessing the implementation of battlefield simulation. The compositional and reusable UWSS is a powerful simulation tool for the study of underwater warfare, which can provide a new perspective for the large-scale joint modeling and simulation.

## References

1. Chen, X., Duan, F., Zhang, Y.C.: Domain ontology driven composable simulation framework. *Journal of Harbin Institute of Technology* 9, 49–54 (2012)
2. Davis, P.K., Tolk, A.: Observations on New Developments in Composability and Multi-Resolution Modeling. In: *Proceedings of the Winter Simulation Conference WSC 2007*, Washington, DC, pp. 859–870 (December 2007)
3. Zeng, Y.Y., Kang, F.J., Xu, P., Li, X.J.: Modeling and real-time scene rendering of underwater warfare visual simulation system. *Systems Engineering and Electronics* 35, 436–440 (2013)
4. Wu, Z.J., Fang, S.L., Zhu, L.: Combined of Electronic Warfare Simulation Model Based on FLAMES. *Shipboard Electronic Countermeasure* 35, 55–60 (2012)
5. Petty, M.D., Weisel, E.W.: A Composability Lexicon. In: *Spring Simulation Interoperability Workshop*, FL: 03S-SIW-023, Kessimmee (2003)
6. Pan, G.H.: *WarShip Integrated Command and Control System Principles*. Northwestern Polytechnical University Press, Xi'an (2010)
7. Bartholet, R.G., Brogan, D.C., Reynolds Jr., P.F., Carnahan, J.C.: In search of the philosopher's stone: Simulation composability versus component-based software design. In: *Proceedings of the 2004 Fall Simulation Interoperability Workshop* (2004)
8. Tolk, A., Muguira, J.A.: The Levels of Conceptual Interoperability Model. In: *Fall Simulation Interoperability Workshop*, FL: 03F-SIW-007, Orlando (2003)

9. Wang, W.G., Wang, W., Zhu, Y., Li, Q.: Service-oriented simulation framework: An overview and unifying methodology. *Simulation* 87, 221–252 (2011)
10. FLAMES Simulation Framework-Ternion Corporation, <http://www.ternion.com>
11. Hu, G.S., Kong, X.H., Wang, R.H.: Design and Implementation of Battlefield Target Simulation Based on FLAMES. *Journal of System Simulation* 22, 2502–2510 (2010)



# Spot Electricity Price Dynamics of Indian Electricity Market

G.P. Girish\* and S. Vijayalakshmi

Department of Finance, IBS Hyderabad, IFHE University,  
Andhra Pradesh, India  
gpgirish.ibs@gmail.com,  
vijayas@ibsindia.org

**Abstract.** Spot Electricity price in a competitive electricity market is determined by the intersection of total demand curve (constructed from aggregated demand bids) and total supply curve (constructed from aggregated supply bids) for a particular hour for each region of the electricity market as bidded in a power/energy exchange. Spot electricity price curve usually exhibit characteristics such as multiple seasonality, volatility, mean reversion and are often characterized by jumps or spikes. In this study we review short term spot electricity price modeling and forecasting techniques inspired by financial econometric literature, electricity spot price pre-processing techniques used and the determinants of spot electricity price in a competitive power market. We also investigate the dynamics of spot electricity prices of Indian Electricity market which has never been done before. The results of the study provide crucial insights for pricing electricity derivatives which will be introduced shortly as announced by Indian Energy Exchange in Competitive Indian power market.

**Keywords:** Spot Electricity Price, Indian Electricity Market, Energy Exchange.

## 1 Introduction

Electricity, when viewed from an economic perspective, is a non-storable good and is the most important man-made commodity till date. Spot Electricity price in a competitive electricity market is determined by the intersection of total demand curve (constructed from aggregated demand bids) and total supply curve (constructed from aggregated supply bids) for a particular hour for each region of the electricity market as bidded in a power/energy exchange. Real time balancing requirements of electric power supply and demand translates itself into seasonal behavior of electricity spot prices in competitive electricity markets thereby making balancing of demand and supply a critical task and hanging at a knife-edge. Minor fluctuation or changes in the amount of electric power generated or demand change/shift can result into large changes in spot electricity prices within few hours in competitive power markets.

---

\* Corresponding author.

Spot electricity market price series usually exhibit certain characteristic behavior such as Seasonality (at daily, monthly, yearly and weekly level), Volatility in spot prices, Mean-Reversion and Jumps or Spikes in electricity spot price due to supply shocks or physical constraints of the power system (see Bierbrauer et al., 2007; Karakatsani and Bunn, 2008; Girish et al., 2013). In this study we review Spot Electricity Price Dynamics of Indian Electricity Market, short term spot electricity price modeling and forecasting techniques, electricity spot price pre-processing techniques and determinants of spot electricity price in a competitive power market. The results of the study provides crucial insights for pricing electricity derivatives which will be introduced shortly as announced by Indian Energy Exchange in Competitive Indian power market. The rest of the paper is structured as follows: We introduce Indian Power market in Section 2. In Section 3 we review modeling and forecasting techniques, electricity spot price pre-processing techniques and the determinants of spot electricity price. We empirically investigate the dynamics of spot electricity prices of Indian electricity market in Section 4 and conclude in Section 5.

## 2 Indian Power Market

The Indian power market is divided into five regions namely Eastern region, Northern region, Southern region, North eastern region and Western region (as seen in Fig. 1). [Girish et al. 2013]



**Fig. 1.** Indian Electricity Market

The Ministry of Power, Government of India, looks after the planning and policy formulation for electricity sector in India and is completely responsible for appropriate administration of the Indian Electricity Act 2003, which has re-organized the electricity sector in India and has introduced competition in power generation, distribution and trading of electricity.

Power Grid Corporation of India Limited (PGCIL) looks after the power system Grid Management and it is the Central Transmission Utility. National Load Dispatch Centre (NLDC), the apex body, was constituted by Ministry of Power in 2005 for ensuring integrated operation of the national power system, supervising

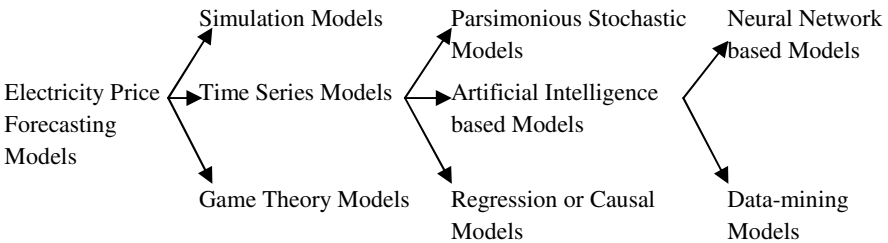
other Regional Load Dispatch Centres (RLDC) for scheduling and dispatch of electricity over the inter-regional links and monitoring operations and grid security.

Indian Electricity Act 2003 has paved way for Electricity trading as a distinct activity with licensing. Electricity Companies which are involved in the distribution of electric power and trading of electricity need to obtain license from the Central Electricity Regulatory Commission (CERC), an independent regulator. CERC is responsible for promotion and development of power market including trading. Indian Electricity Act 2003 has allowed 100% Foreign Direct Investment in power generation, power distribution and electricity trading in India.

India has two energy/power exchanges namely the Indian Energy Exchange (IEX) and the Power Exchange India Limited (PXIL). IEX began its operation on 27<sup>th</sup> June 2008 and has emerged as the most preferred electricity trading platform nationwide covering 80 Members, more than 1600 clients, over 350 Independent/private power generators and more than 1000 direct consumers registered (as on March 31, 2012). IEX is India’s first and leading energy exchange with 92% market share based on Electricity volumes traded for the financial year 2012-13. Products and Services offered by IEX can be divided under Day-Ahead Market (the hourly contracts for day-ahead using double-sided closed auction, the contingency hourly market for Next day having Continuous Trading and the Intra-day continuous Trading for the same day which are all categorized under Spot markets), Term-Ahead Market and Renewable Energy Certificates (including both Solar and Non-Solar).

### 3 Literature Review

Aggarwal et al. (2009) have classified electricity price forecasting techniques for short term into three categories namely simulation models, game theory models and the time Series models.



Source: Aggarwal et. al (2009).

It has been observed that spot electricity prices are pre-processed for modeling and forecasting by a) Using a sinusoidal for deterministic seasonal component (see Keles et al., 2012) b) Using constant piece-wise function for deterministic seasonal component (see Higgs and Worthington, 2008) c) By using a combination of Sinusoidal and Constant piece wise function (see Bierbrauer et al., 2007). d) Fourier analysis and wavelet decomposition (see Janczura and Weron, 2010).

Girish et al. (2013) classify spot electricity price forecasting models based on the financial econometric method used for modeling as shown in table 1. These models have been empirically investigated for forecasting of spot electricity prices for different competitive power markets having energy/power exchanges.

**Table 1.** Models used in finance literature for forecasting spot electricity prices

S.I No	Model Used	Authors
1	Autoregressive Models	Cuaresma et al. (2004)
2	ARIMA Models	Contreras et al. (2003)
3	Dynamic Regression Models and	Karakatsani and Bunn (2008)
4	GARCH Models	Mugele et al. (2005)
5	Jump Diffusion Models	Knittel and Roberts (2005)
6	Regime Switching Models	Weron et al. (2004)
7	Multiple Linear Regression Models	Schmutz and Elkuch (2004)

Source: Girish et al. (2013).

## 4 Empirical Analysis

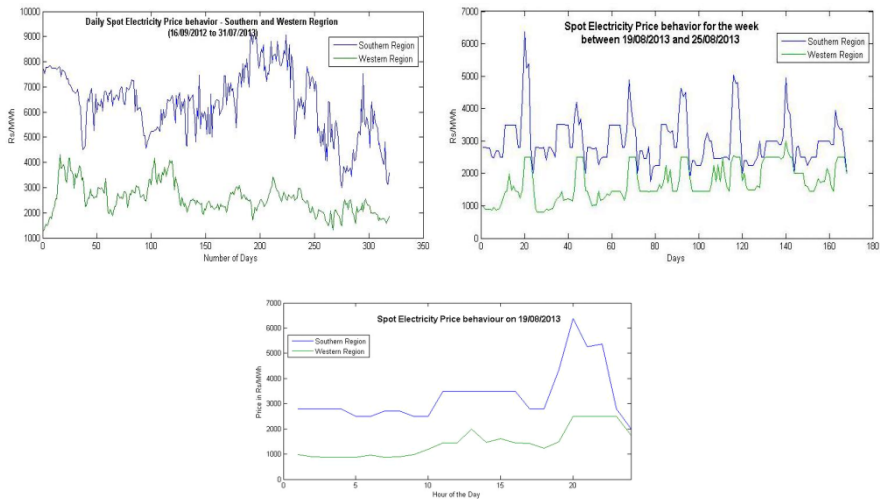
In our study, we use average daily spot electricity prices from the Indian Energy Exchange (IEX) day-ahead market which operates 365 days i.e. 24\*7 throughout the year. The data comprises of average daily spot electricity prices from September 16, 2012 (the date when IEX announced introduction of electricity derivatives in Indian power market) to July 31, 2013 totaling to 319 observations. The descriptive statistics of spot electricity prices is given in table 2.

**Table 2.** Descriptive Statistics of Spot Electricity Price for different regions of Indian Electricity Market

	North East	East	North	South	West
<b>Mean</b>	2744.86	2422.36	2773.64	6342.87	2556.45
<b>Standard Devia-</b>	804.57	667.26	692.83	1298.49	601.75
<b>Kurtosis</b>	2.13	0.39	-0.27	-0.28	0.10
<b>Skewness</b>	1.06	0.42	0.58	-0.27	0.61
<b>Range</b>	4878.79	3852.29	3243.55	6155.08	3034.09
<b>Minimum</b>	1290.57	472.37	1290.57	3000.68	1290.57
<b>Maximum</b>	6169.36	4324.66	4534.12	9155.75	4324.66

The average daily spot electricity price is almost similar for all the regions except Southern region and it ranges from 2744.86 Rs/MWh to 6342.87 Rs/MWh with the highest being in Southern region. Volatility of Spot prices given by Standard deviation is highest for Southern region in absolute spot electricity price terms and lowest for western region. Range of spot electricity price is observed to be highest for

Southern region and lowest for western region. Skewness is a measure of the asymmetry of a distribution. Spot electricity price series for all other regions except Southern region is observed to be positively skewed for the time period considered. Kurtosis is a measure of the extent to which observations cluster around a central point. It is found that spot electricity prices for all regions are platykurtic, dispersed around average mean values with Northern and Southern regions having negative values. The dynamics of the spot electricity prices is as observed in Fig. 2.



**Fig. 2.** Dynamics of Spot Electricity Prices of Indian Electricity Market

Since Spot price series of all other regions is similar (for northern, north eastern, western and eastern) except Southern region, we have considered the case of western and southern region of Indian electricity market as shown in Fig 2. We find mean reversion for Indian electricity spot prices and observe Seasonality at hourly and daily level which could be attributed to seasonal and cyclical demand of electricity by commercial, industrial and household consumers. The spot prices usually tend to peak in the evening/night for all the markets. Prices of spot electricity market are usually higher during weekdays compared to weekends. In instances of increased demand, power generating stations having higher marginal costs will be used (ex: using Oil, diesel) thereby increasing spot electricity prices. But when this demand returns to normal condition, these expensive generating stations will not be required to meet extra demand and are turned off thereby making spot electricity prices to revert back to its mean value. We find that spot electricity prices sometimes show large infrequent spikes or jumps which could be attributed to non-storability of electricity (economically) implying that inventories have no role in smoothening spot electricity price changes. Spikes or Jumps could also be accounted by generation capacity constraint, transmission problems (physical infrastructure limitations) and real time supply-demand balancing.

## 5 Conclusion

Electricity is the most important man-made commodity which is difficult to store economically and the fact that it has to be consumed whenever it is produced makes real time balancing of demand and supply a critical task hanging at a knife-edge. In this study we have reviewed spot electricity price dynamics of Indian Electricity Market in terms of seasonality, volatility, mean-reversion, jumps/spikes. We have introduced Indian Energy Exchange, reviewed short-term spot electricity price modeling and forecasting techniques, electricity spot price pre-processing techniques and determinants of spot electricity price in a competitive power market. The results of the study provides crucial insights for pricing electricity derivatives which will be introduced shortly as announced by Indian Energy Exchange (on 16<sup>th</sup> Sept 2012) in Competitive Indian power market.

## References

1. Aggarwal, S.K., Saini, L.M., Kumar, A.: Electricity price forecasting in deregulated markets: A review and evaluation. *Electrical Power and Energy Systems* 31, 13–22 (2009)
2. Bierbrauer, M., Menn, C., Rachev, S.T., Trueck, S.: Spot and derivative pricing in the EEX power market. *Journal of Banking and Finance* 31, 3462–3485 (2007)
3. Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J.: ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems* 18, 1014–1020 (2003)
4. Cuaresma, J.C., Hlouskova, J., Kossmeier, S., Obersteiner, M.: Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy* 77, 87–106 (2004)
5. Girish, G.P., Panda, A.K., Rath, B.N.: Indian Electricity Market. *Global Business and Economics Anthology* 1, 180–191 (2013)
6. Higgs, H., Worthington, A.: Stochastic price modeling of high volatility, mean reverting, spike-prone commodities: the Australian wholesale spot electricity market. *Energy Economics* 30, 3172–3185 (2008)
7. Janczura, J., Weron, R.: An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics* 32, 1059–1073 (2010)
8. Karakatsani, N.V., Bunn, D.W.: Forecasting Electricity Prices: The Impact of Fundamentals and Time-Varying Coefficients. *International Journal of Forecasting* 24, 764–785 (2008)
9. Keles, D., Hartel, R., Möst, D., Fichtner, W.: Compressed-air energy storage power plant investments under uncertain electricity prices: an evaluation of compressed-air energy storage plants in liberalized energy markets. *Journal of Energy Markets* 5, 53–84 (2012)
10. Knittel, C.R., Roberts, M.R.: An empirical examination of restructured electricity prices. *Energy Economics* 27, 791–817 (2005)
11. Mugele, C., Rachev, S.T., Trueck, S.: Stable Modeling of Different European Power Markets. *Investment Management and Financial Innovations* 3, 65–85 (2005)

12. Schmutz, A., Elkuch, P.: Electricity Price Forecasting: Application and Experience in the European Power Markets. In: Proceedings of the 6th IAEE European Conference, Zurich (2004)
13. Weron, R., Bierbrauer, M., Truck, S.: Modeling Electricity Prices: Jump Diffusion and Regime Switching. *Physica*, 39–48 (2004)

# Guaranteed QoS for UDP and TCP Flows to Measure Throughput in VANETs

Abubakar Aminu Mu'azu<sup>\*</sup>, Low Tang Jung,  
Halabi Hasbullah, Ibrahim A. Lawal, and Peer Azmat Shah

Department of Computer & Information Sciences,  
Universiti Teknologi PETRONAS  
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia  
{abuaminu\_g01797, ibrahim\_g01867}@utp.edu.my,  
{lowtanjung, halabi}@petronas.com.my,  
peer.azmat@yahoo.co.uk

**Abstract.** The level of Quality of Service (QoS) guarantees in vehicular ad hoc networks (VANETs) is much more tasking and challenging as result of rapid topology changing and high mobility of mobile hosts. Thus, making multi-hop communication and as well as contention for channel access more difficult. QoS in VANETs is measured in terms of throughput, connection duration and packet loss. In this paper, UDP and TCP protocols are used as traffic to satisfy bandwidth requirement while optimizing network throughput for providing QoS using clustering approach. The results obtained through NCTUns simulator are used for analysis of throughput for both UDP and TCP traffics.

**Keywords:** QoS Guarantee, UDP/TCP traffic classes, throughput, VANETs.

## 1 Introduction

VANETs are undoubtedly the most favorite network model for Intelligent Transportation Systems (ITS). VANETs are dependent on short-range wireless communication (e.g., IEEE 802.11) among vehicles. The Dedicated Short Range Communication (DSRC) was developed as WLAN standard IEEE802.11p for the wireless access in the vehicular environment (WAVE). The IEEE 802.11p physical layer is comparable to that of IEEE 802.11a standard which is recommended as VANET MAC protocol. This perhaps experiences a huge amount of packet loss rate as a result of collisions as well as access delays viewed as common challenging issues associated with contention-based MAC protocols. Its performance improves by using Time Division Multiple Access (TDMA) scheme [1] to attain restructuring of TDMA slots with no central controls. The bandwidth allocation of 75 MHz supports seven separate channels includes a control channel (CCH) and six other service channels (SCHs) each spanning 10 MHz bandwidth.

---

<sup>\*</sup> Corresponding author.



Ultimately, most of the researches focus on finding quality route and hardly is there any focus on guaranteeing QoS. Therefore our propose design intends to discover and establish the optimal path, as well as guaranteeing QoS. The algorithm for formation as well as maintaining the cluster is not under this research, scope, so we adopt [2] with modifications. Also in our propose scheme, the V2V communication scenario is going to be a clustering-based multi-channel ad hoc network; whereby all the vehicles within the a communication range will self-organized into different clusters each that contains a cluster-head (CH) vehicle chosen as with 3with an extension of CH Link Connectivity Duration. However, such CH is made to have dual transmission powers; that's, whenever a CH needs to communicate with its cluster member, it chooses a short range transmission power. Otherwise, the CH utilizes a long range transmission power while it must exchange information using its neighboring CHs.

The remaining sections of the paper are arranged as follows; Section 2 presents state-of-the-art in guaranteeing QoS over VANETs. Section 3 states related works. Section 4 describes the basic idea of the proposed scheme. Section 5 shows a simulation experiment analysis. Finally, section 6 concludes the paper.

## 2 State-of-the-Art

Supporting QoS in VANETs is really a challenge because of some certain features that led to intermitted link interruptions. In general, the guaranteed service (GS) guarantees that packets are obtainable within the guaranteed delivery time, and does not be discarded due to buffer over flows. Various kinds of multimedia applications typically have extremely diversified QoS requirements in regards to data transfer rates and delay bounds and many others.

The majority of the current MANET routing protocols and QoS designs include difficulties for VANETs in addressing both of these needs due to the next reasons. Several of them fail easily to capture and employ neighbor-availability information. When the route setup fails, re-construction process is needed; and as a result, they can't guarantee a route that may last for an acceptable period of time which results in packets loss. More so, several schemes were suggested [3][4] for QoS guarantees in mobile ad hoc network (MANET). However, there is no proper or appropriate implementation framework to satisfy the QoS needs for the rapid network topology. A distributed cluster-based multi-channel communications scheme in [5] combines the clustering with contention free/- based MAC protocol.

### 2.1 Clustering

The formation of cluster region by gathering VANETs nodes that falls within a radio range. Vehicles within the same direction into are group into the same clusters, with a cluster-head (CH) vehicle elected and some ordinary members (OMs). The cluster concept has successfully been utilized for MANET to get a better delivery ratio as well as to reduce broadcast issue [11].

VANET applications are able to use an extended range,  $Z$ , to utilize the control channel in order that a cluster-head can easily communicate with nearby cluster-heads for safety message disseminations, as well as a shorter range,  $z$ , for a service channel that is utilized for intra-cluster managements as shown in fig. 1 below;

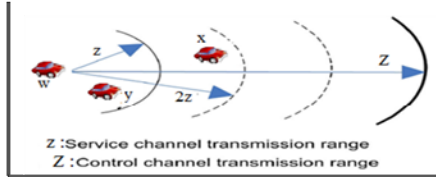


Fig. 1. Neighborhood Relationship

For the reason that nodes exchange their status information through control channel, it will be feasible for node  $w$  to identify that  $x$  is within  $2z$  distance. Conversely, to improve the stability of the CH in [2], we modified the procedures by adding the CH Link Connectivity Duration (LCD) which is the stability of the link. LCD is computed using the formula, motivated from [11]

$$LCD = \frac{\sqrt{(\alpha^2 + \gamma^2)R^2 - (\alpha\delta - \beta\gamma)^2} - (\alpha\beta + \gamma\delta)}{\alpha^2 + \gamma^2}$$

$$\alpha = v_i \cos \theta_i - v_j \cos \theta_j$$

Where  $\beta = a_i - a_j$

$$\gamma = v_i \sin \theta_i - v_j \sin \theta_j$$

$$\delta = b_i - b_j$$

$(a_i, b_i), (a_j, b_j)$ , is the neighboring VANET nodes (vehicles  $i$  and  $j$ ) Cartesian coordinates with the inclination of  $\theta_i$ , respectively  $(0 < \theta_i, \theta_j < 2\Pi)$  depending upon the  $x$ -axis and moving at  $v_i$ , speed.  $R$  is the IEEE 802.11p wireless transmission range. We assume that the CH and source vehicle are adjacent, otherwise LCD is the minimum of  $LCD_i$  ( $1 < i < n$ ), where  $n$  is the number of hops between the source and the CH.

Routing in one node to a different will contain routing in the cluster and routing from cluster to adjacent groupings. Therefore, routing in our cluster-based approach consist of the intra-cluster as well as inter-cluster communications as explained in section 4 below

### 3 Related Work

This section highlights major attempts in applying MANET routing protocols to VANET networks.

Current routing techniques in VANETs incorporate some difficulties in making certain QoS guarantee. However, a QoS routing protocol should therefore need to

guarantee satisfactorily a specific method of measuring the overall performance. The enthusiastic characteristics of VANETs render it challenging to do resource reservation and guarantee QoS in the environment. Some measures to enhance the performance and efficiency in vehicular routing were proposed [8] and provides quality route with higher percentage of throughput. Enhancement of QoS provision in VANET routing protocol with respect to delay, throughput and application response time is shown in [9]. More so, [10] proposes a routing algorithm for obtaining optimal QoS for highly dynamic VANETs.

## 4 Basic Idea of the Proposed Scheme

### 4.1 Protocol Procedures

Within the proposed system, every vehicle node is connected among the three (3) potential condition; Cluster head (CH), gateway and ordinary member (OM).s Within the next, we'll introduce how the algorithms for the source node, CH and immediate node as well as the destination node work as show in the fig.2 below;

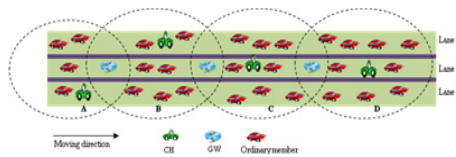


Fig. 2. Network scenario

### 4.2 Intra-cluster and Inter Cluster Communication

Each node sends information to its cluster head for intra-cluster communication. In our scheme as mention in section I, the CH communicates with two transmission range and determines the TDMA frame structure based on the number of OMs within the cluster in addition to number of neighboring CHs as show in fig. 3;

For Inter-cluster communication, each node sends aggregated information to their neighboring clusters. Our design determines the transmissions over two different IEEE 802.11 MAC-based channels as indicated in section 1. The route discovery

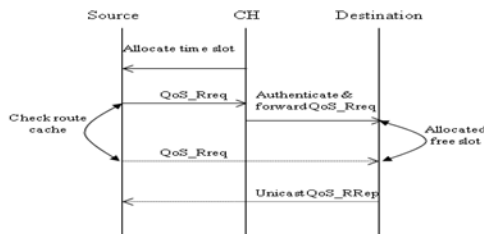


Fig. 3. TDMA-based QoS routing

would be an extension of AODV and location request mechanism of LORA-CBF [8], to which QoS features are embedded in the selection and guaranteeing QoS service in the traffic management. At present, our proposal depends on two stages: neighborhood availability information and QoS\_route request (source node algorithm, CH algorithm as well as destination algorithm) presented in [11].

### 5 Simulation Experiments

The proposed scenarios are implemented in NCTUns-6.0 simulation environment and maintained default values presented table 2 below. Adhere to the DSRC’s seven-channel bandplan, we use Ch174 as the Inter-Cluster Data (ICD) channel as defined in [6]. The simulation parameters are indicated the table below;

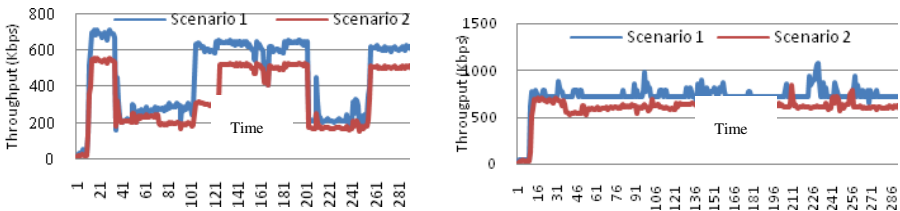
**Table 1.** Simulation Parameter

No..of nodes	Transmission range	Packet size	Simulation time	Node’s speed	MAC Specification	Channel bandwidth	Examined Routing protocol
100	250m	1400bytes	400 Secs	0-18 m/s	IEEE 802.11p	3 Mbps	AODV

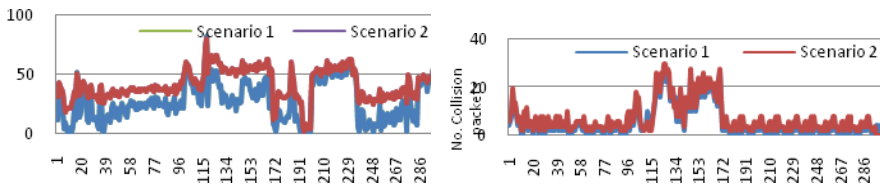
We simulated two scenarios and the performance is measured in terms of network throughput rate packet drop rate and collision rate.

*Scenario 1:* based on a point-to-point transfer of a single message by the suitable vehicle CH (car head) in NCTUns within the transmission range.

*Scenario 2:* Using single hop broadcast (flooding) scheme which results degrading the network throughput, hence affect it performance. The, the overall performance is shown in figure 4 to figure 7 below;



**Fig. 4.** (a) Network Throughput using TCP Connection (b) Network Throughput using UDP Connection



**Fig. 5.** (a) UDP Packet Drop Rate (b) UDP Connection Collision Rate

The simulation performance was analyzed for each scenario is done by generating TCP/UDP packets flows and the throughput of the flows for both scenarios is dispatched in figure 4 (a) & (b). Scenario 1 yields a remarkable gain over scenario 2 because of its capability to prioritize and maintain real-time flows and thus achieved higher network utilization.

For the packet drop and collision, UDP communication is simulated and evaluated against the two scenarios. Subsequently, fig. 6(a) shows the packet drop rate with the increase of the number of mobile nodes contending for the same channel in scenario 2, thus flooding increases in the network which decreases the network throughput. Figure 6 (b), shows less number of packets collide in scenario 1 which allow a reliable communication as the bandwidth is allocated with the defined data channel because burst of packet are generated that are sent with no specific link.

## 6 Conclusion

Adhere to the DSRC's seven-channel bandplan, we use Ch174 as data channel as defined in the appropriate functions of the seven channels spanning 10 MHz bandwidth. UDP and TCP protocols are used as the real-time and best effort traffics to satisfy bandwidth requirement. The results show that using the proposed scheme, the network throughput measured and less packet drop rate are guaranteed for providing the quality of service for nodes running the IEEE802.11 MAC.

## References

1. Jiang And, D., Delgrossi, L.: IEEE 802.11 P: Towards An International Standard For Wireless Access In Vehicular Environments. In: Vehicular Technology Conference, VTC Spring 2008, pp. 2036–2040. IEEE (2008)
2. Ding And, R., Zeng, Q.: A Clustering-Based Multi-Channel Vehicle-To-Vehicle (V2V) Communication System. In: First International Conference on Ubiquitous And Future Networks, ICUFN 2009, pp. 83–88 (2009)
3. Mauve, M., Widmer And, A., Hartenstein, H.: A Survey On Position-Based Routing In Mobile Ad Hoc Networks. *IEEE Network* 15, 30–39 (2001)
4. Zayene, M.A., Tabbane And, N., Elidoudi, R.: Performance Evaluation Of Greedy Perimeter Stateless Routing Protocol In Ad Hoc Networks. In: Fourth International Conference on Computer Sciences And Convergence Information Technology, ICCIT 2009, pp. 907–912 (2009)
5. Su, H., Zhang And, X., Chen, H.: WSN12-6: Cluster-Based DSRC Architecture For Qos Provisioning Over Vehicle Ad Hoc Networks. In: Global Telecommunications Conference, GLOBECOM 2006, pp. 1–5. IEEE (2006)
6. Su And, H., Zhang, X.: Clustering-Based Multichannel MAC Protocols For Qos Provisionings Over Vehicular< Emphasis Emphasistype=. *IEEE Transactions on Vehicular Technology* 56, 3309–3323 (2007)
7. Aquino And, R., Edwards, A.: A Reactive Location Routing Algorithm With Cluster-Based Flooding For Inter-Vehicle Communication. *Computación Y Sistemas* (2006)

8. Mo, Z., Zhu, H., Makki And, K., Pissinou, N.: MURU: A Multi-Hop Routing Protocol For Urban Vehicular Ad Hoc Networks. In: 2006 Third Annual International Conference on Mobile And Ubiquitous Systems: Networking & Services, pp. 1–8 (2006)
9. Gongjun, Y., Rawat And, D., Bista, B.: Provisioning Vehicular Ad Hoc Networks With Qos. In: Int. Conf. on Broadband, Wireless Computing, Communic. And Applications (BWCCA), pp. 102–107 (2010)
10. Subramaniam, P., Thangavelu, A., Venugopal, C.: Qos For Highly Dynamic Vehicular Ad Hoc Network Optimality. In: 2011 11th International Conference on ITS Telecommunications (ITST), pp. 405–411 (2011)
11. Mu'azu, A.A., Low, T.J., Lawal, I.A., Shah, P.A.: Throughput measurement for the guaranteed QoS real-time traffic flows in VANETs. In: International Conference on Innovation Management and Technology Research (ICIMTR 2013), Seremban, Malaysia (2013)

# Detecting People Using Histogram of Oriented Gradients: A Step towards Abnormal Human Activity Detection

Abdul-Lateef Yussiff\*, Suet-Peng Yong, and Baharum B. Baharudin

Department of Computer and Information Sciences  
Universiti Teknologi PETRONAS  
Bandar Seri Iskandar, Tronoh  
Perak, Malaysia  
ayussiff@gmail.com, yongsuetpeng@petronas.com.my

**Abstract.** Human activity understanding is a branch of research in computer vision that has attracted a lot of attention for decades. Accurate identification of humans in video surveillance is fundamental prerequisite towards activities' understanding. Little or no research has been conducted for human detection in financial endpoint premises specifically Automatic Teller Machine (ATM) sceneries. The video surveillance settings have some unique features compared to others applications: static and non-uniform background, low resolution images, and lack of initial background model. The Histogram of oriented gradient technique was used to locate people in each frame of the surveillance video. Our framework achieved a precision of 88.71 and F-score of 56.41.

**Keywords:** Histogram of Oriented Gradient (HOG), people detection, video surveillance, Automatic Teller Machine(ATM) security, Abnormal activity.

## 1 Introduction

Video surveillance is an active area of research especially in computer vision, robotics, gesture recognition and analysis, traffic monitoring, vehicle navigation, and etc. People surveillance emerges as one of the vibrant research area in the last decade especially since the event of September 11 in the USA, and has always been a critical component in guaranteeing security at banks, airport and correctional institutions. The widespread of high quality cheap surveillance cameras and the availability of broadband wireless networks, has made installation of group of cameras to enforce security become technically and economically realistic. Video surveillance's success and response to an event is not determined by the capabilities of the equipment rather determined by the alertness of the operator [1]. The very recent Boston marathon bombing in the USA has shifted the focus to real time detection of abnormal human activities. Human Detection is an integral part of computer vision but many current works lack the real-time performance that is required of practical applications. As stated in [2] "Detecting humans in images and videos is one of the important

---

\* Corresponding author.

challenges in computer vision”. Good abnormal activities detector is directly dependent on the performance of detection and tracking.

The problem in the particular context of Automatic teller machine (ATM) video surveillance is addressed in this paper. The paper focuses mainly on the problem of robust Human detection using Histogram of Oriented Gradient (HOG) [3] in surveillance video. Related works are discussed in section 2, section 3 talked about dataset preparation, methodology is discussed in section 4, and result and conclusion is addressed at the last two sections.

## 2 Related Works

Previous research works on people detection focused mainly on background subtraction. High sensitivity to background changes and illumination, and unsuitability to high density of persons are among the drawbacks of this approach. Background subtraction techniques generally determined foreground object from the video and then group it into categories like human and non-human depending on color (skin color), contour, shape, or motion and others. An object is detected if the difference between the incoming frame and the so called background model is greater than a given threshold as shown in equation 1. Adaptive background model technique was adopted to get an estimate for the background model in the case of non-uniform background.

$$|I_t(x, y) - B_t(x, y)| > \tau \quad (1)$$

$$B_{t+1} = \alpha \times I_t + (1 - \alpha) \times B_t \quad (2)$$

Where,  $\tau$  is user pre-defined threshold. The background image  $B_t$  is updated by the use of a first order recursive filter as shown in equation 2.  $\alpha$  is an adaptation coefficient.  $I_t$  is the image Intensity pixel,  $B_t$  is the background pixel. The idea is to incorporate the new information into the current background image. After that, the new changes in the scene are updated with the background model.

Most current methods are based on machine learning which uses discriminative classifier on images and the technique can be classified in two broad approaches [4]; namely sliding window detection analysis and part-based approach.

Papageorgiou and Poggio [5] adopted Haar-like feature representation coupled with a polynomial SVM as the machine learning classifier. [6] used chamfer distance to compare edge images to an exemplar dataset. Viola et al. [7] built on the Haar-like wavelets in order to handle space time information for human detection.

The other side of human detection approach using machine learning to detect part separately and concludes detection of human if some or all of its parts are presented in a feasible configuration. Corvee and Bremond [8] use hierarchical tree of HOG descriptors coupled with sliding window of  $48 \times 96$  to identify each individual human part. [9] employ pictorial structure approach where by, object is described by its parts, connected with springs, and each part is represented by Gaussian derivative filters of different orientations and scales. Ioffe and Forsyth [10] model parts as projections of straight cylinders and bars, then propose efficient technique to incrementally aggregate these segments into a complete body based on the probability of likelihood.



One of the first early methods with good performances is the cascade of Haar-like features proposed by Viola-Jones [11]. The technique adopted simple Haar wavelet filters as feature and a cascade of weak learner Adaboost. Also, The Histogram of Oriented Gradient (HOG) feature, proposed by Dalal and Triggs [3], proved very effective for person detection. The pioneer used HOG descriptors and linear SVM classifier to detect object in an image. Another variant of Histogram of oriented gradient proposed by Piotr et, al. [12] used integral channel feature. The technique was trained on weak learner cascade of Adaboost classifier.

Covariance features [2] combined with background subtraction technique, was shown to be an effective and fast human detector with promising results in video surveillance. The algorithm uses a cascade of LogitBoost on features mapped from the Riemannian manifold of local region covariance matrices derived from input image features.

### 3 Data Preparation

ATM video surveillance clips were downloaded from YouTube and the relevant ones are manually sorted out from the irrelevant video clips. The downloaded videos were converted and saved in .avi files using ffmpeg [13]. We also downloaded about 176 image data of people using ATM from Google images. Figure 1 shows sample images of people using ATM in the database. Summary of the average image properties is depicted in Table I, while Figure 2 depicts number of people per image sample.

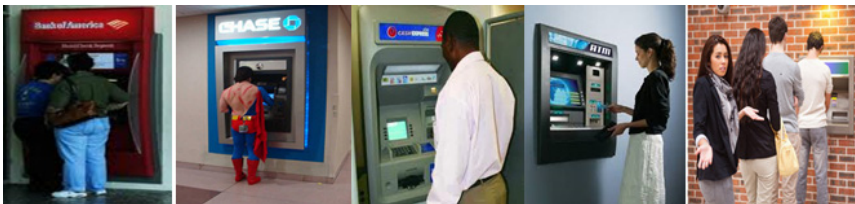


Fig. 1. Sample Images in the Database

Table 1. Average Image Properties

Properties	Value
File Size	7231
Format	jpg
Width	201
Height	251
Bit Per Pixel	24
Color Type	True Color
Number of Samples	3
Coding Method	Huffman
Coding Process	Sequential

## 4 Methodology

There have been several algorithm proposed for people detection. The notable among them are background subtraction and foreground detection, Silhouette matching technique, body-part models, HOG, etc. The technique adopted for this research paper is Histogram of Oriented Gradient (HOG). There are variants of the algorithm to speed the detection but the main core of the algorithm remains intact.

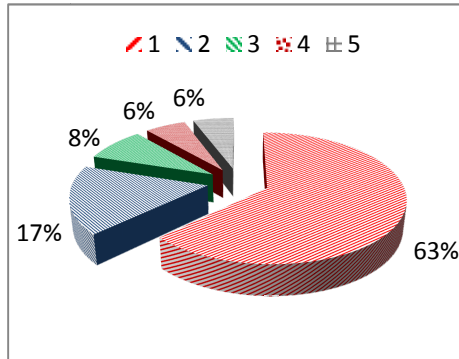


Fig. 2. Pie Chart of the Average Number of People per Image

### 4.1 Histogram of Oriented Gradients Algorithm for People Detection

The algorithm used, is Histogram of Oriented Gradient (HOG) [3] for detecting object in images, took inspiration from SIFT [14] and an Integral channels [12] for fast and efficient calculations of the gradient and orientations. The descriptors are used to classify object into human or non-human using a single sliding window approach, a kind of binary classification using SVM as classifier. The first step is the computation of gradients of an image, centered derivate mask  $[-1 \ 0 \ 1]$  and  $[-1 \ 0 \ 1]^T$  were used on the image object in both x and y direction. Then the gradient magnitude and directions (orientations) are calculated. Each gradients are then classified into a one of the pre-defined nine discrete bins orientation  $(0 \ 180]$  degrees with respect to a localized blocks. A block contributes 36 features, so a  $64 \times 128$  single window with 105 blocks will contributes a total of 3780 descriptors. The detection task is performed by scanning the given input image with a single window at various scales and positions, and classifying each window as human or non-human. The basic theory behind the algorithm is shown in equations 3 to 5. Equation 3 is the center derivatives.

$$f^1(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x-h)}{h} \tag{3}$$

$$G = \sqrt{I_x^2 + I_y^2} \tag{4}$$

$$\Theta = \arctan\left(\frac{I_x}{I_y}\right) \tag{5}$$

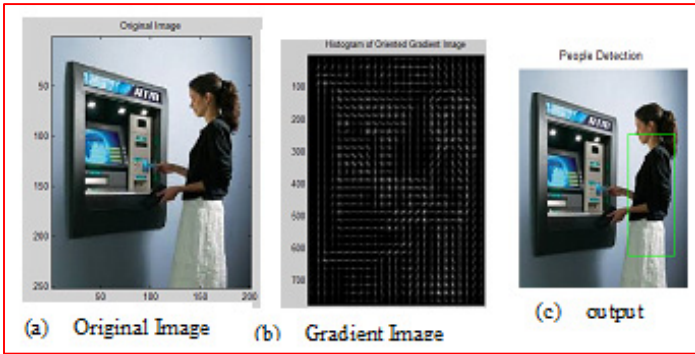
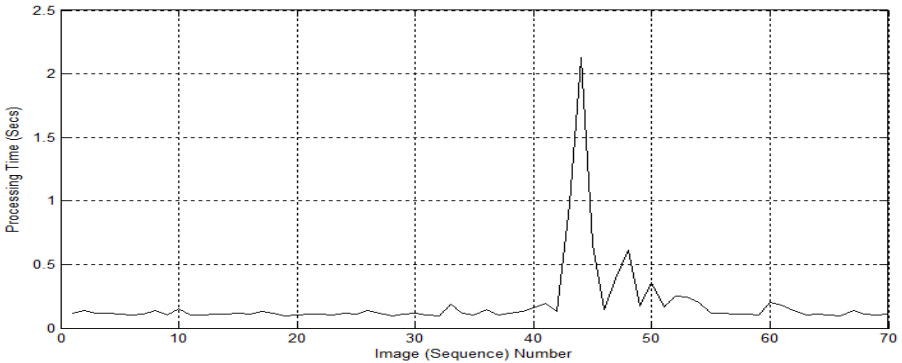


Fig. 3. (a) Original Image (b) HOG Visualization (c) Detected Person in Rectangle

## 5 Result

The algorithm was evaluated against our dataset. The experimental setup consists of 176 Google images with total 326 humans in the test data. Figure 4 depicts the processing time per image. We achieved precision of 88.71, recall of 41.35 and F-score of 56.41 percentage points on the test dataset.



## 6 Conclusion and Future Works

This is an ongoing research work. The algorithm implemented can detect human in both still images and videos. The algorithm produce an effective detection result compared to the background subtraction method. As it is expected, most people prefer to use ATM alone; the fact is that, they don't want people to know their ATM personal identification number (PIN). Effort is being made to improve the detection speed in order to compare it with the state of the art.

## References

- [1] Shah, M., Javed, O., Shafique, K.: Automated visual surveillance in realistic scenarios. *IEEE Multimedia* 14, 30–39 (2007)
- [2] Yao, J., Odobez, J.-M.: Fast human detection from videos using covariance features. In: *The Eighth International Workshop on Visual Surveillance, VS 2008* (2008)
- [3] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, pp. 886–893 (2005)
- [4] Gavrilu, D.M.: The visual analysis of human movement: A survey. *Computer Vision and Image Understanding* 73, 82–98 (1999)
- [5] Papageorgiou, C., Poggio, T.: A trainable system for object detection. *International Journal of Computer Vision* 38, 15–33 (2000)
- [6] Gavrilu, D.M., Philomin, V.: Real-time object detection for “smart” vehicles. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 87–93 (1999)
- [7] Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 734–741 (2003)
- [8] Corvee, E., Bremond, F.: Body parts detection for people tracking using trees of histogram of oriented gradient descriptors. In: *2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS*, pp. 469–475 (2010)
- [9] Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. *International Journal of Computer Vision* 61, 55–79 (2005)
- [10] Ioffe, S., Forsyth, D.A.: Probabilistic methods for finding people. *International Journal of Computer Vision* 43, 45–68 (2001)
- [11] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*, vol. 1, pp. I-511–I-518 (2001)
- [12] Dollár, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. In: *British Machine Vision Conference* (2010)
- [13] Tomar, S.: Converting video formats with FFmpeg. *Linux Journal* 2006, 10 (2006)
- [14] Lowe, D.G.: Object recognition from local scale-invariant features. In: *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 1150–1157 (1999)

# Research of Touchscreen Terminals Gesture Operation Error Based on Kansei Engineering

Rong Qin<sup>1</sup> and Dongxiang Chen<sup>2</sup>

<sup>1</sup> College of Mechanical Engineering, Tianjin University, Tianjin, China  
Pengxiang Apartment, Tianjin University, Tianjin, China

<sup>2</sup> College of Mechanical Engineering, Tianjin University, Tianjin, China  
Xinyuancun, Tianjin University, Tianjin, China  
qinrongwykl@sina.com,  
dxchen@tju.edu.cn

**Abstract.** As mobile touchscreen terminal application is more and more diversification and individuation, gestures gradually become the main way of people interact with touch screen interface. But currently, there is a wide variety of mobile touch type terminal on the market, lead to each brand has its own unique style of gestures with no unified design criteria, confusion, operation error and learning difficulties often appear when in use by the user. In this paper, based on the Kansei Engineering related method, choose the optimal gestures form, to reduce the user operation error, and strengthen the availability of touchscreen terminals.

**Keywords:** gesture operation, mobile touchscreen terminals, wrong operation, Kansei Engineering.

## 1 Introduction

Currently on the market, touch screen terminal equipment type is various, each brand has its own unique style of operation, some even conflict to each other. This complex chaotic scenes usually let users difficult to identify and memory, cause some confusion and wrong operation inevitable. In the face of numerous and large operation specification, user is hard to remember that different platforms using norms and standards, therefore, learning difficulties and wrong operation appear. Some methods can solve such problems: model-based, accuracy is high, the effect is obvious; appearance-based, fast speed, can satisfy the real-time application requirements; Target tracking method and the rough set calculation method also has strong practicability. The results of these methods effectively improve the gesture operation forms of scientific and identifiable degree. But as a result of hand shape change high dimension, lead to these methods and their derivative algorithm great emphasis on virtual reality, machine vision, pattern recognition, human-computer interaction, and other fields of exchange and cooperation in computer science, the demand of user's perceptual study is relatively less, the results of the study disconnect with the user's operation habit. In this paper, from the user's emotional instinct of operation, with gesture as the research object, effectively reduce user's operation error and enhance the availability of touchscreen terminals.

## **2 Gesture-Based Mobile Touchscreen Terminals Process Related Emotional Vocabulary Collection**

Perceptual words collection is the foundation and the key of the Kansei engineering research, transform user emotion into understandable adjectives. Eventually representative perceptual vocabulary selection has a crucial influence to the scientificness of evaluation test and the conclusion, So should be to collect as much as possible to express gestures to the user brings emotional words. So should be to collect adjectives as much as possible that can express emotion that gestures operation brings to the user.

### **2.1 Vocabulary Collection**

With mobile touchscreen terminals as object, through brainstorming, product specification and user manual, information that mobile terminal design professionals provide, magazine, Internet, etc. This case has collected 200 emotional adjectives.

### **2.2 Vocabulary Filtering**

(1) Conduct a preliminary classification and selection to the 200 emotional adjectives according to the degree of similar meaning, eliminate meaning particularly close or relevant enough emotional vocabulary, eventually get 40 emotional adjectives, such as:

Concise, Precise, Clear, Harmonious, Generous, Effective, Striking, geometrical, calm, steady, advanced, individual, sportive, lovely, kind, quick, stereo, coherent, complicated, cursory, plain, unbalanced, stingy, ineffective, fuzzy, vulgar, messy, impulse, lively, base, popular, stillness, mechanical, distant, lengthy, plane, intermittent, strong sense of operation, weak sense of operation.

(2) Divide 40 emotional adjectives into 20 corresponding adjective phrases, making questionnaire, Invited 30 design major students participate in the survey, requiring people to select five most can describe gestures perceptual adjective phrases, analysis questionnaire, select the perceptual phrases that choosed most times, correct analysis the results of the survey, finally get 5 groups of representative adjective vocabulary.

5 groups of emotional adjectives are: Concise and Complicated, Clear and Fuzzy, Effective and Ineffective, Quick and Lengthy, Strong sense of operation and Weak sense of operation.














## **3 Gesture Operate Mobile Touchscreen Terminals Process Related Morphological Analysis**

Final selection of representative sign sample have a crucial impact on the science of evaluation test and the conclusion, in this phase should be as much as possible to collect samples of gestures.

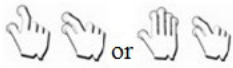
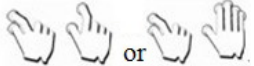

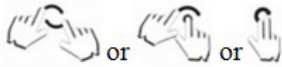
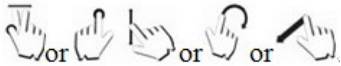
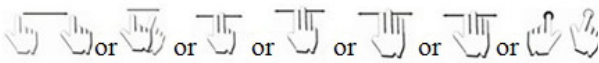
### 3.1 Gestures Form Collection

Through the research of gestures and the sample search from Internet, magazine and product brochure, after deleted some similar gestures samples, work up the pictures as the following form. See from table 1, for the same order, gestures from different operating platform have a variety of forms.

**Table 1.** Gestures table form

Command names	Gestures form
Click	
Double-click	
Press and tap	
Press	
Flick	
Bundle	
Duplicate	
Drag/Delete/Move/Scroll	
Fast scroll	
Pan/Move	
Move through list/Move	
Roll	
Press and drag	

**Table 1.** (continued)

Scale down	
Scale up	
Show	
Rotate/Adjust view	
Aim	
Move	

### 3.2 Shape Analysis

The perceptual cognition of the users need to be quantified. Above summarizes five pairs of adjectives: Concise and Complicated, Clear and Fuzzy, Effective and Ineffective, Quick and Lengthy, Strong sense of operation and Weak sense of operation. The implementation in the form of questionnaire survey, as the site questionnaire and the electronic questionnaire, a total of 30. Participants make 5 attributes evaluation (For example, the most concise, more concise, no obvious bias, more complicated, the most complicated) on the sample, Quantitative standards: -2 represent the most complicated, 1 represent more complicated, 0 represents no obvious bias, 1 represent more concise, 2 represent the most concise. Refer to table 2, the product sample SD perceptual scale.











**Table 2.** The product sample SD perceptual scale

Adjectives	Score					Adjectives
Complicated	-2	-1	0	1	2	Concise
Fuzzy	-2	-1	0	1	2	Clear
Ineffective	-2	-1	0	1	2	Effective
Lengthy	-2	-1	0	1	2	Quick
Weak sense of operation	-2	-1	0	1	2	Strong sense of operation



















With table 2 as scoring criteria, users score perceptual factor,for example,narrow command, can be completed by two types of gestures form, Users score two forms respectively, If the form considered the most concise, get 2 points,if the most complicated,get -2 points. Recycling 30 effective questionnaires. Calculated the total score of each item, the result is divided by 30,get a score is the final score for the commands in the form. The results in table 3.

**Table 3.** Corresponding relationship survey of emotional vocabulary and design combination









Emotional appeal		Concise	Quick	Clear	Effective	Strong sense of operation	Total score	
Factor	Scale down		1.583	1.572	1.675	1.343	0.298	6.471
			0.572	0.772	0.969	1.237	1.263	4.813
	Scale up		1.579	1.583	1.581	1.263	0.352	6.358
			0.593	0.754	1.101	1.336	1.173	4.957
	Press and drag		1.657	1.457	0.265	0.128	-0.195	3.312
			-0.527	0.199	1.754	1.652	1.687	4.765
	Aim		1.201	0.992	0.283	0.643	-0.301	2.818
			-0.677	1.376	1.233	1.675	1.562	5.169
			0.741	1.201	0.913	1.198	0.243	4.296
			1.757	1.689	1.533	1.766	-0.725	6.020

**Table 3.** (continued)

		1.216	1.579	1.385	1.409	-0.289	5.300
		1.263	0.788	0.247	-0.333	-0.509	1.456
		1.346	1.369	1.480	1.298	0.221	5.714
Move		0.765	0.669	0.941	0.887	0.561	3.823
		0.965	0.864	0.782	0.774	0.103	3.488
		0.897	0.846	0.482	0.576	-0.462	2.339
		-0.561	1.119	1.599	1.639	1.487	5.283
		1.401	0.566	1.648	1.521	0.321	5.457
Roll		1.380	1.537	0.723	0.396	-0.109	3.927
		1.378	-0.207	0.137	1.284	0.560	3.152
		-0.254	-0.991	-0.098	1.498	1.279	1.434
Show		0.985	1.253	0.076	-0.372	-0.104	1.838
		1.673	1.518	0.882	-0.116	-0.671	3.286
		0.158	-0.223	0.099	0.477	0.981	1.492
Rotate		-0.103	0.378	0.959	1.338	1.587	4.159
		0.529	0.698	0.554	0.356	0.782	2.919

Through table 3 calculate perceptual factor scores, the highest score tertiary perceptual factor become the secondary emotion factor final result. The results in table 4.

**Table 4.** The highest score perceptual factor

Scale down	Scale up	Press and drag	Aim	Move	Roll	Show	Rotate
							

In Table 4, the form of gestures corresponding operation commands is the result of the highest score, is the gestures form most accord with user perceptual action of instinct, to sum up, research determine the right gesture contained by a set of touchscreen terminals operating system at least by perceptual engineering method. Use mobile touchscreen terminal gestures operation simulation system, input the results as the correct gesture operation form corresponding to the command, a random sample of 100 operators for validation test, the operator operating error rate significantly decreased, prove that the research results are effective.

## 4 Conclusion

In touchscreen terminals gestures, some commands and gestures in the process of operation form cannot determine corresponding, in this paper, Kansei Engineering was used to determine the gesture operation form contained in a set of touchscreen terminals system at least, effectively reduce the user wrong operation happen because of confusion, the research method in this paper is feasible.

## References

1. Jiayuan, Z.: The Small Touch Screen Terminal Interface Usability Research (master degree theses). HuaZhong Science And Technology University, Hubei Province (2011)
2. Sun, C., Feng, Z., Li, Y., Zhang, M., Zhang, W., Pan, Z.: Human-computer Interaction Review based on Gesture Recognition (2010)
3. Liu, Y.: The Study to The Relationship between Gesture Operation and Touch Screen Motion Effects under The Rough Sets (master degree theses). Tianjin University, Tianjin (2012)
4. Wu, D.: Vision based on Gesture Recognition and Human-computer Interaction Research (master degree theses). Nanjing Aeronautics and Astronautics University, Nanjing (2010)
5. Li, Y., Wang, Z., Li, D.: Kansei Theory Research And Product Development of Engineering Application. Wuhan Technology University Journal 01 (2010)
6. Li, Y., Wang, Z., Xu, N.: Kansei Engineering. Ocean Press, Beijing (2009)
7. Li, X.: Design Visual Language Interpretation. Art And Design 11 (2010)
8. Liu, Q.: Thinking About Some Problems in The Design of Man-machine Interface. Luoyang University Journal 12 (2004)

9. Feng, H.: Analyses The Man-machine Interface Design Software. Nanping Teacher College Journal 10 (2006)
10. Li, Y., Guan, Z., Chen, Y., Dai, G.: Gesture-based Human-computer Interaction Research. System Simulation Journal 09 (2000)

# Heart Sound Feature Extraction Based on Wavelet Singular Entropy

Zhang Lu

School of Information Science and Engineering,  
University of Jinan, Jinan 250022, China  
ise\_zhanglu@ujn.edu.cn

**Abstract.** After analyzing the advantages and disadvantages of current methods of heart sound feature extraction, a new method based on wavelet transform, singular value decomposition and information entropy is put forward. In this method, firstly the heart sound is decomposed by wavelet transformation. Then the singular value of the sub-bands containing heart sound information is obtained by decomposition. Finally, according to the constructed wavelet singular entropy, the entropy of the above singular value is obtained. By comparing the wavelet singular entropy of normal heart sound signal with the several heart sound signals with pathological information, wavelet singular entropy can be found a good characterization of heart sound.

**Keywords:** Wavelet Singular Entropy, Heart Sound, SVD.

## 1 Introduction

Heart sound signal is from heart and collected by cardiotelephone. It is caused by mainly diastole and systole of the heart. It's the flow vibration from the body surface, caused by the impact of blood flow against heart valves, wall and large vessels [1]. There is important physiological and pathological information in heart sound signal, so doctors can determine the patient's disease characteristics by his heart sound signal. Normal heart sound is made up of by  $s_1$  and  $s_2$  that can be heard and  $s_3$  and  $s_4$  that can't be heard. If the heart is abnormal, there are murmurs in addition to  $s_1$  and  $s_2$ . For example, Coronary stenosis can cause murmurs in heart sounds diastolic period. This is because the flow caused by the coronary stenosis make surrounding tissues vibration and then sound is produced [2]. Therefore, to analyze the heart sound signals is an important subject about predictor of cardiovascular disease. Many researchers at home and abroad have tried a variety of methods to analyze the heart sound signals. B.El-Asir et al analyzed the heart sound signals with JTFA, and came to the conclusion that there were murmurs with different frequency at different times in different heart diseases [3]. Gauthier et al analyzed diastole with FFT, and draw the conclusion that diastole percentage in coronary heart disease heart sounds was higher than normal heart sound [4]. Above several studies have achieved some results, but there are some deficiencies in them. For example, in reference 4, because heart sounds is a typical non-stationary signals, the effect of differentiate coronary heart disease and healthy people by the method of differentiate coronary heart disease and healthy people is not ideal.

In this paper, based on wavelet transform, singular value decomposition and information entropy, the three theories are organically combined together to form a new method of extraction of heart sound signals wavelet singular entropy. In this method, firstly the heart sound is decomposed by wavelet transformation. Then the singular value of the sub-bands containing heart sound information is obtained by decomposition. Finally, according to the constructed wavelet singular entropy, the entropy of the above singular value is obtained. By comparing the wavelet singular entropy of normal heart sound signal with the several heart sound signals with pathological information, wavelet singular entropy can be found a good characterization of heart sound.

## 2 Wavelet Singular Entropy and Principle

### 2.1 Singular Value Decomposition

Assume that the matrix  $A \in C_r^{m \times n}$  ( $r > 0$ ), then there are unitary matrix  $U$  of order  $m$  and unitary matrix  $V$  of order  $n$ . They have the following equation:

$$U^H A V = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \quad (1)$$

In the above equation 1,  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$ , and  $\sigma_i$  ( $i = 1, 2, \dots, r$ ) are all non-zero singular values of matrix  $A$ . Change the equation 1, Singular Value Decomposition (SVD) of matrix  $A$  is obtained[5]:

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H \quad (2)$$

### 2.2 Singular Entropy

The singular value of the signal is to be described the characteristics of each frequency segment of the signal within the sampling time. The main features of the heart sound signals in various lesions signal appears for the difference of the singular values on the different frequency segment. To quantify the extent of this change, singular entropy is structured:

$$E_i = \frac{\sigma_i}{E} \quad (3)$$

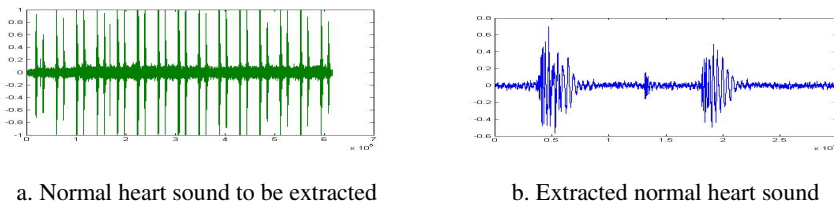
In the above equation 3,  $E = \sigma_1 + \sigma_2 + \dots + \sigma_r$ , so  $\sum_{i=1}^r E_i = 1$ . This is in line with the initial normalization condition of entropy. According to the definition of entropy, singular entropy is calculated as:

$$H = -\sum_{i=1}^r E_i \ln E_i \quad (4)$$

### 3 Heart Sound Feature Extraction Based on Wavelet Singular Entropy

There are five steps in the heart sound feature extraction based on wavelet singular entropy: segmentation of heart sound, wavelet transform, singular value decomposition, calculating the singular entropy, characteristic value extracting.

1) Segmentation of heart sound. Heart sound signals used in the experiment are selected from Texas heart institute. Because the heart sounds are too long, they are envelope extracted. By this process, a complete heart sound signal with not only s1&s2, but also s3&s4. Envelope extraction method is HHT, and the obtained envelope is segmented. In Fig.1, the extracted normal heart sound is segmented, and a complete heart sound signal with not only diastole but also systole is obtained.



**Fig. 1.** Envelope extraction of the normal heart sound

2) Wavelet transformation. In this study, the DB6 heart sound signals are used as the mother wavelet in wavelet transformation. Wavelet decomposition scale and sampling frequency relate to the mother wavelet. According to the sampling theorem of wavelet, the more the decomposition scale, the greater the amount of computation is. So the decomposition scale should be selected according to the actual application and needs. The frequency of the heart sound signal used in this paper is 44100 HZ. Analyzing of heart sound signals by STFT, it's found that the main frequency of s1 is 50 ~150 HZ, and s2 is 250 ~300 HZ. So the decomposition scale is eight. The frequency of eighth order of decomposition detail signal cd8 is 172 ~344 HZ, and the frequency of eighth order of decomposition contour signal ca8 is 0~172 HZ.

3) Singular value decomposition. Because of the frequency of the heart sound signal is mainly in 0 to 300Hz, not only cd8 but also ca8 are decomposed by SVD. Then singular value matrix  $S_d$  and  $S_a$  is obtained.

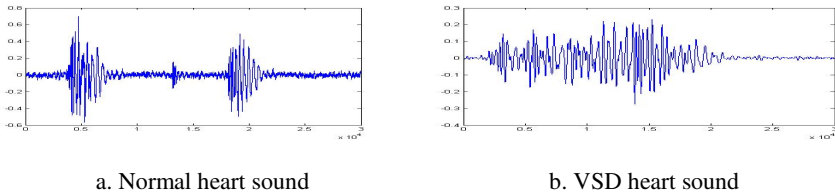
4) Calculate the singular entropy.  $S_d$  and  $S_a$  are respectively substituted into the equation 3&4, the singular entropy  $H_a$  of eighth order of decomposition contour signal ca8 and the singular entropy  $H_d$  of eighth order of decomposition detail signal cd8 are obtained.

### 4 Analysis of Simulation Results

The simulation software is Matlab7.4. There are five simulation signals: a, Normal; b, WSSS; c, HOC; d, VSD; e, LAM.

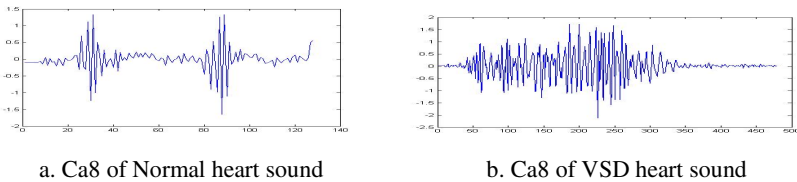
The normal heart sound signal is contrasted with VSD. The Fig.2, Fig.3, Fig.4 respectively show comparison of extracted signal, eighth order of decomposition contour signal ca8 ,eighth order of decomposition detail signal cd8.

Fig.2a is a typical heart sound signal, the front half portion of which is s1 and the latter half portion s2. May also it is mean that the front half portion is systolic and the latter half phase. In Fig.2b, there is a clear difference between the systolic and diastolic. Predictably, there are surely murmurs in Fig.2b, and it shows the signal certainly with lesion signal .It can be judged Fig.2b is signal with lesions signal.



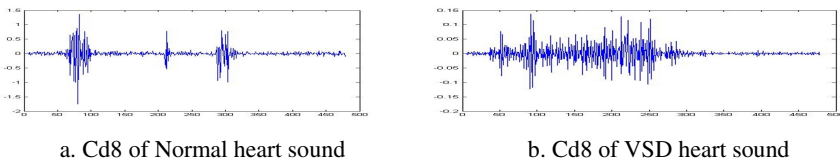
**Fig. 2.** Contrast between Normal and VSD heart sound

Then the Fig.3 is analyzed. The Fig.3 shows their contour signal, that is to say, it shows the difference between 0HZ to 172HZ.In the Fig.3a, two very clear contour signals are found, however in the Fig.3b, the composition of the signal can not be distinguished. So the difference between the normal heart sound and the VSD heart sound is very large.



**Fig. 3.** Contrast between ca8 of Normal and VSD

Finally, the three detail signals are compared. That is to say, it shows the difference between 172HZ to 344HZ.Two signals with large amplitude and one signal with small amplitude are found in the Fig. 4a.Two signals with large amplitude are too found in the Fig. 4b,but signal with small amplitude can not be found. So there is difference between them, but the difference not large. According to reference 1, there is murmur in the whole systolic of VSD and the proportion of metaphase is the largest.



**Fig. 4.** Contrast between cd8 of Normal and VSD



The singular entropy  $H_a$  of eighth order of decomposition contour signal ca8 and the singular entropy  $H_d$  of eighth order of decomposition detail signal cd8 of the five heart sound are shown in the Table 1. Here the normal heart sound and the VSD heart sound are set an example. The  $H_a$  of the normal is 4.3374, while the  $H_a$  of VSD is 7.4331. The difference between the two is 3.0957. The  $H_d$  of the normal is 3.271, while the  $H_d$  of VSD is 4.305. The difference between the two is 1.034. So it's shown that the lesion signals of VSD is mainly between 0 to 172 HZ. In this band, there are mainly the first heart sound s1 and part of the second heart sound s2. The analysis result of the figure and table are unanimous. It is shown that wavelet singular entropy is a good description of the characteristics of heart sound.

**Table 1.** Wavelet singular entropy of heart sound

	$H_a$	$H_d$
Normal	4.3374	3.271
WSSS	9.8094	10.242
HOC	10.662	5.1194
VSD	7.4331	4.305
LAM	5.4886	9.7397

## 5 Conclusion

In this paper, wavelet transform and singular entropy decomposition are used to obtain the feature of heart sound. Through theoretical analysis and simulation test, this method can well express the heart sound signal lesions. What is improved in the method is how to link the singular entropy of the contour signal and detail signal.

**Acknowledgment.** This work was sponsored by Science Fund of University of Jinan (XYK1222), Science Fund of Jinan Young Star (20090206), Project of Shandong Province Higher Education Scientific Research Program (J11LF02).

## References

1. Luo, J.-Z.: Cardiac Auscultation, vol. 3, pp. 32–50. People's Medical Publishing House, BeiJin (2000)
2. Dock, W., Zoneraich, S.: A Diastolic Murmur Arising in A Stenosed Coronary Artery. The American Journal of Medicine 42(4), 617–619 (1967)
3. El-Asir, B., Khadra, L., Al-Abbasi, A.H., et al.: Time-frequency Analysis of Heart Sounds. 1996 IEEE TENCON. Digital Signal Processing Applications 2, 553–558 (1996)
4. Gauthier, D., Akay, Y.M., Paden, R.G., et al.: Spectral Analysis of Heart Sounds Associated with Coronary Occlusions. In: 6th International Special Topic Conference on Information Technology Applications in Biomedicine, pp. 49–52 (2007)
5. Cheng, Y.-P.: Matrix Theory. Xi'An: Northwestern Polytechnical University Publishing House 1, 225–232 (2000)

# Research on MTMP Structure Chlorine Dosing Decoupling Control

Xie Peizhang and Zhou Xingpeng

Key Laboratory of Measurement and Control of CSE,  
School of Automation,  
Southeast University, Ministry of Education,  
Nanjing, China, 210096  
xpzseu@gmail.com

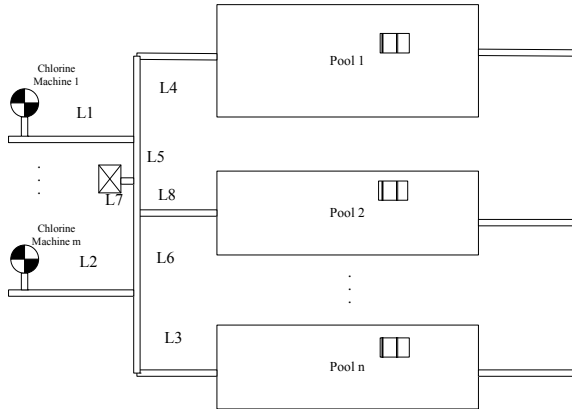
**Abstract.** In this paper, multi tunnels multi pools (MTMP) structure chlorine dosing control method is studied. First residual-chlorine decay model is proposed, then MTMP chlorine dosing process model is acquired. After that, wavelet neural network is introduced to identify the  $\mathcal{O}$ -th order inverse system so that pseudo-linearization system can be obtained. Then Time delay disturbance observer (DOB) control algorithm is designed for each decoupled subsystem, high performance and improved robustness are obtained. Simulation and application in tap-waterworks at Suzhou (China) shows that the algorithm is able to resist the model mismatch, disturbance and time delay, also the lower unit consumption of chlorine is obtained.

**Keywords:** Chlorine dosing, Neural networks inverse system, Pseudo-linearization, Time delay DOB, Multi tunnels multi pools (MTMP).

## 1 Introduction

A major objective of drinking water treatment is to provide water that is both microbiologically and chemically safe for human consumption, so it is a key process in water treatment. While The formation of potentially harmful trihalomethanes (THM) when using chlorine as a sanitizer in potable water supplies has led to tighter regulatory controls and hence a need for better control algorithm. Chlorine dosing is a complicated system with nonlinear, larger time-delay, time-varying and multi models; also the couple is introduced to the system due to the multi tunnels multi pools (MTMP) structure. Residual-chlorine must be controlled smooth and steady so that the THM can be reduced. A typical MTMP sanitizer dosing system is shown in Fig.1.

There are some papers [1,2] on residual-chlorine decay model. Based on these models, some control algorithm is studied. In paper [3, 4] a decentralization structure is proposed in the previous paper for robust model predictive control (MPC) of chlorine residuals in drinking water distribution systems (DWDS). These algorithms focus on the chlorine residuals in water distribution system, little thought is given to tank chlorine residuals, also time-delay, model mismatch and couple is out of consideration.



**Fig. 1.** Multi tunnels multi pools structure sanitizer dosing system

Time-delay DOB-NN [5-7] inverse system is proposed in this paper. Neural network inverse control method can realize the linearization decoupling control for MTMP structure chlorine dosing system. It constructs the  $\alpha$ -th order inverse system using wavelet neural network and then cascades the original system so that the system can be transferred to a kind of normal decoupled system equipped with linear transferring relationship. Then Time delay DOB control algorithm is designed for each decoupled subsystem, high performance and improved robustness are obtained.

## 2 Multi Tunnels Multi Pools Structure Sanitizer Dosing System

There are kinds of sanitizer, chlorine is used the most usually because of its convenient use, storage and simple operation. Residual chlorine is the main parameter of disinfection performance in water treatment. There are some papers about the decay law of residual-chlorine in water [8].

In this paper first-order reaction kinetics equation is chosen, it can be described as follows:  $C_B = C_A \cdot \exp(-kt)$

Where:  $C_A$ ,  $C_B$  is the chlorine concentration of time A and time B, k is the chlorine decay factor.

The process can be divided into two parts, one is the rapid process and the other is the slow process as shown in Fig.2. During the rapid process, chlorine dosing to the clean water, the consumption of chlorine is very large, it is related to the initial dosage and the amount of NH3. Based on the theory, combined to the experiment, the approximate model can be acquired. The Rising Curve of chlorine dosing process is shown in Fig.3.

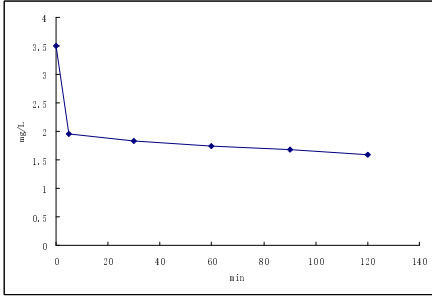


Fig. 2. Residual-chlorine decay curve

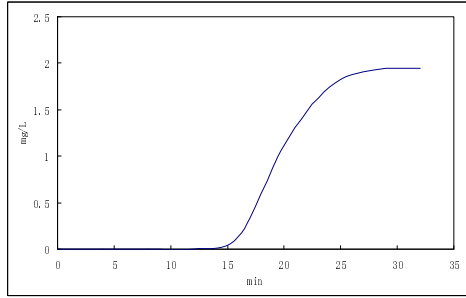


Fig. 3. Rising Curve of chlorine dosing process

In this paper, the model of multi tunnels multi pools structure chlorine dosing system is as follows:

$$\begin{cases} y_1(s) = \eta_{11}(X_1 - C_1) \frac{e^{-\tau_{11}s}}{T_{11}s + 1} + \eta_{21}(X_2 - C_2) \frac{e^{-\tau_{21}s}}{T_{21}s + 1} + \dots + \eta_{m1}(X_m - C_m) \frac{e^{-\tau_{m1}s}}{T_{m1}s + 1} \\ y_2(s) = \eta_{12}(X_1 - C_1) \frac{e^{-\tau_{12}s}}{T_{12}s + 1} + \eta_{22}(X_2 - C_2) \frac{e^{-\tau_{22}s}}{T_{22}s + 1} + \dots + \eta_{m2}(X_m - C_m) \frac{e^{-\tau_{m2}s}}{T_{m2}s + 1} \\ \vdots \\ y_n(s) = \eta_{1n}(X_1 - C_1) \frac{e^{-\tau_{1n}s}}{T_{1n}s + 1} + \eta_{2n}(X_2 - C_2) \frac{e^{-\tau_{2n}s}}{T_{2n}s + 1} + \dots + \eta_{mn}(X_m - C_m) \frac{e^{-\tau_{mn}s}}{T_{mn}s + 1} \end{cases} \quad (1)$$

### 3 Design of Improved Time-delay DOB Based on Pseudo-linearization System

The case study in this paper (waterworks at Suzhou in China) is a complicated MIMO system with large time delay, nonlinear and couple. It is a non-minimum phase system, and the original DOB can not deal with non-minimum phase system. Because of time delay, the inverse system can not be achieved due to advanced arguments. Wavelet Neural Network is proposed to identify the time-delay inverse system, and then this inverse system cascades the original system so that time-delay pseudo-linearization system can be obtained, after that this MIMO system can be transformed to time-delay SISO system without coupling.

#### 1 Block of time-delay DOB based on pseudo-linearization system

The structure of time-delay DOB based on pseudo-linearization system is shown in Fig.4. P is the pseudo-linearization system consists of neural network inverse system and the original system. P is considered as follows,

$$P_i = s^{-n_i} e^{-t_i s}$$

Where  $n_i$  is the order degree of neural network input  $i$ .  $t_i$  is the minimum time delay of Channel  $i$ .

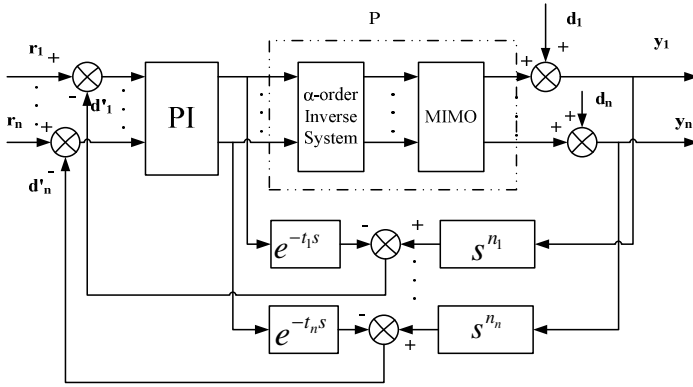


Fig. 4. Block of DOB based on pseudo-linearization system

$$Y_i(s) = G(s)R(s) + G_d(s)D_i(s)$$

$$G(s) = \frac{G_p(s)G_c(s)}{1 + G_p(s)G_n(s)G_c(s) - e^{-t_i s}G_c(s)}$$

$$G_d(s) = \frac{1 - e^{-t_i s}G_c(s)}{1 + G_p(s)G_n(s)G_c(s) - e^{-t_i s}G_c(s)}$$

Where  $G_p(s)$  is the pseudo-linearization system,  $G_n(s) = s^{n_i}$ ,  $G_c(s)$  is the transfer function of PI controller.

Thus,  $d_i(s)$  can be rejected by time delay DOB based on pseudo-linearization system using WNN.

### 4 Case Study

In this paper, the parameters of the case study are shown as following:

$$m=2, n=2. \eta = \begin{bmatrix} \eta_{11} & \eta_{12} \\ \eta_{21} & \eta_{22} \end{bmatrix} = \begin{bmatrix} 0.85 & 0.15 \\ 0.1 & 0.9 \end{bmatrix} \text{ is a matrix of flux ratio into pools}$$

$$\text{from tunnels. } \tau = \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{bmatrix} = \begin{bmatrix} 11 & 12 \\ 12 & 13 \end{bmatrix} \text{ is a matrix of time delay.}$$

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} 1.8 & 2.2 \\ 2.4 & 1.8 \end{bmatrix} \text{ is a matrix of parameters of inertial.}$$

(1) Proof of the exists of the inverse system

2 inputs 2 outputs time-delay chlorine dosing model can be written as following:

$$\begin{cases} y_1(t + \tau_{11}) = x_1(t) = \eta_{11}u_1(t) \cdot \exp(-kt) + \eta_{12}u_2(t) \cdot \exp(-kt) \\ y_2(t + \tau_{12}) = x_2(t) = \eta_{21}u_1(t) \cdot \exp(-kt) + \eta_{22}u_2(t) \cdot \exp(-kt) \end{cases}$$

Remark 1, time delay of each single model can be treated as the same, and then time delay can be separated from the chlorine decay model.

$$\frac{\partial X}{\partial U} = \begin{bmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{bmatrix}, \text{ obviously } \det\left[\frac{\partial X}{\partial U}\right] \neq 0, \text{ so the inverse system exists.}$$

(2) Simulation of no disturbance with the algorithm proposed in this paper

Good results can be obtained with the algorithm using DOB based on wavelet neural network inverse system as is show in Fig.5. Compared with other algorithms, DOB-NN inverse system is much better and no system overshoot is observed.

(3) Simulation of mismatch disturbance

The result is shown in Fig.6, the controller designed using improved DOB based on wavelet neural network is able to resist the disturbance caused by mismatching disturbance or matching disturbance.

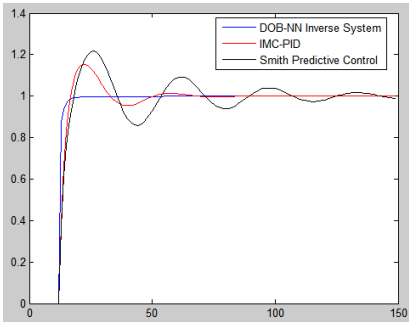


Fig. 5. Simulation of no disturbance

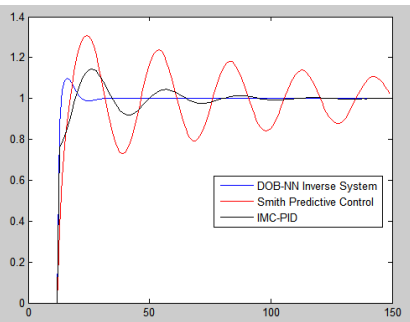


Fig. 6. Simulation of disturbance

## 5 Conclusion

In this paper, residual-chlorine decay law is presented, after that process model of MTMP chlorine dosing system is set up. Then Time delay DOB-NN inverse system control algorithm is proposed in this paper. Simulation and application in tap-waterworks shows that the algorithm is able to resist the model mismatch, disturbance and time delay.

## References

1. Haiyan, C., Xinhua, Z., Yin, Y.: Research on residual chlorine decay model of reclaimed water network. In: 2011 International Conference on Electronics, Communications and Control (ICECC), pp. 3899–3902. IEEE (2011)
2. Clark, R.M., Sivaganesan, M.: Predicting chlorine residuals in drinking water: Second order model. *Journal of Water Resources Planning and Management* 128(2), 152–161 (2002)
3. Chang, T., Brdys, M.A., Duzinkiewicz, K.: Decentralized Robust Model Predictive Control of Chlorine Residuals in Drinking Water Distribution Systems. In: ASCE, pp. 1–10 (2004)
4. Duzinkiewicz, K., Brdys, M.A., Chang, T.: Hierarchical model predictive control of integrated quality and quantity in drinking water distribution systems. *Urban Water Journal* 2(2), 125–137 (2005)
5. Kim, B.K., Chung, W.K., Ohba, K.: Design and performance tuning of sliding-mode controller for high-speed and high-accuracy positioning systems in disturbance observer framework. *IEEE Transactions on Industrial Electronics* 56(10), 3798–3809 (2009)
6. Jung, S.: On the unified approach to the disturbance observer. In: 2012 12th International Conference on Control, Automation and Systems (ICCAS), pp. 573–578. IEEE (2012)
7. Rahman, A.A., Ohnishi, K.: Robust time delayed control system based on communication disturbance observer with inner loop input. In: 36th Annual Conference on IEEE Industrial Electronics Society (IECON 2010), pp. 1621–1626. IEEE (2010)
8. Feben, D., Taras, M.J.: Studies on Chlorine Demand Constants. *Journal of American Water Works Association* 43(11), 922–932 (1951)

# Development of Ontology for the Diseases of Spine

Geun-Hye Kim<sup>1</sup>, Min-Jeoung Kang<sup>1</sup>, Chai Young Jung<sup>1</sup>, Joon-Yong Jung<sup>2</sup>,  
Seung Eun Jung<sup>2</sup>, Jin-Sung Kim<sup>3</sup>, Ji-Seong Jeong<sup>4</sup>, Do-Hyeong Kim<sup>5</sup>,  
Kwan-Hee Yoo<sup>5</sup>, Dongmin Seo<sup>6</sup>, Seungwoo Lee<sup>6</sup>, Seungbock Lee<sup>6</sup>,  
Sangho Lee<sup>6</sup>, and Sukil Kim<sup>1,\*</sup>

<sup>1</sup> College of Medicine, The Catholic University of Korea

<sup>2</sup> Department of Radiology, Seoul St. Mary's Hospital, College of Medicine,  
The Catholic University of Korea

<sup>3</sup> Spine Center, Department of Neurosurgery, Seoul St. Mary's Hospital,  
College of Medicine, The Catholic University of Korea

<sup>4</sup> Department of Information Industry, Chungbuk National University of Korea

<sup>5</sup> Department of Digital Information Convergence,  
Chungbuk National University of Korea

<sup>6</sup> Software Research Center,  
Korea Institute of Science and Technology Information,  
Daejeon, Korea

**Abstract.** KISTI is carrying out an e-Spine project for spinal diseases to prepare for the aged society, so called NAP.

The purpose of the study is to build a spine ontology that represents the anatomical structure and disease information which is compatible with Simulation Model of KISTI. The final use of the ontology includes diagnosis of diseases and setting treatment directions and by the clinicians. The ontology was represented using a 3D software.

Twenty diseases were selected to be represented after discussions with a spine specialist and two medical imaging specialists. Several ontology studies were reviewed before modeling the ontology. Reference books were selected for each disease by the spine specialist and were organized in MS Excel by a trained nurse. The contents were then reviewed by the specialists. Altova SemanticWorks and Protégé were used to code spine ontology with OWL Full model. Links to the images from KISTI and sample images of diseases were included in the ontology. The OWL ontology was also reviewed by the specialists again with Protégé.

We represented uni-directional ontology from anatomical structure to disease, images and treatment. The ontology was human understandable. It would be useful for the education of medical students or residents studying diseases of spine. But in order for the computer to understand the ontology, a new model with OWL DL or Lite is needed.

**Keywords:** Ontology, OWL ontology, Spinal disease.

---

\* Corresponding author.



## 1 Introduction

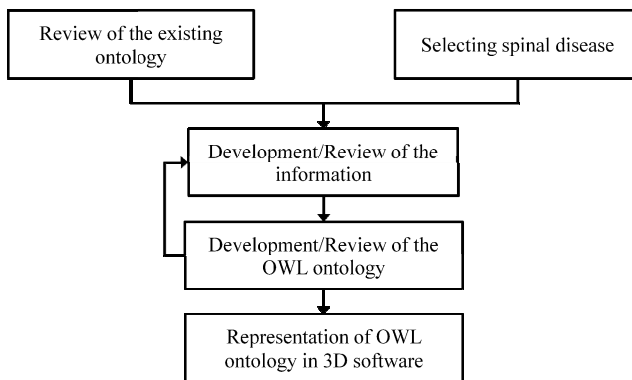
KISTI(Korean Institute of Science and Technology information) has been studying a National Agenda Project(NAP) for developing elderly human body model for treatment and rehabilitation of age-related spinal disorders. The purpose of the study is to build a virtual human spine as a simulation model through mathematical modeling to use in virtual experiment instead of real human spine. Accurate diagnosis and treatment of spinal diseases were expected from the project. We tried to develop the spinal ontology which contains information on spine and the related diseases for the success of the main project.

Ontology is systemized process accomplished by using computers for building a model which present the recognizable concepts and the relations between them [1]. Neches(1991) defined ontology as “basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary”[2] and Gruber(1993) defined it as “an explicit specification of a conceptualization” [3].

This study focused on developing spinal ontology with frequently occurring spinal diseases in Koreans. It contains anatomy of spine, method of treatment, cause, classification information related with spine. Further, the spinal ontology can be linked to the simulation model for education of medical students and for the physicians and biomedical engineers by offering the necessary information in their fields.

## 2 Method

The authors are composed of a wide range of professional researchers; medical informatics, computer professionals and clinical experts such as nurses, a neurosurgeon, and 2 imaging specialists. The research was conducted in 5 phases (Figure 1); (1) Review related to exiting ontology for the construction of the model (2) Selecting



**Fig. 1.** Process of building spine ontology

the spine related diseases and the subject of the research at the same time (3) Developing/Reviewing spinal ontology (4) Creating OWL ontology in accordance to the clinician's feedback (5) Review of the OWL ontology by the specialists. In addition, the ontology was represented by 3 D image software for the easy understanding.

## 2.1 Review of the Existing Ontology

We reviewed results of ontology project using Protégé which were led in 1987 as a public project at Stanford University[4]. Among their results, we gathered information that is necessary for spine ontology after analyzing the ontology of Rat anatomy and classification of diseases.

## 2.2 Selecting Spinal Diseases

The spinal diseases were selected based on the following three criteria. First, the disease has to be one of the highly occurring spinal diseases among Koreans. Second, the disease must occur in a specific area of the spine rather than throughout the whole spine. This makes it possible for the ontology to provide information that are suitable to the characteristics of the diseases among Koreans. Also, it is much easier to link with the simulation model of KISTI which is made according to Korean human being. Last, the OWL ontology of the disease should be able to be expressed on the computer so that it can be used in clinics or medical schools for education.

## 2.3 Development/Review of the Information

The developed ontology was organized according to the diseases with Microsoft Excel. The anatomic definitions that are consisted of ontology was referenced from a medical dictionary [5] and the disease related information was extracted from publications recommended by clinicians[6]. We classified anatomical information into two categories; anatomical location and anatomic properties were represented in OWL ontology (Table 1).

**Table 1.** Anatomical information represented in the ontology

Entry	Ontology	OWL expression
Location	Anatomical location	spine:isPartOf
Properties	Part Name	rdfs:label
	Anatomical classification	rdfs:subClassOf
	Standard code for the structure	spine:KOSTOM
	Definition and description of the structure	spine:definition
spine:description		

Disease related information was classified into five categories as shown in Table 2; anatomical location, property of the disease, symptom/sign, method of treatment, image were represented in OWL ontology.

The spinal ontology was reviewed by a neurosurgeon and two imaging specialists. Sample images of diseases, CT or M.R.I, were collected during the study period in Seoul St. Mary's hospital and linked to the diseases in the ontology.

**Table 2.** Diseases related information represented in ontology

Entry	Ontology	OWL expression	
Location	Anatomical disease location	spine:hasSite	
Disease Properties	Name of disease	rdfs:label	
	Classification of Diseases	rdfs:subClassOf	
	Apply the standard code	spine:KOSTOM	
	Definition of disease		spine:definition
			spine:description
	Clinical diagnosis	spine:diagnosis	
	Cause	spine:hasCause	
	Concomitant diseases	spine:hasConcomitantDisease	
Complication	spine:hasComplication		
Symptom /sign	Symptom	spine:hasSymptom	
		spine:causeOfSymptom	
	Sign	spine:hasSign	
Treatment	Surgical Treatment	spine:hasSurgicalTreatment	
	Non-surgical treatment	spine:hasNonSurgicalTreatment	
	Conservative Treatment	spine:hasTreatmentConservative	
Image	Preoperative image	spine:hasImageBeforeTx	
	Postoperative image	spine:hasImageAfterTx	

## 2.4 Development/Review of the OWL Ontology

Spinal OWL ontology was built based on OWL Full model which is a standard ontology language developed by W3C(World Wide Web Consortium). Both Altova SemanticWorks and Protégé were used to build and review the OWL ontology.

In the process of OWL representation, we tried to determine the level of expression in classes (Resource object) or individuals (Literal object). For example, if the disease related information is 'herniated nucleus pulposus', <herniated nucleus pulposus> is identified as superclass and <cervical intervertebral disc herniation>, <lumbar

intervertebral disc herniation>, <thoracic intervertebral disc herniation> which are classified under herniated nucleus pulposus are identified as subclass according to their location. Each class contains additional information related to their occurred region.

## 2.5 Representation of OWL Ontology in 3D Software

The contents of OWL ontology on spine was represented by using 3D image S/W. The software has three modules: 3D rendering module, OWL query module and the module for showing disease information that comes from the ontology. The users may select a part of spine image of question. Then a list of the diseases from the OWL spine ontology file will appear through OWL query operation. When a disease among the list is selected, the query module searches the disease related information such as causes, symptoms, diagnoses, treatment, complication, and image of the disease.

## 3 Result

### 3.1 Selected Spinal Diseases

The list of 20 selected diseases is presented in Table 3. If scientific papers were referred in addition to text book, they were added as references in the table.

**Table 3.** The twenty selected diseases

Atlas fracture	Ossification of ligament flavum
Degenerative marrow change (Modic type change)	Osteoarthritis in facet joint (Pfirman grade) [10] [11]
Grading of lumbar disc degeneration [7]	Osteoporosis
Hangman's fracture	Osteoporotic Compression Fracture
HNP(Herniation of Nucleus Pulposus) [8]	Scoliosis
Infectious spondylitis	Spinal stenosis
Kyphosis	Spondyloarthropathy
Meningocele[9]	Spondylolisthesis
Odontoid process fracture	Subaxial fracture( fractures in C3~C7) [12]
OPLL(Ossification of Posterior Longitudinal Ligament)	Thoracolumbar spine fracture

### 3.2 Development of the Ontology

The Figure 2 presents Protégé OWL ontology graph created from the spinal ontology of the twenty selected diseases. The Anatomical class represents the entire structure that composes the spine. There are 50 classes; 1 vertebral column, 5 vertebrae, 33 vertebrae and other 11 material of spine. Each of class has 6 properties. To express a

sentence 'C1 cervical is part of cervical vertebra' in OWL Full model, we defined anatomical structures 'C1', 'cervical vertebra' as a class and <is a part of> as a property representing a predicate. <Disease> was defined as a superclass and the 20 selected diseases were treated as classes. As a result, 21 classes were formed for the concept of diseases and each class has 18 properties.

More than 100 images that were collected were linked to 20 diseases, one per each. The predicates that link images to the other part of ontology include <hasImageBeforeTx> and <hasImageAfterTx>.

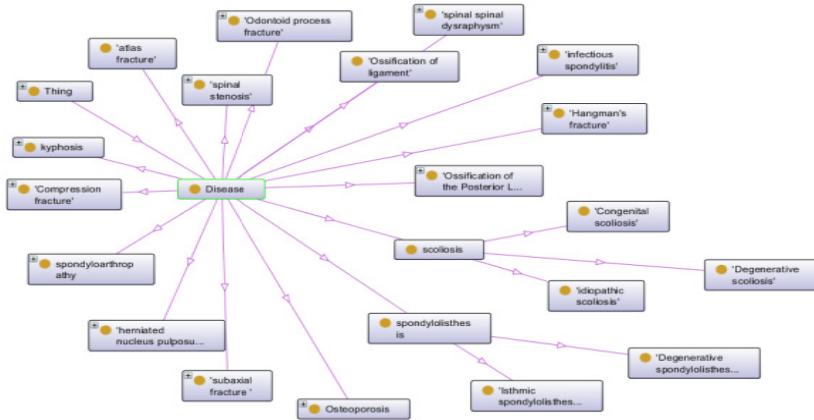


Fig. 2. List of the spinal diseases shown in Protégé

T1		
Osteoporotic Compression Fractures	Thoracolumbar-spine fracture	Osteoporosis
HNP	spondyloarthropathy	Ankylosing spondylitis
Ossification of ligament	OPLL	TB spondylitis
infectious spondylitis	kyphosis	scoliosis

Osteoporotic Compression Fractures		
Cause	Symptoms / Sign	Diagnosis
Complications /ConcomitantDisease	treatment	Image

Treatment

SurgicalTreatment  
 척추성형술(vertebroplasty)과 척추후골종전복원술 (kyphoplasty) : 경피적으로 시행할 수 있는 최소침습적 수술방법 -> 골다공증성 압박골절의 환자에서 통증의 완화가 빠르게 나타남  
 골다공증 척추체의 고정술(instrumentation of osteoporotic spine) : 분봉 경추형 접근법을 이용하여 전후의 기둥은 척추체절상핵 경방 피질골( anterior cortical bone)까지 고정내시가 도달할도록 하는 것이 중요함 -> 최근에는 골다공증 환자의 척추체형 골시멘트를 시행하여 고정외 효과를 극대화하는 방법 등도 쓰이고 있음.

Fig. 3. User interface of the proposed system

### 3.3 Representation Spine Ontology in 3D Software

The user interface is composed of three parts (Figure 3). (1) 3D rendering part for the whole spine where the users can rotate and move, the spine and zoom in and out of it on the left upper part and the part for the selected anatomical structure in a large scale at the bottom left. (2) The list of diseases selected by users on the right upper part. (3) The details of the disease specific information at the bottom right.

## 4 Conclusions

We built the ontology of spine with links to the cause, symptoms, method of treatment of highly occurred spinal disease among Koreans, and anatomical information.

The completed spinal ontology expresses anatomical connection of the parts of spine and their vertical relationships as well as information on the diseases in the spine. It is easy to understand the structure and the diseases of spine by conceptualizing the anatomical structure of spine and show them in 3D images.

This study was completed by the use of literal object of OWL Full model by expressing the contents of the main reference dictionary and publications about spine literally. But in order for the computer to interpret the ontology, a new model with OWL DL or Lite is needed. Further studies need to include the process of the transformation of literal object into resource object through the structuralization process of items completed by literal object, further systematizing the concept. In addition, the review of class and property is necessary to show the anatomical information of spine and information of diseases specifically. Also, the studies about methods which offer visual information are related to simulation model of KISTI.

**Acknowledgement.** This work was supported by 2013 National Agenda Project (NAP) funded by Korea Research Council of Fundamental Science & Technology (NAP-09-2) and National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST)(No. 2008-0062611).

## References

1. Mizoguchi, R.: The next generation web and critical technology of knowledge processing: dooyangsa (2009)
2. Neches, R., Fikes, R.E., et al.: Enabling technology for knowledge sharing. *AI Magazine* 12(3), 36 (1991)
3. Gruber, T.: A translation approach to portable ontology specifications. *Knowledge Acquisition* 5(2), 199–220 (1993)
4. Gennari, J.H., Musen, M.A., et al.: The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies* 58(1), 89–123 (2003)
5. Ji, J.: *Stedman's Medical Dictionary*. Koonja Publishing (2006)
6. Society TKS. *The textbook of Spine* (2008)

7. Pfirrmann, C.W., Metzdorf, A., et al.: Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)* 26(17), 1873–1878 (2001)
8. Modic, M.T., Ross, J.: Lumbar degenerative disk disease. *Radiology* 245(1), 43–61 (2007)
9. Ross Krm, J.S., Borg, B., et al.: *Diagnostic Imaging*. Amirsys, Inc, Spine (2010)
10. Pathria, M., Sartoris, D.J., et al.: Osteoarthritis of the facet joints: accuracy of oblique radiographic assessment. *Radiology* 164(1), 227–230 (1987)
11. Weishaupt, D., Zanetti, M., et al.: MR imaging and CT in osteoarthritis of the lumbar facet joints. *Skeletal Radiol.* 28(4), 215–219 (1999)
12. Johan, W.M., Gothem, V., et al.: *Spinal Imaging: Diagnostic Imaging of the Spine and Spinal Cord*. Springer (2007)

# Implementation of Information Retrieval Service for Korean Spine Database with Degenerative Spinal Disease

Dongmin Seo, Seungwoo Lee, Seungbock Lee, Sangho Lee,  
Hanmin Jung, and Won-Kyung Sung

Software Research Center, Korean Institute of Science and Technology Information,  
Daejeon, South Korea

{dmseo, swlee, sblee, shlee, jhm, wksung}@kisti.re.kr

**Abstract.** Many people are interested in medical information because they want to take care of their health by themselves. Thus, some institutions provide medical term search services to help their understanding about difficult medical terms. However, those services are suffered from the low quality of search results because simple keyword search technologies are applied to medical term databases and Web sites. To prevent unnecessary spinal surgery and to support scientific diagnosis of spinal diseases and systematic prediction of treatment effects, we have been developing *e-spine*, which is a computerized simulation model of human spines. In this paper, as a background data for realizing *e-spine*, we have collected spine data from 77 cadavers and 298 patients with normal spine or degenerative spinal diseases. The spine data consists of 2D images such as CT, MRI, or X-ray, 3D shapes, geometry data and property data. Especially, we propose our database and the bioinformatics linked data based spine information retrieval service that provides more user-friendly services and very precise results. To provide abundant medical knowledge, our service provides search results of MeSH, OMIM, UniProt, GeneOntology and DBpedia. Also, the search results of PubMed are displayed to provide the related papers and books. As a result, our database will offer great value and utility in the diagnosis, treatment, and rehabilitation of patients suffering from spinal diseases.

**Keywords:** *e-Spine*, Korean Spine, Spine Database, Spinal Diseases.

## 1 Introduction

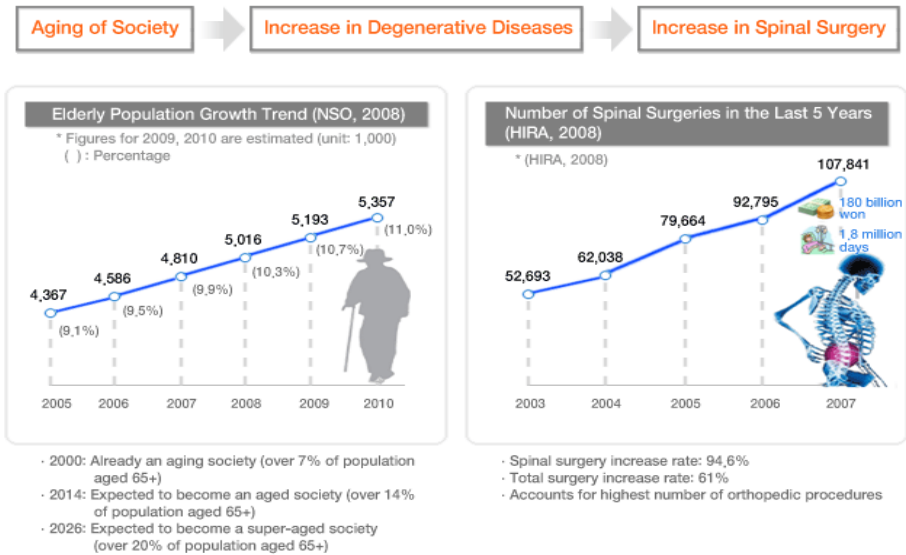
By 2026, Korea is expected to surpass the UN's definition of an aged society and reaches the level of a "Super-aged society". As a result, degenerative spinal diseases and related surgical procedures will increase exponentially. As of 2007, medical expenses incurred due to spinal surgery in Korea totaled 178.6 billion won/year, and the treatment duration reached 1.82 million days/year as shown in figure 1. The resulting medical burden and economic loss are increasing at a rapid rate. Spinal diseases make everyday life of people impossible and impede economic activities, resulting in a compromised quality of life. Between 2002 and 2004, spinal surgery



increased at a particularly high rate among the older demographic, by 68.2% among ages 60-69 and by 94.6% among those aged 70 and older. Among the leading causes of hospitalization for ages 65 and older, spinal diseases ranked at No. 2 with over 65,000 instances [1].

To prevent unnecessary spinal surgery resulting from over-treatment, systematic prediction of treatment effects is required, including scientific diagnosis, scientific effect analysis, and analysis of spinal rehabilitation exercises. Computer simulations have been utilized in biomechanical research for the past three decades. Today, advancements in computer hardware and software are bringing continually increasing simulation accuracy. We have been developing *e-spine*, which is a computerized simulation model of human spines created by mathematically calculating images, geometries and properties of human spines and will allow virtual testing without using a real human spine. However, the high quality spine data are essential for realizing *e-Spine*. Therefore, as a background data for realizing *e-spine*, we have produced and collected many images, geometry and property data of spines from Korean cadavers and patients with normal spine or degenerative spinal diseases [2,3]. In particular, we implemented a meta-data based spine data retrieval service for efficiently searching and managing the data. Also, we proposed our data and the bioinformatics linked data based spine information retrieval service to provide abundant medical knowledge [4].

In this paper, we present our spine database and spine information retrieval service in detail. The rest of this paper is organized as follows. Section 2 explores previous work related to *e-Spine* and digital human data. Section 3 explains spine data obtained from cadavers and patients. Section 4 describes our spine information retrieval service on Korean spine with degenerative spinal disease. Finally, section 5 presents the conclusion.



**Fig. 1.** Status of an aged society and spinal surgeries in Korea

## 2 Related Works

### 2.1 *e*-Spine

*e*-Spine is a computer-run simulation model created by mathematically modeling collected human spinal image data, which allows virtual testing without using a real spine. Figure 2 shows the comparison with a vehicle's navigation system. For optimal and safe driving, a navigation collects map information and models a map and predicts a route. Similarly, for optimal treatment, *e*-Spine collects spine images and models 3D spine models and predicts virtual testing and results. The expected effect of *e*-Spine is as follows. However, the high quality spine data are essential for realizing *e*-Spine.

- Acquisition of reliable, economic, advanced IT-based medical support technologies that can be used in the diagnosis and treatment of degenerative spinal diseases
- Strengthened market competitiveness for Korea's medical equipment industry through the utilization of *e*-Spine
- Reduction of medical expenses and improvement of the quality of life during old age by making available reliable, affordable IT-based medical technologies

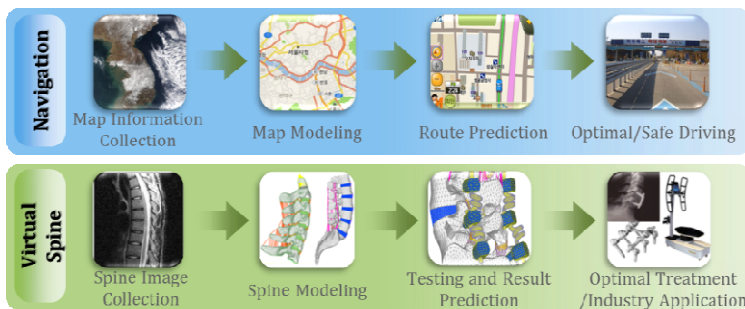


Fig. 2. Comparison between a vehicle's navigation and *e*-Spine

### 2.2 Digital Human Data

Digital human data such as CT, MRI, or X-ray collected from cadavers and patients are used in academic, clinical and industrial researchers to improve their understanding of human physiology and pathology and diagnosing of human diseases. There have been several research projects which construct digital human data in China and EU as well as U.S. The first project was VHP (Visible Human Project) [5] which was planned by U.S. NLM (National Library of Medicine) in 1989. This project had developed and opened digital human data sets including continuously sectioned color images, CT (Computerized Tomography), and MRI (Magnetic Resonance Imaging) of whole bodies of a man and a woman. China started its own digital human project, CVH (Chinese Visible Human) [6], in 2002. They constructed

digital human data sets consisting of continuously-sectioned color images, CT and MRI of whole bodies of Chinese male and female. The sectioned images have higher resolution and narrower intervals in part than those of VKH. EU developed VPH (Virtual Physiological Human), a framework which aims to enable collaborative investigation of the human body and was inspired from Physiome project. EU also developed a biomedical data management and sharing service, called Physiome Space, for collecting and sharing very large collections of biomedical data between researchers, through LHDL (Living Human Digital Library) project [7]. However, various human data such as 3D shape model, geometric and physical data are required to derive predictive hypotheses and simulations for developing and testing new therapies.

### 3 Korean Spine Database with Degenerative Spinal Diseases

Korean physical function is different from foreign physical function. Unlike these previous works, we focus on aged Korean spines with degenerative spinal diseases to construct database for *e*-spine. Therefore, we collected various images produced by CT, MRI or X-ray from many aged Korean cadavers and patients having degenerative diseases on cervical and lumbar vertebrae or inter-vertebral discs. Also, to support abundant information about Korean spine, we made 3D shapes from series of CT images and measured geometry and property data of spines. The construction process of our database is as follows.

- (a) Select degenerative spinal disease patients and cadavers with an age of not less than 60 and not more than 80 without serious spinal damage in an accident.
- (b) Product cross-sectional images such as CT, MRI or X-ray on spinal areas. It is ultra-high-resolution images obtained by 128-channel dual source MDCT.
- (c) Product 3D spinal models that are generated with 3D models of C1-C7 (neck bone), T1-T12 (hucklebone), L-S1 (backbone and sacrum). We use Mimics software to product 3D spinal models and ensure connection between CT images and 3D spinal models.
- (d) Measure physical and clinical bone mineral densities. For physical bone mineral density, we calculate BMD of sponge bone areas at CT images through the HU No. conversion equation ( $r=1.122*HU+47$ ). Clinical bone mineral density is measured by DEXA equipment.
- (e) Measure motion property and compression strength for a spine in case of cadavers. The measurement of motion properties for functional spine performs functionality test on load-motion curve. And, the measurement of compression strength for vertebral body performs compression strength on load-displacement curve. The measurement processes of motion property and compression strength for a spine are as follows. Firstly, we extract a spine from a cadaver and divide that into neck bone, upper hucklebone, lower hucklebone and backbone. Then, we measure motion property and compression strength on the divided bones through MTS spine simulator.

To date, we have collected Korean spine data from 77 cadavers and 298 patients with normal spine or degenerative spinal diseases. The collected data covers seven diseases on cervical vertebrae and discs and ten diseases on lumbar vertebrae and discs. Types of diseases in each vertebra are as follows.

- Cervical : disc degeneration, disc height reduction, disc herniation, endplate sclerosis, OPLL, ossification, osteophyte
- Lumbar : compression fracture, disc degeneration, disc height reduction, disc herniation, endplate sclerosis, facet joint degeneration, LDK, osteophyte, osteoporosis, spondylolisthesis

## 4 Korean Spine Information Retrieval Service

We implemented the meta-data based Korean spine data retrieval service. As shown in figure 3(a), the service finds Korean spine specimens that have the specimen identifier, sex, age, height, weight, clinical bone density, type and grade of the spinal disease entered by user. Figure 3(b) shows the statistical information on Korean spine data in our database. So, users can understand the spine and spinal disease classification in our database easily. Figure 3(c) shows the spine information navigation. The spine information navigation supports all spine information on Korean spine specimen selected in figure 3(a). The spine information consists of cross-sectional images, 3D geometric model, reference dimensions, physical and clinical measurements of bone mineral density and measurements of motion properties and compression strength on spinal areas. Korean spine data serves to analyze the correlation between the image data and the physical properties data on Korean spine data. Also, our service supports CT and STL images on Korean spinal specimens. CT images are obtained by ultra-high-resolution images with the interval of 0.6mm and the pixel dimension of less than 0.5mm on spine areas. STL images are obtained by 3D modeling on CT. Existing medical viewers support the alternative of CT viewer or STL viewer. As shown in figure 3(d), our service supports CT & STL viewer. Therefore, user can analyze the correlation between cross-sectional images and 3D geometric model efficiently.

In particular, we implemented the bioinformatics linked data based spine information retrieval service that provides more user-friendly services and very precise results. Figure 4 shows the our service that gets url and linked data for the entered keyword in MeSH, OMIM, UniProt and GeneOntology databases. The databases are the representative databases on bioinformatics linked data and provide 654,198, 765,384, 338,602,962 and 338,602,962 triples, respectively [8]. Also, the searched linked data are stored in our database to improve the search performance by recycling of that. Our service displays the searched linked data on figure 4(2). To provide an abundance of medical knowledge, the search results of DBpedia and Wikipedia are displayed on figure 4(1) and 4(3). Also, to provide the related papers and books, the search results of PubMed are displayed on figure 4(4).

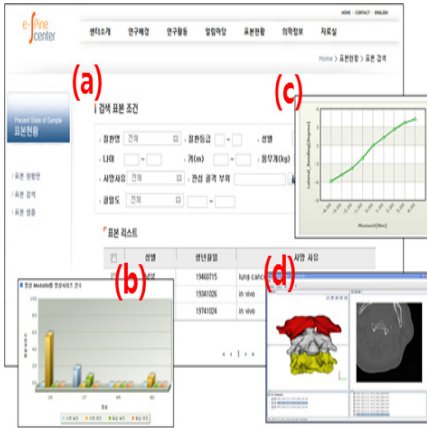


Fig. 3. Meta-data based spine data retrieval service



Fig. 4. Spine information retrieval service

## 5 Conclusions

We collected various Korean spine data from 77 cadavers and 298 patients with normal spine or degenerative spinal diseases. Also, we implemented a meta-data based spine data and spine information retrieval services to provide an abundance of spinal knowledge. We will offer our spine data and services to many researcher and doctors for a spine research development.

**Acknowledgement.** This work was supported by 2013 National Agenda Project (NAP) funded by Korea Research Council of Fundamental Science & Technology (NAP-09-2/P-13-JC-LU01).

## References

1. National Health Insurance, “2009” major surgery statistics. Press Release 12(7), 1–15 (2010)Release
2. e-Spine project, <http://www.espine.re.kr>
3. Seo, D.M., Lee, S.W., Lee, S.B., Lee, S.H., Jung, H.M., Sung, W.K.: Implementation of spine database system on Korean patient with degenerative spinal diseases. In: Proc. of International Smart Media Application (SMS), pp. 1–3 (2012)
4. Belleau, F., Nolin, M.A., Tourigny, N., Rigault, P., Morissette, J.: Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J. of Biomed. Inform. 41, 706–716 (2008)
5. Visible Human Project (VHP), <http://www.nlm.nih.gov/research/visible>
6. Chinese Visible Human (CVH) project, <http://www.chinesevisiblehuman.com>
7. Living Human Digital Library (LHDL) project, <http://www.livinghuman.org>
8. Nolin, M.A., Ansell, P., Belleau, F., Idehen, K., Rigault, P., et al.: Bio2RDF network of linked data (2008), <http://www.cs.vu.nl/~pmika/swc-2008>

# Automatic Surface Mesh Intersection Algorithm of Spine and Implant FEM Models

Dukyun Nam<sup>1</sup>, Ikhwang Choi<sup>2</sup>, Younho Kim<sup>2</sup>, Wanho Jeon<sup>2</sup>, and Kum Won Cho<sup>1</sup>

<sup>1</sup> Korea Institute of Science and Technology Information (KISTI)  
245 Daehak-ro, Yuseong-gu, Daejeon, 305-806, Korea  
{dynam,ckw}@kisti.re.kr

<sup>2</sup> CEDIC Co. Ltd., 371-50 Gasan-dong, Kumcheon-gu, Seoul, Korea  
{ihchoi,roseho,whjeon}@cedic.biz

**Abstract.** We propose the automatic surface mesh intersection algorithm of spine and implant models. The spine-implant intersection has unique constraint: the implant model should not change as much as possible because spine-implant structure analysis simulates the process to insert a strong durable implant into a relatively weak durable spine. Through the actual simulation, we show there is no abnormal results due to automatically generated intersection regions.

**Keywords:** Automatic surface mesh intersection, spine, implant, FEM.

## 1 Introduction

The human spine is important structure to bear the diverse and complex load in everyday life. If we have a spinal damage or deformation caused by accident or aging factor, we can't keep up the normal activities. As one of the ways of improving the spinal condition, e.g., spinal deformity, fractures, degenerative disease, etc., there is the spinal surgery with implant. If spinal surgery is incorrectly performed, it can lead to a very serious situation. So preliminary surgical planning should be preceded.

The computer simulation using the finite element method (FEM) model of spine and implant is one of methodologies for the preliminary surgical planning. To get the FEM models, researchers basically conduct the pre-processing work. They build computer-aided design (CAD) models of spine and implant, i.e., three-dimensional structural models, and get FEM models from the CAD models. Here, for the spine-implant simulation, spine and implant FEM models need to be merged into one FEM model. This is not only manually done so far, but also time-consuming work.

In this paper, we propose the automatic surface mesh intersection algorithm of spine and implant models. There have been many researches of surface mesh intersection with multiple models [1, 2, 3, 4]. However, the spine-implant intersection has unique constraint, i.e., the implant model should not change as much as possible because the most of implant is solid materials such as iron, aluminum, etc. We use and modify the most methodologies with minimal changes in the existing intersection algorithms [1, 5]. In the intersection tracking algorithm, we exploit the intersection position in implant's surface as default positions in order not to change the implant model. Based on these positions, we automatically re-mesh the spine-implant intersection model to be

valid for finite element analysis (FEA). This makes it possible to run the FEA using the spine-implant mesh model without any manual effort.

The remainder of the paper is structured as follows. Section 2 presents a description of the data structure. Section 3 explains the proposed intersection algorithm. Section 4 shows the simulation result with the mesh models generated by the proposed algorithm. The conclusion is presented in Section 5.

## 2 Data Structure Definition

To implement an intersection algorithm, we need to define the mesh storage data structure. The data structure can access neighbor vertices, edges, and faces which are linked each other [5]. We basically use the existing data structure in [1] and add one factor which can distinguish between spine and implant meshes.

### 2.1 Mesh Entities

A mesh is composed of vertices, edges, and faces. Each component in the mesh has the information of neighbor components to track neighbors. A vertex  $V$  has the coordinate such as  $(x, y, z)$  in 3 dimensional space and the face containing the vertices. The reason why we call the face, not the triangle, is that the proposed algorithm can be directly applied into the rectangle. For the sake of the simplicity, however, we use the triangle even the face.

$$V = (x, y, z, F)$$

A segment is the unit of edge, the connection of two vertices  $(V_0, V_1)$ , and has the information of each connected neighboring segment  $(S_0, S_1)$  and the face which the segment is located.

$$S = (V_0, V_1, S_0, S_1, F)$$

A face has three vertices  $(V_0, V_1, V_2)$ , three segments  $(S_0, S_1, S_2)$ , and three neighbor faces  $(F_0, F_1, F_2)$  containing the segments.

$$F = (V_0, V_1, V_2, S_0, S_1, S_2, F_0, F_1, F_2)$$

Although a vertex and a segment are located in multiple faces, they have only one face, respectively. Meanwhile the face has neighbor faces. So we can track the neighbor vertices and segments as well.

### 2.2 Intersection Entities

To find the intersection curve between spine and implant models, we need to define the intersection point  $P$ . In the process of rearranging points on the intersection curve, the property of points varies according to types of points or faces containing the intersection point. The topological information  $T$  about  $P$  within  $F$  is as follows.

$$T(P, F) = (F, r, n)$$

P is the intersection point on the face F. r and n are the information of the intersection point. r = 0 (resp. 1, 2) if P lies on a vertex (resp. on an edge, or inside T). n is the index of the entity r in Figure 1.(b).

When two faces  $F_0$  and  $F_1$  are intersected, the intersection structure has two of topological information because the intersection point is located in two faces.

$$I_p = (P, T(P, F_1), T(P, F_2))$$

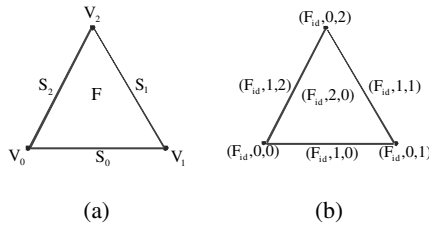


Fig. 1. Data structure. (a) Mesh entities. (b) Intersection entities.

### 3 Intersection Algorithm

We applied the tracking algorithm in order to rapidly and accurately explore an intersection. The tracking algorithm is to find the intersection regions along the intersection-curve from a valid tracking point [1, 3]. After finding an initial intersection point, the algorithm starts at the initial intersection point and creates intersection-curve along the direction to new intersection points. While this happens, the algorithm also searches for intersection region.

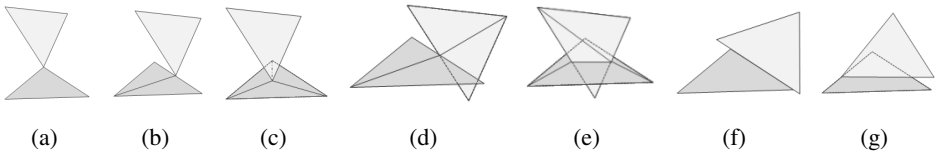
#### 3.1 Finding Intersection Point

##### 1) Searching intersection region

A plane equation is derived from the outer product of the points of mesh elements. An intersection point is calculated by using topological relation between two intersection points. After that, we derive the angle from the inner product of three points of a triangle element. If the sum of the derived angles is  $2\pi$ , the intersection point is inside the triangle element.

The intersection of planes on three-dimensional space has various cases. When the tracking algorithm searches an intersection, we determine the intersection from an interrelation of a line and a side and there are cases of intersection between intersected triangles in Figure 2.





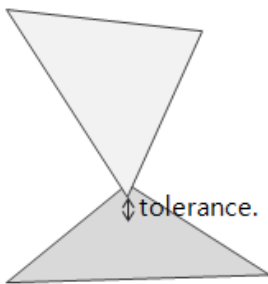
**Fig. 2.** Searching intersection regions. (a) Coincident nodes intersection. (b) Nodes along edge intersection. (c) Nodes onto face intersection. (d) Edges cutting other edge intersection. (e) Edges cutting face intersection. (f) Edges overlapping other edge intersection. (g) Edges overlapping face intersection.

**2) Tolerance**

The intersection cannot be mathematically determined because the numerical calculation of a computer does not work on consecutive space. We have to define a tolerance to calculate an intersection.

Suppose that the gap has the difference  $d$  like Figure 3. If  $d \geq \text{tolerance}$ , two triangles are not intersected. Otherwise they are intersected. When two meshes are intersected, the shape of intersected meshes changes. This is because the intersection points are derived from the tolerance by moving existing shapes.

The shape of an implant in a spine-implant insertion model should not change as using the tolerance. Algorithm 1 defines an intersection generation procedure with the tolerance. In case of an intersection point between edge  $E$  and face  $F$ , if  $E$  is an element of an implant mesh, the intersection point is either  $p$  or  $q$  which is start or end points of an edge  $E$ . On the other hand, if  $F$  is an element of an implant mesh, the intersection point is generated by intersecting  $F$  and the extension of  $E$ . The tolerance is automatically set up depending on the size of a model or by user’s configuration.



**Fig. 3.** Node-Face intersection with tolerance

```

Input : E : edge , F : surface mesh, tolerance
Output : P : intersection point
p, q : point of E | a, b, c : point of F;

if  $E \in M_{\text{implant}}$  ,  $F \in M_{\text{spine}}$ 
  if Distance : p to abc < tolerance
    Return P = p;
  else if Distance : q to abc < tolerance
    Return P = q;
else if  $E \in M_{\text{spine}}$  ,  $F \in M_{\text{implant}}$ 
  if Distance : p to abc < tolerance
    Return P = Create new point (qp, F);
  else if Distance : q to abc < tolerance
    Return P = Create new point (pq, F);
    
```

**Algorithm 1.** Tolerance

### 3.2 Tracking Algorithm

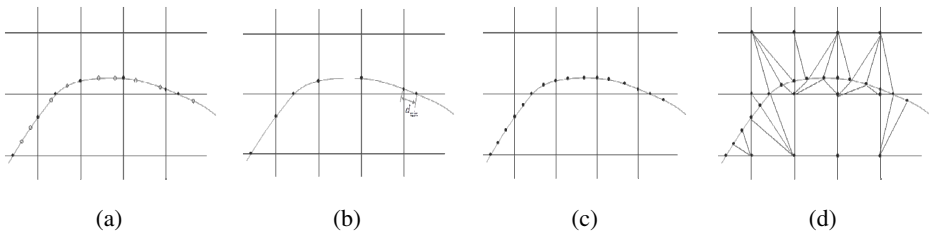
In the model for volume finite element, an intersection curve is always closed. So tracking an intersection can process starting from an initial intersection point,  $P_0$  to the intersection progress direction. After finding an initial intersection point, the intersection points continuously searched along the intersection progressing direction [1, 3, 4]. An intersection curve is continuously generated by inputting the continuously generated intersection points to an intersection curve  $C$ . By using the intersection curve data, we retrieve intersection regions and generate new mesh of the intersection region.

### 3.3 Meshing

To generate a spine-implant finite element intersection model, the generated mesh on an intersection region should satisfy three following conditions.

1. The shape of spine model can be changed but the shape of the implant cannot be changed. This is because this spine-implant structure analysis simulates the process to insert a strong durable implant into a relatively weak durable spine. If we allow the shape change of the implant, the non-intended stress concentration can occur. So the implant shape does not allow the change.
2. We do not generate more meshes than user's needs. The size and the shape of a mesh work as an important factor. So a small size of mesh is densely formed in a complicated shape. However, if the number of meshes is blindly large, we waste more analysis resources than needs.
3. A mesh should have the right quality for the finite element analysis. The intersection point, which can only be created by an intersection search, has an unbalanced gap. A mesh with the intersection points is not enough for finite element analysis. So we need to add the fixed points along the intersection curve to create a mesh for finite element analysis.

When generating the intersected mesh model, we have to satisfy the above three requirements. First of all, the points of the intersection curves should be reorganized. Algorithm 2 reorganizes the points of the intersection curve. Figure 4 (b) shows the points to define the shape of an implant among the points of an intersection curve.  $d_{\min}$  is defined as the minimum gap of all other points. The unequally distributed points  $P_n$  on the cross curve are reorganized equally based on the  $d_{\min}$  standard.



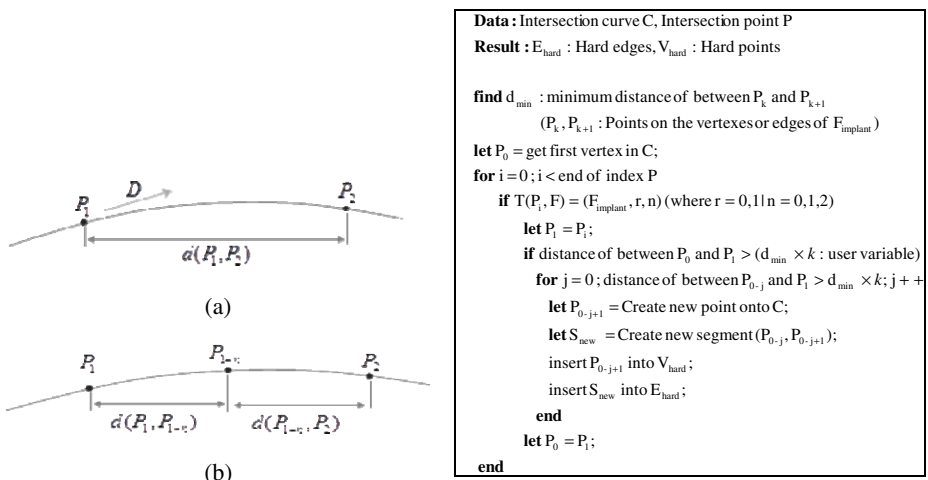
**Fig. 4.** Meshing algorithm. (a) Intersection curve. (b) Intersection points of  $M_{\text{implant}}$ . (c) Intersection hard edges. (d) Triangulation.

The intersection points created on the points or the lines of an implant element are presented as followings.

$$T(P, F_{\text{implant}}) = (F_{\text{implant}}, r, n) \quad (r = 0, 1 \mid n = 0, 1, 2)$$

The algorithm automatically create new element network to keep the shape of an original model by using Delaunay Triangulation in Figure 4 (d). The element created by Delaunay Triangulation may be not suitable for finite element analysis. Therefore the algorithm reorganizes the triangulation element in a spine model though re-mesh process. In addition, we re-mesh the intersection element with neighbor elements to prevent a sharp form.

The tracking algorithm searches the continuous two points,  $P_1$  and  $P_2$  to satisfy this condition. In Figure 5, we generate the points,  $P_{1-n}$  with an uniformed gap. The algorithm stores the points on the new generated curve and constructs the line,  $S_{\text{new}}$ , between newly generated points. New organized line is sequentially stored as a hard edge,  $E_{\text{hard}}$ . The hard edge is fixed and a base form on mesh or re-mesh processes.



**Fig. 5.** Remeshing algorithm. (a)  $P_1$  and  $P_2$  with distance  $d$  on the intersection curve. (b) New generated point  $P_{n-1}$  between  $P_1$  and  $P_2$ .

**Algorithm 2.** Node generation algorithm on the intersection curve to prevent the change of implant model

### 3.4 Inserting Implant

After processing triangulation and re-mesh, we remove the existing elements of intersection to insert reorganized elements. The algorithm intersects two finite element models by creating the inserted part inside a target element model. We search the intersection part with the outer product of surface elements along intersection curve.

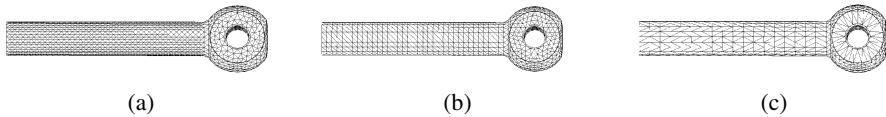


**Fig. 6.** Inserting implant. (a) Divided regions depending on the intersection curve. (b) Combined regions to insert implant.

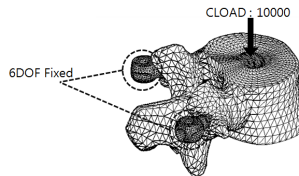
In Figure 6, the algorithm removes the region where an implant is inserted in the spine model after searching the part of the implant inserted to the spine model. It finishes an automatic intersection processing of an implant and a spine by inserting the implant depending on the direction to the spine model.

#### 4 Structural Analysis Application

A mesh is closely related to an analysis result and the intersection processing in the intersection regions influences the existing shape. Therefore, in most cases, a mesh is manually created to guarantee the mesh quality. We show that the proposed algorithm create an appropriate mesh for structural mesh. The algorithm creates a mesh by automatic intersecting of spine and implant models. As the spine model, we use three layered spine model which consists of vertebral arch, outer vertebral body, and inner vertebral body. To evaluate the usefulness of the automatic intersection algorithm, we prepare three automatic intersection models with different sizes of implants such as mesh size 1.0, 1.5, and 2.0 in Figure 7.

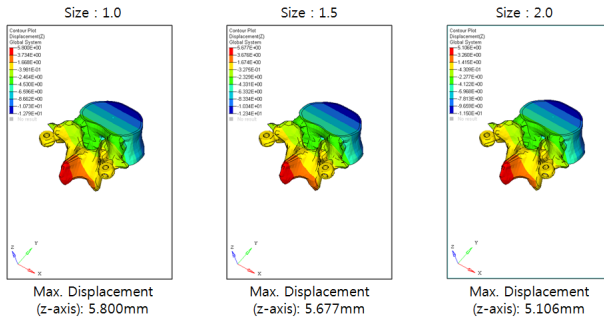


**Fig. 7.** Three different implant models. (a) Mesh size 1.0. (b) Size 1.5. (c) Size 2.0.



**Fig. 8.** Boundary and load condition for evaluation

In Figure 8, we set up the analysis condition: 10,000 load to  $-Z$  direction on top of spine with fixed 6 degrees of freedom (DOF). We performed the analysis by using ABAQUS solver.



**Fig. 9.** Analysis results

Figure 9 shows the common patterns even though there is a little derivation depending on the element size. We cannot find the abnormal results due to intersection regions. Therefore we expect the proposed automatic intersection algorithm be used without human intervention.

## 5 Concluding Remark

In this paper, we propose the automatic surface mesh intersection algorithm of spine and implant models. There have been many researches of surface mesh intersection with multiple models. However, the spine-implant intersection has unique constraint, i.e., the implant model should not change as much as possible because the most of implant is solid materials such as iron, aluminum, etc. We use and modify the most methodologies with minimal changes in the existing intersection algorithms. In the intersection tracking algorithm, we exploit the intersection position in implant's surface as default positions in order not to change the implant model. Based on these positions, we automatically re-mesh the spine-implant intersection model to be valid for finite element analysis (FEA). This makes it possible to run the FEA using the spine-implant mesh model without any manual effort. In the near future, we plan to perform an additional test for a complicated model to validate the proposed algorithm.

**Acknowledgement.** This work was supported by 2013 National Agenda Project (NAP) funded by Korea Research Council of Fundamental Science & Technology (NAP-09-2).

## References

1. Elsheikh, A.H., Elsheikh, M.: A reliable triangular mesh intersection algorithm and its application in geological modeling. *Engineering with Computers* (2012)
2. Lo, S.H., Wang, W.X.: An algorithm for the intersection of quadrilateral surfaces by tracing of neighbours. *Computer Methods in Applied Mechanics and Engineering* 192, 2319–2338 (2003)
3. Lo, S.H., Wang, W.X.: A fast robust algorithm for the intersection of triangulated surfaces. *Engineering with Computers* 20, 11–21 (2004)
4. Lo, S.H.: Automatic mesh generation over intersecting surfaces. *International Journal for Numerical Methods in Engineering* 38, 943–954 (1995)
5. Mucke, E.P.: Shapes and implementations in three-dimensional geometry. PhD thesis, Department of Computer Science, University of Illinois at Urbana-Champaign (1993)

# Simulation Program Binding Technology Based on Supercomputing

Sun-Rae Park<sup>1</sup>, DuSeok Jin<sup>1</sup>, Kyu-Chul Lee<sup>2</sup>, and Kum Won Cho<sup>1,\*</sup>

<sup>1</sup> Korea Institute of Science and Technology Information,  
Supercomputing Center, Rep. of Korea  
{srpark, dsjin, ckw}@kisti.re.kr

<sup>2</sup> Chungnam National University, Computer Engineering, Rep. of Korea  
kclee@cnu.ac.kr

**Abstract.** There has been a rapid improvement every year in the performance of supercomputer, as 10 times in the past three years, as 1000 times in the past ten years. Computational science simulation is on the rise to the third method as successor of theory and experiment in medical field as well as physics and engineering. However, despite demands on a variety of fields such as medical treatment, computational engineering, researchers should spend enormous cost and time developing computer simulation program. Also, simulation program developed on this way usually dies out only after performing single separate function. To improve this problem, in this paper, we manage informal processes which produce new procedure using unstructured scientific workflow to share information about simulation program developed by users through storage and to provide optimal integrated environment for problem solving demand.

**Keywords:** Supercomputing, Simulation, Science workflow, Medical Science.

## 1 Introduction

There has been a rapid improvement in the performance of supercomputer every year, as 10 times in the past three years, as 1000 times in the past ten years[1]. For this reason, the industry world expects an innovation in a variety of fields, such as medical field, CAE, using computational science simulation with high-end supercomputing resource speed and which performs widely differing from original simulation qualitatively and quantitatively.

Computational science is on the rise to the third method as successor of theory and experiment and is expected to get the largest improvement in the 21<sup>st</sup> century.

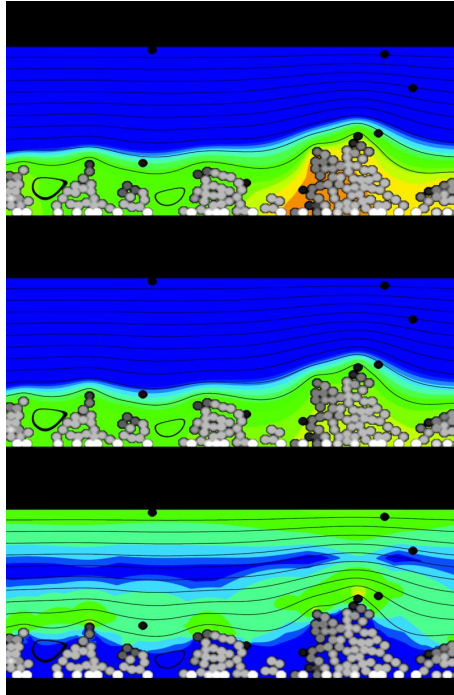
For example, through computer simulation, virtual-copy data about the patient's blood can be used by doctors as important references. Using this information, like Fig. 1. we can compute the possibility of heart attack from blood coagulation with hurt coronary artery, and test whether drugs like aspirin are effective to relieve blood

---

\* Corresponding author.

clotting[2]. But, patient customized simulation with multi-scale for blood function is just one illustration of the system biology field advancing at fast-growing rate.

To complete multi-scale model, doctors or researchers should understand thousands of intracellular messages related to damaged vein, as well as specific computation about blood flow and molecular diffusion. Also, they should understand the origin about complicated and various situations as doing air-flow simulation for weather prediction and aircraft design.



**Fig. 1.** A visualization of simulated platelets under flow conditions. (Credit: Image courtesy of University of Pennsylvania)

In this way, large-scale computer simulation can be a chance to provide individual customized medical solution. However, though there are lots of demands for computer simulation program in many fields, it consumes enormous cost and time. Also, simulation program developed on this way usually dies out only after performing single separate function. To improve this problem, in this paper, we propose simulation program binding technology which can provide new service converging simulation programs based on user's problem solving demand.

## 2 Scientific Workflow

As the application of computation science becomes more and more complicated, to make complicated procedures simple and automatic, researches about scientific



workflow go through several computational science fields[3-5]. Also, workflow is usually used to automatize (re)structured business process as one of process re-engineering tools[6].

Pegasus-WMS(Pegasus Workflow Management Service)[3] is the workflow management system which has been used applying to a variety of science fields such as astronomy, biology, and it manages implement of complicated workflow on bulks of distributed computing resources. When workflow is made up, Pegasus-WMS executes work included in workflow in appropriate order defined by user, through connecting processes to work on computing resources which can use workflow. Also, included component provides function to make workflow asked by user feasible form, which enables to find appropriate resources and to connect them each together. Pegasus-WMS provides effective management service in workflow execution and management, but, actually for researchers in various science fields to use the system, researchers need special set-up in their own computing environment and they should use specific work scheduler, Condor, which makes it possible to communicate with Pegasus-WMS in their own cluster. Therefore, it comes up to provide additional interface which makes researchers use these system easily and actively.

Taverna[4] is scientific workflow management tool focused on biology. By supporting execution based on web-service, it connects with web-service directly through provided workbench or makes user use workflow through web-site. User can design workflow by composing each processor and link using panel interface expressed visually, then it provides results to user in text file form or through connecting with visualization processing software directly.

Triana[5] is open source problem solving environment software which supports to compose scientific workflow on astronomy and physics. It enables grid service connection and provides constructible and installable components based on visual interface which guarantees user-friendly tool for science researchers. However, three researches looking previously have a disadvantage for beginner not to use easily because of demands of additional set-up.

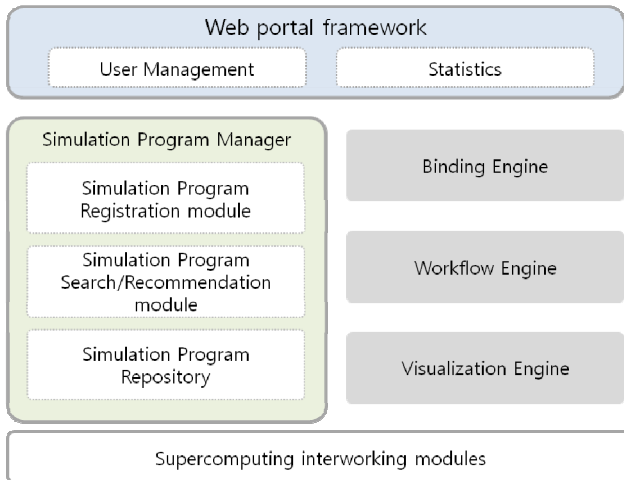
### **3 Design of Simulation Program Binding Technology**

Previous scientific workflow is linked with original desktop-application program or host program successively. This paper, however, presents simulation program binding technology, which helps users share information about simulation program developed by users through storage. It also manages informal processes that produce new procedure using unstructured workflow to provide optimal integrated environment for problem solving demand.

The important thing in providing computational science simulation environment is to provide a variety of simulation programs to users.

Simulation program binding technology is provided with web-based form to increase access convenience of user and consists of a simulation program manager, binding engine, workflow engine, visualization engine and supercomputing linkage model for link with resources. The simulation manager provides registration environment for

simulation program to offer used or new research result through web-based simulation execution environment. Its main functions are code store which help researchers register and search simulation program through web and recommendation function on simulation program users want. Also, provides IDE based on private virtual server which includes library/tool/API for developing and testing simulation program.



**Fig. 2.** Shows the structure of simulation program binding technology based on supercompute

Additionally, it presents automatic function producing web-interface to use developed simulation program through web-based environment easily, and metadata to integrate through binding engine.

Simulation workflow engine supports workflow process with preprocessor, simulation tool and post-processor. Therefore, the system provides search method, such as keyword based search, simulation structure based search, for users to find appropriate preprocessor, simulation tool, post-processor easily and exactly, and concurrently presents simulation tool filtering to specialize each section. On the other hand, when executing workflow, we use branch-and-cut[6] algorithm to minimize makespan. Branch-and-cut algorithm is revised version of original branch-and-bound[7] algorithm to strengthen linear programming relaxation among IP problem to add new valid equalities before branching partial task.

The whole task speedup is generated by applying branch-and-cut algorithm to tree or nodes of branch-and-bound algorithm, because branch-and-cut algorithm presents cutting plane and reduce the size of tree.

The system defines workflow along user's problem solving demand by using metadata provided from simulation program manager and does binding based on defined workflow. Simulation program result after binding is visualized for user to check it through web.

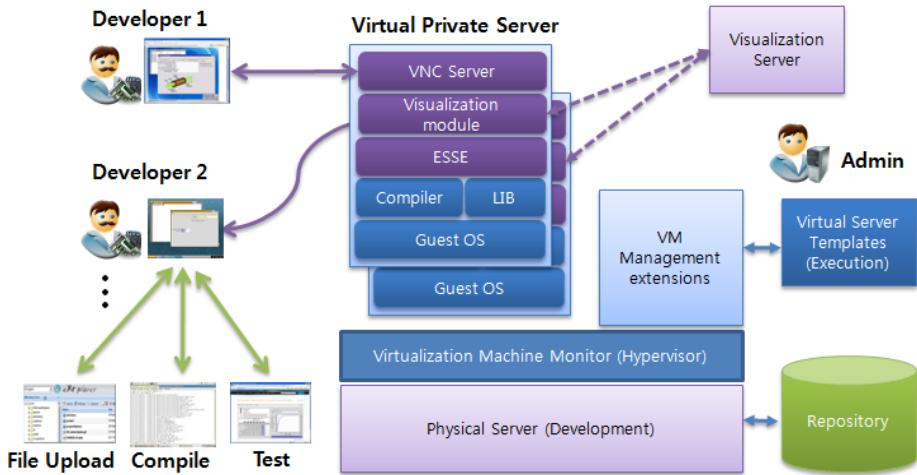


Fig. 3. Structure of Virtual Private Server

## 4 Conclusion

We presented simulation program binding technology, which enables users to share information of simulation program developed by themselves through storage. In addition, we designed informal processes which produce new user customized procedure using unstructured scientific workflow to provide optimal integrated environment for problem solving demand. However, in the worst case, branch-and-cut algorithm has exponential running time and this causes problem when there are many tasks to deal with. For this reason, an improvement in applying simulation binding at many fields effectively will be a future work.

**Acknowledgment.** This work was supported by 2013 National Agenda Project (NAP) funded by Korea Research Council of Fundamental Science & Technology (NAP-09-2).

## References

1. Japan, the promotion of production in the field of supercomputing (2011), <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-21-h135-2.pdf>
2. Large-scale computer simulations, the essential requirements of personalized medicine (2012), <http://www.sciencedaily.com/releases/2012/05/120502092040.htm>
3. Pegasus-WMS, <http://Pegasus.isi.edu/>
4. Oini, T., et al.: Taverna: A tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* (2004)

5. Majithia, S., et al.: Triana: A graphical web service composition and execution toolkit. In: IEEE International Conference on Web Services, ICWS (2004)
6. Branch and cut, [http://homepages.rpi.edu/~mitchj/handouts/bc\\_eg/](http://homepages.rpi.edu/~mitchj/handouts/bc_eg/)
7. Branch and Bound, [http://www.aistudy.com/heuristic/branch\\_and\\_bound.htm](http://www.aistudy.com/heuristic/branch_and_bound.htm)

# Data Security in Cloud for Health Care Applications

R. Anitha and Saswati Mukherjee

Department of Information Science and Technology,  
Anna University, Chennai, 600025  
{anitabalajim, msaswati}@yahoo.com

**Abstract.** With the proliferation of data in medical field and health care systems towards cloud, maintaining the huge volume of sensitive data becomes mandatory. This paper describes into the benefit of cloud computing for healthcare organizations and examines the availability and security considerations that healthcare data requires. The proposed model strengthens the availability and security of data using metadata. The metadata created based on DCMI standards provides easy access of data by locating the server and secure the data resting in the cloud. In this paper security is enforced by cipher key which is generated from the attributes of metadata by providing two novel features. 1. Security is provided, where the encryption and decryption keys cannot be compromised without the involvement of data owner and health care organization, hence makes the data secured 2. The cipher key generated using modified feistel network increases the complexity of the key which strengthens the security effect.

**Keywords:** Security, HealthCare, Metadata, Cloud, Feistel function.

## 1 Introduction

In medical field as changes sweep through the healthcare industry to reduce costs and improve clinical outcomes, healthcare organizations are turning to streamline and manage patient and other medical data in cloud environment. The increased amount of data in the data storage makes use of the recent technique of cloud computing, due to the utilization of the software and the hardware with less investment [4]. As the efficiency of storing and retrieving the data in cloud is improved for the past few years but it also faces major challenge is to achieve two important aims concerning information security in healthcare. The first is to reach a high level of information privacy about the patient, the second is to provide consistent information to the patient that is making sure that the data integrity is maintained. We propose a new methodology which provides ease of access and security to the patient information based on the information of the patient. In general a recent survey regarding the use of cloud services made by IDC, highlights that the security is the greatest challenge for the adoption of cloud computing technology [5]. The four key components of data

security in cloud computing are data availability, data integrity, data confidentiality, and data traceability. Data traceability means that the data transactions and data communication are genuine and that the parties involved are said to be the authorized persons [6]. Several studies shows that data traceability mechanism have been introduced, ranging from data encryption to intrusion detection or role-based access control, doing a great work in protecting sensitive information. However, the majority of these concepts are centrally controlled by administrators, who are one of the major threats to security [8]. In some modern distributed file systems, data is stored on devices that can be accessed through the metadata, which is managed separately by one or more specialized metadata servers [1]. Metadata is a data about data and it is structured information that describes, explains, locates, and makes easier to retrieve, use, or manage an information resource. The metadata file holds the information about a file stored in the data servers. The data owners i.e. the doctors or the patients in cloud computing environment want to make sure that their data are kept confidential to outsiders, including the cloud service provider which will be the major data security requirement. Several studies have been carried out relating to security issues in cloud computing in health domain but the proposed work presents a detailed analysis of the data security in cloud environment through metadata and challenges focusing on security of data through metadata. The Novel scheme supports security through metadata attributes and the access control of the data is limited within the keys used in this model. Furthermore, the ability to generate and compare the keys between the user and the metadata server plays a major role. The security policy in this proposed model depends on 1) User's key which is used to encrypt the original data 2) Cipher key generated using metadata attributes stored in the MDS.

Our contributions can be summarized as follows:

1. We propose a model to create a metadata using DCMI standard along with specification from medical records for easy access of data.
2. We have also proposed a model to create a cipher key Cmxn based on the information of the patient which are provided as an attribute of metadata stored using a modified feistel network and support user to access the data in a secured mode.
3. We have also proposed a novel security policy which involves the patient information by means of key creation and sharing policies. Hence the model prevents unauthorized access of data and also makes data ease of access.

The rest of the paper is organized as follows: Section 2 summarizes the related work and the problem statement. Section 3 describes the system architecture model and discusses the detailed design of the system model. Section 4 describes the modified feistel network structure design in Health care system and issues of the proposed model. The performance evaluation based on the prototype implementation is given in Section 5 and Section 6 concludes the paper.

## 2 Related Works

The Related work discusses about the previous work carried out in the area of cloud security and we have also discussed about how metadata is used in cloud computing environment.

### 2.1 Health Care in Cloud Computing Storage System

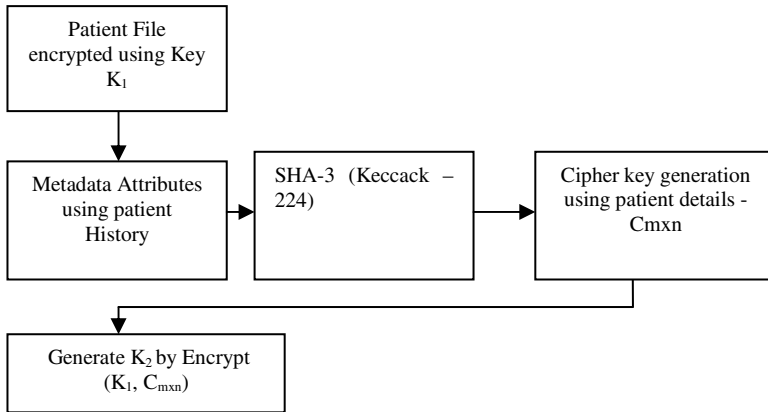
Healthcare has a challenging task to deal with an increasing volume of data [3]. There are reports, showing that the data in health care in the society is growing day to day hence needs more resources to succeed with the future's need of care. Hence, there is a need of effectiveness with cloud computing as a support. The healthcare business claims that cloud computing technique support to improve the service, by providing everything as a service. i.e. Storage is provided as a service and security is also provided as a service. According to SEMA [12] the need of research and studies in this area is obvious.

### 2.2 Security Schemes

Johannes Heurix et al. [7] presents a security protocol for data privacy that is strictly controlled by the data owner, where the PERiMETER makes use of a layer-based security model which protects the secret cryptographic key used to encrypt each user's metadata. Chirag Modi et al. [11] discussed a survey paper where they discussed about the factors affecting cloud computing storage adoption, vulnerabilities and attacks, and identify relevant solution directives to strengthen security and privacy in the Cloud environment. They discuss about the various threats like abusive use of cloud computing, insecure interfaces, data loss and leakage, identity theft and metadata spoofing attack. J. Ravi Kumar et al. [10] shows that third party auditor is used periodically to verify the data integrity stored at cloud service provider without retrieving original data. In this model, the user sends a request to the cloud service provider and receives the original data. If data is in encrypted form then it can be decrypted using his secret key. However, the data stored in cloud is vulnerable to malicious attacks and it would bring irretrievable losses to the users, since their data is stored at an untrusted storage servers.

## 3 System Architecture

The architecture diagram of the proposed system model is shown in figure 1. Each block in the architecture explains about how the patient record is stored safely in the cloud scenario and how securely the record is downloaded.



**Fig. 1.** Cloud Security Architecture for Health Care Application

The initial step of architecture is creation of metadata using PMR based on Dublin Core Metadata Initiative standard. When a PMR is uploaded, the information of the record has been analyzed and based on DCMI design the attributes are extracted along with keyword extraction and stored as a metadata file. In this model the user uploads the encrypted file using the key  $K_1$ . The next step of architecture proposes security to the data using modified feistel function where the metadata attributes are taken as input in the form of matrices. Based on the attributes of metadata the cipher key  $C_{mxn}$  is created. The metadata server sends the cipher key  $C_{mxn}$  to the user. Using the cipher key  $C_{mxn}$  the user encrypts the key  $K_1$  and generates  $K_2$ . Now the user holds  $K_2$  as the key. While downloading the file, the user provides key  $K_2$  and MDS provides  $C_{mxn}$  and are used to retrieve  $K_1$ . The downloaded file is decrypted using  $k_1$  and provides security from unauthorized users accessing the PMR. Thus the process of decryption cannot be compromised without the involvement of the user and MDS. This model proposes a modified feistel function  $F$  which introduces the matrix operations like transpose, shuffle, addition and multiplication along with the key matrix. The cryptanalysis carried out in this paper clearly indicates that this cipher cannot be broken by the brute force attack. This model provides high strength to the cipher, as the encryption key induces a significant amount of matrix obfuscation into the cipher. The avalanche effect discussed shows the strength of the cipher  $C_{mxn}$ . The proposed system model also ensures that the data is identically maintained by making use of the cipher key  $C_{mxn}$  during any operation like transfer, storage, or retrieval. This proposed model also ensures that the data integrity is also maintained.

**File Access:** When a user sends request for data stored on the cloud environment, the request is given to the metadata server which provides the cipher key  $C_{mxn}$  to the user. Using  $C_{mxn}$  user decrypts the key  $K_2$  and gets  $K_1$ . Using  $K_1$  the encrypted file from the cloud storage is decrypted to get the original data. By providing the recent cipher key  $C_{mxn}$  the data integrity is also verified. Our system methodology uses the



functionalities 1. Creation of metadata using Patient Medical Record and uploading PMR to data server. 2. Data Pre-processing. 3. Construction of modified feistel network function. 4. Generation of Cipher key C using the patient details.

## 4 Modified Feistel Network

Feistel ciphers are a special class of iterated block ciphers where the cipher text is calculated from the attributes of metadata by repeated application of the same transformation or round function.

### 4.1 Development of the Cipher Key “ $C_{m \times n}$ ” Using Modified Feistel Function

In this paper we propose a complex procedure for generating the cipher key “ $C_{m \times n}$ ” based on matrix manipulations, which could be introduced in symmetric ciphers. The proposed cipher key generation model explained in algorithm 1 offers two advantages. First, the procedure is simple to implement and has complexity in determining the key through crypt analysis. Secondly, the procedure produces a strong avalanche effect making many values in the output block of a cipher to undergo changes with one value change in the secret key. As a case study, matrix based cipher key generation procedure has been introduced in this cloud security model and key avalanche have been observed. Thus the cloud security model is improved by providing a novel mechanism using modified Feistel network where the cipher key  $C_{m \times n}$  is generated with the matrix based cipher key generation procedure.

**Procedure for generating Cipher Key  $C_{m \times n}$ .** The Cipher key generation procedure is based on a matrix initialized using secret key and the modified feistel function F. The input values used in various feistel rounds are taken from the previous round. The selection of rows and columns for the creation of matrix is based on the number of attributes of the metadata and the secret key matrix “ $K_{m \times n}$ ” and the other functional logic as explained in the following subsections.

#### 4.1.1 Data Preprocessing

Data preprocessing is a model for converting the metadata attributes into matrix form using the SHA-3 cryptographic algorithm, containing m rows and n columns, where m is the number of attributes of the metadata and n takes the size of the SHA-3 output. The matrix is splitted into 4 equal matrix say  $m_1$ ,  $m_2$ ,  $m_3$  and  $m_4$ . The matrix obfuscation is carried out in order to make the hacker opaque. The matrices  $m_1$ ,  $m_3$  and  $m_2$ ,  $m_4$  are concatenated. This obfuscated matrix is fed as input to the feistel network structure where concatenated value of  $m_1$ ,  $m_3$  will be the left value and  $m_2$ ,  $m_4$  be the right value of the feistel network.

#### 4.1.2 Modified Feistel Network Structure

Figure 2 below represents the one round modified feistel network. The Matrix  $L_{m \times n}$  which is a concatenated value of  $m_1 \parallel m_3$  is considered as the left value of the feistel

network structure and Matrix  $R_{m \times n} = m_2 \parallel m_4$  is considered as the right value of the feistel network structure. Using MD5 cryptographic hash algorithm the key matrix  $K_{m \times n}$  is generated whose size is  $m \times n$  where “m” is the number of attributes of metadata and “n” is the size of the MD5 algorithm. The development of the cipher key in the feistel network is done through the number of rounds until the condition is satisfied. In this symmetric block ciphers, matrix obfuscation operations are performed in multiple rounds using the key matrix and the right side value of the feistel network structure. The function F plays a very important role in deciding the security of block ciphers. The concatenated value of  $L_{m \times n}$  and  $R_{m \times n}$  in the last round will be the cipher key  $C_{m \times n}$ .

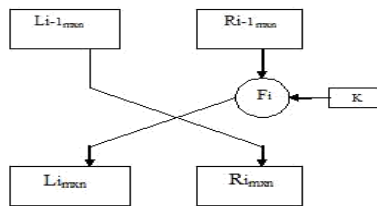


Fig. 2. One Round of Modified Feistel Network

**Definition of Feistel Function F.** Let R be a function variable and let K be a hidden random seed, then the function f is defined as,  $F(R, K) = F_K(R)$  where F is a modified feistel function. The procedure for developing the function is described below. Each round has its own feistel function F. The function f is considered to be varied based on the right side value of the feistel network i.e. the function F is indexed by the matrix  $R_{m \times n}$  for that round. In this modified feistel network structure, the function for each round depends on the previous round

$$\text{Round}_i (L_i, R_i) = ( R_{i-1}, F( K, L_{i-2} ) ) \tag{1}$$

The above formula shows that a small change in one round affects the entire feistel network. For each round as the value of R of the network gets compressed at some point in time the feistel round automatically stops based on the size of the attributes. The procedure for deriving function F is explained in steps as follows:

**Algorithm 1: Creation of cipher key  $C_{m \times n}$**

---

```

Input: Patient Medical Record
Output: Cipher Key  $C_{m \times n}$ 

Begin
Read: Metadata attribute
Apply: SHA-3
Generate: Matrix  $M_{m \times n}$ , split the matrix into  $M_1, M_2, M_3$ 
and  $M_4$ 
Concatenate:  $M_1, M_3$  and  $M_2, M_4$ 
    
```

```

Generate:  $M_1, M_3 = L_{m \times n}$ 
Generate:  $M_2, M_4 = R_{m \times n}$ 
Left value of Feistel =  $L_{m \times n}$ 
Right value of Feistel =  $R_{m \times n}$ 
For i = 1 to n Repeat till n / 2 = 1
Begin
Split:  $R_{m \times n}$  into equal matrix,  $R_{1 \times m \times n}, R_{2 \times m \times n}$ 
Transpose:  $R_{1 \times m \times n}, R_{2 \times m \times n}$  as  $R_{1 \times n \times m}, R_{2 \times n \times m}$ 
Apply matrix addition of  $R_{1 \times n \times m}, R_{2 \times n \times m} = T_{m \times n}$ 
Transpose:  $T_{m \times n}$  Matrix multiplication of  $T_{n \times m} * K_{m \times n} =$ 
 $RV_{m \times n}$ 
/ *condition for multiplication is verified */
New  $L_{m \times n} = RV_{m \times n}$ 
New  $R_{m \times n} =$  Old value of  $L_{m \times n}$ 
End
Repeat the step till n takes odd value
Write(C) Cipher key  $C_{m \times n} = L_{m \times n} || R_{m \times n} / * ||$ 
                                represents concatenation */
End

```

## 4.2 Analysis of the Proposed Model

As the proposed model involves in generating the cipher key  $C_{m \times n}$  the major analysis leads to the analysis of strength of key  $C_{m \times n}$  based on avalanche effect and is compared with the existing algorithms. Avalanche effect is an important characteristic for encryption algorithm. This characteristic is seen that a small change in the metadata attribute will have the effect on its cipher key which shows the efficacy of the cipher key i.e. when changing one bit in plaintext and will change the outcome of at least half of the bits in the cipher text. The need for the discussion of avalanche effect is that by changing only one bit leads to a large change in the key value, hence it is hard to perform an analysis of cipher text, when trying to come up with an attack. From the experimental analysis explained in figure 5 we prove that the modified feistel network holds good for the avalanche effect as each round depends on the previous round value. The avalanche effect is calculated by the formula,

$$\text{Avalanche Effect} = \frac{\text{Number of values changed in the cipher Key } C_{m \times n}}{\text{Total Number of values in the cipher key } C_{m \times n}} \quad (2)$$

## 5 Implementation and Results

The experiments have been carried out in a cloud setup using eucalyptus which contains cloud controller and walrus as storage controller. In the experimental model, the patient medical records are uploaded into the cloud storage server and downloaded based on the user's requirement with and without metadata. In our experiment we have used the medical dataset and investigated the effect of Record

access performance using metadata with respect to the response time. Figure.3 compares the response time for accessing the file with metadata and without using metadata. Figure. 4 and 5 illustrates the performance of the proposed key generation model.

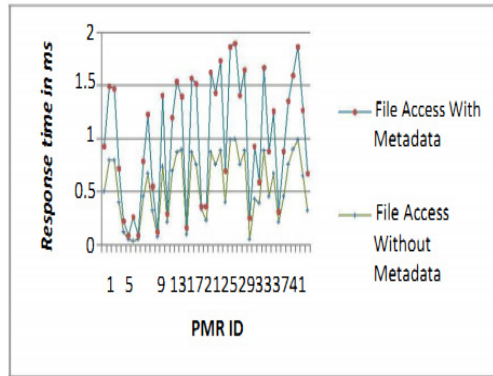


Fig. 3. Comparison with respect to response time with and without using metadata

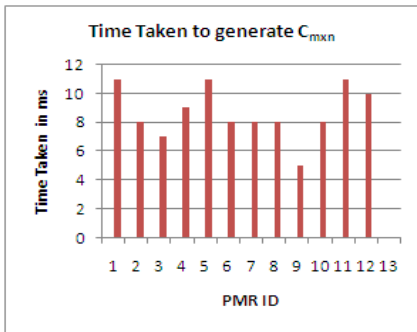


Fig. 4. Time taken to generate  $C_{m \times n}$

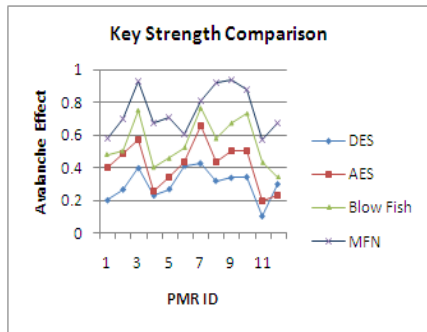


Fig. 5. Avalanche Effect of Modified Feistel Algorithm.

## 6 Conclusion

This paper investigates the problem of data security in health care applications in cloud where the data is stored away from the user. We have studied the problem of privacy of data stored and proposed an efficient and secured protocol to store data at the cloud storage servers. Our method provides privacy to the data stored and the challenges in constructing the security policy and key which involves the data owner and the MDS in order to retrieve the original data. Security is provided by the proposed model, where the encryption and decryption keys cannot be compromised without the involvement of patient or the health care organization and the MDS hence

makes patient or the health care organization feels comfortable about the data stored. The data residing in the cloud is also prone to a number of threats and various issues like confidentiality and integrity. The proposed model addresses all these issues in an efficient way.

## References

1. Cammert, M., Kramer, J., Seeger, B.: Dynamic Metadata Management for Scalable Stream Processing Systems. In: IEEE International Conference on Data Engineering Workshop, pp. 644–653 (2007)
2. Wu, J.-J., Liu, P., Chung, Y.-C.: Metadata Partitioning for Large-scale Distributed Storage Systems. In: IEEE International Conference on Cloud Computing (2010)
3. Blobel, B., Roger-France, F.: A systematic approach for analysis and design of secure health information systems. *International Journal of Medical Informatics*, 51–78 (2001)
4. Anitha, R., Mukherjee, S.: A Dynamic Metadata Model in Cloud Computing. In: Krishna, P.V., Babu, M.R., Ariwa, E. (eds.) *ObCom 2011, Part II. CCIS*, vol. 270, pp. 13–21. Springer, Heidelberg (2012)
5. Kuyoro, S.O., Ibikunle, F., Awodele, O.: Cloud Computing Security Issues and Challenges. *International Journal of Computer Networks* 3(5), 247–255 (2011)
6. Mathew, A.: Survey Paper on Security & Privacy Issues in Cloud Storage Systems. In: *Electrical Engineering Seminar and Special Problems 571B* (2012)
7. Heurix, J., Karlinger, M., Neubauer, T.: Perimeter – Pseudonymization and Personal Metadata Encryption for Privacy-Preserving Searchable Documents. In: *International Conference on Health Systems*, vol. 1(1), pp. 46–57 (2012)
8. Tang, Y., Lee, P.P.C., Lui, J.C.S., Perlman, R.: Secure Overlay Cloud Storage with Access Control and Assured Deletion. *IEEE Transactions on Dependable and Secure Computing* 9(6) (2012)
9. Wang, Q., Wang, C., Ren, K., Lou, W., Li, J.: Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. *IEEE Transactions on Parallel and Distributed Systems* 22(5) (2011)
10. Ravi Kumar, J., Revati, M.: Efficient Data Storage and Security in Cloud. *International Journal of Emerging trends In Engineering and Development* 6(2) (2012)
11. Modi, C., Patel, D., Borisaniya, B., Patel, A., Rajarajan, M.: A survey on security issues and solutions at different layers of Cloud computing. *Journal of SuperComputers*, 561–592 (2013)
12. Arai, Y., Watanabe, C.: Query Log Perturbation Method for Privacy Preserving Query. In: *4th International Conference on Ubiquitous Information Management and Communication* (2010)

# Web Service Selection Using Decision Tree Analysis in a Risky Environment

P. Sandhya<sup>1</sup> and M. Lakshmi<sup>2</sup>

<sup>1</sup>Department of Information Technology, Sathyabama University,  
Chennai, Tamil Nadu, India  
{catherinejeremiah@gmail.com}

<sup>2</sup>Department of Computer Science and Engineering, Sathyabama University,  
Chennai, Tamil Nadu, India  
{laks@icadsindia.com}

**Abstract.** In real world scenario managerial decisions are made to maximize profit and minimize loss even under circumstances of risk. Web service composition is the process of aggregation of services to create virtual enterprises. During composition the aim of service providers is to offer good service to the client and ensure good return of investment. Software agent composers have several strategies to compose services automatically. However most of the composition algorithms aim at serving good service to the client without considering the corporate metrics and hence regarded as a toy model. To promote web service composition as a feasible business model we have proposed a service provider collaboration stack that considers the service providers metrics during composition. The modules for time planning, profit management, native intelligence and user adoption of SLAKY composition stack have already been implemented. In this paper we focus on selecting services with maximum profit for the composition under environmental metric of risk using decision theory. The software agent uses decision tree analysis to identify the best course of action to be chosen till the end under various possible outcomes as a whole.

**Keywords:** Web service composition, SLAKY Composer, Decision of composition under condition of risk, Decision tree analysis, Expected Money Value, Expected Opportunity Loss.

## 1 Introduction

Web service composition is the task of combining and linking existing web services to create new web processes and add value to the collection of services. Dynamic composition is a type of composition done at runtime and can be automated by a software agent. Several algorithms have been proposed to perform automatic web service composition. However most of the composition algorithms deal with providing apt services to the client. The deficiency of works on service composition based on service provider metrics results in lack of adoption of web service

composition in real world business. In this paper we introduce SLAKY Composition stack which describes service provider metrics for realistic selection of business partner services on the fly under environmental condition of risk

## 2 Related Works

Though several works have been done on web service composition most of them focus on assigning an apt service to the client. Paper [1] discusses that research in automating service composition is rarely concerned with service providers, apart from work in quality guarantees and contracts. Hence in our approach we introduce SLAKY stack that considers provider specific metrics for business oriented selection of services. In SLAKY composer profit management [2], time planning [3], native intelligence [4] and user adoption [5] metrics are already implemented. In this paper we evaluate all possible criterions based on decision theory to select service for composition in environmental conditions of risk using decision tree analysis.

## 3 SLAKY Collaboration Stack

SLAKY is a realistic model for choosing business service partners by considering service partner collaboration metrics including vision, time planning, environmental context, user adoption, usage policies, trust management, risk management, market scenario, native intelligence and competitive profit management apart from functionality satisfaction of client’s requirements. In this paper we implement selection of service for composition based on environmental condition of risk using decision tree analysis.

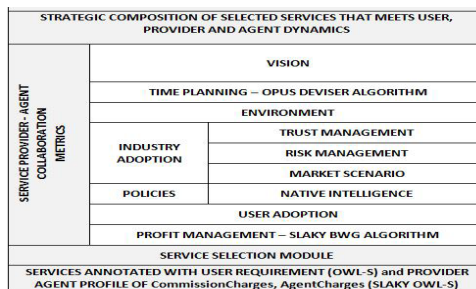


Fig. 1. SLAKY Collaboration Stack

## 4 Decision Making of Service Selection under Conditions of Risk

When the availability of information for decision is partial then a decision under such environment is called decision under risk. The decision criteria under circumstance of risk are as follows [6]:

### 4.1 Expected Money Value Criterion

Consider that there are five services offering the same functionality requested by the client. Let the services be S1, S2, S3, S4 and S5. Let the demand for the services be 10, 11, 12, 13 and 14. The assigned probabilities are 0.10, 0.15, 0.20, 0.25 and 0.30 respectively. Let the actual cost of the service is 30 units and the offering price be 50 units.

$$\text{The payoff} = \text{profit X service hits} - \text{actual cost X count of service missed} \tag{1}$$

$$\text{In this case the: Payoff} = 20 \text{ X Service hits} - 30 \text{ X service missed} \tag{2}$$

**Table 1.** Payoff table

Possible demand	Probability	State of nature				
		10 S1	11 S2	12 S3	13 S4	14 S5
10	0.10	200	170	140	110	80
11	0.15	200	220	190	160	130
12	0.20	200	220	240	210	180
13	0.25	200	220	240	260	230
14	0.30	200	220	240	260	280

The expected value of each service alternative is computed by multiplying the computational profit with the associated probability and adding the resulting values.

**Table 2.** Expected Profit table

Possible demand	Probability	State of nature				
		10 S1	11 S2	12 S3	13 S4	14 S5
10	0.10	20	17	14	11	8
11	0.15	30	33	28.5	24	19.5
12	0.20	40	44	48	42	36
13	0.25	50	55	60	65	57.5
14	0.30	60	66	72	78	84
Total expected profit		200	215	222.5	220	205

From the above we conclude that service3 is the service that yields maximum profit of 222.5 even under circumstance of risk.

### 4.2 Expected Opportunity Loss (EOL)

This approach minimizes the expected opportunity loss. EOL or expected value of regrets represents the amount by which maximum possible profit will be reduced under various possible actions. Consider that there are four services offering the same functionality requested by the client. Let the services be S1, S2, S3 and S4. Let the demand for the services be 15, 16, 17 and 18. Let the assigned probabilities be 0.2, 0.4, 0.3 and 0.1 respectively. Let the actual cost of the service is 2 units and the offering price be 2.50.

$$\text{The payoff is: demand X offering price} - \text{availability X actual price} \tag{1}$$

$$\text{In this case the: Payoff} = \text{Demand X 2.50} - \text{Availability X 2.0} \tag{2}$$



Now we populate the conditional profit in the payoff table. Next we compute the conditional probability loss for each of the alternative services. Subtract each payoff in a row from the maximum payoff of that row. Finally compute the expected opportunity loss for each alternative service by multiplying conditional loss values by the associated probabilities.

**Table 3.** Conditional profit

Possible demand	Probability	Availability				
		15 S1	16 S2	17 S3	18 S4	
15	0.2	7.5	5.5	3.5	1.5	
16	0.4	7.5	8	6	4	
17	0.3	7.5	8	8.5	6.5	
18	0.1	7.5	8	8.5	9	

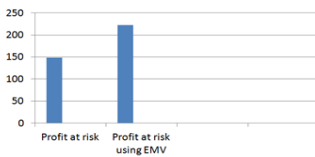
**Table 4.** Conditional loss

Possible demand	Probability	Availability				
		15 S1	16 S2	17 S3	18 S4	
15	0.2	0	2	4	6	
16	0.4	0.5	0	2	4	
17	0.3	1	0.5	0	2	
18	0.1	1.5	1	0.5	0	

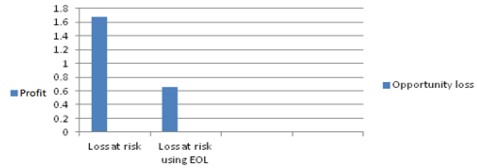
**Table 5.** Expected loss

Possible demand	Probability	Availability				
		15 S1	16 S2	17 S3	18 S4	
15	0.2	0	0.4	0.8	1.2	
16	0.4	0.20	0	0.8	1.6	
17	0.3	0.3	0.15	0.0	0.6	
18	0.1	.15	0.1	0.05	0	
Expected opportunity loss		0.65	0.65	1.65	3.4	

The expected opportunity loss is 0.65. Hence services S1 or S2 will be selected to minimize the opportunity loss in a risky environment. In case of service selection based on client’s metrics the opportunity loss is 1.68. The comparative results of EMV and EOL is given below:



**Fig. 2.** Comparison of service selection without and with EVM



**Fig. 3.** Comparison of service selection without and with EOL

## 5 Implementation of Decision Tree Analysis

A decision tree is a graphical representation of decision process indicating decision alternatives, states of nature, probabilities associated with the states of nature and conditional profits and losses. The steps in decision tree analysis is to identify the decision points and alternatives (Services), at each decision point, determine the probability and associated payoff (of service) with each course of action, compute expected payoff for each course of action, start from extreme right and move towards left, choose the course of action that yields the best payoff for each of the decisions (Service selection) and proceed backward to the next stage of decision points. Repeat the above step till the first decision point is reached. Identify the best course of action (Service selection) to be adopted from the beginning to the end, under various possible outcomes as a whole. Let us consider that there are three alternative services A1, A2, A3 that offer the same functionality with charges of 3000, 2200 and 900 points respectively. Let the unit price of the product be 25. The product and the cost of units associated with respective decision alternatives are 9, 5 and 11. The expected demand is:

**Table 6.** Demand and Probability

Demand	200	300	400	500
Probability	0.3	0.2	0.4	0.1

To decide the best alternatives formulate a payoff table.

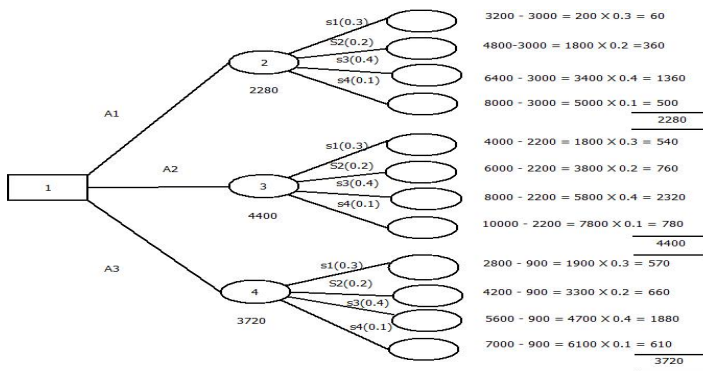
$$\text{Profit} = (\text{Fixed price} - \text{Cost of units}) \times \text{Demand} \tag{1}$$

**Table 7.** Payoff Table

Alternative	States of Nature (demand)			
	200	300	400	500
A1(9)	(25-9)X200=3200	(25-9)X300=4800	(25-9)X400=6400	(25-9)X500=8000
A2(5)	(25-5)X200=4000	(25-5)X300=6000	(25-5)X400=8000	(25-5)X500=10000
A3(11)	(25-11)X200=2800	(25-11)X300=4200	(25-11)X400=5600	(25-11)X500=7000

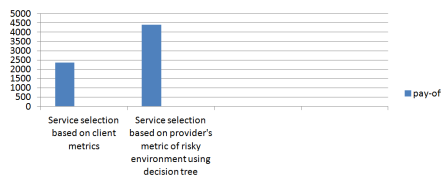
The next step is to construct decision tree by using the above pay-off values and their associated probabilities. The equation is:

$$\text{payoff} = (\text{profit of alternative } i - \text{charges of alternative } i) \times \text{probability} \tag{2}$$



**Fig. 4.** Decision Tree Analysis in a risky environment

The alternative with maximum payoff is selected as the best alternative. Alternative service A2 has maximum payoff and it is chosen as the best alternative in a risky environment. The comparison graph is plotted below:



**Fig. 5.** Comparison of service selection using and not using decision tree analysis

## 6 Process of Service Selection

Based on client’s request a lookup is done on the UDDI to select appropriate services for the client. These services are then fed to the agent. The agent checks the environment to decide whether it is risky based on states of nature. Now the service selection process can be done using EMV or EOL and graphically using decision tree. For each course of actions the probability and payoff of the service alternatives are fed to the agent are determined. Traverse the tree from right to left. We then select the alternative service that has the highest payoff.

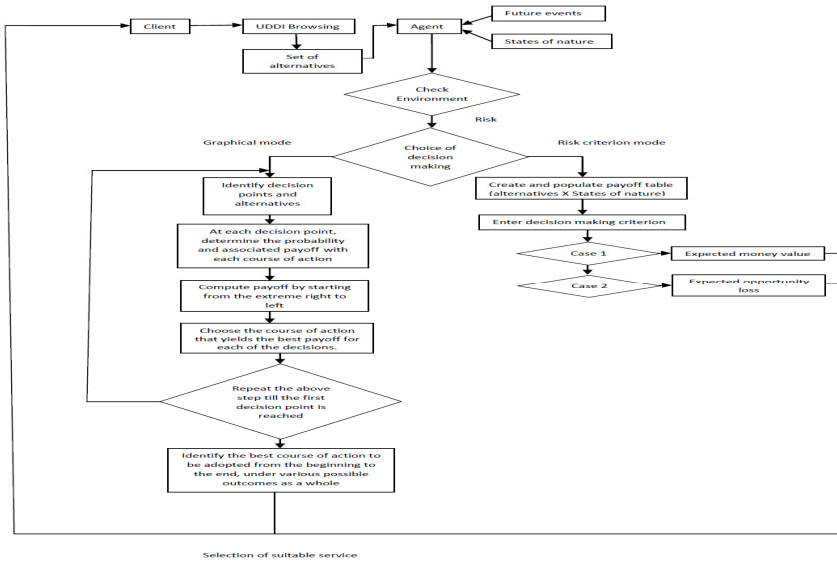


Fig. 6. Process of Service Selection

## 7 Analysis of Service Selection using Decision Tree Approach

The advantages are that service structures decision making in a sequential order. It is mainly useful in situations where the initial decision and its outcome affect the subsequent decision. It displays the logical relationship between the parts of a complex decision and identifies the time sequence in which various actions subsequent event would occur.

## 8 Conclusion and Future Enhancement

In this paper we design a composer agent that selects services of given domain in an environment of risk considering the criterion of expected money value and expected money loss. We have also extended the risk management metric using decision tree

analysis for graphical mode. The disadvantages of decision tree including complexity when there are numerous alternatives and inconsistency in assigning probabilities which can be dealt as future enhancement.

## References

1. Mehandjiev, N., Lécué, F., Wajid, U.: Provider-Composer Negotiations for Semantic Robustness in Service Compositions. In: Baresi, L., Chi, C.-H., Suzuki, J. (eds.) ICSOC-ServiceWave 2009. LNCS, vol. 5900, pp. 205–220. Springer, Heidelberg (2009)
2. Sandhya, P., Lakshmi, M.: Strategic Composition of semantic web services using SLAKY composer. In: Meghanathan, N., Nagamalai, D., Chaki, N. (eds.) Advances in Computing & Inf. Technology. AISC, vol. 178, pp. 411–420. Springer, Heidelberg (2012)
3. Sandhya, P., Lakshmi, M.: A Novel Approach for Realizing Business Agility through Temporally Planned Automatic Web Service Composition Using Network Analysis. In: Worldcomp 2011 (SWWS 2011), July 18-21, pp. 24–31. CSREA Press, LasVegas (2011), <http://www.lidi.info.unlp.edu.ar/worldcomp2011-mirror/sww2789.pdf> ISBN: 1-60132-202-X
4. Sandhya, P., Lakshmi, M.: Modeling Native Intelligence Semantics For Indigenous Selection Of Web Services Using Slaky Composer. International Journal on Cloud Computing: Services and Architecture (IJCCSA) 2(5) (October 2012)
5. Sandhya, P., Lakshmi, M.: Enhancing User Adoption For Service Oriented Business Model Through Semantically Annotated And Extended Association Rule Mining In Slaky Composer System. International Journal of Computer Science Engineering and Information Technology Research (IJCEITR) 2(4), 5–18 (2012) ISSN 2249-6831
6. Kalavathy, S.: Operations Research with C Programs, 3rd edn. (2010)

# An Enhanced e-Voting System in Cloud Using Fingerprint Authentication

Gandhi Usha Devi, Kannan Anusha, and G.V. Rajyalakshmi

School of Information Technology and Engineering  
VIT University, Vellore, Tamil Nadu, India  
{ushadevi.g, anusha.k}@vit.ac.in,  
rajyalakshmigv@yahoo.com

**Abstract.** Authentication and privacy are considered to be the central issues of any e-voting system. Citizens cast their votes in the polling station as a protocol where the duplicate and black votes are difficult to identify by the national commission of a country who has the responsibility to release the counted votes. This paper based voting system is inefficient and takes a delay in counting the votes. In order to achieve an efficient voting system, we have formulated an accustomed e-voting system with fingerprint deployed in the public cloud which gives reliable and quicker results.

**Keywords:** Fingerprint, voting, e-governance, cloud computing, database efficiency.

## 1 Introduction

Cloud computing is currently regarded as an external resource (the public cloud) which appears to solve the problem of over provisioning and manage the privileges of data centres [4]. However, the cloud tends to produce its own complexities and management challenges. Irrespective of the challenges, the cloud has the ability to bring a new way of doing things and a whole new set of opportunities.

The cloud's hard-edged, warehouse-sized data centres accessible on the Internet are filled with seven-foot-tall racks of pizza-box servers. As a result, this seems to be concrete enough. However, the scenario is completely different in case of access by individuals. When an individual end user accesses a server in the cloud, the server has the ability to take on or shared processing cycles from CPUs and uses more memory or less, as needed. The user's cloud machine expands according to his/her needs and shrinks when peak processing is over. It may be on one side of the data centre on one moment and on the opposite at the next moment.

## 2 Related Works

For a democratic country, public opinion is the most important determinant to establish a government. To obtain this opinion, the government has launched the voting system. Therefore, it is highly important for the voting system to be reliable, accurate

and transparent [1-3]. The system that exists currently in our country is totally paper based. This paves way for the occurrence of more number of human errors and is a time consuming process. The voters are registered just before the poll so the election commission gets some time in hand for making all the necessary arrangements within this short period of time. They add the new voters with the previous voters list so that the people who have deceased by this time may be considered as the existing voter if they are not informed. So, people may not bestow their faith on the voters list as it contains numerous fake voters [5-6].

If any voter stays abroad or misses the registration processes somehow due to prior obligations or unavoidable circumstances, he or she wouldn't be considered as a voter unless or until she/he informs the authority. Any voter may change his place of residence between two elections. If the authority is not informed, they are not considered as the voter of that area though he is a voter as per the constitution. Therefore, he misses the opportunity to confer his opinion. Even if he is a registered voter of his new locality, it is often seen that he is still existing voter of his old area. Thereby he can vote twice which is illegal. Sometimes, people ruin their votes by stamping on two or more signs mistakenly. While casting the votes, the acting officer present in the centres marks the voter with a black ink on his or her nail but it is not always irremovable. So there is a chance for casting illegal votes. Again, these votes are counted manually so that the process becomes a time-consuming one which may be inaccurate as well. All these problems together made people think about inventing a new system that reduces corruption, increase accuracy and fast paced.

### 3 Proposed Model

In the existing system, we define an e-voting system as a system that can resolve the problems in enrolment and counting the votes. It is feasible to ignore redundant votes and gain a higher degree of control on the correctness of the voter's background in the database. Accuracy of the voters is the vital parameter that needs to be ensured by the e-voting system. It is believed that this can be achieved with the database servers' communication with its security. As a result, this system is proposed to overcome all the demerits of the paper based system and produce a quality system with no practical issues. The proposed system works based on the following Fig. 1. From the technical viewpoint, the elections comprise of the following components: Calling of elections, Registration of candidates, Preparation of polling list, Voting (The subset of voting is e-voting), Counting of votes. The proposed e-voting system adopts architecture in the Fig. 2. The architecture separates the common people from the administrators. Additionally, the storage process happens in a secured manner with SQL Azure.

Windows Azure is a promising technology that promises numerous advantages such as data maintenance, flexibility, and reliability, data filtering and controlled access to data. Security is yet another parameter that focuses throughout this paper as cloud technology lacks on security aspects. As a result of the analysis of loopholes in the paper based system, the proposed e-voting system is devised with certain crucial features that eliminate the key issues and safeguard information and processes associated with the implementation in the real time.

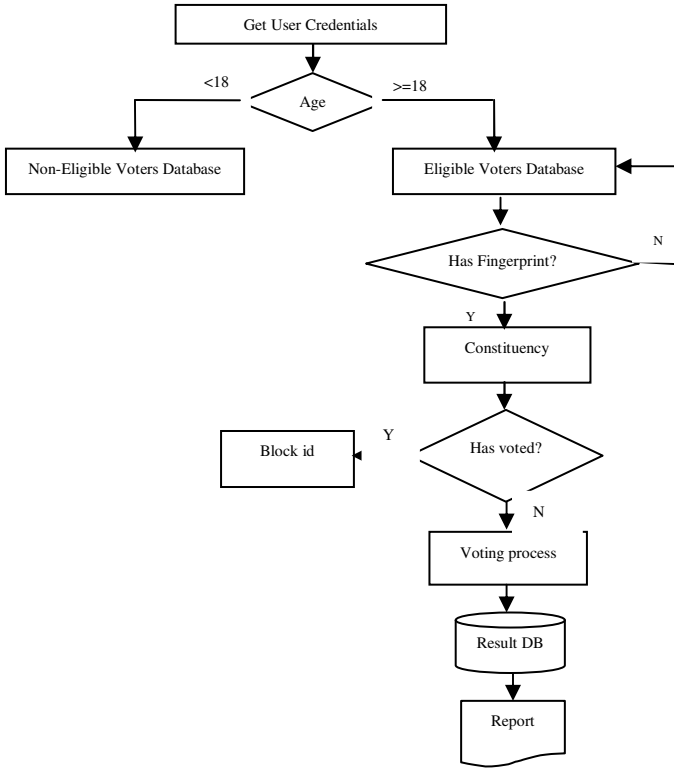


Fig. 1. E-voting system

### 3.1 Maintenance of Database System

In this section, a computerized registration form for each voter is introduced. This form collects voter details including name, area, etc. and their corresponding fingerprint and the system generates a unique id for every voter. Population database is created to register the citizens. The information of the people is enrolled along with the fingerprint in the SQL Server and maintained. This database is classified into three segments namely Eligible and non-eligible databases. Non-Eligible database is exclusively for citizens between 1 year old and less than 18 years old. Eligible Database maintains the data of the people who are at least 18 years old and above 18 of the existing voters who are alive and would cast their votes. Additionally, people who have expired won't be migrated to the voter database but they are in the population database and their status is dead. This approach does not save the space of the voter database but it also improves the accuracy of the system. The inputs of the primary database collects from the 'X' (who has the information of all aged people of country) and the information of death people is collected from 'Y' (who has the records of the death people of the country).The information about the people who are eligible for voting along with the fingerprint stores in the cloud. Whenever the local server

undergoes updates, it is essential to update the cloud database of e-voting system in a parallel manner so that the website is up-to-date. The authentication process becomes easier with the inclusion of cloud technology in the proposed system to authenticate earlier to vote in the cloud. The database in the Windows Azure is maintained with all local host data of all the citizens for voting process.

### **3.2 Enrolment Phase**

The election commission authority collects the details as well as finger prints from the people who are at least 17 years. Birth certificate, H.S.C or S.S.L.C certificate are the documents used to verify the age. People who don't have this certificate may use chairman certificate or commissioner certificate to prove their ages. For people who failed to give their details and finger prints to the authority, their status does not register and they migrate to the database automatically again. As a result, they won't be migrated to the voter database. The registration performs later and migrated to voter database instantly. The authority collects the details from even 17 years old people so that if any person turns 18 years old between the time of collecting data and the election, he or she is able to cast their vote. Even the people who live outside the country is able to cast his or her vote. It doesn't matter whether he has the birth certificate or H.S.C or S.S.C certificate. The passport proves his/her legality for being the citizenship of India and that also serves as the proof of his/her age. This is the critical advantage of this proposed e-voting system. Amidst these advantages, it is important for the citizen to register fingerprints although they don't reside in the country at that moment. These processes would enable the citizen to cast their vote on the day of election.

### **3.3 Fingerprint Matching**

The crucial part of this system is fingerprint. Fingerprint is a unique identification for any voter. All the information about voter will be preserved in accordance with the fingerprint. The system generates an id for the registration of every individual voter. This id is protected by his/her fingerprint. If anyone tries to make double entry in the voter database, it is possible because of fingerprint as the redundancy is overcome by the proposed system. As a result, the system ensures single entry for an individual and the system does not transfer the entry until the corresponding fingerprint is provided. The system is highly secured to the extent that even administrator cannot cause modifications to the voter's information. Eventually, all the information is strongly preserved in the database. When a voter casts his/her votes, the system initially locates his/her fingerprint at the database and find out the specific area from where the vote is cast. Subsequently, the system can check whether the specific id is blocked or unblocked. Blocked denotes that the person has already voted and unblocked represents nil votes from that individual. The system automatically rejects the vote when an id is blocked whereas it accepts provided it is not blocked and the fingerprint



matching is successful. It is not feasible to give duplicate votes too. In the cloud, we follow an authentication phase before we vote. In this phase, the fingerprint of the individuals are stored in the database and verified in terms of location. Franchise voting service is permitted if the above mentioned processes yield positive results.

### 4 Results

The experiment is conducted and tested with 50 samples of votes from different places. The fingerprint authentication done in the enrollment process is demonstrated in Fig. 3. The enrollment process receives all the personal information about the voters and it is validated with the fingerprint matching. This is the authentication procedure followed in the evoting system during the registration process. The Azure platform being used in this system is depicted in Fig. 4 in terms of its credentials. The variations of the cost difference between e-voting system and paper based voting over a three years is shown in Fig.5.

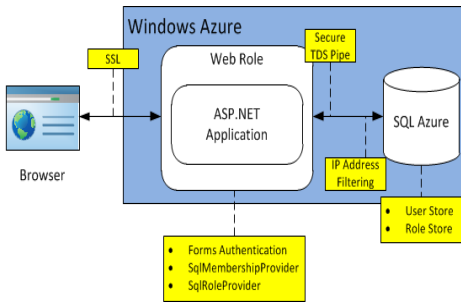


Fig. 2. E-voting system in cloud

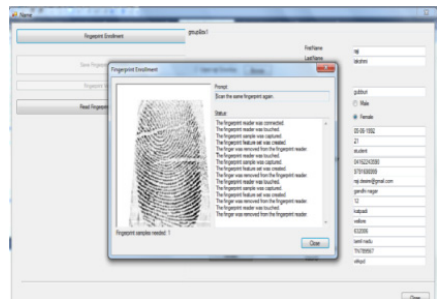


Fig. 3. Fingerprint authentication

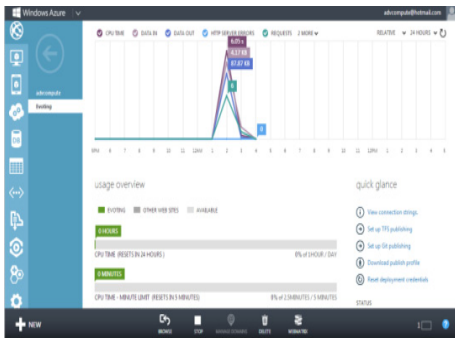


Fig. 4. Credentials of windows azure

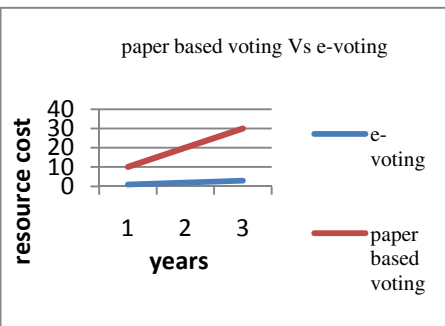


Fig. 5. Variations of the cost

## 5 Conclusion

An Enhanced e-voting system provides an authenticated fingerprint of each eligible voter. The voter verification and audit trails can also be achieved with the cloud environment. The proposed e-voting system appears to eliminate the manual errors and conduct an effective and qualitative voting process that will no way involve the presence of illegal votes or inappropriate inclusion of voters. Any paperless electronic voting system might suffer similar flaws irrespective of certification that it would have received. The reliability of the votes is achieved with the audit trail efficiency during the e-voting system where it is portable in all the platforms deployed in environment. With these results, scalability is obtained. Data partitions for the security and speed of transaction can be included in the near future. Cost effective service provisioning requires modeling thereby taking scalability into account. Suitable and fast data encryption techniques need to be studied.

## References

1. Kremer, S., Ryan, M.D.: Analysis of an electronic voting protocol in the applied pi calculus. In: Sagiv, M. (ed.) ESOP 2005. LNCS, vol. 3444, pp. 186–200. Springer, Heidelberg (2005)
2. Li, C., Hwang, M., Liu, C.: An electronic voting protocol with deniable authentication for mobile ad hoc networks. *Computer communications* 31, 2534–2540 (2008)
3. Caarls, S.: E-voting handbook: Key steps in the implementation of e-enabled elections. Council of Europe Publishing, Europe (2010)
4. Zissis, D., Lekkas, D.: Securing e-government and e-voting with an open cloud computing architecture. *Government Information Quarterly* 28, 239–251 (2011)
5. Rexha, B., Neziri, V., Dervishi, R.: Improving authentication and transparency of e-voting system – Kosovo case. *International Journal of Computers and Communications* 6(1), 84–91 (2012)
6. Brown, J., Dickinson, D., Steinebach, C., Zhang, J.: E-voting system: Specification and design document (2003)

# Local Minima Jump PSO for Workflow Scheduling in Cloud Computing Environments

S. Chitra<sup>1</sup>, B. Madhusudhanan<sup>2</sup>, and G.R. Sakthidharan<sup>3</sup>, and P. Saravanan<sup>4</sup>

<sup>1</sup> Er. Perumal Manimekalai College of Engineering, Principal, Hosur, India  
schitra3@gmail.com

<sup>2</sup> Er. Perumal Manimekalai College of Engineering,  
Computer Science and Engineering, Hosur, India  
phdmadhu@gmail.com

<sup>3</sup> Kalaignar Karunanidhi Institute of Technology,  
Computer Science and Engineering, Coimbatore, India  
grsdharan@gmail.com

<sup>4</sup> M. Kumarasamy College of Engineering Computer Science and Engineering,  
Karur, India  
saravancse@gmail.com

**Abstract.** Earlier Grids implemented workflows, but the former's reduced performance has resulted in workflows being implemented in cloud. Cloud computing the latest in distributed computing, facilitates virtualized resources for applications. Cloud computing environment workflows enable use of varied cloud services facilitating workflow execution. Good workflow examples are online banking, insurance claim processing, e-business, and e-government scenarios. As workflow scheduling is NP hard, meta-heuristic based methods solve issues. This paper attempts to locate a suitable workflow schedule where Particle Swarm Optimization (PSO) is used to optimize load balancing, speedup ratio, and makespan. Experimental results demonstrate the effectiveness of the proposed algorithm. The proposed algorithm is more effective with higher number of tasks.

**Keywords:** Cloud computing, Workflow scheduling, Particle Swarm Optimization (PSO).

## 1 Introduction

Cloud computing is a distributed computing model which attracts researchers in Distributed/Parallel Computing [1, 2] and Service Oriented Computing [3]. Though there is no proper definition of a Cloud, its distinctive characteristics as proposed by Ian Foster include large-scale distributed computing paradigm extending on demand services to external users over the net. Compared to conventional computing paradigm definitions like cluster [5], grid [6], and peer-to-peer [7], "economies" is the cloud computing key word which is not seen in other paradigm. "Economies" denotes that cloud computing adopts a market-oriented business model where users are charged for the cloud services. Cloud services include computing, network services and

storage [8]. The cloud service providers must ensure satisfactory QoS (quality of service) while fulfilling business contracts.

Cloud users' benefit from not having to invest up-front for their computing, network services or storage necessities; thus, reducing operating and maintenance costs of their business operation. Cloud features provide flexibility to a user's computing environment, adapting computer systems to user's needs. Cloud computing has many advantages: it lowers entry cost for small firms which benefit from computer-intensive business analytics and access to hardware resources, with no capital investments. Cloud computing reduces IT innovation barriers as evidenced from startups and online applications like Face book and YouTube. Cloud computing ensures easy service up scaling enterprises on demand. It also ensures new application classes and delivers services like mobile interactive applications which are location/environment /context-aware and parallel batch processing [9].

Three core cloud computing technologies used are – virtualization, multi-tenancy and Web services. Virtualization prevents a user from seeing a computing platform's physical characteristics providing an abstract, emulated computing platform [10] that behaves independently which can be configured on demand. Computing infrastructure is better utilized by reducing upfront/operational costs. While virtualization is available from 1960s, only recent computing power/networking resources deliver seamless performance in emulated systems that users use on personal computers. A related concept, multi tenancy where a single application software instance benefits many clients by ensuring improved system's resources (regarding memory/processing overhead) use, which might require more if software instance is duplicated for all clients. W3C defines a Web service including varied systems, but it usually refers to clients/servers communications over web used HTTP protocol [11]. Web services standardize intra interface applications by ensuring access to server applications for clients over networks. For the schedulers to be efficient with a cost effective schedule where tasks/applications data are loaded onto cloud computing environments, schedulers are based on different policies. Most of the policies are designed to optimize one or more particular function of the cloud computing environment, such as minimizing execution time, execution cost, load balancing of the resources used, deadline constraints and others. Earlier grids implemented workflows, but diminishing grid performance lead to cloud workflows with the added benefits of scalability. Cloud resources scalability ensures real time resources provisioning to meet application requirements. This ensures workflow management systems to meet Quality of Service applications requirements. At reduced costs, cloud services compute, store and bandwidth resources are available substantially, as other workflow applications need complex execution environments. In Cloud computing environment, workflows enable the usage of different cloud services use by ensuring workflow execution [12]. Good workflow examples are online banking, insurance claim processing, e-business, and e-government scenarios.

Existing workflow scheduling algorithms established in clouds are as follows [13]:

- Optimized Resource Scheduling Algorithm.
- Improved Cost-Based Algorithm for Task Scheduling.
- Innovative transaction intensive cost-constraint scheduling algorithm.

- A Compromised-Time-Cost Scheduling Algorithm.
- PSO-based Heuristic for Scheduling Workflow Applications.
- Scalable-Heterogeneous-Earliest-Finish-Time Algorithm (SHEFT).
- Market-Oriented-Hierarchical Scheduling.
- Multiple QoS Constrained Scheduling Strategy of Multi-Workflows (MQMW).
- Optimal Workflow based Scheduling (OWS) algorithm.
- Resource-Aware-Scheduling algorithm (RASA).
- Heterogeneous-Earliest-Finish-Time algorithm (HEFT).

Some cloud established workflow scheduling algorithm mentioned above is as follows:

- ✓ PSO-based Heuristic for Scheduling Workflow Applications: PSO optimizes application's scheduling of cloud resources considering computation and data transmission costs which is used by varying computation/communication costs for all workflow applications. Experimental results reveal that PSO achieves cost savings and good workload distribution onto resources [14, 15, 16].
- ✓ Optimized Resource Scheduling algorithm: An Improved Genetic Algorithm (GA) is used for automated scheduling to achieve optimization/sub-optimization for cloud scheduling and to increase utilization resources rate and speed [17, 18].
- ✓ Optimal Workflow based Scheduling (OWS) algorithm: The OWS algorithm schedules workflows in cloud environments. It finds a solution meeting all users preferred QoS constraints thereby achieving great improvement in CPU use [19].

As workflow scheduling is NP hard, meta-heuristic methods like: particle swarm optimization (PSO) [14], tabu search (TS) [20], simulated annealing (SA) [21] and genetic algorithm (GA) [22, 23] solve issues. This paper locates a suitable workflow schedule for speedup ratio, workflow, load balancing and makespan which are optimized through use of PSO.

## 2 Related Works

A PSO based heuristic to schedule cloud resources applications which consider computation and data transmission costs was presented by Pandey, et al., [14]. Heuristics minimized the total execution cost of Cloud computing environments application workflows. Total execution costs was obtained by changing communication cost between resources and complete resources execution cost. Varying computation and communication costs investigated work flow applications. Cost savings through use of PSO and existing "Best Resource Selection" (BRS) algorithm was compared. Experiments revealed that PSO achieved thrice the cost savings when compared to BRS, and a more efficient workload distribution on resources. The proposed heuristic was

generic and could be used for many tasks and resources performed with dimension particles increase and resources number, respectively.

A “Revised Discrete Particle Swarm Optimization” (RDPSO) to schedule cloud service application that considered computation costs and data transmission were proposed by Wu, et al., [15]. Makespan and cost optimization ratio were compared as also cost savings with RDPSO, BRS algorithm and PSO. Experiments demonstrated that proposed RDPSO algorithm achieved cost savings with improved performance on cost optimization and makespan. The market-oriented business model of cloud workflow system is different from the conventional workflow scheduling strategies. Wu, et al., [24] proposed a market-oriented hierarchical scheduling strategy in cloud workflow systems, specifically for service-level scheduling. The proposed method dealt with Task-to-Service assignment, mapping individual workflow instances tasks to cloud services in global cloud markets. The assigning was based on task-level scheduling, QoS requirements and optimizing of Task-to-VM (virtual machine) in local cloud data centres. The cloud workflow system’s overall running cost was minimised by satisfying QoS constraints for individual tasks. A package based random scheduling algorithm, based on hierarchical scheduling strategy was presented as candidate service-level scheduling algorithm and 3 representative meta-heuristic based scheduling algorithms including GA, ant colony optimisation (ACO), and PSO which were implemented. The proposed hierarchical scheduling strategy was implemented in SwinDeW-C cloud workflow system and evaluation demonstrated satisfactory performance. Experiments revealed that ACO based scheduling algorithm’s overall performance was better than that of others for CPU time, optimisation rate on makespan, and rate on cost.

A task scheduling model was formulated by Guo, et al., [16] which in turn proposed a PSO algorithm based on small position value rule. The model was reduced with cost of problem task scheduling solved by a PSO algorithm. A PSO algorithm with crossover, mutation and local search algorithm were analysed on particle swarm basis and it was proposed that PS algorithm was embedded with SPV representing better performance. Experiments manifested that PSO algorithm in both gained optimal solution and converged faster in big tasks than the other alternatives. Also, running time was lesser than the other algorithms. It was obvious that PSO suited cloud computing. Experiments proved that PSO algorithm was highly suited for cloud computing. Merging of public and a private clouds result in a hybrid cloud. Bittencourt and Madeira [25] presented “Hybrid Cloud Optimized Cost scheduling algorithm” (HCOC) which decides the resources to be leased from public cloud and those which are to be aggregated to private cloud to ensure processing power for workflow execution in given execution time. Simulation results were presented showing that HCOC reduces costs and achieved desired execution time.

### 3 Methodology

A workflow cloud application is modeled as a Directed Acyclic Graph (DAG) with  $G = (V, A)$ . Set of nodes,  $v = \{T_1, T_2, \dots, T_n\}$  represents workflow application tasks

and set of arcs are represented as precedence constraints and data dependencies between tasks. An arc can be  $d_i, j = (T_i, T_j) \in A$  where  $T_i$  is parent task and  $T_j$  is child task of  $T_i$ ,  $d_{i,j}$  is the data produced by  $T_i$  and consumed by  $T_j$  respectively. During workflow execution, it is assumed a child task can be executed only when all parent tasks are completed.

Figure 1 depicts an example of a workflow with 9 tasks and figure 2 depicting data storage and service instances. In the example,  $n$  tasks are scheduled to  $m$  service instances. For every task  $T_i$  in the workflow,  $s_i = (s_i^1, \dots, s_i^{m_i})$  set of candidate service instances where  $s_i^j (1 \leq m_i)$  and  $m_i$  is total number of service instance for task  $T_i$  and  $D = (D_1, \dots, D_m)$  storage sites are available. The service instance  $s_i^j$  properties such as execution time and cost are represented as variables  $(s_i^j.t, s_i^j.c)$ .

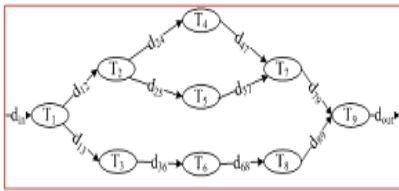


Fig. 1. Cloud workflow (9 tasks)

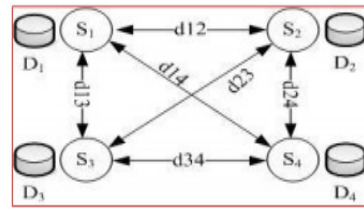


Fig. 2. Data storage ( $D_i$ ) and Service instances ( $S_i$ )

In the above mentioned figures, each task is implemented by 4 service instances and they are connected to the data storage and are also symmetric respectively.

### 3.1 Particle Swarm Optimization (PSO)

PSO is a self-adaptive global search based optimization [26] which is similar to other population-based algorithms like the GA. PSO has no direct population individual's re-combination and depends on particles social behavior. In PSO algorithm, a particle adjusts trajectory based on local best position and position of global best particle of the population during each generation. These iterative adjustments increase the particle's stochastic nature and help to converge to global minimum with a reasonably good solution. PSO is widely used due to its ease of use, efficacy, low computational cost and its use in many applications. PSO was used to solve NP-Hard problems like Scheduling [27] and task allocation [28].

PSO maximizes objectives of locating parameters by exploring search space for a problem. This technique is from swarm intelligence and evolutionary computation [26]. Swarm intelligence is based on bird/fish, swarming habits and evolutionary computation by locating a local/global maximum. PSO algorithm represents every

solution as a ‘bird’ in search space calling it a ‘particle’. Candidate solutions evaluate objective functions by operating on resulting fitness values. Candidate solution/estimated fitness/velocity ensure the particle’s location. It remembers best fitness value it achieved during its operation and which are called individual best fitness. The candidate solution it achieved is called individual best position ‘pbest’. Best fitness value among swarm particles is known as global best fitness, and candidate solution that reached this fitness is called global best position/global best candidate solution ‘gbest’. PSO algorithm’s 3 steps reiterated till stopping criteria is met [26] are as follows:

- Evaluating every particle’s fitness.
- Individual/global best fitness and updating positions
- Updating velocity/position of each particle.

The PSO algorithm used here is summarized by 5 steps as follows.

1. First initialize swarm particles so that position  $\bar{x}_{ij}(t=0)$  of each particle is random in hyperspace.
2. Comparing each particle’s fitness function,  $F(\bar{x}_{ij}(t))$ , is packet delivery ratio for individuals in current time period, for best performance till then,  $pbest_{ij}$ : if  $F(\bar{x}_{ij}(t)) < pbest_{ij}$ , then

$$pbest_{ij} = F(\bar{x}_{ij}(t)), \bar{x}_{pbest_{ij}} = \bar{x}_{ij}(t)$$

3. With comparison of  $F(\bar{x}_{ij}(t))$  to the global best particle,  $gbest_j$ : if  $F(\bar{x}_{ij}(t)) < gbest_j$ , then

$$gbest_j = F(\bar{x}_{ij}(t)), \bar{x}_{gbest_j} = \bar{x}_{ij}(t)$$

4. To revise the velocity for each particle:

$$\begin{aligned} \bar{v}_{ij}(t) = & \bar{v}_{ij}(t-1) + c_1 \cdot r_1 \cdot (\bar{x}_{pbest_{ij}}(t) - \bar{x}_{ij}(t)) \\ & + c_2 \cdot r_2 \cdot (\bar{x}_{gbest_j}(t) - \bar{x}_{ij}(t)) \end{aligned}$$

where  $r_1$  and  $r_2$  are random numbers between 0 and 1, and  $c_1$  and  $c_2$  are positive acceleration constants.

5. To move each particle to new position:

$$\bar{x}_{ij}(t) = \bar{x}_{ij}(t-1) + \bar{v}_{ij}(t) \quad t = t+1$$

Repeat steps (2) through (5) till it converges.

### 3.2 Enhancements in the Proposed PSO

When the value of gbest does not change much with each iteration, it has been found that convergence is poorer as gbest gets stuck with the local minima problem. In



order to overcome this problem faced by regular PSO algorithms, an improved technique for determination of gbest is proposed in this paper. The modified PSO algorithm Jump PSO (JPSO) for identifying gbest the following steps are added to step 3 of the PSO algorithm.

For N number of initial population

Rank obtained pbest as

$$p(i) = \text{sort}(pbest_{i,j}, \text{desc})$$

Select  $k = \text{round}(\alpha.N)$  pbest from  $p(i)$

$$\text{Compute } gbest = \frac{1}{k} \sum_{i=1}^k p(i)$$

where  $0.1 \leq \alpha \leq 0.5$  is the particle jumping factor . Lower value of  $\alpha$  ensures faster convergence and higher value of  $\alpha$  ensures exiting the local minima value faster.

### 4 Results and Discussion

The initial population was selected using random work flow scheduling for alpha value of 0.2. Experiments were conducted using Genetic Algorithm, PSO and JPSO. The experiments are conducted to evaluate speedup, loading balancing of resources, and makespan. Figure 3 shows the speed up obtained. The proposed algorithm improves the overall speed up by about 3.8%. Genetic algorithm shows better speed up compared to PSO when the number of tasks is lower, however the speed up is almost the same as the number of tasks increases.

Figure 4 shows the load balancing rate for different number of tasks. When the number of tasks is lower, the load balancing rate is higher though the number of available resources is higher.

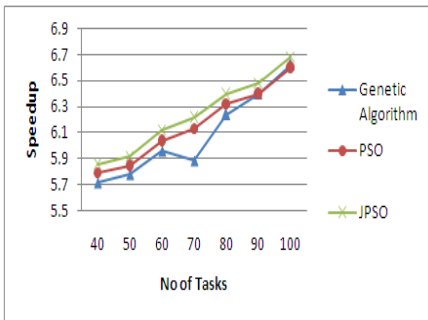


Fig. 3. Speedup for varying tasks

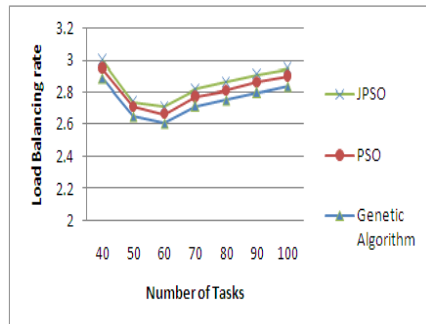


Fig. 4. Load Balancing Rate for varying tasks

Figure 5 shows the Makespan of the three techniques studied. The proposed PSO improves the makespan by 9.98% compared to GA and 6.45 % compared PSO when the number of tasks to be scheduled is 40. Similarly PSO outperforms GA by 3.77% when the number of tasks is 40. Averaging the makespan over the number of tasks the overall improvement of the proposed PSO is 6.25% compared to GA and 4.18% compared to PSO. This clearly shows that workflow scheduling problem convergence is faster when gbest is aided to get out of local minima.

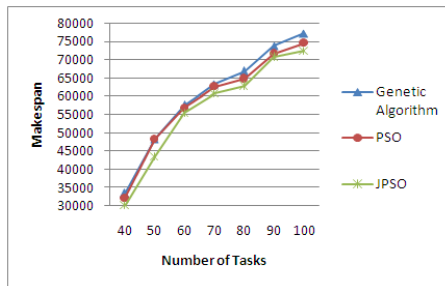


Fig. 5. Makespan for varying tasks

## 5 Conclusion

For the schedulers to be efficient with a cost effective schedule where tasks/applications data are loaded onto cloud computing environments, schedulers are based on different policies. Minimizing execution time, execution cost, load balancing of the resources used, deadline constraints are optimized by the schedulers. As the workflow scheduling is NP hard problem, meta-heuristic based methods are used to solve the issue. In this paper to improve the workflow scheduling an improved PSO is proposed. The experiments are conducted to evaluate speedup, loading balancing of resources and makespan for varying number of tasks. Experimental results demonstrate the effectiveness of the proposed method when compared to existing GA and PSO methods.

## References

1. Raghavan, B., Ramabhadran, S., Yocum, K., Snoeren, A.C.: Cloud control with distributed rate limiting. In: Proc. 2007 ACM SIGCOMM, Kyoto, Japan, pp. 337–348 (2007)
2. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Zaharia, M.: A view of cloud computing. *Communications of the ACM* 53(4), 50–58 (2010)
3. Ardagna, D., Pernici, B.: Adaptive service composition in flexible processes. *IEEE Trans. Softw. Eng.* 33(6), 369–384 (2007)
4. Foster, I., Yong, Z., Raicu, I., Lu, S.: Cloud computing and grid computing 360-degree compared. In: Proc. 2008 Grid Computing Environments Workshop, pp. 1–10 (2008)

5. Martinez, A., Alfaro, F.J., Sanchez, J.L., Quiles, F.J., Duato, J.: A new cost-effective technique for QoS support in clusters. *IEEE Trans. Parallel Distrib. Syst.* 18(12), 1714–1726 (2007)
6. Foster, I., Kesselman, C.: *The grid: blueprint for a new computing infrastructure*, 2nd edn. Morgan Kaufmann, San Mateo (2004)
7. Yu, J., Buyya, R.: Scheduling scientific workflow applications with deadline and budget constraints using genetic algorithms. *Sci. Program.* 14(3-4), 217–230 (2006)
8. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. *Future Gener. Comput. Syst.* 25(6), 599–616 (2009)
9. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A.: Cloud computing—The business perspective. *Decision Support Systems* 51(1), 176–189 (2011)
10. Vouk, M.A.: Cloud computing — issues, research and implementations. *Journal of Computing and Information Technology* 16(4), 235–246 (2008)
11. W3C, Web Services Glossary, <http://www.w3.org/TR/ws-gloss/> (cited April 3, 2009)
12. Lin, C., Lu, S.: Scheduling scientific workflows elastically for cloud computing. In: 2011 IEEE International Conference on Cloud Computing (CLOUD), pp. 746–747. IEEE (July 2011)
13. Bala, A.: A Survey of Various Workflow Scheduling Algorithms in Cloud Environment. In: 2nd National Conference on Information and Communication Technology, NCICT (2011)
14. Pandey, S., Wu, L., Guru, S.M., Buyya, R.: A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In: 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 400–407. IEEE (April 2010)
15. Wu, Z., Ni, Z., Gu, L., Liu, X.: A revised discrete particle swarm optimization for cloud workflow scheduling. In: 2010 International Conference on Computational Intelligence and Security (CIS), pp. 184–188. IEEE (December 2010)
16. Guo, L., Zhao, S., Shen, S., Jiang, C.: Task scheduling optimization in cloud computing based on heuristic algorithm. *Journal of Networks* 7(3), 547–553 (2012)
17. Zhong, H., Tao, K., Zhang, X.: An Approach to Optimized Resource Scheduling Algorithm for Open-source Cloud Systems. In: 2010 Fifth Annual ChinaGrid Conference (ChinaGrid), pp. 124–129. IEEE (July 2010)
18. Yang, C.-T., Cheng, H.-Y., Huang, K.-L.: A dynamic resource allocation model for virtual machine management on cloud. In: Kim, T.-H., Adeli, H., Cho, H.-S., Gervasi, O., Yau, S.S., Kang, B.-H., Villalba, J.G. (eds.) GDC 2011. CCIST, vol. 261, pp. 581–590. Springer, Heidelberg (2011)
19. Varalakshmi, P., Ramaswamy, A., Balasubramanian, A., Vijaykumar, P.: An optimal workflow based scheduling and resource allocation in cloud. In: Abraham, A., Lloret Mauri, J., Buford, J.F., Suzuki, J., Thampi, S.M. (eds.) ACC 2011, Part I. CCIST, vol. 190, pp. 411–420. Springer, Heidelberg (2011)
20. Porto, S., Ribeiro, C.: A tabu search approach to task scheduling on heterogeneous processors under precedence constraints. *Int. J. High Speed Comput.* 7, 45–72 (1995)
21. Kalashnikov, A., Kostenko, V.: A parallel algorithm of simulated annealing for multiprocessor scheduling. *J. Comput. Syst. Sci. Int.* 47, 455–463 (2008)
22. Yoo, M.: Real-time task scheduling by multi objective genetic algorithm. *J. Syst. Softw.* 82, 619–628 (2009)

23. Yu, J., Buyya, R., Ramamohanarao, K.: Workflow scheduling algorithms for grid computing. Department of Computer Science and Software Engineering. The University of Melbourne, VIC 3010, Australia (2009)
24. Wu, Z., Liu, X., Ni, Z., Yuan, D., Yang, Y.: A market-oriented hierarchical scheduling strategy in cloud workflow systems. *The Journal of Supercomputing* 63(1), 256–293 (2013)
25. Bittencourt, L.F., Madeira, E.R.M.: HCOC: a cost optimization algorithm for workflow scheduling in hybrid clouds. *Journal of Internet Services and Applications* 2(3), 207–227 (2011)
26. Kennedy, J., Eberhart, R.: Particle swarm optimization. *IEEE International Conference on Neural Networks* 4, 1942–1948 (1995)
27. Yu, B., Yuan, X., Wang, J.: Short-term hydro-thermal scheduling using particle swarm optimisation method. *Energy Conversion and Management* 48(7), 1902–1908 (2007)
28. Zavala, A.E.M., Aguirre, A.H., Villa Diharce, E.R., Rionda, S.B.: Constrained optimisation with an improved particle swarm optimisation algorithm. *Intl. Journal of Intelligent Computing and Cybernetics* 1(3), 425–453 (2008)

# Respiratory Rate Estimation by Extracted PPG Signals from Embedded Smart Attire of Operation Strategists

Kathir Deivanai

VIT University, Chennai, India

**Abstract.** In this paper, we have proposed a simple procedure to identify the respiratory variations of humans during the strive conditions from the extracted PPG signals for clinical and intervention treatments. Protection of people who involve in dangerous and critical tasks is necessary to save their life using life guard systems. PPG is being used to recognize and interpret this common symptoms of the human body, using heart rate variability and respiration rate. In this experiment the deviations of respiratory rate from the normal to stressed state of a person is measured using the derived RIIV parameter extracted from the potential PPG signals. Significant amount of data taken for analysis from five different healthy subjects and are used in this research to ensure the coherence of raw PPG signal with the RIIV and additional experiments were done to ensure the stressful conditions has an impact on the respiratory rate.

**Keywords:** ATAFOS-Ambulatory Technology aid for Operation Strategist, PPG-Photo-plethysmography, RIIV-Respiratory induced intensity variations.

## 1 Introduction

In clinical and life-saving procedures, monitoring a patient or an operation strategist in a reasonable frequency is an essential component. However, Health monitoring 24 X 7 is a wearisome process and to be attentive around the clock is tiresome and humanly impossible. Similarly, for an operation strategist, the monitoring by conventional methods is impossible due to the nature of the task environment. Bringing the strategist back to a place of treatment or sending a medical aid to the spot of operation is too expensive and too risky. Hence for good and effective care, the anxious users rely on support systems inclusive of medical equipment [1]. The proposed Ambulatory Technology Aid For Operation Strategists (ATAFOS) is an experimental setup designed suitably to support operation strategist like mine laborers, sea divers, military commandos etc., To aid ATAFOS, wearable smart attire which is an infrastructure [2], is desired, which is automatically get activated through inbuilt sensors. The Photoplethysmogram (PPG) is an optical measurement technique that detects blood volume changes in the micro vascular bed of tissue. It has been used in a wide range of clinical monitoring processes through commercially available medical devices by measuring oxygen saturation, blood pressure and cardiac output accessing autonomic function of a person. Respiratory sinus arrhythmia is the

phenomenon by which respiration modulates heart rate in normal humans and in many animals. Based on the Human Anatomy, the PPG signals are collected from forearm, forefinger, and ear lobe etc., the real time feedback mechanism on such smart attire can provide timely medical inspection data [3].

Here our primary goal is to make health monitoring online for operation strategists say mine labors, military commandos, sea divers who are in need of precautionary measures [2]

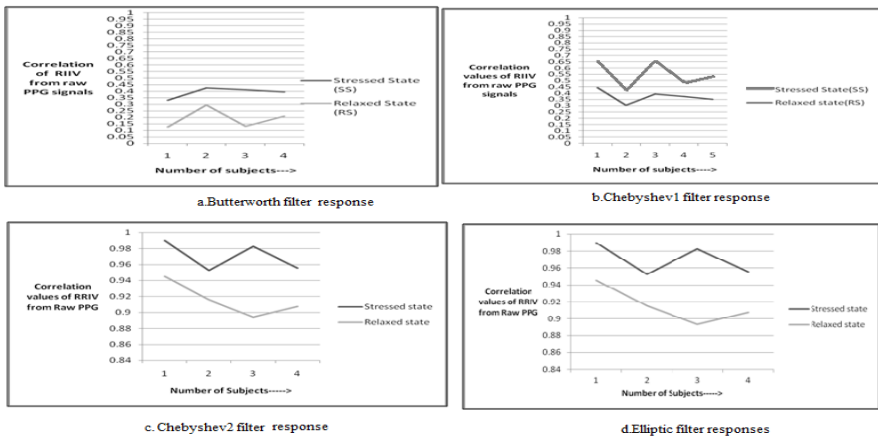
## 2 Materials and Methods

In general, investigation to respiratory system usually falls under three categories [10]. The development of a mathematical model relating respiration to those variations that it causes in heart rate. The use of digital filtering techniques to attenuate fluctuations in heart rate which are due to respiration. The development of methods that use only heart rate to get information about respiration. In this paper, our works adopt the second category of methodology filtering process to ensure the suitability of PPG to estimate the respiratory system function. The acquisition procedures of Electrocardiogram- ECG, The Electroencephalogram EEG, are complex when compared to the PPG signal. That is why PPG is found suitable for embedding into the smart attire for operation strategists than the other bio signals. More over the estimation of the respiratory rate from the PPG signals is technically possible [12]. The reason is that the PPG signals are applied in pulse oximetry, where the component of the signals synchronizes with pulse which is used to monitor arterial oxygen saturation and heart rate. The exhale and inhale ventilation, changes occur in peripheral venous circulation makes variation in respiratory system. These variations are influenced by many parameters. To make more precise measurements, PPG signals were collected from five subjects. All the signals captured from the 5 subjects were pre-processed by means of low-pass filtering. A three point moving average filter was implemented non-recursively. The signal processing procedures were implemented in Mat-lab. The advantage of using the moving average filter is that it reduces random noise [11]. The 3 point moving average filter can be expressed as

$$y(n) = \frac{1}{3} [x(n) + x(n-1) + x(n-2)] \quad (1)$$

From the above equation (1), the average is taken from the current and two previous inputs i.e. to compute the current output  $y(n)$ , two previous inputs  $x(n-1)$  and  $x(n-2)$  are needed [11]. The RIV is extracted from the collected PPG signals, for monitoring respiratory rate for any instance to provide ambulatory services for our operation strategist to ensure whether the clinical condition is normal. Sometimes necessity due to sudden crisis requiring action then immediate care will be provided by giving adequate oxygen to the person. This proposed system is clinically essential for saving the lives of operation strategists by better intervention methods. Mainly our work will help to decide treatment methods to prevent serious life-threatening complications during strategic operations. Filters were used to extract certain specific region of the original PPG signals acquired [7,11]. We have chosen appropriate

filters tuned to the exact measurement of respiratory fluctuations. Here we propose the use of band pass filters of four different types of filters to verify the correlation among them. Butterworth filters are usually flat pass band, chebyshev 1 filters are all pole filter that exhibit equi-ripple behavior in the pass band, chebhshev2 that contains both pole and zero exhibits monotonic behavior in the pass band and equi ripple in the stop band. The credibility of the extracted RIIV from PPG is ensured by means of the coherence arrived from all the results of different filter responses. Analyze with sample raw PPG signals, here we detected respiratory by RIIV of the PPG signals which are close to each other [7,8]. Operation strategists are prevented from fatality by giving first aid adequately through the reliable real time monitoring experimental setup. In order to provide accurate results and reduce false alarms high specification is desirable. Feature extraction process is to help trace out the correlation between two different signals and analysis of PPG will have new scope in bio-medical signal processing[7]. The Correlation with respiratory to raw PPG signals is estimated. The observation records steady increase of respiration in stressed state for all subjects and is proved by the RIIV extracted using different types of filters. The results of this correlation value are given in the Table 1. The merits of PPG signals over ECG signals are to minimize efforts in data acquisition methods and additional materials in ambulatory respiration monitoring for operation strategists. The efficiency or health condition of a strategist or the patient drastically descend and is fatal if not intervention takes place. The PPG signals are captured from 5 healthy subjects by reusable finger clip and ear pulse oximeter embedded at appropriate positions [9]. The performance and uniformity of 4 different filters for 5 healthy subjects, producing the RIIV from the PPG signals in stressed and relaxed state were analyzed.



**Fig. 1.** Represents the all four filter responses of raw PPG signals and RIIV for stressed and relaxed states

### 3 Results

In our system, respiratory induced intensity variations were extracted from raw PPG signals and compared for stressed and relaxed states which results to be statistically distinct too. The work is done with four different filters as Butterworth, Chebyshev1, Chebyshev2 and Elliptic filters. The results are found to be coherent with each other, confirming the reflection of respiratory functions on the captured PPG signals. The filtering process applied to raw PPG signals, results in Respiratory induced intensity variations signal which is approximately equal to the respiration rate . With this, the stressed state is easily predicted and thus explicitly the operational strategist is in need of immediate care to recover. After the experiments, the results were analyzed and interpretations are listed in the figures.

**Table 1.** Statistical Methods Analyzed With Four Different Filters

Filters VS Subjects	Butterworth filter		Chebyshev1 filter		Chebyshev2 Filter		ELLIPTIC FILTER	
	SS	RS	SS	SS	SS	RS	SS	RS
<i>Subject1</i>	0.485	0.236	0.442	0.118	0.935	0.860	0.97	0.85
<i>Subject2</i>	0.331	0.126	0.301	0.261	0.989	0.942	0.91	0.79
<i>Subject 3</i>	0.424	0.294	0.392	0.116	0.952	0.915	0.98	0.92
<i>Subject 4</i>	0.410	0.130	0.379	0.178	0.982	0.893	0.97	0.82
<i>Subject 5</i>	0.394	0.208	0.349	0.165	0.956	0.907	0.91	0.85

*SS-Stressed state and RS-Relaxed state*

Subjects	Difference in states using Butterworth filter	Difference in states using Chebyshev1 filter	Difference in states using Chebyshev2 Filter	DIFFERENCE IN STATES USING ELLIPTIC FILTER
<i>Subject1</i>	0.249	0.324	0.075	0.12
<i>Subject2</i>	0.205	0.040	0.047	0.12
<i>Subject 3</i>	0.130	0.276	0.037	0.06
<i>Subject 4</i>	0.280	0.201	0.089	0.09
<i>Subject 5</i>	0.186	0.184	0.049	0.06
<i>Average</i>	<b>0.210</b>	<b>0.205</b>	<b>0.059</b>	<b>0.09</b>



## 4 Discussions

Here, the stress is recognized from the respiratory rate estimated from the PPG signals. The recognition of stress in the ambulatory service is a major challenge. The implicit relation between PPG and RIIV is found in two conditional states as stressed and relaxed. The respiratory rate calculated from the RIIV helps in taking many decisions in the health condition of a person [6]. The ultimate aim is to overcome the respiratory problems caused due to allergic reactions, oxygen deficiency, and memory loss and bleeding due to damages in lungs, nose and sinus inflammation and also wheezing problem.

## 5 Conclusion

In this work, the respiratory rhythm in two states – stressed and relaxed from the PPG signals are captured which is in betterment with the conventionally used ECG signals. With RIIV the comparison is made in the respiratory rate and results are tabulated. results in finding the respiratory rhythm of different subjects produced using different filters are found to be coherent. This significant difference between two states will be helpful in intervention treatment and life saving strategies for operational strategist. It can also be used by the e – textile technology to incorporate optical PPG sensors in replacement with the complex hardware.

## References

- [1] Dipietro, L., Sabatini, A.M., Dario, P.: A survey of Glove Based Systems and their Applications. *IEEE Transactions on Systems, Man and Cybernetics* 38(4) (July 2008)
- [2] Park, Jayaraman: Enhance the quality of life through Wearable technology 22 (June 2003)
- [3] Grillet, A., Kinet, D., Witt, J., Schukar, M., Krebber, K., Pirotte, F., Depre, A.: Optical Fiber Sensors Embedded into medical textiles for Healthcare monitoring. *IEEE Sensors Journal* 8(7) (July 2008)
- [4] Dilmaghani, R.S., Bobarshad, H., Ghavami, M., Choobkar, S., Wolfe, C.: Wireless Sensor Networks for monitoring Physiological signals of multiple patients. *IEEE Transactions on Biomedical Circuits and Systems* 5(4) (August 2011)
- [5] Folke, M., Cernerud, L., Ekstrom, M., Hok, B.: Critical review of non-invasive respiratory monitoring in medical care. *Med. Biol. Eng. Comput.* 41, 377–383 (2003)
- [6] Madhav, K.V., Ram, M.R., Krishna, E.H., Reddy, K.N., Reddy, K.A.: Estimation of respiratory rate from principal components of photoplethysmographic signals. In: *IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2010*, pp. 311–314. IEEE (December 2010)
- [7] Nilsson, L., Goscinski, T., Kaman, S., Lindberg, L.-G., Johansson, A.: Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects, vol. 51(9), pp. 1250–1257. Blackwell, Publishing
- [8] Lindberg, L.-G., Ugnell, H., Oberg, P.: Monitoring of respiratory and heart rates using a fiber optic sensor. *Med. Biol. Eng. Computing* 30, 533–537 (1992)

- [9] A.: Quantification of Emotional features of Photoplethysmographic waveforms using box-counting method of fractal dimension. In: Recent Advances in Circuits, Systems, Electronics Control and Signal Processing-G, Ugnell, H., Oberg, P.: Monitoring of respiratory and heart rates using a fiber optic sensor. *Med. Biol. Eng. Computing* 30, 533–537 (1992), Sezgin, M.C., Dokur, Z., Olmez, T., Korurek, M.: Classification of respiratory sounds by using an artificial neural network. *National Journal of Pattern Recognition and Artificial Intelligence* 1, 697–699 (2002)
- [10] Sezgin, M.C., Dokur, Z., Olmez, T., Korurek, M.: Classification of respiratory sounds by using an artificial neural network. *National Journal of Pattern Recognition and Artificial Intelligence* 1, 697–699 (2002)
- [11] Womack, B.F.: The Analysis of Respiratory Sinus Arrhythmia Using Spectral Analysis and Digital filtering. *IEEE Transactions on Biomedical Engineering BME-18*(6) (November 1971)
- [12] Lin, Y.-D., Liu, W.-T., Tsai, C.-C., Chen, W.: Coherence analysis between respiration and PPG signal by Bivariate AR model. *World Academy Science, Engineering and Technology* 53, 847–851 (2009)

# The Assessment of Information and Communication Technology (ICT) Policy in South Korea

Young-Hyun Yeo<sup>1</sup>, Sung-Ki Kim<sup>2</sup>, Ji-Hye Bae<sup>2</sup>, and Byung-Gyu Kim<sup>3</sup>

<sup>1</sup>Dept. of Public Administration, SunMoon University,  
A-san city, Rep. of Korea

<sup>2</sup>Div. of IT Education, SunMoon University,  
A-san city, Rep. of Korea

<sup>3</sup>Dept. of Computer Engineering, SunMoon University,  
A-san city, Rep. of Korea

{yeosun, skkim, angdoo98}@sunmoon.ac.kr,  
bg.kim@mpcl.sunmoon.ac.kr

**Abstract.** The economic development in South Korea has been heavily relied on ICT industries. As the 10th largest economic scale in the world, Korea is ranked high in her ICT infrastructure and competitiveness. The main factors of this accomplishment are successful leaderships of Korean governments, intensive investment on public infrastructure, direct support to the ICT industry, competition-based industrial structure, and ICT convergent policies. This paper evaluates Korean ICT policy in terms of government leaderships and consequent infrastructure and performances. As a policy implication, we suggest that ICT policy in Korea focus more on strategic perspectives, development of software and source technology, and securing high-brains in ICT industries. Further, we conclude that the cyber count-intelligence capability is needed, which is accomplished through building the capacity in high-edge convergent industries, supporting software industries and small firms, focusing on local ICT industries as a new driving force in the future.

**Keywords:** Information and Communication Technology Policy, ICT Policy, Korean ICT Industry, Assessment of Policy. Korean ICT, IT.

## 1 Introduction

Despite recent global crises such as the financial crisis in the U.S and the Recession in European countries, growth in ICT industries has increased steadily up to the trade scale of 3.7 trillion U.S. dollars in 2011. The global market of ICT industry consists of communication services (46.5%), ICT services (23.1%), communication devices (12.1%), computer hardware (11.0%), and software (7.3%). Recently, ICT industry expands its role into convergence with other industries, alternatives in solving global challenges such as climate change and the improvement of healthcare and welfare.

Most developed countries regard ICT industry as a core part to elevate their global position into leading country in economic development and industrial competitiveness, which leads those into expanding national investments on ICT policy.

In her history of industrial development, Korea shows dramatic shifts from one of the lowest income countries in 1950s to one of leading countries in ICT industry in 2010s. Since 1980s when the Master-plan for Development of Electric Industry was established, ICT industry has been taken the place of light industry in 1960s and heavy industry in 1970s. In the area of ICT policy, Korean government has been leading her industrial development through subsidizing the export of ICT products, direct funding R&D programs and basic studies. As a result, in the term between 2007 and 2011, the production in Korean ICT industry increased as 8.8% annually, whereas average growth in global market was 3.6% during the same period.

In this light, this paper investigates the changes and characteristics of ICT policies in Korea. Also, we suggest policy implications indicating limitations and alternative solutions in future ICT policy.

## **2 Overview of Korean ICT**

### **2.1 Global Indicator for Korean ICT**

According to [9], Korea is a global leader in ICT infrastructure and competitiveness. In 2012, the International Telecommunication Union (ITU) ranks Korea the 1st in the area of development of telecommunication, e-government and civic participation in online. Also, other global bodies such as the World Economy Forum (WEF) scores Korea high in their indicators for global ICT industry. For instance, according to WEF, Korea marks the 12th in the preparedness of network and the 18th in the initiatives of technology acceptance. The International Institute for Management Development (IMD) ranks World Competitiveness of Korea the 14th in the area of technology infrastructure, while EIU of the British Economist puts the 2nd highest score on governmental broadband in Korea. Finally, Waseda University in Japan ranks Korea the 3rd in global e-government.

### **2.2 Overview of Korean ICT Industry**

According to [6], as the 10th largest economic scale, Korea is highly relied on ICT industry in her economic development. In 2011, total amount of ICT production is estimated 326.5 billiondollars, which includes 58.7 billiondollars in ICT services, 241.9billion dollars in ICT devices, and 26 billion dollars in software. While growth of GDP and ICT industry in Korea was reported 6.3% and 17.7% in 2010, those slightly decreased as 3.6% and 7.9% in 2011 mainly due to the global crises.

In the Korean economic structure, proportion of ICT export is unconditionally high. In 2011, for instance, ICT industry takes 25.1% of whole national export and 13.4% of import. Also, in 2011, Korea is the 2nd largest economy in ICT export and the 5th largest in the global ICT trade balance. Top ten products in export include semiconductor (3rd), flat-panel display (5th), wireless communication devices (6th), and computer (10th). Further, these products share the high proportion of global market; semiconductor industry shares 51.3% of global market, while LCD panel industry shares 53.3%.

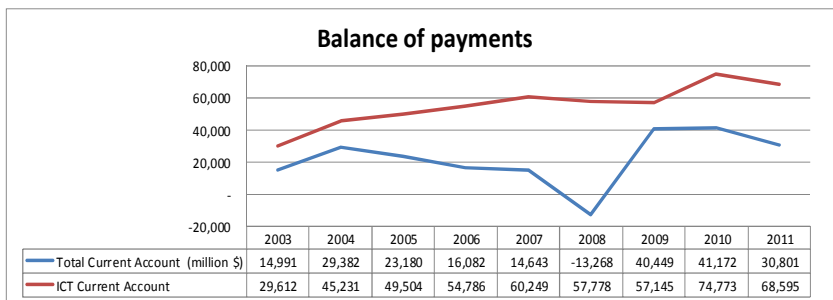


Fig. 1. Trade balance and ICT Industry in Korea

Table 1. Trade in Total Industries and ICT in Korea(Source : Ministry of Science, ICT, and Future Planning [8])

(Unit: Million US\$, %)

Years		2003	2004	2005	2006	2007	2008	2009	2010	2011
Export	Total	193,817	253,845	284,419	325,465	371,489	422,007	365,534	466,384	555,214
	ICT	68,394	90,872	99,097	109,029	121,005	121,538	111,254	140,546	139,103
	%	35.2	35.8	34.8	33.5	32.6	28.8	30.6	30.1	25.1
Import	Total	178,827	224,463	261,238	309,383	356,846	435,275	323,085	425,212	524,413
	ICT	38,782	45,641	49,593	54,244	60,756	63,760	54,109	65,772	70,507
	%	21.7	20.3	19.0	17.5	17.0	14.6	16.7	15.5	13.4

### 3 ICT Policies in Other Countries

According to ICT Annual Report [6], major countries increase government funds and subsidiary programs on ICT industry, which is regarded as a fundamental element in developing whole national economy. In 2010, for instance, the National Telecommunications & Information Administration of the U.S. Department of Commerce initiates ‘The National Broadband Plan’. This initiative aims at economic recovery via increasing government expenditure in ICT investment and expanding broadband.

EU focuses on socio-economic potential of ICT industry through which EU countries address the emerging challenges. In this light, the Digital Agenda released in 2010, sets primary goals of preparedness of ICT related institutions, establishment of service standardization, online security, networks among business groups in communication services, R&D budget and civic participation.

Among Asian countries, Japan and China as well as Korea lead ICT industry. Japan emphasizes the human-friendly ICT strategy, shifted from provider oriented one, during the period of 2009-2015. Japanese ICT strategy aims at increasing ICT convenience in various areas such as e-government, healthcare, and ICT education.

In 2010, Chinese government identifies 7 primary industries in purpose of expanding the proportion of these strategic industries in GDP up to 8% in 2015 and 15% in 2020. These industries are new generation of ICT, novel material, bio, electronic vehicle, cutting-edge devices, alternative energy, energy efficiency and environmental industry.

As described above, major countries such as U.S., EU, Japan and China dominate global market of ICT industry. The common features of these countries are that they regard ICT as a fundamental element in national economy, since ICT industry is likely to combine with other related industries. Also, these governments focus on the capability of ICT in addressing social problems. Indeed, via internet-based networks, global collaborations increase in order to deal with common problems such as global climate changes. Moreover, development of e-government contributes to better public services. Most of developed countries establish e-government system in public services area such as healthcare, welfare and education. Finally, these major countries support ICT industry with indirect policies as well as direct ones, which ranges from conventional regulations such as standardization of ICT services and information security to supportive policies such as public procurement and tax redemptions.

In accordance with the opinion of [10], we characterized and compared the evolution of the ICT policy of the most important global players in ICT. In absolute values, the top world ICT manufacturing economies are the USA, China and Japan, while the economies that invest most in ICT manufacturing business R&D are the USA, Japan and the EU. But the USA is also the world leader, followed by the EU in ICT services.

Korea and Japan are the most “ICT-specialized” economies. Especially the ICT industry of Korea is more specialized in manufacturing, while it is strongly oriented toward services in the EU, the USA and Japan.

The policy maker of Korea focused on ICT hardware like as infrastructure and promotion to export-oriented enterprises. On the other hand, the U.S. and the EU focused on ICT software in addressing social problems, intellectual properties, capitalized software and other intangible.

Interestingly, ICT market in China is shifting from hardware oriented to software oriented services.

## 4 ICT Policies in Korea

Ref [11] provides a brief map to understand ICT policies in Korea. In 1994, Korea has a turning point in ICT policy, as the Korean Ministry of Information and Communication is established. As a control tower in ICT policy, KMIC integrates ICT related policies in other government bodies into comprehensive institutions, which results in the establishment of the Comprehensive Plan for ICT Infrastructure. In 1999, The National Information Infrastructure Protection Act is established as a basic law, which contains 3 phases of master plans for stimulating ICT industry. In 2000s, Korean government develops diverse ICT policies such as 31 roadmaps for e-government, 10 driving forces for next generation, and “839(8 services, 3 infrastructures, 9 industries)

strategies” for ICT, which aims at intensive investment in ICT industry. Particularly, in 2008, Korean government decentralizes ICT authority into 4 related authorities, responding to multi-convergence in ICT industry. As a result, 4 government agencies such as the Korean Ministry of Public Administration and Security (KMOPAS) for infomatization at national level, the Korean Ministry of Knowledge Economy (KMKE) for industrial policy, the Korean Ministry of Culture, Sports and Tourism (KMCST) for digital contents, and the Korea Communications Committee (KCC) for communications and media collaborate in national ICT policies (See Fig. 2.).

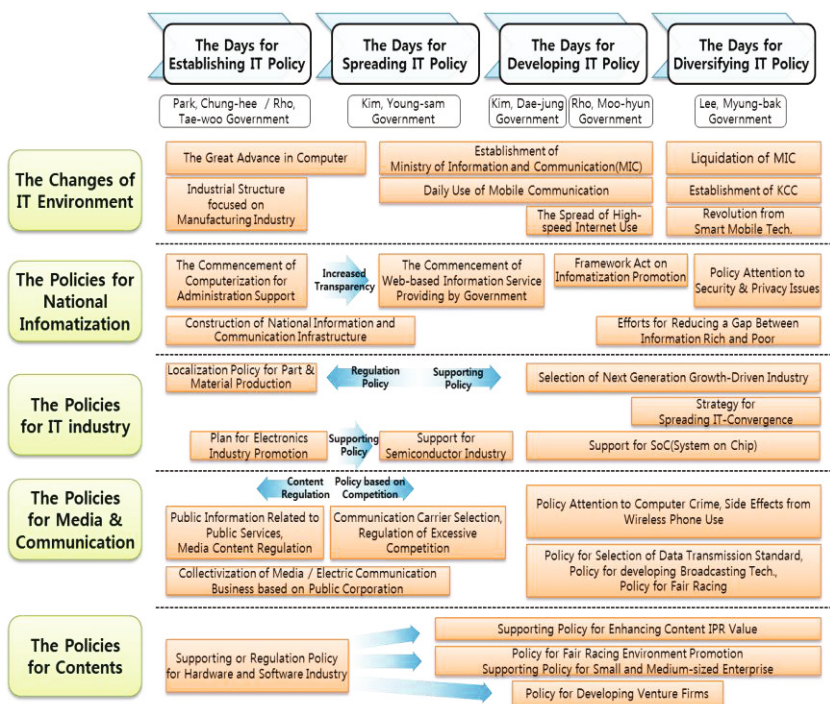


Fig. 2. Trends of ICT Policies in Korea (Source: Park, Soo-yong [11])

The main streams of Korean ICT policies can be summarized as below:

First, government plays a critical role in ICT policy with strong leaderships. For instance, former Kim (Kim, Young-sam) administration controls and coordinates ICT policy through the KMIC. Based on such an established system, another Kim (Kim, Dae-jung) administration focuses more on the long-term strategy in ICT policy, the 1st phase of ICT initiative selecting e-government project as one of 11 national priorities. Also, the Rho (Rho, Moo-hyun) administration continues the long-term platform in ICT policy in the sense that it strengthens support on building public ICT infrastructure, the 2nd phase of ICT initiative (31 roadmaps for ICT). In this light, the development of ICT industry in Korea is driven by government in terms of its development of technology and industrialization. According to [2], for instance,

government organizes and controls whole process of R&D and production of TDMA and CDMA. In such a system, R&D process in ICT industry is crucially relied on government funding and government-funded research institutes.

Second, intensive investment on public infrastructure enables Korea to catch up the first comers in ICT industry. Under the national priority of building information society, Korean government keeps intensive supports on ICT industry. In the last decade, for instance, total investment on national informatization is up to 28.8 billion dollars, which is 2.9 billion dollars annually. Since 1995 when government establishes intra-networks among government agencies, ICT infrastructure has been extended into education and business. Particularly [3] point out that the Dae-duk Technopark, benchmarking technopolos like the Silicon Valley in the U.S., provides a locus to stimulating innovative collaborations among various actors including national labs, universities and business entities, which lead to the world first rate of spread of high-speed internet and consequent outcomes such as development of e-commerce, communication and entertainment.

Third, the development of ICT industry in Korea is attributed to direct supports from government. Conventionally, the industrial structure of Korea is oriented toward export by government. In this circumstance, government driven ICT industry grows as 14% in 2010, which shares 9% of national GDP. In this light, increase of export surplus in ICT industry is regarded as a national growth factor in economic development.

Fourth, ICT policy drives a shift towards competition-based institutions. Privatization of government utilities in ICT industry such as Korea Telecom (KT) stimulates market-based competitions. According to [2], deregulations on entry barriers in telecommunication business are successful in preventing from monopoly and oligopoly in telecommunication market.

Fifth, recent ICT policies in Korea target convergence with other major industries. As the ICT technology and market reach the mature stage, ICT strategy is changed from unified control by government to multi-dimensional supports by multiple governments. Since 2005, diverse policies provide extensive opportunities to converging ICT with other major technologies and industries, which includes the Strategy for developing convergent technology in 2005, the Strategy for New ICT in 2008, ICT Korea in 2009, and the Strategy for Diffusion of ICT Convergence in 2010. These strategic changes stimulate the convergence of ICT and conventional major industries such as automobile, ship, construction and textile.

## 5 Assessment of ICT Policy in Korea

Korean model of ICT policy shows that balance between strong leaderships from government and competition-based market leads ICT industry into a main driving force in national economy. However, this paper indicates several limitations of Korean ICT policy as below:

First, despite the world top-tier ICT infrastructure, attention to the utility of ICT in addressing complicated social problems and creating added-values has been less



taken. Although ICT policies have accomplished well the given goals of industrialization and productivity, in terms of cyber count-intelligence capability[4]<sup>1</sup>, those policies are limited to dealing with complicated social problems in education, disaster management, security.

Second, competitiveness in software industry is still needed, while the hardware industry takes a world-best position in ICT industry. Particularly, the shift towards convergence also focuses more on hardware industry rather than software. But we think that software technology and industry are becoming more important. There are examples in [1], software production was virtually non-existent in the early 1980s. The Indian, Irish and Israeli software industries export a substantial fraction of their output (and services) to advanced economies, particularly the U.S. Today software employs more than 450,000 employees, sustaining annual growth rates of 30-40% in revenues and employment over more than 10 years. Although less remarkable than India, countries like Ireland and Israel have also had double digit growth. Korean government already recognizes such problems in the sense that it has a long-term plan to support individual and small business to develop the platform for new generation software. The Ministry of Science, ICT and Future Planning should organize a software cluster, vitalize start-up software companies, and strengthen education curriculum from elementary to secondary schools.

Nonetheless, it is still far from desirable circumstances in which innovative small firms are to be protected and stimulated by indirect supports from government, and in which open culture in business as well as government system cultivates innovative atmosphere in development of ICT. If the ICT policy of hardware industry lied on public sector, now this is the time to develop business based all the established platforms.

Third, ICT industry in Korea is still lagged behind global leaders in terms of source technology and knowledge. Although large companies such as Samsung and LG lead global ICT industry in their commercialization capability, they pay the large amount of loyalty for the core and critical technology and knowledge. In such a business circumstance as large firm orientation, it is important to improve the structural problems in industry and business.

Fourth, Korea is suffered from lack of high brains in ICT as well as decrease of employment. Recently, as the ICT market reaches the mature stage, employment in ICT industry keeps decreasing despite the increase of production. As a result, the elasticity of employment in ICT industry decreases from 3.26% in 2005 to 0.25% in 2011. Therefore, regarding such a limitation of growth in ICT industry, a new education and training policy is needed to prepare for convergent ICT industry such as big data and cloud-based industries.

---

<sup>1</sup> According to [4], Habermas said "the global community seeks to enhance the quality of its members' lives". It struggles to uphold shared communitarian values and humanistic ideals: clean air; fresh water; biodiversity; unadulterated food; health care; education; child/elder care and productive work. Habermas, J. (1987). *The theory of communicative action*, Boston, MA: Beacon Press..

## 6 Conclusion

This paper indicates the main features of Korean ICT policy as a strong leadership of government and the changes towards competition-based industrial structure. However, regarding the stagnated growth of the industry and the lack of source technology and knowledge, it is strongly needed to restructuring ICT industry and reinforcing supports upon small firms, as well as focusing on the cyber count-intelligence capability. The ICT policy needs to find new direction. According to [5], [12], most European countries have either conducted national technology foresight exercises or have given serious consideration to do so. However Korea has not yet thought these issues such as labor productivity, education, environment, inequality related to ICT. ICT policy should be utilized more creatively to create new value and opportunity. Also ICT policy should assist to develop source technology and software industry.

Policy maker should establish new competitive strategies for high-tech areas like green ICT and human infrastructures and they should organize a software cluster, vitalize start-up software companies, and strengthen education curriculum.

A new role of ICT policy should rely less on state intervention by deregulation, empowerment, decentralization, and rely more on social forces and private firm.

These values and ideals are reflected and acted on collaborative governance between government and private firm.

**Acknowledgement.** This work was supported by the Sun Moon University Research Grant of 2009.

## References

1. Arora, A., Gambardella, A.: The globalization of the software industry: Perspectives and opportunities for developed and developing countries. National Bureau of Economic Research. Working Paper 10538 (2004)
2. Cheong, I.-S.: Telecommunications policy and market performance. Korean Association for Telecommunication Policies. Seminar Report of 2007, 11–39 (2007)
3. Castells, M., Hall, P.: Technopoles of the world: The making of twenty-first-century industrial complexes. Routledge, London (1994)
4. Brooke, M.: A Critical Analysis of Selected Policy Making Decisions in the US and the UK with Regard to the Implementation of Information and Communication Technology (ICT) in National State Primary and Secondary School Education Systems. Open Journal of Modern Linguistics 3(1), 94–99 (2013)
5. Martin, B.R.: Foresight in science and technology. Technology Analysis & Strategic Management 7(2), 139–168 (1995)
6. Ministry of Knowledge and Economy, Annual report on the promotion of ICT industry (2012)
7. Ministry of Knowledge and Economy, Information and Communication Technology Implementation Plan (2013)
8. Ministry of Science, ICT, and Future Planning, The ICT amount of import and export (2013)

9. Ministry of Science, ICT, and Future Planning, The Ranking of the international telecommunications index2013
10. Desruelle, P., Stancik, J.: Characterizing and Comparing the Evolution of the Major Global Players. *Information and Communications Technologies* (2013) SSRN 2242013
11. Park, S.-Y.: Policy direction of the telecommunications industry convergence era. *Information & Communications Magazine*. Korea Institute of Information and Communication Engineering 30(1), 17–23 (2012)
12. Shim, Y.-H., Kim, K.-Y., Cho, J.-Y., Park, J.K., Lee, B.-G.: Strategic Priority of Green ICT Policy in Korea: Applying Analytic Hierarchy Process. *World Academy of Science, Engineering and Technology* 58, 2009 (2009)

# Computer Graphical Score and Music Education: Application to Music Animation Machine MIDI Player

Soon-Hak Hwang<sup>1</sup> and Se-Hak Chun<sup>2,\*</sup>

<sup>1</sup> Department of Basic Education, College of Art and Social Science,  
Seoul National University of Science and Technology,  
Kongneung-gil 138 Nowon-gu, Seoul 139-743  
ricarrdo@seoultech.ac.kr

<sup>2</sup> Department of Business Administration, College of Business and Technology,  
Seoul National University of Science and Technology,  
Kongneung-gil 138 Nowon-gu, Seoul 139-743  
shchun@seoultech.ac.kr

**Abstract.** This paper shows how a graphical score has greater effect on music education than the traditional score in general courses at a university level. Generally students have difficulty in understanding music scores, thus music courses as a general selective at the university level need to provide them efficient ways to understand music. In this regard, a computer graphical score using a MIDI player is a very effective way to make students understand music structure and movement fast and accurately overcoming limitations in music class through only listening audio. Thus, this paper investigates how this computer technique can be used for music classes and discusses various cases a MIDI player plays a role in music education.

**Keywords:** Music programs, Music Animation Machine, MIDI Player, Graphical score, Music Education.

## 1 Introduction

This paper focuses on the fact that, since polyphonic (polyphony) started in western musical history, it is difficult for students whose major is music to analyze and follow songs by only using music listening and traditional scores. Therefore, it is very interesting that a recent computer application, Music Animation Machine MIDI player (hereafter MAM player) can be a effective tool for students by playing music through audio and visual monitor. MAM player is a freeware program made by Stephen Malinowski, and it has been uploaded to 'musanim.com'. This program can convert sounds to many kinds of visual graphics for people to follow songs easily, and it can play only midi files. This paper investigates how this computer technique can be used for music classes and discusses various cases a MIDI player plays a role in music education.

---

\* Corresponding author.

The remainder of this paper is organized as follows: Section 2 presents explanation about visual effect models of ‘MAM player’. Section 3 shows the examples using MAM player. In section 4, conclusions are discussed.

## 2 Literature Review and Visual Pattern of MIDI Player

Traditional music education focuses on individual practice assisted by an instructor. [1]. Due to time and financial constraints most students only have one lesson per week [2]. Some interfaces such as Vuzik can be used for creating and visualizing music through painting gestures on a large interactive surface [3]. MIROR, music notation is used as the trace of both the user and the system activity, produced from MIDI instruments [4]. In the basic bar-graph of Music Animation Machine [5, 6], each note is represented by a bar with its length corresponding to the exact duration as performed as shown in Figure 1. Music Animation Machine MIDI Player is a graphical score converter computer program, which can open midi file and shows the graphical score like pictures below and the graphics from outputs form MIDI player is very simple to be used for beginner's music classes. Here are pictures which has visual effect expressed to graphical score in playing music. For example, top left-hand corner in Figure 1 shows the basic bar-graph.

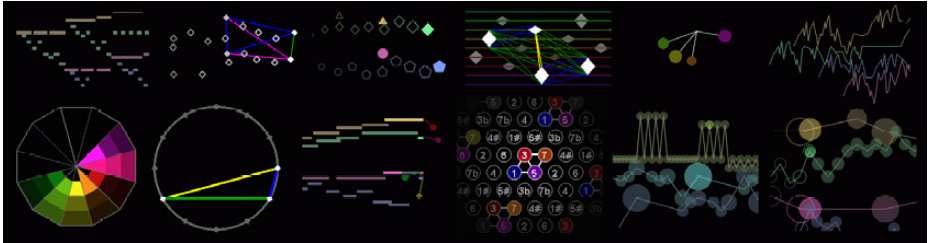


Fig. 1. Different graphical modules available on MIDI Player

## 3 Application Examples of Music Animation Machine MIDI Player

### 3.1 The Music before the 16<sup>th</sup>

In the western musical history, catholic traditional music, Gregorian chant (Fig. 2) is easy to be followed by general score, because it is a simple song using monophony and accapella.



Fig. 2. General score: Gregorian chant(mono voice style)

**Fig. 3.** General score: Josquin des Prez(3voices style)

Figure 3 shows polyphony pattern, which is difficult for beginner to follow. In musical history, the outstanding skill, Melody Imitation made by Josquin des Prez of Flandre school has been used to make many songs polyphony. After then, it is very hard to teach and follow polyphony songs by using general score.

### 3.2 The Music after the 16<sup>th</sup> with Animated Graphical Score

In listening classical music, especially symphony is the core to understand the constitution of song, but its limitation is in difficulty to follow the constitution of songs by using general score. However, using a graphical score, songs show simple visual forms effectively. So, beginners can easily follow music by using a graphical score. Figure 4 shows how Ludwig van Beethoven's Fifth Symphony, first movement, can be represented by animated graphical score.

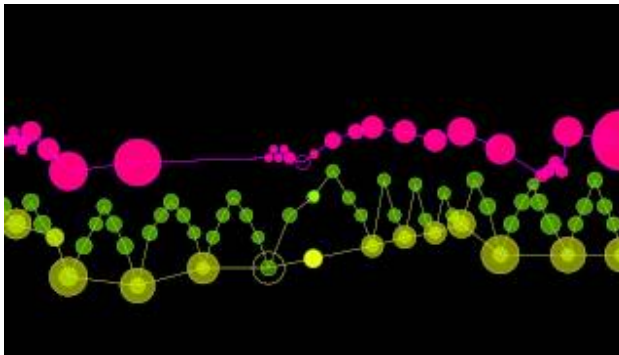


**Fig. 4.** Beethoven's Fifth Symphony with animated graphical score

Principles of sectional structure, particularly in sonata form, were firmly established in the late 18<sup>th</sup> century. Phrase structure was characteristically clear with well-defined cadences, and phrases were shorter (most commonly four measures) than in the Baroque. Especially, in studying Classical Music, students can understand the songs which has the style like Choral (all the parts in song are start and finish in the

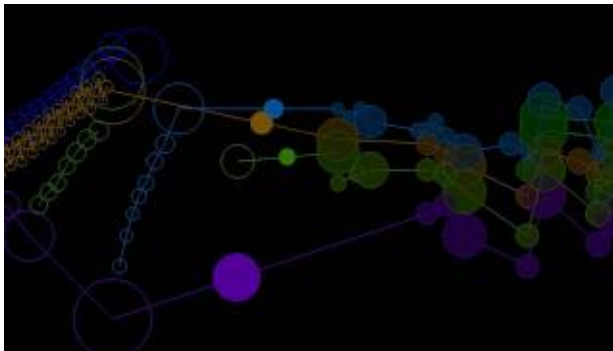
same time), which is different from the past forms, Melody Imitation and Basso Continuo Style.

Figure 5 shows Fryderyk Franciszek Chopin's 'Fantasie-Impromptu', op. 66 with animated graphical score. In Romanticism Music, songs have a point which can effectively transfer the emotional mind. To maximize emotional expression, Romanticism composers used Tempo Rubato which often stops metronome from tic-toc. Especially, Chopin used Tempo Rubato so many times, and his graphical score shows the timely distances from circle and linear graph. Romantic texture, as in classical music, were still basically homophonic. Counterpoint, when used, was of secondary importance. in terms of sonority, 19th-century music was notable for a marked increase of sound.



**Fig. 5.** Chopin's Fantasie-Impromptu with animated graphical score

Figure 6 shows Claude Debussy 'Arabesque' with animated graphical score. Impressionism music shows a point which are using new 'Impressionism 5 notes scale' different from 7 notes scale.



**Fig. 6.** Debussy's Arabesque with animated graphical score

The program can show the point by graph which shows the adding by many sounds. The first important trend toward 20th-century modernism in music was impressionism. In the hands of Claude Debussy, it paralleled movements in French painting, sculpture, and poetry.

Figure 7 shows Igor Stravinsky's 'The Rite of Spring' with animated graphical score. In Stravinsky's 'The Rite of Spring', Graphical Score shows well many harmonies including discordant sounds and drum sounds which have explosive energy. Neoclassicism is a very extensive and pervasive trend. Beginning about 1920, it continues to be a dominant trend today. In a general sense it implies a return to pre-romantic ideals of objectivity and clarity of texture, but it is not confined to 18th-century classicism.



**Fig. 7.** Stravinsky's The Rite of Spring with animated graphical score

### 3.3 Result and Discussion

Table 1 summarizes some results and implications drawn by applying to computer graphical score according to various types of music. As shown in table 1, how graphical score can overcome limitations in existing music class and shows solutions for future music class. Knowing constitution of classical music make it easy to listen songs. Graphical score's visual effect can help many people who did not studied, majoring in music, to understand the constitution and characteristic of songs. This visual graphical score with MIDI player is a very effective tool for some music class, especially in such as a music listening class.

**Table 1.** Types of Music and Findings

Music types and Characteristics	Problem: in existing education	Solutions: The Effect of Graphical Score, and each Implication.
Monophony music(Gregorian chant)	Not difficult to be followed.	Not essential to be used by Visual effect and Graphical score.
Polyphony music	Difficult for beginners to follow plural sounds.	Essential to be used by Visual effect and Graphical score.



**Table 1.** (continued)

Understanding Symphonic constitution Beethoven's Fifth Symphony	Difficult to understand constitution of symphonies in Sonata form just in listening.	Graphical Score is helpful for people to understand the sonata form.
Understanding Tempo rubato: Chopin 'Fantasie-Impromptu'	Difficult for beginners to understand rhythmical differences just in listening.	Expressing the Tempo rubato well, Romanticism songs by circle and linear graphs.
Debussy's Impressionism music. Claude Debussy 'Arabesque'	Difficult for beginners to follow Impressionistic melody just in listening.	Graphical Score shows Impressionistic melody effectively which has been free from vertical and horizontal melody.
Igor Stravinsky's 'The Rite of Spring'	Difficult to follow discordant sounds by general score.	Graphical score shows the cut-offs by discordant sounds well.

## 4 Conclusion

Each famous composers represents unique characteristics according to musical ages. Therefore, playing music through the only radio machine is not effective to teach people not having professional knowledge. Music Animation Machine MIDI Player which has fused acoustic and visual points has creative item that people can approach to this program easily and raise themselves to the professional level. The program could be applauded for the point that it gives professional group delicate analytical skills. It is a surprising 'innovation' to be able to use a MAM program in Listening Music Class. It is a good point that the MAM program can convert musical constitution to visual graphic, and help people follow music. MAM program solves problem that analogue listening without professional knowledge can't be made easily, therefore, it made basis to enjoy complex classical music songs.

## References

- [1] Chow, J., Feng, H., Amor, R., Wünsche, B.C.: Music Education using Augmented Reality with a Head Mounted Display. In: Proceedings of the Fourteenth Australasian User Interface Conference (AUI 2013), Adelaide, Australia (2013)
- [2] Percival, G., Wang, Y., Tzanetakis, G.: Effective use of multimedia for computer-assisted musical instrument tutoring. In: Proceedings of the International Workshop on Educational Multimedia and Multimedia Education, Emme 2007, pp. 67–76. ACM, New York (2007)
- [3] Pon, A., Ichino, J., Sharlin, E., Eagle, D., d'Alessandro, N., Carpendale, S.: VUZYK: A Painting Graphic Score Interface for Composing and Control of Sound Generation. In: ICMC 2012, September 9-14, Non-Cochlear Sound (2012)
- [4] Fober, D., Kilian, J.F., Pachet, F.: Real-Time Score Notation from Raw MIDI Inputs. Technical report n 2013-1 SONY Computer Science Laboratory Paris 6 rue Amyot, 75005 Paris (July 2013)

- [5] Malinowski, S.: Music animation machine,  
[http://en.wikipedia.org/wiki/Stephen\\_Malinowski](http://en.wikipedia.org/wiki/Stephen_Malinowski)
- [6] Alessandra, A.: Music Animation Machine: Intervista a Stephen Anthony Malinowski.  
*Musica Domani* 42(163), 32–38 (2012)

# A Newton's Universal Gravitation Inspired Firefly Algorithm for Document Clustering

Athraa Jasim Mohammed, Yuhanis Yusof, and Husniza Husni

School of Computing, Universiti Utara Malaysia,  
06010 Sintok, Kedah, Malaysia  
autoathraa@yahoo.com,  
{yuhanis, husniza}@uum.edu.my

**Abstract.** The divisive clustering has the advantage to build a hierarchical structure that is more efficient to represent documents in search engines. Its operation employs one of the partition clustering algorithms that leads to being trapped in a local optima. This paper proposes a Firefly algorithm that is based on Newton's law of universal gravitation, known as Gravitation Firefly Algorithm (GFA), for document clustering. GFA is used to find centers of clusters based on objective function that maximizes the force between each document and an initial center. Upon identification of a center, the algorithm then locates documents that are similar to the center using cosine similarity function. The process of finding centers for new clusters continues by sorting the light intensity values of the balance documents. Experimental results on Reuters datasets showed that the proposed Newton inspired Firefly algorithm is suitable to be used for document clustering in text mining.

**Keywords:** Firefly algorithm, Newton's universal gravitation, divisive clustering, document clustering, text mining.

## 1 Introduction

Clustering is grouping similar objects into a cluster and dissimilar objects in another cluster [1]. In general, the clustering algorithms are classified into two categories: Hierarchical clustering algorithms and Partition clustering algorithms [2]. Hierarchical clustering algorithm is a technique to build a hierarchy of clusters [1, 2] which is very efficient for document clustering. There are two approaches of this technique agglomerative hierarchical clustering and divisive hierarchical clustering. Agglomerative hierarchical clustering processes the objects from bottom to top, which means every object is a single cluster, and later are merged based on some criteria [3]. The divisive hierarchical clustering processes the objects from top to bottom. Such a clustering initially locate all objects in one cluster and the cluster is later divided using one of the partition clustering algorithms. The partition clustering algorithms have some drawbacks which includes the local optimum problem in selecting initial centers of clusters.

This paper proposes the use of Firefly algorithm (FA), a nature inspired meta-heuristic algorithm that was developed by Xin-She Yang in 2007 at Cambridge University. FA has two important issues: (a) the light intensity; and (b): the attractiveness. The light intensity,  $I$ , of a Firefly is related with the objective function and the attractiveness is relative. It changes depending on the distance between two fireflies [11]. The advantage of FA is the effectiveness in locating global optima [7]. In practice, the objective function of FA is to minimize the distance between objects and the center of a cluster. In this paper, we propose to employ the Newton's law of universal gravitation formula as the objective function. The law states that "Every point mass attracts every single other point mass by a force pointing along the line intersecting both points. The force is proportional to the product of the two masses and inversely proportional to the square of the distance between them." [9]. The formula is as shown in equation 1:

$$F = G \frac{M_1 * M_2}{R^2} \quad (1)$$

Where:  $F$  is the force between two masses,  $G$  is the gravitational constant,  $M1$  is the first mass,  $M2$  is the second mass,  $R$  is the distance between two masses.

## 2 Related Works

One of the shortcomings of partition clustering is that it causes the solution to be trapped at local optima based on the initial centroids selection. K-means is one of the most favorable methods that is used to partition datasets into clusters. K-means selects initial centers randomly and assigns each object to similar and nearest centers. Hence, clustering performance is highly depending on initial centers selection. Various work have been done to solve the problem of initial centers and local optima. Optimized K-means clustering has been proposed to select initial centroids by calculating the midpoint for each term [4]. The proposed algorithm produced less intra-cluster and high inter-cluster but the number of clusters was pre-defined. Another approach was proposed to find initial centroids [5]. The farthest distance pair of two objects was selected as initial center and objects are assigned to the nearest center. The result of Dunn's index (DI) and the Davies-Bouldin's index (DBI) in far efficient K-means was better than K-means. Additionally, a constrained K-means clustering method, S3-Kmeans, was proposed in [6]. The findings showed that the S3-Kmeans is better than three comparable methods which are K-means, S-Kmeans and Cop-Kmeans.

The FA performance in clustering has been investigated in [2]. The obtained result indicates that the proposed FA was efficient, robust, and reliable in generating optimal cluster centers compared to the Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC). On the other hand, a hybrid of FA and K-means for data clustering (KFA) has also been proposed in [7]. Similarly, five nature inspired optimization algorithms (Firefly, Cuckoo, Bat, Ant and Wolf) have been combined with K-means [8]. The result indicated that C-bat, C-cuckoo and C-wolf performed the best in objective values and the C-bat was the fastest execution algorithm. Furthermore, a hybrid

gravitational search algorithm with K-means, GSA-KM, for data clustering was presented in [9]. Gravitational search algorithm prevents K-means from falling into local optima and also speeds up the convergence of GSA. In addition, the problem of randomly selected initial cluster centers of K-means is also solved by integrating with Harmony Search optimization [10]. Five clustering algorithms, i.e. K-means, HSCLUST, Sequential hybrid, interleaved hybrid, and GA were compared. The result indicated that the proposed hybrid Harmony Search optimization with K-means performs better in terms of F-measure, Entropy, Purity and ADDC.

In this paper we use a Firefly algorithm to find the initial cluster center using the new objective function based on the Newton’s law of universal gravitation. The objective function is maximizing the force between each documents and center of cluster which then leads to less intra distance between documents and center of cluster.

### 3 The Proposed Gravitation Firefly Algorithm

Document clustering includes three phases; first is the preprocessing, second is the vector space construction and is followed by the clustering. The preprocessing phase extracts only the title and body of each document and cleaning the extracted documents from digit, special characters, as well as removing the selected words. Furthermore, each term in the document is stemmed and counted for its frequency [15]. The vector space construction [10] phase will then create a normalized term frequency matrix, TF, which contains the frequency of each term in the documents. This is followed by producing a TF-IDF [10] matrix which contains the weight of each term in the documents. By using the TF-IDF matrix, we calculate the total weight of each document. The total weight of one document is obtained by the summation of all term weight in the document using equation 2 [16]:

$$total\ weight_{d_j} = \sum_{i=1}^m tf - idf_{t_i,d_j} \tag{2}$$

Where:  $j$  is the number of document,  $i$  is the term number.

The proposed GFA later calculates the distance between any two documents,  $i$  and  $j$ , at  $D_i$  and  $D_j$  using Euclidean distance [7]. The formula is shown in equation 3:

$$Euclidean\ distance(D_i, D_j) = \sqrt{(D_i - D_j)^2} \tag{3}$$

Similarity between two documents,  $i$  and  $j$ , at  $D_i$  and  $D_j$  is identified based on cosine similarity. The value of vectors,  $D_i$  and  $D_j$ , are obtained from the normalized matrix and the similarity formula becomes as in equation 4:

$$similarity = \sum_{j=1}^m (D_j * V_j) \tag{4}$$

Where:  $j$  is the number of terms in the collection,  $D_j$ ,  $V_j$  is the documents in collection.

The force between any two documents,  $i$  and  $j$ , at  $D_i$  and  $D_j$  is later calculated using Newton’s law of universal gravitation. In this paper, we assume that  $F$  is the force

between two documents and that  $G$  is the similarity between any two documents  $i$  and  $j$ . Suppose that  $M1$  and  $M2$  are the total weight of any two documents,  $D1$  and  $D2$ , and suppose  $R$  is the distance between any two documents,  $i$  and  $j$ , the value of force is can be obtained via equation 5:

$$F(D_i, D_j) = \text{similarity} * \frac{D_i * D_j}{R^2} \quad (5)$$

In this paper, each document is represented by a single firefly. Hence, the initial population of fireflies is given by  $D = (D1, \dots, Dn)^T$  where  $n$  is the number of documents in the collection. GFA later constructs a matrix of  $D_i * D_j$  where row is the centers of documents and column is the documents. The intersection between row and column is the values of force between any two documents that is obtained using equation 5. The summation of each row in matrix  $D_i * D_j$  is calculated using formula 6 and is used as the initial light intensity. In addition, the objective function,  $f(D_i)$ , is treated as a maximization problem.

$$\text{light intensity centers}(I_i) = f(D_i) = \sum_{j=1}^n F(D_i, D_j) \quad (6)$$

The proposed GFA is used to obtain the global best document in the collection and this is done by identifying document with the highest brightness. The identified document is later known as the center of cluster. Upon the identification, clustering process continues by calculating intra similarity between documents to the center using cosine similarity (as in equation 4). Documents with similarity value that exceed a pre-defined threshold will be located in the first cluster while the rest are located in another cluster. On the other hand, the threshold value is of experimental basis. The documents in the second cluster will then undergo the ranking process of  $I_i$  values in order to find a new center for a new cluster. The process of finding centers and clusters will continue until the very last document.

## 4 The Experiment

A standard benchmark text classification dataset, called Reuters-21578 [12] is utilized in evaluating the proposed GFA. The selected documents are of two collections, RE0 and RE1. The datasets are split into three parts, two parts are used for training and one is used for testing. The description of the collections is provided in Table 1.

**Table 1.** Description of Data

Data Set	No. of Documents	Classes	No. of Training data	No. of Testing data	No. of Terms
RE0	201	13	134	67	2149
RE1	192	25	128	64	2156

### 4.1 Performance Evaluation

The performance of the proposed algorithm is evaluated using Classification Error Percentage (CEP) [2] and F-measure [13]. After clustering documents from the training dataset using the proposed GFA, the label is manually assigned to the clusters. The CEP is then calculated by finding the similarity between each produced center of clusters and documents in the test dataset. The value of CEP is obtained by checking on the known document class with the produced class as suggested by GFA. If it is not in the same class then the document is considered as misclassified otherwise correctly classified. The equation of CEP is shown in equation 7 [2]:

$$CEP = \frac{\text{number of misclassified documents}}{\text{total number of documents in test data set}} * 100 \tag{7}$$

In addition, we also employ the F-measure to measure the accuracy. The equation of F-measure is shown in equation 8 and 9 [13]:

$$F(\Omega_k) = C_j \in \{C_1, \dots, C_k\} \left( \frac{2 * R(\Omega_k, C_j) * P(\Omega_k, C_j)}{R(\Omega_k, C_j) + P(\Omega_k, C_j)} \right) \tag{8}$$

$$Total\ F - measure = - \sum_{k=1}^c \frac{|\Omega_k|}{N} * \max(F(\Omega_k)) \tag{9}$$

## 5 The Results

The proposed GFA is executed on both dataset as shown in Table 1 and the results are shown in Table II. The experiment result shows that the obtained CEP is 28.35 for RE0 and 23.43 for RE1 while the F-measure is 0.4387 and 0.6626 respectively. Based on literature [13], it is learned that a good clustering is when the CEP value is low (less than 30) and the F-measure is high (larger than 0.5). Hence, the obtained result indicates that the proposed Firefly algorithm may be promising to be used for documents clustering.

**Table 2.** Result of GFA

Data Sets	CEP	F-measure
RE0	28.35	0.4387
RE1	23.43	0.6626

## 6 Conclusion

In this paper, we proposed a new approach of FA for document clustering that uses Newton’s law of universal gravitation as the objective function. The objective function maximizes the force between each document and a center. In the proposed algorithm, each document is treated as a single firefly and the force between two

documents is calculated using the proposed force function. The proposed GFA is used to find document with the highest brightness, which later is identified as the center of cluster (centroid). Assuming one document represents the centroid, the force of this document is the summation of force between the document and the remaining documents. In the evaluation, the obtained results indicated that the proposed GFA is promising in document clustering of text mining. Future work may later be undertaken to evaluate the GFA on larger datasets.

## References

1. Das, S., Abraham, A., Konar, A.: *Metaheuristic Clustering*. Studies in Computational Intelligence, vol. 178. Springer, Heidelberg (2009)
2. Senthilnath, J., Omkar, S.N., Mani, V.: Clustering Using Firefly Algorithm: Performance Study. *Elsevier, Swarm and Evolutionary Computation* 1(3), 164–171 (2011)
3. Wilson, H., Boots, B., Millward, A.A.: A Comparison of Hierarchical and Partitional Clustering. *Techniques for Multispectral Image Classification* 3, 1624–1626 (2002)
4. Poomagal, S., Hamsapriya, T.: Optimized K-means Clustering with Intelligent Initial Centroids selection for web search using URL and Tag contents. In: *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*. ACM (2011)
5. Mishra, B.K., Nayak, N.R., Rath, A., Swain, S.: Far Efficient K-means Clustering Algorithm. In: *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, pp. 106–110. ACM (2012)
6. Hu, G., Zhou, S., Guan, J., Hu, X.: Towards Effective Document Clustering: A Constrained K-means Based Approach. *Information Processing and Management* 44(4), 1397–1409 (2008)
7. Hassanzadeh, T., Meybodi, M.R.: A New Hybrid Approach for Data Clustering Using Firefly Algorithm and K-means. In: *Proceedings of the 16th IEEECSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, pp. 007–011 (2012)
8. Tang, R., Fong, S., Yang, X.S., Deb, S.: Integrating Nature-Inspired Optimization Algorithms to K-means Clustering. In: *Proceedings of the 7th International Conference on Digital Information Management (ICDIM)*, pp. 116–123. IEEE, Macau (2012)
9. Hatamlou, A., Abdullah, S., Nezamabdi-pour, H.: A Combined Approach for Clustering Based on K-means and Gravitational Search Algorithms. *Elsevier, Swarm and Evolutionary Computation* 6, 47–52 (2012)
10. Forsati, R., Mahdavi, M., Shamsfar, M., Meybodi, M.R.: Efficient stochastic algorithms for document clustering. *Elsevier, Information Sciences* 220, 269–29 (2013)
11. Yang, X.S.: *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, United Kingdom (2011)
12. Lewis, D.: The reuters-21578 text categorization test collection (1999), <http://kdd.ics.uci.edu/database/reuters21578/reuters21578.html>
13. Murugesan, K., Zhang, J.: Hybrid Bisect K-means Clustering Algorithm. In: *Proceedings of the IEEE International Conference on Business Computing and Global Information*, pp. 216–219. IEEE, Shanghai (2011)
14. Michael, S., George, K., Vipin, K.: A Comparison of Document Clustering Techniques. In: *Proceedings of the KDD Workshop on Text Mining* (2000)
15. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*, 1st edn. Cambridge University Press (2008)
16. Mohammed, A.J., Yusof, Y., Husni, H.: Weight-Based Firefly Algorithm for Document Clustering. In: *Proceedings of the First International Conference on Advanced Data Engineering (DaEng 2013)* (2013) (accepted for publication)



# A Study on the Correlation between the Customers' Perceived Risks and Online Shopping Tendencies

Chiung En Huang, Ping Kuo Chen, and Chih Chung Chen

Aletheia University, No.70-11, Beishiliao, Madou Dist., Tainan City 721, Taiwan (R.O.C.)  
a3126747@gmail.com, a1104100@ms23.hinet.net,  
jason556@mail.au.edu.tw

**Abstract.** With the growing popularity of the internet infrastructure, the online shopping patterns have been gradually accepted by the public. This study aims to integrate relevant theoretical literatures related to the online shopping tendencies and perceived risks etc. into the theoretical frameworks and attempts to empirically identify the relationship between the customers' perceived risks and online shopping tendencies and thus, to have a understanding of consumers' shopping behaviors. In this study, the related software of SPSS 19.0 was adopted as a data analysis tool, and the analytical methods included in this study were the reliability analysis, factor analysis, correlation analysis and regression analysis etc. The research results found that there was a correlation between the online shopping tendencies and the perceived risks, and the online shopping tendencies had a significant influence on the perceived risks.

**Keywords:** Online shopping tendencies, Perceived risks.

## 1 Introduction

Nowadays, the development of e-Commerce has greatly changed consumers' shopping behaviors. The most obvious impact on shopping behavior would be that almost every shopper performs preliminary price checks over the internet before actually buying them. Sufficient information must be obtained before the purchase. The internet provides a convenient and instant exchange of information, having sufficient information not only reduces the chance of info asymmetry, but provides consumers with the opportunity to make more beneficial purchase decisions. However, the e-Commerce environment is not perfect, and may sometimes receive excessive information (information overload), which could be a burden for customers. In an e-Commerce environment, lack of assistance from the staff, consumers must procure the means of self-service in order to complete the transaction. At this point, consumers must undergo product evaluation themselves. What is the difference between consumers' decision-making processes on e-Commerce sites and traditional shopping environments; it's an important issue worth understanding particularly when facing development boom over e-Commerce.

According to the report on "A Survey of Taiwanese Internet Users' Living Behaviors" co-conducted by the Insight Xplorer Limited and ECRC-FIND of Institute

for Information Industry, it was pointed out that security mechanisms were the most important factor for selecting online shopping websites, wherein, up to 87.6% of the online purchasers admitted that security mechanism is a critical factor; and in addition, the "Survey on Taiwan Internet Usage" also revealed the issue which most internet users are concerned about when doing the online shopping: personal data confidentiality, which is concerned by 50% of total internet users, followed by "safe and secure transactions" (31%). Both factors accounted for 81% of total internet user concerns, indicating that information security over the net is still doubted by most users. Therefore, when e-Commerce websites are trying to increase the consumers' willingness to make new purchases, the most important issue will be "how to enhance the convenience of online shopping and to reduce the doubts on the online transaction security".

The result of a research showed that consumers might perceive more purchasing risks in non-store environments than in normal stores [1]; the consumer's online perceived risk was a major obstacle for developing the electronic business [2]. Therefore, the objective of this study is to survey the correlation between the online shopping tendencies and online consumers' perceived risks in order to explore the relationship as well as its influence between their attitudes towards online shopping (website security, availability, convenience, sense of trust and logistics) and Online consumers' perceived risks (financial risk, performance risk, privacy risk, psychological risk, and time risk).

## **2 Online Shopping Tendencies**

The online shopping mechanism is constructed on the basis of the virtual stores over the Internet, in which, the consumers use the Internet to enter the virtual stores to purchase goods, hence, the online shopping is being defined as a website that is able to provide goods or services through the Internet, and meanwhile accepting the customers to place online orders directly. From the suppliers' perspective, the Internet shopping is to set up an online shopping website in order to create business profits by providing all kinds of goods or services through the Internet; for consumers, online shopping means that he or she can access the website through Internet for purchasing goods or services.

## **3 Perceived Risks**

The perceived risk is a potential level of risk a consumer believes exists while making a specific purchase decision. [3] Pointed out when a consumer was making a purchase decision, it was not often completely assured that he/she could fully achieve the purpose of online shopping, so the perceived risk could be defined as follows: When consumers making their purchase decisions, they might encounter a series of unpleasant results. Thus, the perceived risk can be defined as: The consumers might perceive the uncertainty and the possibility of adverse and undesirable events during the purchase of goods and services [4], wherein, it was a subjective expected loss [5].

Online shopping not only provides a convenient way for consumers to find and shop for the goods with many options, but also overcomes the limitation and obstacles which may occur during physical transaction. However, online shopping is conducted virtually, which may result in problems such as security of payment methods, unavailability to check products, and issues of personal privacy etc., which leads to more uncertainty and the increase in consumers' perceived risks, as well reduces consumer's' willingness toward online purchase. Therefore, to understand what factors contribute to the risk of online shopping, and to find out whether the perceived risks can influence the consumers' shopping behaviors or not have become critical issues. According to [6], it was found that the consumer's perception of risk, a valuable cognition and trust would have a direct impact on the willingness to purchase online, and in which, the perceived risk is an important factor affecting the consumer's willingness to accept e-Commerce. Thus, the perceived risk is a critical factor affecting the consumers' online shopping behaviors, and therefore, how to reduce the consumers' perceived risks as well as increase the willingness of online shopping have become the major concerns for most of online shopping websites.

The perceived risks could be affected by the personal, situational and cultural factors [7]. [8] Revealed that the risk was associated with "trade-offs". [9] Considered that a purchase decision might contain multiple types of risk patterns, however, it was still associated with different levels of risks; for instance, the purchase of a luxury sedan might have a high financial risk, but it had a lower performance risk and social risk. In the relevant studies of consumers' behaviors, the perceived risks had been classified into six dimensions comprising: financial, performance, social, psychological, physical and time risk etc. [10] proposed that the perceived risks that affected the purchasing behavior on the Internet comprising: performance risk, financial risk, social risk, psychological risk and time risk. In this study, the perceived risks are classified into five categories, which include the financial risk, performance risk, psychological risk, time risk and privacy risk.

#### **4 Online Shopping Tendencies and Perceived Risks**

Internet shopping was not only convenient and time-saving, but also it had the abundance of free online reference information, and the consumers could even easily compare the prices and product characteristics among different suppliers [11]. Although significant progress has been made on B2C e-Commerce transactions in Taiwan in recent years, but while comparing with its scale of global markets or B2C e-Commerce markets in the United States, there is still great potential for Taiwan's online shopping market, and the main reasons are that it's not easy to convince the elderly people to change their buying habits and the issue of online transaction security (such as, personal identity information has been misused or stolen, or the customers cannot get the satisfactory products or services). Apparently, the perceived risk was an important variable affecting the dimensions of customers' perceived values [12], [13] revealed that the customers would conduct a comprehensive evaluation their perceptions of profits, costs and transaction risks. [14] also suggested

that to reduce the customers' perceived risks was a critical method to strengthen their perceived values. Therefore, Hypotheses were proposed as follows.

- H1: Online shopping tendency has a significant correlation with the perceived risk.
- H1-1: The convenience of online shopping tendency has a strong positive correlation with the perceived risks, which include the financial risk, performance risk, time risk, psychological risk, and privacy risk;
- H1-2: The logistics of online shopping tendency has a strong positive correlation with the perceived risks, comprising the financial risk, performance risk, time risk, psychological risk, and privacy risk;
- H1-3: A sense of trust in online shopping tendency has a strong positive correlation with the perceived risks, comprising the financial risk, performance risk, time risk, psychological risk, and privacy risk;
- H1-4: The sense of security in online shopping tendency has a strong positive correlation with the perceived risks, which include the financial risk, performance risk, time risk, psychological risk, and privacy risk; and
- H1-5: The availability of online shopping tendency has a strong positive correlation with the perceived risks, comprising the financial risk, performance risk, time risk, psychological risk, and privacy risk.
- H2: Online shopping tendency has a significant influence on the perceived risk.

## 5 Method

### 5.1 Data Collection and Sampling

The online shopping population was deemed to be the subject of this study, the survey had been conducted by distributing questionnaires via the Internet, after removing the invalid questionnaires, and a total of 2090 valid questionnaires were obtained.

### 5.2 Measures

The questionnaire has been considered as the primary research tool in this study, which contains the online shopping tendency scale and the perceived risk scale, and the contents in each scale was measured by using a 5-point Likert Scale, and there were five ordered response levels including: 1 - Strongly Disagree, 2 - Disagree, 3 - Neither Agree nor Disagree, 4 - Agree, and 5 - Strongly Agree, in which, a higher score reflected a higher level of agreement of each item.

## 6 Analysis and Results

### 6.1 Reliability Analysis and Factor Analysis

As for the reliability analysis of online shopping tendency, Cronbach's Alpha value was 0.955 overall; each Cronbach's  $\alpha$  value for online shopping tendencies were 0.925, 0.899, 0.807, 0.763, and 0.839 respectively. [15] indicated that the 0.7 and

above was an acceptable reliability coefficient, and the results were greater than 0.9, which indicated that internal reliability of the questionnaires adopted by this study had reached a certain level, therefore, test results must be stable and reliable. There were five factors generated after performing a factor analysis, which was respectively referred to the convenience, logistics, a sense of trust, a sense of security and availability, and in which, there were total 20 items included in our survey questionnaire; the characteristic value was then calculated for each factor, which was 4.769, 3.910, 2.686, 2.388, and 1.835 respectively; cumulative percentage of total variance explained was of 19.075%, 34.717%, 45.460%, 55.014% and 70.247%, respectively.

Regarding the reliability analysis of perceived risk, overall Cronbach's  $\alpha$  coefficient was 0.907, and the Cronbach's  $\alpha$  value for the perceived risks was 0.835, 0.815, 0.835, 0.850, and 0.790, indicating that test results were stable and reliable. Through the factor analysis, there were five factors available, which represented the financial risk, performance risk, psychological risk, time, risk, and privacy risk respectively, wherein, there were total 17 items included in our survey questionnaire; the characteristic value was then calculated for each factor, which was 2.700, 2.528, 2.499, 2.498, and 2.205 respectively; cumulative percentage of total variance explained was of 15.881%, 30.749%, 45.446%, 60.023%, and 72.995% respectively.

## 6.2 Correlation Analysis

In this study, a Pearson's correlation was used to measure the degree of relationship between variables, the correlation coefficient related to the convenience of online shopping tendencies associated with the financial risk, performance risk, time risk, psychological risk, and privacy risk included in the perceived risks was 0.122, 0.261, 0.087, -0.014, and 0.232, indicating that the more convenience of online shopping, the relationship between online shopping tendencies and performance risk as well as privacy risk is becoming more positive, and thus, hypothesis H1-1 is supported.

The correlation coefficient related to the logistics of online shopping tendencies associated with the financial risk, performance risk, time risk, psychological risk, and privacy risk included in the perceived risks was 0.186, 0.380, 0.062, -0.115, and 0.248, which indicated that indicating that the more convenience of online shopping, the relationship between the logistics of online shopping tendencies and performance risk as well as privacy risk is becoming more positive, and thus, hypothesis H1-2 is supported.

The correlation coefficient related to a sense of trust of online shopping tendencies associated with the financial risk, performance risk, time risk, psychological risk, and privacy risk included in the perceived risks was 0.022, 0.137, 0.015, -0.035, and 0.170, indicating that the higher degree of trust of online shopping, the relationship between online shopping tendencies and performance risk as well as privacy risk is becoming more positive, and hence, hypothesis H1-3 is supported.

The correlation coefficient related to a sense of security of online shopping tendencies associated with the financial risk, performance risk, time risk, psychological risk, and privacy risk included in the perceived risks was 0.203, 0.411,

0.046, 0.001, and 0.330, and it indicated that the higher level of security in the online shopping, the relationship between online shopping tendencies and performance risk as well as privacy risk is becoming more positive, and therefore, hypothesis H1-4 is supported. Additionally, the availability of online shopping has a significant relationship with the psychological and privacy risks, and thus, H1-5 is supported.

### 6.3 Regression Analysis

The results of online shopping tendencies affecting on the perceived risks is shown in Table 1. Wherein, F value is 5.892, and the significance test  $p=0.016<0.05$ , which indicates that total variance explained by the regression model has reached a significant level.  $R^2$  is 0.028, which indicated that the variance explained by the online shopping tendency and perceived risk is 2.8%,  $p=0.016<0.05$ , indicating that online shopping tendency has a significant positive influence on the perceived risk, and hence, H2 is supported.

**Table 1.** The regression analysis of online shopping tendencies and perceived risks

Independent variables	Dependent variables	$R^2$	$\beta$ coefficient	t-value	F-value	Significance
Online shopping tendency	Perceived risk	0.028	0.166*	2.427	5.892	$p<0.05$

\* is used to indicate:  $P < 0.05$ ,

\*\*are used to indicate:  $p < 0.01$ ,

\*\*\*are used to indicate:  $p < 0.001$ .

## 7 Conclusions

The development of the Internet has a significant impact on our daily lives, which has completely changed various aspects for all mankind. When traditional industries are moving their businesses into e-commerce, as a result, people's consumption patterns will follow the trend of electronic business operation, and thereby, consumers do not always visit traditional retail stores, instead, patterns have been transferred to conform the virtual stores; powerful Internet search capabilities allows consumers to instantly grasp the latest merchandises. Plus, more stores are offering systems for comparing specs and prices, where consumers are able to purchase with lowest prices. In this case, security for online transactions has just become the first priority.

In the correlation between online shopping tendencies and perceived risks, the privacy risk has the most positive correlation; followed by performance risk, financial risk, and time-related risk is the most irrelevant risk.

This study only focuses on the survey of online shopping tendencies, due to a wide variety of potential applications in dimensions of the tendencies and perceived risks, it is not only applying to online shopping tendencies, any department who utilizes

e-Commerce may conduct the follow-up studies aiming at the same issues, or conducting empirical studies in different industries.

This study suggests that the electronics retailers must try their best to reduce the customers' perceived risks by enhancing the quality of products; the only way to earn a customer's trust is to effectively reduce the customer's uncertainty.

## References

1. Tan, S.J.: Strategies for reducing consumers' risk aversion in Internet shopping. *Journal of Consumer Marketing* 16, 163–180 (1999)
2. Miyazaki, F.: Consumer perception of privacy and security risks for online shopping. *Journal of Consumer Affairs* 35, 27–44 (2001)
3. Cox, D.F., Rich, S.J.: Perceived Risk and Consumer Decision Making. *Journal of Market Research* 1, 9–32 (1964)
4. Dowling, G.R., Staelin, R.: A Model of Perceived Risk and Intended Risk-handling Activity. *Journal of Consumer Research* 21, 119–133 (1994)
5. Sweeney, J.C., Soutar, G.N., Johnson, L.W.: The Role of Perceived Risk in the Quality-Value Relationship: A Study in a Retail Environment. *Journal of Retailing* 75, 77–105 (1999)
6. Pavlou, P.A.: Consumer acceptance of electronic commerce integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce* 7, 101–134 (2003)
7. Weber, E.U., Hsee, C.: Cross-cultural differences in risk perception, but cross-cultural similarities in attitudes toward perceived risk. *Management Science* 44, 1205–1217 (1998)
8. Jacoby, J., Kaplan, L.B.: The Components of Perceived Risk. In: Venkatesan, M. (ed.) *Advance in Consumer Research*, pp. 383–393. Association for Consumer Research (1972)
9. Mitchell, V.-W.: Consumer Perceived Risk: Conceptualizations and Models *European Journal of Marketing* 33, 163–195 (1999)
10. Brian, J.C., Theerasak, T., Han, Y.: Trust and E-commerce: A Study of Consumer Perceptions. *Electronic Commerce Research and Applications* 2, 203–215 (2003)
11. Chen, Z., Dubinsky, A.J.: A conceptual model of perceived customer value in e-commerce: a preliminary investigation. *Psychology and Marketing* 20, 323–347 (2003)
12. Teas, R.K., Sanjeev, A.: The Effect of Extrinsic Product Cues on Consumers' Perceptions of Quality, Sacrifice, and Value. *Journal of the Academy of Marketing Science* 28, 278–290 (2000)
13. Wood, C.M., Scheer, L.K.: Incorporating Perceived Risk Into Models of Consumer Deal Assessment and Purchase Intent. *Advances in Consumer Research* 23, 399–404 (1996)
14. Broydrick, S.C.: Seven Laws of Customer Value. *Executive Excellence* 15, 15 (1998)
15. Nunnally, J.C., Bernstein, I.H.: *Psychometric theory*, 3rd edn. McGraw-Hill, New York (1994)

# Model Checking Probabilistic Timed Systems against Timed Automata Specification

Junhua Zhang

Ningbo City College of Vocational Technology, Ningbo 315100, China  
zhangjunhua\_cn@163.com

**Abstract.** We study a model checking problem on a complex probabilistic timed system. In such a system, components are interacting with each other under time constraints, and their behaviors exhibit some uncertainties. People always want the system to run along with some demands. We model the system as a network of probabilistic timed automata, and use a timed automaton to denote the demands about the interactions among the system's components. To solve the problem, we use the method of "observe" which let the requirement automaton to observe the running of the system, and also the probabilistic model checker PRISM is used to calculate the probability reflecting the satisfying level of the original system to the demands.

**Keywords:** Probabilistic Timed System, Network of Probabilistic Timed Automata, Timed Automata, Model Checking.

## 1 Introduction

Systems involving time constraints and reliability exist prevalently. In the sophisticated cases, such as an "Internet of things", there are several components, every component runs independently and interacts with each other if needed. Probabilistic timed automaton is a popular model to describe bodies involving time and probability [1]. We can use probabilistic timed automaton to model the above component in a complex probabilistic timed system, and further more, use a network of probabilistic timed automata to model the whole system. To denote the property people want the system to obey, there are some selections. Several kinds of logics, such as PCTL, PTCTL, PCTL\*, etc. are used popularly. Besides that, [2] using a Living Sequence Chart enriched with Time (TLSC) to describe the scenario the system running. [3, 4] using timed automata directly to denote the system property. In this paper, we select timed automata as the tool to represent the property a system should obey. Since timed logic or TLSC can be translated into timed automata [2], the study in this paper has a wide adaptability. Model checking is a strict mathematical method to prove the correctness and reliability of a system running along with some properties. In this paper, model checking is used to check if the system satisfies the demands of the property at a level (probability).



## 2 Definitions

Let  $x$  and  $y$  be variables of non-negative real number recording time elapsing.  $X$  is the set of clock. A zone  $Z$  is a set of about clock variables as follows:

$$Z ::= x \sim d \parallel x - y \sim d \parallel Z \wedge Z \parallel \text{true},$$

where  $x, y \in X$ ,  $\sim \in \{<, \leq, \geq, >\}$ , and  $d$  is non-negative integer. Semantically, zone  $Z$  is the set of all clock valuations satisfying  $Z$ , and can be shown as  $\llbracket Z \rrbracket$ .  $Zones(X)$  denotes the set of  $Z$ . A (discrete probability) distribution over a finite set  $Q$  is a function  $\mu: Q \rightarrow [0,1]$  such that  $\sum_{q \in Q} \mu(q) = 1$ . For any  $q \in Q$ , the point distribution  $\mu_q$  means the distribution which assigns probability 1 to  $q$ . We use  $Dist(Q)$  represent the set of probabilistic distributions over  $Q$ .

Timed automata can model the character of a real time system. A **timed automaton** is a tuple  $(L, l_0, X, Act, Inv, E)$ , where  $L$  is the set of locations,  $l_0$  is the initial location,  $X$  is the set of clocks,  $Inv$  denotes the clock constraints while times keep elapsing in a location. For location  $l$ ,  $Inv(l) \in Z$ .  $E \subseteq L \times \{Act \cup \tau\} \times Zones(X) \times 2^X \times L$ , denotes the transitions between locations, where  $Act$  label the synchronizing actions accompanying with the transitions. While one automaton sends a message, the other automata receive that message. These actions are denoted by  $Act$ .  $\tau$  label the non-synchronizing actions occurring only within one automaton.  $Zones(X)$  represent the guards in a transition. A transition  $l_1 \rightarrow l_2$  are permitted to occur only when  $inv(l_1)$  are satisfied with the guard in the transition. In a transition, some clocks can be reset to 0,  $r$  is the set of such clocks. UPPAAL [5] is a successful tool for model checking timed automata. UPPAAL extends above definition about timed automata in some notations, such as urgent location, committed location, urgent channel, committed channel, broadcast channel (vs. binary channel).

In a real time system, transitions may occur in uncertain way. Actions and delays happen in a probability distribution. To denote these cases, we can extend timed automata with probability, and use the model of probabilistic timed automata. A **probabilistic timed automaton** (PTA) is a tuple  $(L, l_0, X, Act, inv, pE)$ , where  $L, l_0, X, inv$  are the same as in the timed automata.  $pE \subseteq L \times Zones(X) \times \{Act \cup \tau\} \times Dist(2^X \times L)$  is a probabilistic edge relation, denotes that for every  $l'$  and  $r \subseteq X$ ,  $p(r, l') > 0$  and  $(l, g, a, p) \in pE$ ,  $a \in Act \cup \tau$ , there exists  $\llbracket g \rrbracket \subseteq \llbracket [r:=0] inv(l') \rrbracket$ .

Timed automata can be seen as a special kind of probabilistic timed automata, where the probabilistic distributions are only point distributions. For probabilistic timed automaton, we can also extend it as UPPAAL done for timed automaton.

### 3 A Model Checking Question

Consider a probabilistic timed system with hierarchical architecture. Every component is denoted by a probabilistic timed automaton. Several interacting probabilistic timed automata construct a whole system. We can use a network of PTAs defined as above to denote the system. To describe the correctness and reliability of above system, a timed automaton can be used as a model for the system’s property. In this situation, the timed automaton is paid more attention to the interactions under time constraints among the components made up of the whole system.

An interesting question is, does a probabilistic timed system run, obeying the demands expressing with a timed automaton, and keeping a satisfied reliability (probability)?

### 4 The Concrete Steps to Solve the Question

#### 4.1 The Rules to Modify Specification Automaton and System Automata

In the following, we present the rules to modify system automata  $S$  and specification automaton  $O$ , in order for the specification automaton to observe the system running, similar to [6]. In a PTA as a part of the system  $S$ , for each probabilistic branch  $pb = (l_1, ch!, g, r, prob, l_2)$  where  $prob = p(l_1, r, l_2)$  in the probabilistic edges, we add an intermediate committed location  $l'_1$  between  $l_1$  and  $l_2$ , and a probabilistic branch between  $l'_1$  and  $l_2$ . The probabilistic branch between  $l_1$  and  $l'_1$  is the same as  $pb$  except that the target location name is changed to  $l'_1$ . The probabilistic branch between  $l'_1$  and  $l_2$  is assigned as  $pb' = (l'_1, cho!, true, null, 1, l_2)$ .

Fig.1 is a probabilistic timed system composed of two interactive probabilistic timed automata A and B. PTA A models a message sending process. A message can be sent from location  $m_1$ . With probability 0.9, the message is sent successfully. But with probability 0.1, message sending fails. Accompanying with the success, there is a transition to  $m_2$  with action  $SendSucc!$  under a condition  $x \geq 1$ . Accompanying with the failure, there is a transition to  $m_3$  with action  $SendFail!$ . After the latter

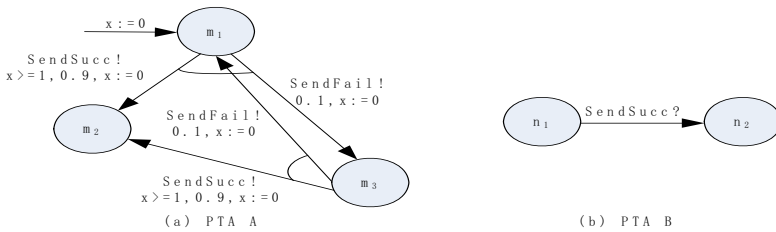


Fig. 1. A probabilistic timed system S (a network of PTAs)

transition, clock  $x$  is reset. In location  $m_3$ , the same message will be sent again as in location  $m_1$ . PTA B models a message receiving process. While a message is received, there is a transition with action  $SendSucc?$  from location  $n_1$  to  $n_2$  occurring.

Fig.2 is the modified version of fig. 1 according to the above adjusting rule for probabilistic timed automata. Here we don't do any modification for the action  $SendFail!$ . That is because it is an internal action of PTA A, no modification will not affect the final result of probability calculation in the following.

In  $O$  composed of a timed automaton, the name of the action in each edge is changed from  $ch?$  to  $cho?$ . Fig.3 is a timed automaton  $O$  that expresses a requirement on the probabilistic timed system in fig.1. Fig.4 is the modified result of fig.3 according to the above rules.

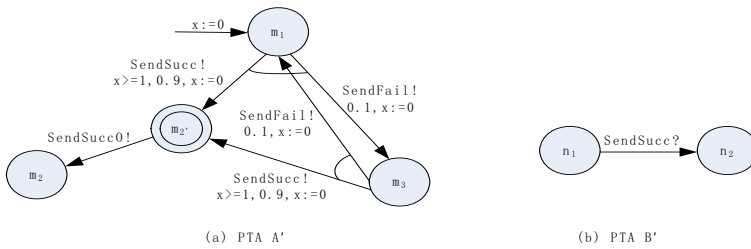


Fig. 2. The modified probabilistic timed system  $S'$  of Fig.1

Fig. 3. A timed automaton  $O$  that expresses a requirement on the probabilistic timed system in Fig. 1

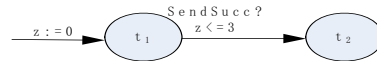
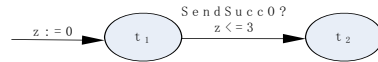


Fig. 4. A modified timed automaton  $O'$  of Fig. 3



## 4.2 The Rules to Compose Above Automata

### 1. Composition of a Network of Probabilistic Timed Automata

To model a complex probabilistic timed system (or to performance evaluate a complex timed system), composing of interactive components, we can use a network of probabilistic timed automata. In the description for PTA,  $Act$  and  $\tau$  has the same meanings as in the timed automaton.  $a \in Act$ . Furthermore, for the action  $a$  of sending a message, we can use  $a!$ ; for the corresponding action of receiving the message,  $a?$  can be used.  $\tau$  is a set of internal actions in a PTA, which don't affect other PTAs.

Let PTA  $W_i = (L_i, I_{0i}, X_i, Act_i, Inv_i, pE_i)$  ( $i \in \mathbb{N}_{z2}$ ) and Assume  $X_i \cap X_j = \emptyset$ . Refer to [7], we can give the definition of the network of probabilistic timed

automata. The parallel composition of two probabilistic timed automata  $W_1$  and  $W_2$  is the probabilistic timed automaton

$$W_1 \parallel W_2 = (L_1 \times L_2, (l_{01}, l_{02}), X_1 \cup X_2, Act_1 \cup Act_2, inv, pE)$$

such that

1)  $Act \in Act_1 \cup Act_2$  is declared as committed if and only if they are committed in at least one of PTAs;

2)  $inv(l, l') = inv_1(l) \wedge inv_2(l')$  for all  $(l, l') \in L_1 \times L_2$ ;

3)  $((l_1, l_2), g, a, p) \in pE$  if and only if one of the following conditions holds:

I.  $a \in \mathbb{R}_{\geq 0}$ , and there exists  $v_i + a \in Inv(l_i)$  such that  $g = g_i$ ,  $p = p_i \times \mu(\phi, l_j)$ ;

II.  $a \in Act$ , there exist  $(l_i, g_i, a!, p_i) \in pE_i$  and  $(l_j, g_j, a?, p_j) \in pE_j$  such that  $g = g_i \wedge g_j$ ,  $p = p_i \otimes p_j$ ;

III.  $a \in Act \cup \tau_1 \cup \tau_2$ , there exist  $(l_i, g_i, a, p_i) \in pE_i$ , such that  $g = g_i$  and  $p = p_i \times \mu(\phi, l_j)$ , but do not satisfy neither case I or case II.

In these cases,  $i, j = \{1, 2\}$ ,  $i \neq j$ , and for any  $l_1 \in L_1$ ,  $l_2 \in L_2$ ,  $\chi_1 \subseteq X_1$ ,  $\chi_2 \subseteq X_2$ :

$$p_1 \otimes p_2(\chi_1 \cup \chi_2, (l_1, l_2)) = p_1(\chi_1, l_1) \cdot p_2(\chi_2, l_2).$$

### 2. The Parallel Composition of a Network of PTAs with a Timed Automaton

From the definitions in section 2, we can see that a timed automaton can be considered as the reduction of a probabilistic timed automaton, where all of the probability distributions are point distributions. So we can obtain the parallel composition of a network of PTAs with a timed automaton just using the definitions from above. After the composition, we can obtain a standard PTA.

Fig. 5 is the composition result of  $A' \parallel B' \parallel O'$  in fig. 2 and fig. 4. In fig. 5, there are some transitions which can be realized only when synchronizing notifications are

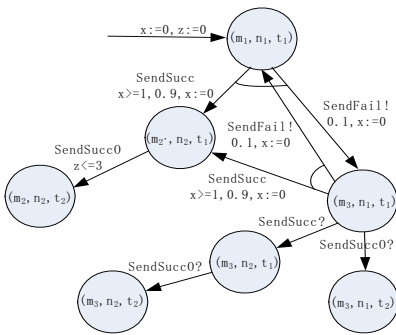


Fig. 5. The composition of  $A' \parallel B' \parallel O'$

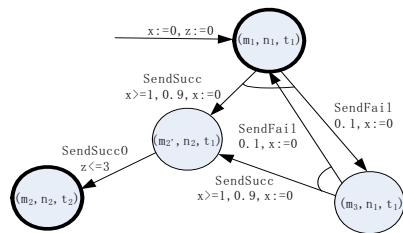


Fig. 6. The modified version of Fig. 5

received. Since this is the final composition result, these notifications will never be received, which means these transitions will not occur really. So we can cut these transitions. Fig. 6 is the result after these cuts. We can ignore the sending label in the final model, and just write *SendFail!* as *SendFail* since the composition has finished.

### 4.3 The Verification for the Composed Automaton

We first present a theorem:

$$\text{Prob}(S \triangleright O) \equiv \text{Prob}((S' \parallel O') \triangleright (l_{\min} \rightarrow l_{\max})),$$

where  $S$  and  $O$  express a network of probabilistic timed automata and a timed automaton denoting the property of the network.  $S'$  and  $O'$  are the modified results.  $l_{\min}$  and  $l_{\max}$  represent the initial location and the final location in  $O'$ . PRISM [8] is a notable probabilistic model checker developed by the Universities of Birmingham and Oxford. We can use PRISM to calculate the probability of  $\text{Prob}((S' \parallel O') \triangleright (l_{\min} \rightarrow l_{\max}))$ .

**Acknowledgement.** This work is supported by Natural Scientific Foundation of Ningbo, China (Grant No. 2012A610063).

## References

1. Kwiatkowska, M., et al.: Symbolic model checking for probabilistic timed automata. *Information and Computation* 205(7), 1027–1077 (2007)
2. Li, S., et al.: Scenario-based verification of real-time systems using UPPAAL. *Formal Methods in System Design* 37(2-3), 200–264 (2010)
3. Alur, R., et al.: Verifying automata specifications of probabilistic real-time systems. In: Huizing, C., de Bakker, J.W., Rozenberg, G., de Roever, W.-P. (eds.) REX 1991. LNCS, vol. 600, pp. 28–44. Springer, Heidelberg (1992)
4. Chen, T., et al.: Model checking of continuous-time markov chains against timed automata specifications. *Logical Methods in Computer Science* 7(1:12), 1–34 (2011)
5. Larsen, K.G., et al.: UPPAAL in a nutshell. *Int. Journal on Software Tools for Technology Transfer* 1(1-2), 134–152 (1997)
6. Firley, T., et al.: Timed sequence diagrams and tool-based analysis - a case study. In: *Proceeding of 2nd International Conference on the Unified Modeling Language*. Fort Collins, Colorado (1999)
7. Kwiatkowska, M., et al.: Performance Analysis of Probabilistic Timed Automata using Digital Clocks. *Formal Methods in System Design* 29, 33–78 (2006)
8. PRISM web site, <http://www.prismodelchecker.org/>

# Particle Filter Parallel of Improved Algorithm Based on OpenMp

Yang Zhang

Software College  
Jilin University  
Daqing, Heilongjiang  
dear\_zhang@163.com

**Abstract.** Particle filter was not limited by the system model and noise distribution, it was more consistent with the requirements of the actual filtering task, therefore, it received extensive attention in dynamic system problems of non-linear and non-Gaussian filtering. However, pointed at the complexity of the filter tracking and the increasing requirements for accuracy, the traditional non-linear filtering method was difficult to meet the needs of practical application. To solve this problem, this paper proposed a particle filter parallel of improved algorithm based on OpenMp, by parallelization improvement of traditional particle filter algorithm, mapping on each processors to run simultaneously of each stages in the model in the form of threads in parallelization, thus, made the video frame to be pipelined. The simulation showed that the proposed algorithm could effectively improve the advantages of performance in programs, made full use of computing resources, improved the filtering accuracy, and made the particle filter to be more widely used.

**Keywords:** OpenMp, particle filter, parallel technology, target tracking.

## 1 Introduction

Particle filter based on sequential importance sampling process, however, a serious drawback of SIS was the "weight degradation" problem, that was the the variance importance of the weight increased over time, that showed particle degradation was inevitable [1]. The particle degradation seriously affected the estimation performance of particle filter, after several iterations only a few particles had a particle concentration of great weight, but the rest were negligible, which made the particle set could not effectively express the density distribution of posterior probability of status, resulting in degradation of performance and even filter divergence [2]. In order to alleviate the degradation phenomenon, generally there were two ways: selected appropriate importance density function and introduce resampling steps.

Currently, the mainly algorithm for particle filter were: classical resampling algorithm based on random number optimization[3], particle resampling algorithm based on particle weight analysis[4], Markov Chain Monte Carlo resampling transfer optimization algorithm [5] and adaptive resampling algorithm [6]. The estimate

performance of four resampling algorithms was not different, wherein the difference of estimate result between residuals, systems and stratified were less, the latter two were slightly better, the polynomial resampling was poor.

To solve this problem, this paper proposed a particle filter parallel of improved algorithm based on OpenMp, by parallelization improvement of traditional particle filter algorithm, mapping on each processors to run simultaneously of each stages in the model in the form of threads in parallelization, thus, made the video frame to be pipelined. The simulation showed that the proposed algorithm could effectively improve the advantages of performance in programs, made full use of computing resources, improved the filtering accuracy, and made the particle filter to be more widely used.

## 2 Introduction of OpenMP

OpenMP was based on the fork-join parallel execution model, it divided program into parallel and serial area, different processors complete the data exchange through memory share[7]. As shown in Figure 1, when the program started executing, the main thread was the executing thread, as executing the first parallel region, there would derive multiple threads forming a team of thread, all codes that surrounded by parallel regions in program were executing in parallel, until all tasks that responsible to sub-thread of region were completed, the main thread continue executing parallel region in real program can be specified more than once, before encountering a parallel region, the program was still executing orderly, so the codes of serial region were executed by the main thread, while the parallel region codes were executed by the other derived threads[8].

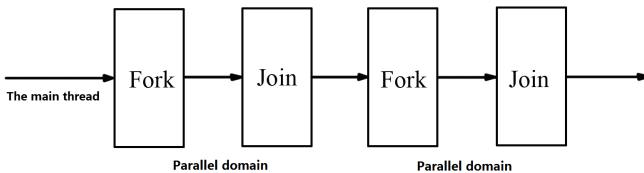


Fig. 1. OpenMP model of Fork-Join

## 3 The Theoretical Basis Of Particle Filter Algorithm

### 3.1 Bayesian Estimation

Set the dynamic system state space model as:

$$\begin{cases} x_k = f_k(x_{k-1}, v_k) \\ y_k = h_k(x_k, v_k) \end{cases} \quad (1)$$

$x_k \in R^n$  : system state,  $y_k \in R^n$  : Measured value,  $v_k \in R^n$ ,  $w_k \in R^n$  : Independent identically distributed system noise and measurement noise.

The theory of Bayesian filtering was to estimate the posteriori probability density of system state variables using all known information, which was, predicted prior probability density of state using system model, then using the latest measured value to correct, and finally got a posteriori probability density.

First step--- prediction:

Assumed that the system initial state probability density as:  $p(x_0|y_0) = p(x_0)$ , the probability density of  $k-1$  time was  $p(x_{k-1}|y_{k-1})$ , the first-order process of kov( $k-1$  time state was only related with  $k-2$  time state), we could know the prediction equation of  $k$  time by Chapman-Kolmogorov equation:

$$p(x_k | y_{1:k-1}) = \int p(x_k | y_{k-1})p(x_{k-1} | y_{1:k-1})dx_{k-1} \tag{2}$$

So we could obtain the prior probability of observed value of  $k$  time, and was calculated by  $p(x_k | x_{k-1})$ , the state transition probability of the system.

Second step---update:

This step was a correction process, using the latest observed value  $y_k$ , and the prior probability  $p(x_{k-1} | y_{1:k-1})$  of  $k$  time to achieve the derivation of  $p(x_k | y_{1:k})$ .

After obtaining the latest observed value, using Bayes equation:

$$p(a|b) = \frac{p(b|a)p(a)}{p(b)}$$

$$p(x_k | y_{1:k}) = \frac{p(y_{1:k} | x_k)p(x_k)}{p(y_{1:k})} \tag{3}$$

To divided  $y_k$ ,  $p(y_{1:k} | x_k)$  and  $p(y_{1:k})$  could expressed as:

$$p(y_{1:k} | x_k) = p(y_k, y_{1:k-1} | x_k) \tag{4}$$

$$p(y_{1:k}) = p(y_k, y_{1:k-1}) \tag{5}$$

Plugged(3)and(4)into(2):

$$p(x_k | y_{1:k}) = \frac{p(y_k, y_{1:k-1} | x_k)p(x_k)}{p(y_k, y_{1:k-1})} \tag{6}$$

The conditional probability density equation  $p(a|b) = p(b|a)p(a)$  :

$$p(y_k, y_{1:k-1}) = p(y_k | y_{1:k-1})p(y_{1:k-1}) \tag{7}$$

The Joint probability density formula  $p(a, b|c) = p(a|b, c)p(b|c)$  :

$$p(y_k, y_{1:k-1} | x_k) = p(y_k | y_{1:k-1}, x_k)p(y_{1:k-1} | x_k) \tag{8}$$



The Bayes formula:

$$p(y_{1:k-1} | x_k) = \frac{p(y_k | y_{1:k-1})p(y_{1:k-1})}{p(x_k)} \tag{9}$$

Plugged (6),(7),(8)into(5):

$$p(x_k | y_{1:k}) = \frac{p(y_k | y_{1:k-1})p(x_k | y_{1:k-1})p(y_{1:k-1})p(x_k)}{p(y_k | y_{1:k-1})p(y_{1:k-1} | x_k)} \tag{10}$$

Because each various observed values were independent, then we could get:

$$p(y_k | y_{1:k-1}, x_k) = p(y_k | x_k) \tag{11}$$

Plugged(10)into(9):

$$p(x_k | y_{1:k}) = \frac{p(y_k | x_k)p(x_k | y_{1:k-1})}{p(y_k | y_{1:k-1})} \tag{12}$$

We could get recursive process that the prior probability density to achieve a posteriori probability density by (2) to (12).

### 4 Parallelization of Particle Filter

The hot spot of application could be parallelized to speed up the running of the program in particle filter. In particle filter tracking, the more each target using one particle filter to track when there were multiple targets. The current number of simultaneously running threads should be equal to the processors of system, if the number of threads was less than processors, processor utilization had not been fully utilized, and if the number of threads was greater than the processor, processor would schedule threads frequently so that the processor utilization rate would decline. This also made it easy to achieve better load balance on each processing unit. In addition, this kind of coarse-grained level parallel was simpler than parallel particle filter algorithm, and the processor utilization could be improved.

SIS sequential importance sampling algorithm pseudo code could be described as:

$$(1) \quad \left[ \{x_k^i, w_k^i\}_{i=1}^{N_s} \right] = SIS \left[ \{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, Z_k \right]$$

$$(2) \quad i = 1, 2, \dots, N_s,$$

Calculated sample particle  $x_k^i \sim q(x_k | x_{k-1}^i, Z_k)$ ,

$$(3) \text{Set the normalized weights as } \sum_{i=1}^{N_s} w_k^i = 1$$

And the resampling algorithm for pseudo code descriptions was:

$$(1) \quad \left[ \{x_k^i, w_k^i, i^j\}_{j=1}^{N_s} \right] = \text{Re.sample} \left[ \{x_k^i, w_k^i\}_{i=1}^{N_s} \right]$$

(2)Initialize the CDF probability cumulative distribution function:  $C_1 = 0$

(3)  $i = 2 : N_s$ , calculate CDF:  $C_1 = C_{i=1} + w_k^i$

(4) Among CDF:  $i = 1$

(5) Take a starting point  $u_1 \sim U[0, N_s^{-1}]$  that obey uniform distribution

(6) For  $j = 1 : N_s$ , moving on CDF:  $u_j = u_1 + N_s^{-1}(j-1)$ , when  $u_j > c_i$ ,

\* $i = i + 1$ , gave a value to sample:  $x_k^{j^*} = x_k^i$ , gave a value to weight:  $w_k^j = N_s^{-1}$ , gave a value to index:  $i^j = i$ .

Finally, complete particle filter algorithm steps could be expressed as:

Initialization: when  $k = 0$ , the target sample N uniform distributed particle set  $S_{k-1}$ , established a target model:

$$\hat{q} = f \sum_{i=1}^I k \left( \left\| \frac{x_0 - x_i}{a} \right\|^2 \right) \delta[h(x_i) - u] \quad (14)$$

The tracing result  $S_{k-1}(E)$  of  $k-1$  time was known, particle set  $S_{k-1}(i)$ , weight of each particle  $w_{k-1}(i)$ , cumulative probability  $C_{k-1}(i)$ ,  $i = 1 \dots N_{k-1}$ , the number of particle was  $N_{k-1}$ , color template of  $k$  time was  $q_N^k$ .

Perform the following steps:

(1) Re-sampling  $S_{k-1}$  according to the weights  $w_{k-1}^{(n)}$  of each particles.

① Calculated normalized cumulative probability function  $C_{k-1}$ ;

② Uniformly distributed random Numbers  $r \in [0, 1]$ ;

③ searching the minimum  $j$  that meet  $C_{k-1}^{(j)} \geq r$ ;

④ then got  $S_{k-1}^{(n)} = S_{k-1}^{(j)}$ .

(2) Particle set  $S_{k-1}^{(n)}$  from  $k-1$  through dynamic equation, calculate particle set  $S_k^{(n)} = AS_{k-1}^{(n)} + v_{k-1}^{(n)}$  of  $k$  time, wherein:  $v_{k-1}^{(n)} \sim N(0, \sigma^2)$ .

(3) State estimation

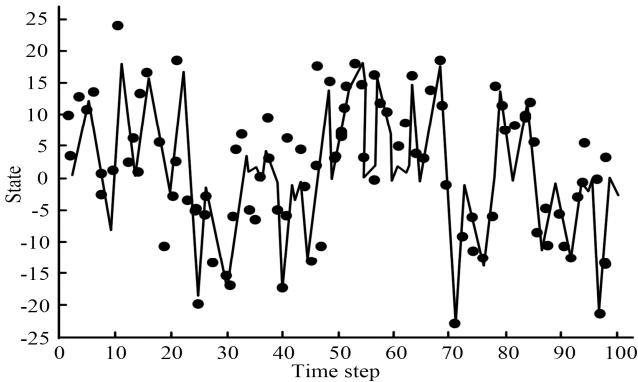
Calculated a weighted average of the state  $S_k(E) = \sum_{n=1}^N w_k^{(n)} S_k^{(n)}$ .

Now, all the steps of the parallelization for particle filter were completed.

## 5 Algorithm Emulation

To compare the different state estimation performances of the dissimilar resampling algorithm in nonlinear system, simulation experiments using non-static growth model were conducted, of which model is widely taken advantaged as standard authentication model with bimodal posterior distribution and strong-nonlinearity.

To explore the influences on the performance of particle filter algorithm after introducing the resampling step, the presented algorithm was adopted to the models mentioned above for state estimation in the first place. Set up the simulation time step  $T = 100$ , and the state estimation results were revealed in figure 2.



**Fig. 2.** State estimation results

On algorithm of particle filter, the paper puts forward is led to a faster convergence to the nonlinear systems, giving an effective system state estimation.

In view of the average elapsed time for four types of resampling algorithm in Table 1, which were calculated by Monte Carlo simulation for 100 times, respectively, show little differences, with slightly less time in computing and resampling, while in polynomial resampling with the longest execution time.

**Table 1.** The average running time of four algorithms (s)

Particle Number	100	200	500
PF-Mul	0.0576	0.1154	0.2548
PF-Str	0.0303	0.0621	0.1297
PF-Sys	0.0228	0.0468	0.0897
PF-Res	0.0433	0.0877	0.1724

## 6 Result

The simulations demonstrate that the proposed modified particle filter upon OpenMp parallel algorithm, which optimized the traditional particle filter algorithm parallelization, mapped parallel threads at each stage of the model on each processor, following with simultaneous execution so that the video frame to be processed in assembly line, can effectively enhance the advantages in terms of performance, and take full use of the computing resources, increasing the filtering accuracy.

## References

- [1] Wu, X.-T., Deng, J.-X., Ren, Y.-L., Yang, Y.: Multiple description of compressed sensing parallel processing algorithm based on OpenMP. *Application Research of Computers* 30(4), 1278–1280 (2013)

- [2] Zhu, X.-M., Pan, J.-S., Sun, Z.-Q., Gu, W.-D.: Parallel Implementation and Optimization of Two Basic Geo-Spatial-Analysis Algorithms Based on OpenMP. *Computer Science* 40(2), 8–11 (2013)
- [3] Fu, H.-Y., Ding, Y., Song, W., Yang, X.-J.: Fault Tolerance Scheme Using Parallel Recomputing for OpenMP Programs. *Journal of Software* 23(2), 411–427 (2012)
- [4] Pan, X.-M., Pi, W.-C., Sheng, X.-Q.: Efficient Parallelization of Multilevel Fast Multipole Algorithm Based on OpenMP. *Transactions of Beijing Institute of Technology* 32(2), 164–169 (2012)
- [5] Ren, X.-X., Tang, L., Li, R.-F., Ling, C.-Q.: Study and Implementation of OpenMP Multi-Thread Load Balance Scheduling Scheme. *Computer Science* 32(2), 164–169 (2012)
- [6] Zhang, J., Deng, J.: Target Tracking Algorithm based on Gray System Theory and Particle Filter. *Computer Application and Software* 30(4), 131–134 (2013)
- [7] Yi, L., Ya, E.: Study of Color Image Contour Extraction Algorithm based on Particle Filter. *Computer Simulation* 30(3), 384–388 (2013)
- [8] Chan, T.-Q., Li, Y., Liu, Z.-R., Dong, T.-Z.: An Improved Resampling Particle Filtering Algorithm. *Application Research of Computers* 30(3), 748–750 (2013)
- [9] Wan, Y., Wang, S.-Y.: Particle Filter method based on Recursive Bayes Model. *Journal of Signal Processing* (2), 152–158 (2013)
- [10] Zhang, D.-W., Sun, L., Liu, D.: Track-before-Detect Algorithm Based on Particle Filter. *Computer Simulation* 29(11), 264–267 (2012)

# An E-commerce-System-Based Research on Trust Risk Assessment Models

Mei Ting<sup>\*</sup>, Zhang Yi, and Dai Qun

School of Humanities and Information Management Chengdu Medical College  
Chengdu, China, 610500

School of Electronic Engineering and Optoelectronic Technology, Chengdu University  
of Information Technology Chengdu, China, 610225

{meiting,daiqun}@cuit.edu.cn, 657291678@qq.com

**Abstract.** Taking e-commerce in complex network as research background, this paper studies and proposes the risk assessment model based on trust. This model can effectively assess transaction risk in complex networks. Simulation results show that the risk assessment model researched in this paper can predict the risk of e-commerce transactions, and further also confirm the feasibility and rationality of the risk assessment model researched in this paper. Finally, we design and realize a risk assessment system, the system can be applied to existing e-commerce systems, and can carry out a risk assessment of e-commerce transactions, which will provide reference for buyers' trading, and also provide a new promising method for research of risk assessment in complex networks.

**Keywords:** Trust, Risk Assessment Model, E-commerce, Trust Evaluation.

## 1 Introduction

With the rapid development of network applications, e-commerce has also been fast in popularity. But there is virtual, non-face-to-face, online payment and other safety problems in the e-commerce transactions [1], these problems is the so-called the risk of e-commerce transactions [2]. The premise and foundation of e-commerce transactions is trust, therefore, in order to achieve the security of e-commerce, we are currently facing the main problem is trust, risk and other issues [3]. But the trust and risk assessment model is still at the exploratory stage, there are still many problems to be solved in the present [6]. For example, how to reasonably carry out formal description of trust relationship, how to develop trust evaluation mechanism, how to judge the credibility or not, how to design the trust evaluation system, how to use trust policy, how to determine the relationship between trust and risk, and how to achieve the trust integration of the next generation network and so on. At present, e-commerce does change the business model of many companies, enabling companies to save costs and create more profits [5], but having to face the new trust and risk. Therefore, e-commerce companies must fully understand the risks of e-commerce, to take

---

<sup>\*</sup> Corresponding author.

appropriate precautionary measures to minimize the harm caused by the e-commerce risk, to reduce unnecessary losses. Taking e-commerce in complex network as research background, drawing on existed research work, and the achievement of literature [6], and this paper proposes and researches a risk assessment model based on trust.

## 2 Research on Issues Related

### 2.1 Research and Analysis of Risk

Risk refers to the entity may be subjected to the expectations of loss in order to achieve the desired results. Risk has some unique characteristics: (1) Objective possibility (Risk is an objective reality, but risk can not be completely eliminated.); (2) Contingency (The loss resulted by risk may be having uncertainty, there is a great contingency, and is difficult to determine in advance.); (3) Testability (The occurrence of risk can be assessed by use of probability.). The risk can be expressed a function of the probability and consequences of adverse events occurred, namely:

$$R = (P_f, C_f) \quad (1)$$

Where,  $R$  represents risk,  $P_f$  is the probability of occurrence of adverse events, and  $C_f$  is the consequences resulted by the event of adverse events.

### 2.2 Analysis of Risk Factors

In the complex network, it is difficult to identify the authenticity of the interactive information. For example, according to iResearch survey, 48.4% of respondents believe that e-commerce credit is the primary concern of people, which is the so-called credit risk. In the online trading process, transaction information is difficult to grasp, the information insecurity prior to transaction, loss of value and moral hazard in the transaction process will bring out a lot of risk. By analyzing the trading process of C2C, we conclude that the risks of online transactions involve the following aspects.

- **Commodity prices**

Because the buyer doesn't know the actual commodity of online transactions, merchants may shoddy. The risk of high-value commodities trading is much higher than that of the low-price goods. Therefore, the average price of the same goods can be used as an important indicator of the risk assessment.

- **Sellers' credibility**

The grade of seller's credibility reflects the trust grade of the buyer to the seller. The seller's credibility includes the following factors: the grade of seller's credibility, the number of transactions, the number of failed transactions, the total historical transaction volume, and etc.

## 3 Risk Assessment Model Based on Trust

The risk assessment model based on trust studied in this section is based on "Trust Evaluation Approach Based on Cloud Model Theory" studied in the lecture [6],

combines the other factors that affect risk, and comprehensively assesses the final grade of risk, as shown in Figure 1.

### 3.1 Composition of Risk Assessment Model

As shown in Figure 1, risk assessment model based on trust is consisted of the following components.

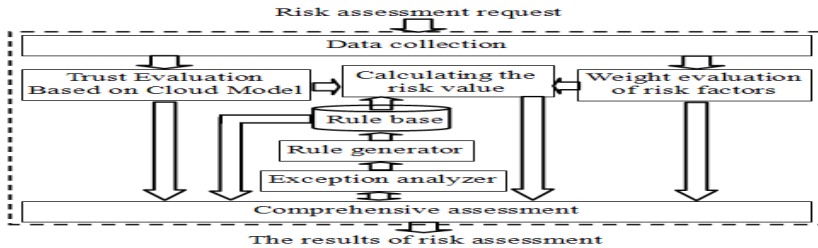


Fig. 1. The risk assessment model based on trust

Data collection modules: Its mission is responsible for collecting required assessing data, such as assessment data obtained by the entity, the trust value of evaluator, the interactive content between the entities and so on.

Trust evaluation module based on cloud model [6]: Using the attributes or characteristics of entities to evaluate the trust degree of entity, this objectively reflects the randomness and fuzziness of subjective trust.

Weight evaluation module of risk factors: The weight of each risk factor is assessed according to the rules in the rule base. Affecting factors of transaction risk includes the trust grade, transaction amount, the historical average amount per transaction, the transaction failure rate, the risk of online payment, the rate of operational errors and the probability of malicious attack.

### 3.2 Calculating Risk Value

According to the preceded discussing, risk assessment need to consider the various risk factors of the entity. According to the risk factors of practical application for a given weight coefficient, the calculation formula of risk value is as follows:

$$R = (P_f, C_f = \sum_{i=1}^n F_i \times \lambda_i) \tag{2}$$

Where,  $P_f$  is the occurring probability of risk event,  $C_f$  is the consequences which risk event occurs,  $F_i$  is the assessment value of the  $i$ th risk factor,  $\lambda_i$  is weight,

and  $\sum_{i=1} \lambda_i = 1$ .

### 3.3 Risk Assessment Process Based on Trust

As shown in Figure 1, the risk assessment processes are as follows:

Collecting data after receiving the risk assessment request, preprocessing collected data, which mainly format the assessment data of various attributes;

By means of trust evaluation module based on cloud model [6], the trust value is calculated;

Reading from the information of rules base, identifying the weight coefficient of risk factor, and calculating risk value;

Exception analysis: if the analysis result has exception, then executing step (5), otherwise executing step (6);

Rules generator generates new rules, and which write into rule base;

Comprehensive assessment of risk: according to the calculated risk value, trust value and the weight of risk factor, carrying out comprehensive assessment, and then determining the risk grade of current transaction by means of the rule base. Finally, the reference conclusion (such as trading or terminating trade etc.) is given, namely the comprehensive assessment results of risk.

### 3.4 Simulation Experiment and Analysis

The simulation experimental data comes from the transitive records of an electronic products business (2000 records) in Taobao website, the content of each record: transaction time, commodity name, commodity prices, and reviews of buyers to seller, the trust grade of sellers.

- Data preprocessing

**Table 1.** The score value of the attribute intervals

Assessment Attribute The score value	Very poor	Poor	General	Good	Very good
	[0,2]	[2,4]	[4,6]	[6,8]	[8,10]
The commodity's description match	1.9	3.6	5.6	6.6	8.2
The seller's service attitude	1.8	3.5	5.5	6.6	8.2
The seller's shipped speed	1.9	3.8	5.8	6.8	8.3

**Table 2.** The score value of intervals of commodity's price

Intervals of price	[10,150]	[150,350]	[350,800]	[800,1500]	[1500,2540]
	[4930,5070]	[4730,4930]	[4280,4730]	[3580,4280]	[2540,3580]
The grade intervals	[0,2]	[2,4]	[4,6]	[6,8]	[8,10]
The score value	1.9	3.6	5.6	6.4	8.1

- Defining trust intervals

Suppose the trust values is in the interval [0,10], and then this interval is divided into five small interval: [0, 1.5] (Very untrusted), [1.5,3.5] (untrusted), [3.5,6.5] (low trusted), [6.5,8.5] (trusted), [8.5,10] (very trusted).



- Calculating trust value

The historical trust values and the current trust value respectively accounts for the weight of 50%. E.g. the trust value of previous assessment is 6, so the final trust value is  $6 * 50\% + 6.2 * 50\% = 6.1$ .

- Calculating the risk value by use of the formula (2)

$$R = (1-t) * \lambda_1 + w * \lambda_2 + d * \lambda_3 + p * \lambda_4 + e * \lambda_5 + m * \lambda_6 \text{ Where, } \sum_{i=1}^6 \lambda_i = 1,$$

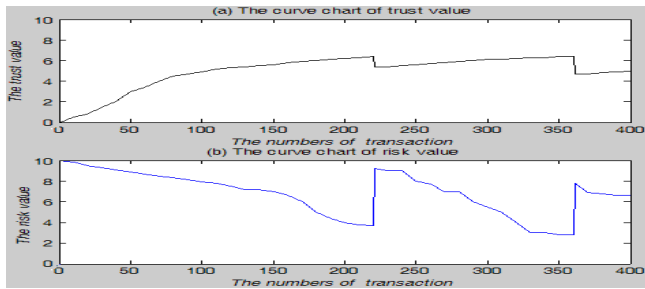


Fig. 2. The cure chart of risk and trust

Analysis of experimental results. Figure 2 shows the curve of trust and risk in the previous 400 transactions of a business, the prediction accuracy of risk is shown in Table 3. In Figure 2, the risk of a transaction suddenly increasing between 200 and 250 times, 350 and 400 times, because the business has abnormal behavior in these transactions, while the corresponding trust value is suddenly decreased after these transactions, which shows there is failure transactions, and the buyer is given a poor evaluation. The results of this experiment are consistent with the actual situation.

Table 3. The accuracy statistics of risk prediction

Risk prediction Item		Percentag
Accuracy		85.6%
Error rate	Risk prediction is too large	7.8%
	Risk prediction is too small	6.6%

In table 3, While the predict risk value is greater than 0.8 and the danger does not occur, which is called the risk prediction is too large. While the predict risk value is less than 0.4 and the danger occurs, which is called the risk prediction is too small.

## 4 Implement of Risk Assessment Model Based on Trust

### 4.1 Design of Risk Assessment Model Based on Trust

Frame structure of risk assessment system based on trust. The frame structure of risk assessment system based on trust is shown in Figure 3.

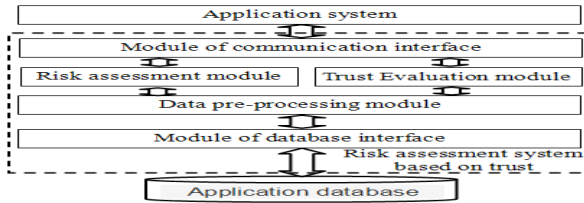


Fig. 3. The frame structure of risk assessment system

- The process of trust evaluation

Application system sends a trust evaluation request to the application system;  
 If the request received by the application system is trust evaluation request, this request message will be sent to the trust evaluation module;

The user's ID in the trust evaluation module is sent to the data pre-processing module;

According to the user's ID, data pre-processing module extracts relevant data from the database, then carries out pre-processing, and then send to the trust evaluation module;

The result of trust evaluation evaluated by trust evaluation module returns to the application interface module;

The application interface module sends the results to the application system.

- The process of risk assessment

Application system sends a risk assessment request to the application system;

If the request received by the application system is risk assessment request, this request message will be sent to the risk assessment module;

The user's ID and the order's ID in the risk assessment module are sent to the data pre-processing module;

According to the user's ID and the order's ID, data pre-processing module extracts relevant data from the database, then carries out pre-processing, and then send to the risk assessment module;

The result of risk assessment assessed by risk assessment module returns to the application interface module;

The application interface module sends the results to the application system.

- Module of communication interface

This module is responsible for communicating with application system; the message format of communicating is shown in Figure 4.

Message is divided into two parts of message header and message body, each message contains the same structure of message header, the length of message body is variable, as shown in Figure 4 (a).

Message header	Message body		
(a) The structure of message			
The serial number of message	Time	Message types	Message length
(b) The structure of message header			
The type of request	Requestor's ID	The person's ID being evaluated	The order's ID
(c) The message structure of request type			
Response time	Response type	The trust evaluation result	The risk assessment result
(d) The message structure of respond type			

Fig. 4. The message format

### 4.2 Implement of the Risk Assessment System Based on Trust

The main function of this system is processing the request message, carries out trust evaluation and risk assessment according to message type, and then returns the assessment result to the application system. The total flow chart of the assessment system is shown in Figure 5. The following is brief introduction to some key functions.

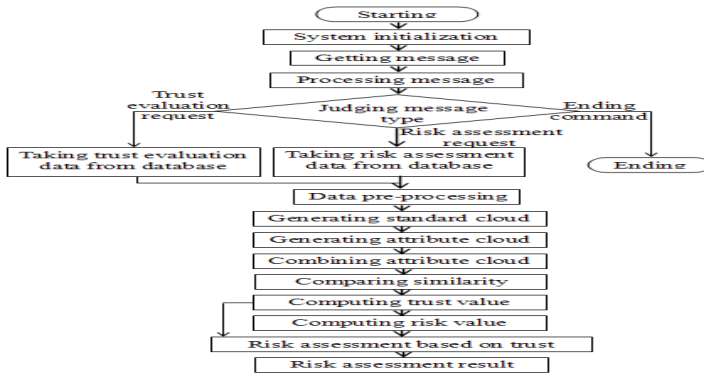


Fig. 5. The total flow chart of the assessment system

### 4.3 Simulation Results

The risk assessment system is developed by C language, and is respectively tested in the Windows XP SP3 and Suse Linux 10 environment. Testing data obtains from Taobao website. Figure 6 is a transaction when the risk is relatively low. It can be seen that the trust value of the business is 7.5, the risk value of transaction is 3.2, and is low-risk from Figure 6. Figure 7 is a transaction when the risk is relatively high. It can be seen that the risk value of transaction is 8.2, and is high-risk from Figure 7. The system gives a high-risk warning, reminding buyers cautious trading.


Transaction time: 2012-05-22 16:35			Order number: 172570263053943			
Businesses: Li Ning store			Trust value: 7.5			
Commodity	Price (¥)	Quantity	Total (¥)	Transaction status	Operating	Transaction risk
 Male sports shoes	199.00	1	199.00	Awaiting payment	<input type="button" value="Payment"/>	3.2 Low risk

Fig. 6. Low-risk transaction


Transaction time: 2012-05-22 18:23			Order number: 172570263053980			
Businesses: Creative Digital			Trust value: 5.5			
Commodity	Price (¥)	Quantity	Total (¥)	Transaction status	Operating	Transaction risk
 Panda cartoon audio	358.00	1	358.00	Awaiting payment	<input type="button" value="Payment"/> High risk Proposal to terminate transaction	8.2 High risk

Fig. 7. High-risk transaction

## 5 Summary

In complex network, there is some risk in the process of online transactions. Because online transaction is virtual, the role of trust in the online transactions is far more than that of the traditional interaction. Taking e-commerce in complex network as research background in this paper, the risk assessment model based on trust is studied and proposed. Simulation results show that the risk assessment model researched in this paper can predict the risk of e-commerce transactions. Finally, we design and realize a risk assessment system; the system can be applied to existing e-commerce systems, and can carry out a risk assessment of e-commerce transactions. At present, although we has conducted some research in the field of trusted access, and also achieved some stage results, but there are a lot of technology is still stuck in the lab, and many practical application problems need further to be studied.

**Acknowledgments.** This work is partially supported by Postdoctoral Fund of University of Electronic Science and Technology of China with Grant #73984 and the research fund of middle-aged academic leaders of CUIT with Grant #J201107; the authors sincerely express the gratitude to them.

## References

1. Xi, R.-R., Yun, X.-C., Jin, S.-Y., Zhang, Y.-Z.: Research survey of network security situation awareness. *Journal of Computer Applications* 32(1), 1–4 (2012)
2. Wang, S.-X., Zhang, L., Li, H.-S.: Evaluation Approach of Subjective Trust Based on Cloud Model. *Journal of Software* 21(6), 1341–1352 (2010)
3. Liu, Y., Du, R., Feng, J., Tian, J.: Trust model of software behaviors based on check point risk assessment. *Journal of Xidian University* 39(1), 179–185 (2012)
4. Yang, L., Lu, Y.-H.: An Evaluation Model For Network Risk Based on Cloud Theory. *Computer Simulation* 27(10), 95–98 (2010)
5. Gan, Z.-B., Zeng, C., Li, K., Han, J.-J.: Construction and Optimization of Trust Network in E-Commerce Environment. *Chinese Journal of Computers* 35(1), 27–37 (2012)
6. Zhang, S., Xu, C., Li, W.: Study on Trust Evaluation Approach Based on Cloud Model Theory. *Computer Science Applications and Education* 2(7), 14–22 (2012)

# An Analysis about the Security of the Operating System Trusted Platform

Zhang Yi

School of Humanities and Information Management Chengdu Medical College  
Xindu Avenue No. 783, Xindu District, Chengdu, China, 610500  
657291678@qq.com

**Abstract.** To enhance the security of the detection system of core files, it is suggested in this article a trusted computing platform to protect the confidentiality, integrity and availability of the resources of the system. This article first gives an analysis to the structure and organizational relationships of trusted computing module in Vista operating system. Then it suggests a general structure of typical trusted computation platform, whose major function module also receives a detailed analysis. This article attaches much importance to the analysis about TPM, and gives a detailed analysis to its internal structure and its support for system safety. Through the analysis of the security of the TPM's internal structure, it is proven that the security of TPM can be safeguarded, and a set of complete, reliable trusted transmission mechanism is necessary for this.

**Keywords:** trusted computing, Trusted Platform Module, chain of trust, safety.

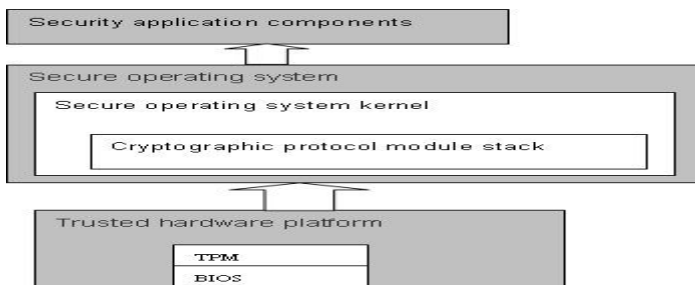
## 1 Introduction

In order to enhance the security of the terminal, "trusted computing" technology is being more and more attention. So far, about the "credibility" has not yet formed a unified definition. Currently, people trusted computing basic from a number of research institutions, organizations or products given definition. Overall, the current domestic and international recognized concept of trusted computing should at least meet the following conditions: trusted computing refers to a set of hardware and software environment, the user be able to use to improve safety, reliability, and can prove to be trusted operating environment; trusted computing hardware environment mainly refers to the basic calculations and operations platform, software environment is created by software on the platform to achieve a variety of features and requirements of the user environment, such as the operating system real users, user procedures, and so on; trusted computing environment to identification and authentication of the identity of the user to ensure that the use of computer users really expect; trusted computing environment should provide a set of managers to the user or environment safety management methods and available operating, management and configuration interface. Trusted computing environment with a complete set of the trust transfer mechanism, to ensure trusted hardware platform and credibility on the basis of

software, the establishment of a trusted user environment, not because the user program loaded and destroyed the original The trusted computing environment.

## 2 Analysis of Available Computing Environment Framework

Conditions of trusted computing environment based on the foregoing analysis, we can see that the first need to establish the related software and hardware platform to build a trusted computing environment. Although the facilities of its hardware implementation is still a conventional computer hardware entities, but, in the trusted computing environment organization and run function of these hardware entity will undergo a substantial change. Since the hardware infrastructure is the basis of any one of the computing platform, a trusted computing its trusted root often is based on the trusted hardware environments, based on this re-design and development of associated hardware and software interface program and system startup program, and then further on is the operating system and applications. Mainly due to a combination with hardware interface and startup programs hardware module key computational procedures and system configuration information and start the program BIOS. These two modules are often designed in achieving curing in the hardware environment, respectively the password calculation module TPM, and with a digital certificate BIOS program. Through these two modules to achieve soft, the combination between the hardware and call each other, which is the core component of Trusted Computing hardware platform. Security based on trusted computing TPM module as a credible roots, using step by step verification to verify the credibility of the system hardware and software, in order to achieve the file protection, Digital signature, reliability, and the like. A trusted computing environment overall framework is shown in Figure 1:



**Fig. 1.** Trusted platform architecture

Trusted computing environment, the TPM is the key modules of the system security, system security support module. The chip built into almost all of the commonly used cryptographic algorithms and parameters provided by the user interface, the flexibility to be calculated. The internal component of the TPM module structure is shown in Figure 2. Main component modules in its internal random number generator, the cryptographic computation processor dedicated module for the operation of the HMAC

in the nonvolatile memory in computing engines, SHA-1, the key generation module, the control module, the configuration register and so on. These modules cryptographic processor is a dedicated processor designed specifically for cryptographic operations, can greatly improve the efficiency and speed of the cryptographic operations. Meanwhile, in order to improve the speed of operation of the entire TPM is also equipped with two HMAC calculation engine and SHA-1 engine, further optimization and realization of such common computing functions, improve the efficiency of the TPM.

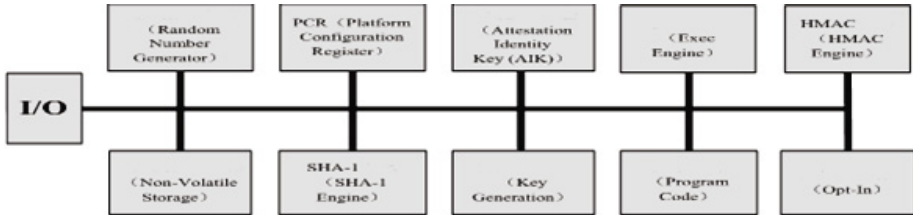


Fig. 2. Trusted Platform Module TPM structure

### 3 Trusted Computing Environment Internal Security Attribute Analysis

The TCG specifications defined in the trusted computing platform to achieve the minimum module composition and performance standards. According to the needs of the application, a reasonable trusted computing platform in addition to a full containing computer operating system should have the ability to protect, to prove the integrity measurement, storage and reporting features. Trusted computing environment capable of theoretical proof, reliable security trusted computing features, quantitative analysis and description of the trusted computing environment, degree of safety.

#### 3.1 Capable of Theoretical Proof of Trusted Computing Environment

- TPM credible proof

TPM proved to be a credibility to provide TPM data validation operation, which is done by using the digital signature of AIK (Attestation Identity Key) on a PCR value TPM internal. AIK is obtained through the only secret private key EK (Endorsement Key) can be uniquely identified.

- platform identity

The platform identification certificate by using the platform to provide evidence that the platform can be trusted to make integrity metrics report.

- between the state and the state migration prove

Between migration in the state and status of the trusted computing environment in which the proof essence to create a state transition diagram of the trusted computing environment to prove and analysis, in each state in the state transition diagram to give

each state expressly Meaning . On this basis, and then analyze the relationship between the state and status, migration constraint analysis, etc., get credible proof state transition. Identify the various states of the trusted computing environment, the TPM module register PCR preserve relevant information, and complete user and trusted computing platform parameter exchange function.

### **3.2 Has Reliable Security Trusted Computing Features**

A hardware platform in the trusted computing environment, the core of the stored information is the TPM chip TPM chip is a dedicated integrated chip, however, its design and implementation costs are high, therefore, can not be stored too much information. In order to improve the efficiency of information storage, trusted computing environment mainly in the TPM chip stores only EK, SRK and TPM Proof, and other relevant information. These three types of information are trusted computing environment running in relying important core information, its security is a decisive impact on the overall security of trusted computing environment, storage must be stored in the TPM chip interior. Some other non-core of the information is stored in the storage space outside the TPM chip. In fact, in order to facilitate management and guarantee system security, the information stored in the TPM outside is not messy and disorderly storage but managed by RTS (system root trusted storage mechanism) system. In this system, all the user keys and other critical information is organized into a tree structure to be saved. And information on all nodes in the tree are on a layer of its node encryption to encrypt all stored information that the child nodes of the parent node. If you need to obtain the child nodes must be decrypted by the parent node, the information can be obtained. All the information of the user in this way all the secret information of the user up the security association with a root node of confidence, thereby protecting the trusted computing environment can not malicious tampering. On this basis, the development of related applications naturally in a trusted computing environment, therefore, to ensure a safe and reliable computing environment to the user.

### **3.3 Quantitative Analysis and Description of Trusted Computing the Degree of Environmental Safety**

Quantitative analysis and description of the trusted computing environment level of security refers to these characteristic values of the platform impact the integrity of the hardware and software configuration and status information for the integrity metric obtained, and safe storage, and to generate the relevant recorded information, the subsequent analysis to the user the security of the system is to provide the basis.

Trusted Platform same three trusted root:

- metric trusted root (RTM) the basis for the trust chain transfer root;
- storage root of trust (RTS) is used to save the file integrity digest value and summary of the order;



- report credible root (RTR) is used to report the reliability of the RTS supports file integrity metrics summary.

In addition, the degree of security of trusted computing environment description is an important indicator of the trusted computing environment security analysis, only the quantitative description and analysis of the security of trusted computing environment, to be able to credibility compare calculated design and demonstration, select the highest degree of safety, the most suitable for the user's trusted computing environment. Trusted computing environment safety description, sampling of some of the characteristics of the trusted computing environment centrally stored in the TPM parameter register brings together the value as a description of a state of trusted computing environment identification information. If the user identification information of trusted computing environment contrast, found them a status equal to the identity value, to determine the degree of security of trusted computing environment, the safety of the state identification word. Storage achieved by this parameter register identifies the degree of safety of the trusted computing environment.

#### 4 Trusted Computing Platform in the Trust Chain Transfer Process

Trusted root above the system's BIOS program that is related to the security of the system startup. System during the boot process, none of the various security soft and checking mechanism established, therefore, the moment the system is most vulnerable moment. BIOS program highly execute permissions, can achieve the highest system execute permissions, the modified operation management, configuration, and so on all the hardware and software of the system. If the legitimacy of the BIOS program itself can not be guaranteed, then run the program on this basis is difficult to ensure security. Trusted computing environment in order to solve this problem, from the root of trust to the BIOS, the BIOS program introduces a digital signature mechanisms, signature verification information is stored with the TPM chip, only the signature verification was only loaded into the system BIOS run, thus avoiding some illegal BIOS program is loaded, and to achieve a trusted environment to the safety of the upper program directly from the trusted root, trusted loading process.

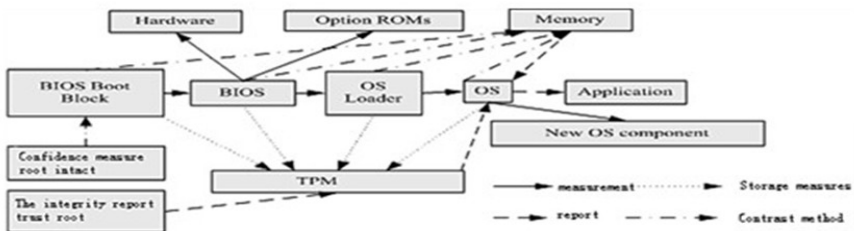


Fig. 3. Trusted computing environment complete chain of trust the process of establishing

Specifically, the first system initialization trusted environment in accordance with the trust of the need to verify that the file select a hash algorithm the hashed file loaded, which can be a fixed-length hash value sign the hash value, while the private key signature algorithm, and finally get the signature value and public key together with write verify file header. Private key signature algorithm used by TPM chip in the key tree management. The establishment of the signature header file is shown in Figure 4.

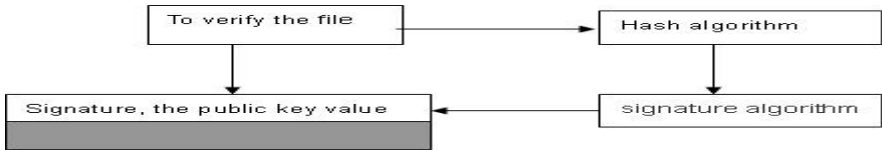


Fig. 4. Signature header file to establish

When after the first run to complete the creation of the signature head every time you boot the system while loading the file will be read in the file header signature value and the public key and decrypt the signature value with the public key signature algorithm to obtain the hash value of a fixed length. At the same time, the system will be the header of the file to extract hashed get a length of the public key to decrypt the value of the same hash value. Finally, the two hash values are compared, if they are the same, indicating that the file has not been tampered with or destroyed; different, indicating that the file has been tampered with or destroyed. The system in accordance with a result of comparison of the two hash values, the user is prompted to file is damaged, if it is damaged, will ask the user whether to continue to load the file. The file is loaded, the comparison of the hash value shown in Figure 5.

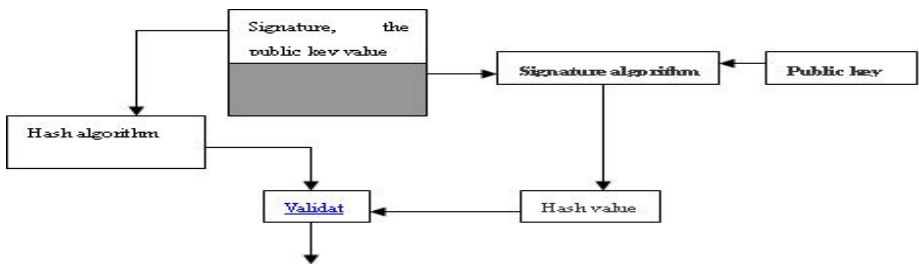


Fig. 5. Verify the hash value

## 5 Summary

The Design and Implementation of trusted computing environment is to improve computer security is an important way by series cryptography-based security theory and algorithms, design a theoretical proof security trusted computing environment can significantly reduce the current risks facing computer running tampering, forgery, deception.

According to the study,concluded as follows:

- Trusted computing module on trusted computing platform show that the system can start up quickly and return to normal after the damage;
- Effective organization of trusted computing module data to ensure the integrity and security of the operating system core files;
- Trusted hardware development is still not mature enough and more time-consuming operation of the TPM, single-core environment at the same time only one process can TPM access to multicore environment thread scheduling method can effectively improve system performance and thus reduce the time-consuming ;

## References

1. Hu, H., Zhang, M., Feng, D.: Metric based on the information flow trusted operating system architecture. *Journal of the Graduate School of the Chinese Academy of Sciences* 26, 522–529 (2009)
2. Qiu, G., Wang, Y., Zhou, L.: Based on Trusted Computing DRM interoperability. *The Computer Science* 36, 77–80 (2009)
3. Yin, Z., Wu, H.: Operating system security verification formal analysis framework. *The Computer Engineering the Scientific* 31, 24–26 (2009)
4. Li, C., Wang, H., Chen, J., Wide: Trusted computing-based computer security system design. *Computer Security* 1, 41–43 (2009)
5. Chen, W., Huang, W., Xie, C., et al.: Based virtualization platform trusted computing base. *Zhejiang University the Engineering Science* 43, 276–282 (2009)
6. Tang, W., Gao, L.: Pang double and credible based on the Universal Mobile Trusted Platform Module PC system enhanced. *Wuhan University: Science Edition* 55, 22–26 (2009)
7. Single, Z., Shi, W.: STBAC: a new operating system access control model. *To Computer Research and Development* 45, 758–764 (2008)

# Technology Venture Startup Invigoration Strategy for Building Infrastructures for the Business Startup Ecosystem

Hye-Sun Kim<sup>1</sup>, Yunho Lee<sup>2,\*</sup>, and Hyoung-Ro Kim<sup>1</sup>

<sup>1</sup>Dept. of Business Incubation Agency, Induk University, 12 Choansan-ro, Nowon-gu Seoul, 139-749, Korea

<sup>2</sup>Dept. of Social Welfare, Kyung Hee Cyber University, 1 Hoegi-Dong, dongdaemun-Gu, Seoul, 130-701, Korea  
daisyhsun@hanmail.net, anne6@khcu.ac.kr, hrkim@induk.ac.kr

**Abstract.** The people who wish to establish a business should have excellent relevant knowledge, technology, and entrepreneurship skills, and should be able to address the issue of job creation accompanied by employment. Startup companies accounted for about 6.6% of the country's gross domestic product (GDP) in 2009 and 9.8% of the total employment nationwide, according to Korea Institute of Startup & Entrepreneurship Development, and are highly likely to create job opportunities. Technology venture startups apply the research outcomes from enterprises, universities, research institutes, etc. to real businesses and can provide new products through technology innovations, can create new markets, and can vitalize regional economies. As such, it is necessary to conduct continuous studies, pursue policy development, and make proposals on the technology venture startup ecosystem.

**Keywords:** Startup, startup ecosystem, technology venture, startup environment, technology venture invigoration strategy.

## 1 Introduction

The Ministry of Science, ICT, & Future Planning, newly established by the South Korean government, announced that key to achieving a creative economy is business startups, and emphasized that the key subjects for realizing a creative economy are the construction of a business startup ecosystem where even students, housewives, etc. could establish a business. The ministry added that a virtuous circulation structure established by creating new jobs through business startups and rearing businesses through the enterprise welfare expansion policy will become a foothold for national economic growth. In this regard, the government is interested in Israel, an advanced country in terms of cultivating venture companies, and shows great concern about the policy of Technion Israel Institute of Technology, the cradle of venture companies

---

\* Corresponding author.

[1, 2]. Israel boasts the establishment of some 5,000 venture companies over the last 15 years, over 100 of which have been listed on the NASDAQ. Moreover, the USA has 30-fold more medicine universities as Israel does, but Israel has the biggest number of business startups in the bio healthcare sector (about 40%) in the whole world.

Business startups, therefore, are very important as they can serve as the foundation and driving force of national competitiveness for creating a new, 21<sup>st</sup>-century economy. In an environment where the industrial paradigms are rapidly changing, the utilization of external innovation through business ventures is considered integral and can contribute to addressing the unemployment problems of the highly educated as it can contribute to economic growth and high-quality job creation.

Therefore, in this paper, for continuous policy development and proposals on the technology venture startup ecosystem, the related policies are investigated in Chapter 2, theoretical considerations are studied in Chapter 3, and a conclusion is formulated in Chapter 4 through strategy proposals for utilizing technology venture startups.

## 2 Related Policies

The Ministry of Science, ICT, & Future Planning recently announced that universities can provide education for business startups; can expand the technology-holding companies that provide tailored support to business founders; can utilize business startups by collecting ideas, cultivating talented individuals, commercializing technologies presented by universities, linking the industry-academy-research sectors, creating new large-scale markets and demands, and reinforcing protection measures and financial support for intellectual properties; and can support the growth of venture and small- and medium-sized companies. In addition, the ministry plans to install an “infinite imagination room,” which collects various ideas from the citizens and links them with researches and business startups at two places in all cities, districts, and boroughs by the end of 2017.

The ministry maintains its thrust of reinforcing thorough education and system maintenance in the primary and middle schools for technology venture startups. In addition, the ministry has a plan to cultivate a high-quality workforce in strategic ICT sectors, such as software (SW) and big data, and to develop special workforces to meet the private demands by nominating/supporting 40 information communication academies led by the private sector by 2017.

Small & Medium Business Administration has declared that it intends to build a merger and acquisition (M&A) network that mediates merging and acquisitions among small- and medium-sized companies, and is promoting a plan to extend the benefits for small- and medium-sized companies from three to ten years. In addition, the administration is studying the introduction of a cloud funding system by revising the Support for Small and Medium Enterprise Establishment Act.

In particular, the culture and intellectual knowledge service industry sector expects utilization effects in funding and investments. Besides, through the introduction of Israel’s venture investment system described in the beginning of this paper, the

administration is making progress in establishing a system that will link the government with and will support matching-type research and development (R&D) funds for companies with venture capital investments as establishment and operation funds.

### 3 Theoretical Considerations

#### 3.1 Business Startup Ecosystem

**Table 1.** Ecological Stages of a Technology Venture Startup

Ecological Stage	Conditions	Induction of a Business Startup
Satisfaction Period	<ul style="list-style-type: none"> <li>✓ Accumulating a certain size of technology/manpower (necessary condition)</li> <li>✓ Maintaining a support system for a technology venture starter (sufficient condition)</li> </ul>	<ul style="list-style-type: none"> <li>✓ Helping an expected business starter with technologies at companies, universities, institutions, etc. to become a venture starter</li> </ul>
Formation Period	<ul style="list-style-type: none"> <li>✓ Converting to a large-scale technology venture startup and initiating growth through a combination of the necessary and sufficient conditions</li> </ul>	<ul style="list-style-type: none"> <li>✓ Increasing technology venture startups by supplying funds, manpower, etc.</li> <li>✓ Supporting information, facilities, etc. to nurture technology venture companies</li> <li>✓ Continuously inducing technology venture startups through public relations, etc.</li> </ul>
Establishment Period	<ul style="list-style-type: none"> <li>✓ Building a new high-tech industry through the accumulation of exceptional technology companies</li> </ul>	<ul style="list-style-type: none"> <li>✓ A large-scale technology venture startup is completed.</li> <li>✓ A connective cooperation system with successful companies is established, and a virtuous circulation ecosystem of technology-based venture startups is built.</li> </ul>

Source: Shin Chang-Ho and Kim Mook-Han (2012), *A Plan for Building a Virtuous Circulation Ecosystem of Technology Venture Startups in Seoul*.

The term *ecosystem* refers to a combination of organisms interacting with one another and its surrounding inorganic environments that affect such organisms. The business startup ecosystem in technology venture establishment is used with the concept of a virtuous circulation ecosystem, in which a virtuous cycle is created through associations with various organizations and institutions in the industrial, academic, and research fields that enable business startups, growth and accumulation, and interactions of knowledge, talented individuals, and funds flow that support such associations. The startup ecosystem of exceptional technology ventures can be largely divided into a period of satisfaction of the ecological conditions, an ecosystem formation period, and an ecosystem establishment period [2].

### 3.2 Technology Venture Startup

#### (1) Definition

The term *technology venture startup* refers to the establishment of a company that creates innovative technologies and knowledge, and is differently defined by institutions as the common standard for defining a technology venture group is practically ambiguous [1, 6]. Korea Technology Finance Corporation (KTFC) and Korea Institute of Startup & Entrepreneurship Development (KISED) have selected the cutting-edge high-technology business line as the key business line for technology venture startups, define the establishment of a business by the relevant industry as the real technology venture startup, and investigate the trends of technology venture business establishment using venture companies and other companies that have obtained technology innovation certification as an InnoBiz company as a surrogate variable of technology venture startups.

#### (2) Scope

The scope of the technology venture startup includes the establishment of a business involving not only cutting-edge, high-tech technologies but also knowledge as the motive power in a broad sense. KTFC cites the manufacturing, specialized service, and knowledge culture business types as the companies that conduct production and product/service sales activities with new technologies or ideas, and that can become middle-standing enterprises characterized by high-risk, high-revenue, and high-speed growth.

**Table 2.** Scope of the Technology Venture Startup

Classification	Business Line
Cutting-Edge Technology Business	Computer and office appliances, electronic parts and components, image/sound/communication equipment, medical/precision/optical appliances and clocks, aircraft and space shuttle parts, pharmaceutical products
High-Tech Business	Chemical compounds and products, other machineries and equipment, electricity conversion equipment, automobiles and trailers, other transportation equipment

Source: Korea Institute of Startup & Entrepreneurship Development (KISED).

Korea Research Institute for Vocational Education & Training (KRIVET, 2011) limits the knowledge technology venture startup to the contents and software of the IT fusion sector, manufacturing fusion sector, and knowledge-based service business in the middle classification of Korean Standard Industrial Classification Method, and excludes the livelihood-type business and the traditional manufacturing sector. KISED (2011) classifies five cutting-edge technology businesses and five high-tech businesses in accordance with the technological levels and regards them as the scope of the technology venture setup [1].

## 4 Conclusion: Proposed Invigoration Strategy for Technology Venture Startup

Recently, the government expressed concerns about the construction of a virtuous circulation structure in the venture business startup ecosystem, and announced that it would pour its efforts into improving the systems and building a startup ecosystem so that businesses can be started without government support in the long term.

In an effort to invigorate technology venture startups by building infrastructures for the business startup ecosystem investigated earlier, this paper proposes the following:

(1) that systematic technology startup education be offered for nurturing and developing the manpower needed to build the startup infrastructures

The designing of a business-startup-friendly education system will help improve the interested parties' understanding of business startups. Education on business startups can be offered in primary, middle, and high school, under the slogan "Business startup is a part of education."

There is a need to enable the recognition of business establishments and professions from a young age, to reinforce the education on entrepreneurship, and to consider a plan for R&BD participating companies to obligatorily complete the business startup education.

(2) that the operation of the government-driven business incubator be converted and extended to the private-centered operation system

Through consignments, the incubating system should provide a wide range of support for selecting business startup items, research and development targets, manufacturing trial products, commercializing, developing sales routes, management, managing human resources and others, building infrastructures, and inducing business startups specialized by region.

(3) that the importance of intellectual property rights be recognized, and that education be reinforced through the related business practitioners, including business founders, potential business founders, experts, universities, etc.

Lawsuits not only on leakages of cutting-edge technologies and infringements of others' rights through the globalized free trade agreements (FTAs), such as the recent legal battle on patents and designs between Apple and Samsung, but also on violations of the trademarks and designs of small and medium-sized companies, individuals, etc. are on the rise day by day. As more and more people are suffering in such cases, many of whom do not know their relevant rights, education on how to respond to such situations needs to be offered. As South Korea, however, restricts the individual possession of intellectual property rights to university professors, few research institutes are established. Through the Office of University-Industry Cooperation's revision of the provisions for the across-the-board possession of intellectual property rights, the system can be improved so that the researchers' intellectual property rights can be upheld.

To achieve such, there is a need to create a globalized competition landscape by maintaining the intellectual property rights system and establishing acquisition strategies for the intellectual property rights of the expected business founders; construct environments where people can receive compensation for their various



ideas, technologies, and designs; and build a sustainable business startup culture by fostering the trade and investment environments where ideas can be sold and bought.

(4) that the support funds be converted from loan-centered to investment-centered to systematically reduce the risk factors of credit rating deterioration if the business fails. In addition, the investment mentoring system should be introduced, and the mentors should be continuously educated and managed.

Further, the clouding funding scheme promoted by the government should be settled so that the investments can be vitalized by the private sectors by expanding the tax deduction benefits for the angel investments, or extending the period. If the investments were made by the mother companies or by large companies, the tax deductions should be increased to return the benefits to the investing companies. In this case, an investment mentoring system should be introduced to successfully lead business startups through systematic mentoring by a group of experts rather than having such startups solve all their problems by themselves, and the mentors should manage experts through DB management and continuous education. Innopoli in Finland introduced specialized mentoring management by experts, and as a result, 90% of its venture companies have survived.

(5) that M&As be vitalized through the acquisition of venture companies by large companies

The large ICT companies in the USA are particularly aggressive in the acquisition of venture companies, and a virtuous circulation system is naturally formed in the order of business startup-growth-sale-reestablishment. In South Korea, however, large companies are inactive in terms of venture company acquisition, and a social environment where the initial founders of venture companies are not well treated exists. Therefore, encouragement is necessary to establish the growth-sale-reestablishment system. A numerical investigation showed that Google has taken over 69 companies from 2006 to 2011, Microsoft 52 companies, and Samsung only 17 companies.

(6) that a cultural maintenance scheme or system be provided to turn failures into assets so as to exhibit creativity for the reinforcement of the venture ecosystem

The personal guaranty responsibility of the representative director should be relaxed for the exploration of another venture by turning a failure experience into a medicine rather than a poison. Further, through such system, the exit barriers should be eased so that such representative directors could retire from the business while minimizing the losses if failure is expected.

(7) that global competitiveness be reinforced as the domestic market is limited in size.

The growth of domestic companies is restricted due to the limitations of the domestic market. Due to their lack of knowledge about export/import, experience, etc., however, companies often fail without attempting to penetrate overseas markets. To overcome such problems, export/import experts should be trained to lend practical assistance to such companies through appropriate human resource expansion in international trade, the social standing of such experts should be upgraded, and the recognition of many talented individuals (i.e., first-class citizens) as global trading experts should be expanded. Of course, as the cases where companies return to the domestic market after achieving success in the USA or European markets are currently increasing, such is considered definitely possible if the language barrier will be removed.

(8) that a business startup space hub be established for use as a business startup space, where the expected business starter simply brings a notebook for convenience, by utilizing spaces such as a library of a local autonomous entity. The space should be arranged to sufficiently play the role of a preparation space before participating in the incubating system, and a sustainable support system should be introduced to induce a virtuous circulation of business startups by providing a cooperative space, networking opportunities, etc. to people who subscribed at the site or online homepage. The space should include a concept of small investment by the government and/or the private sectors, and should be used as a cooperative, incubating, and multi-purpose space. It will be better, for instance, if additional points can be given to users who borrow business-startup-related books at this place.

## References

1. Shin, C.-H., Kim, M.-H.: A Plan for Building a Virtuous Circulation Ecosystem of Technology Venture Startup in Seoul. The Seoul Institute (2013)
2. Yang, H.-B., Park, J.-B.: Construction of Business Startup Ecosystem for Youth. Korea Institute of Engineering and Technology (2011)
3. Kang, H.-H.: To Build 'Busan Good for Business' by Upgrading Business Startup Ecosystem. Busan Development Institute (2011)
4. Lee, Y.-J.: An Approach for Building Infrastructures for Knowledge Management. Graduate School of Information & Science, Sungsil University (Master) 2 (2006)
5. Lee, S.-G., Kim, Y.-S.: Utilization of Technology Venture Startup through Industry-Academia Cooperation: Around University Technology Business. Industry-Academia Support Center Affiliated with Seoul Development Institute (2010)
6. The Small and Medium Business Administration, Supplement Projects for Utilization of Technology Venture Startup. The Small and Medium Business Administration (Outside Edition) (2008)
7. Daeduk Research & Development Area. Step Up for Utilization of Technology Venture Startup. The Ministry of Science & Technology (2008)
8. Kim, W.-M., Lee, J.-H.: The Role of Universities for Utilization of Technology Concentrated Business Startup. Journal of the Korean Management Association 18(5), Serial 52, 2079–2106 (2005)
9. The Ministry of Science & Technology. Supports for Utilization of Business Startup Incubation of Gwangju Science & Technology Institute (2001)
10. Park, J.-H., et al.: Status of Business Startup Policy and Utilization Plan for Business Startup for Youth at Business Startup Ecosystem View. Korea Association of Business Education 10 (2012)
11. Kim, G.-S.: Implications from Policy Discussion Data: Construction Plan for Venture Ecosystem for Rearing Up the Small and Medium Venture Companies. Office of a Member of Parliament Kim Gi-Sik (2012)
12. Park, Y.-R., Jeon, S.-J., Joo, J.-W.: Structural Problems and Improvement Plan of ICT Venture Ecosystem. Gwacheon Information Society Development Institute (2012)
13. Han, J.-H., Shin, J.-G.: Development Process and Implications of Venture Ecosystem in South Korea. Korea Industrial Technology Foundation
14. Rae, D.: Mid-Career Entrepreneurial Learning, The Derbyshire Business School, pp. 562–574. University of Derby, Derby (2005)
15. GERA, The Global Entrepreneurship Monitor, Global Report (2011)

# Smart-Contents Visualization of Vehicle Big Data Using Vehicle Navigation Status Information

Hae-Jong Joo<sup>1</sup>, Suck-Joo Hong<sup>2,\*</sup>, and Dong-Su Park<sup>3</sup>

<sup>1</sup> Dept. of LINC, Dongguk University, 82-1 Pil-dong 2-ga, Jung-gu, Seoul, 100-272, Korea

<sup>2</sup> Dept. of CIEP, Mokwon University, Doanbookro 88, Seo-ki, Daejeon, 302-729, Korea

<sup>3</sup> NOVOSYS, 5 Fl., LIFE Tool Center, 614-33 Guro 2-dong, Guro-gu, Seoul 152-865, Korea  
hjjo@dongguk.edu, sjhong@mokwonu.ac.kr, toward21c@empas.com

**Abstract.** The recent technological advances have engrafted information technology (IT) on vehicles, making such vehicles more convenient and safer, and enabling them to meet the consumers' needs. As traffic congestion due to the increased number of vehicles has become a social issue, however, various other connected vehicle technologies for traffic and safety services using the data from the surrounding vehicles are being developed, standard industrial technologies connecting vehicles' infotainment system with mobile devices are under way, and many automobile and IT companies are participating in the promotion of their standardization. This paper discusses an effective method of managing and applying vehicle navigation information by deducing a method of collecting vehicle navigation information and the analysis factors of such information. If such analysis factors and collection technologies of vehicle navigation information manage to accommodate and rapidly apply the customers' demands using precise, highly reliable, and effective visualization methods, they can serve as important factors in creating company profits and in dominating the market in advance. In addition, the transmission of vehicle navigation information in real time is likely to provide an effective way of managing the problem occurrence factors and of instilling effective driving habits, etc. for safe driving.

**Keywords:** Vehicle Navigation Status Information, Vehicle Big Data, Smart-Contents Visualization.

## 1 Introduction

The technological advances and standardization in the automobile sector that started in earnest in 2004 have gone far, aiming to provide external information and multimedia environments to the vehicle space, and their actual services have been focused on the provision of traffic information, road guides, and various life convenience information.

The recent technological advances have engrafted various cutting-edge equipment and information technology (IT) on vehicles, and various sensors and state-of-the-art

---

\* Corresponding author.

IT parts and components are largely being attached to improve the convenience and safety of vehicles. Accordingly, the percentage of IT parts and components in vehicle is likely to climb from 25% in 2010 to 40% in 2015 [1].

The trends of the telematics technological advances and services not only in South Korea but in the whole world are heading towards safety and driving convenience rather than infotainment. As the concerns about the telematics services for vehicle diagnosis and management, vehicle convenience equipment control, etc. are rising, studies on and the commercial-service development of a terminal service platform are under way.

Many services that display basic vehicle data for the driver after collecting the internal vehicle data from an external equipment (smartphone, pad, telematics terminal, etc.) via Bluetooth have been introduced, using the Onboard Diagnostic System II (OBD-II) ports originally developed for vehicle emission control. In this study, the OBD-II ports were used to collect vehicle information and other data, and the data were transmitted to the information center using wireless communication, depending on the data type. Service directions on fuel saving and vehicle failure prediction are then suggested in this paper based on the results of the analysis of the data regarding each driver's driving habits and the travel distances, traffic information, etc.

## **2 Vehicle Navigation Status Information Factors**

In this study, the collection of vehicle navigation information using the OBD-II standard was considered to collect and save vehicle navigation information, and vehicles supported by a standard other than OBD-II were excluded.

### **2.1 OBD-II Standard**

The Society of Automotive Engineers (SAE) in the USA established a standard for the plug connectors that process diagnostic test signals and the onboard diagnosis program OBD in 1988. The OBD standard was thereafter developed under the names OBD-1.5 and OBD-II through supplements [2].

### **2.2 OBD-II Protocol**

All OBD-II-supported vehicles largely use three or five different detailed standard signal methods, such as VPW-PWM (SAE-J1850), CAN communication (ISO 15765, SAE-J2234), or the ISO methods (ISO 1941-2, ISO 14230-4) [3]. The signal methods differ by vehicle manufacturer, and different signal methods are used by vehicle model. As such, the OBD-II interface should be taken into account considering the three different methods cited above.

The vehicle navigation information can be monitored and saved through real-time communication with the electronic control unit (ECU) using the OBD-II standard protocol, and if vehicle failure occurs, the failure details are identified through the standardized five-digit failure diagnostic code.

### 1) OBD PIDs (Onboard Diagnostics Parameter IDs)

OBD PID is the code used for demanding information from the vehicle for the failure diagnosis. The latest OBD-II standard, SAE J1979, stipulates ten different modes, as follows [2]:

- 0x01. Show current data
- 0x02. Show freeze frame data
- 0x03. Show stored Diagnostic Trouble Codes|
- 0x04. Clear Diagnostic Trouble Codes and stored values
- 0x05. Test results, oxygen sensor monitoring (non CAN only)
- 0x06. Test results, other component/system monitoring (Test results, oxygen sensor monitoring for CAN only)
- 0x07. Show pending Diagnostic Trouble Codes (detected during current or last driving cycle)
- 0x08. Control operation of on-board component/system
- 0x09. Request vehicle information
- 0x0A. Permanent DTC's (Cleared DTC's)

In this paper, the mode 01 data are collected and analyzed for monitoring the vehicle driving status. The user data collected and used at mode 01 are provided in the following types:

- Vehicle status inspection
  - Analyzing the codes generated from the vehicle's ECU about problems with various parts and components of the vehicle
- Vehicle navigation record management and statistical information
  - Vehicle navigation starting/finishing time, navigation time, travel distance, average speed, highest speed, number of rapid accelerations, fuel injection time, etc.
  - Vehicle navigation records by weekday, week, month, and year
- Real-time vehicle information display
  - Providing real-time information on the vehicle speed, dashboard, coolant temperature, battery voltage, etc., as with the real vehicle dashboard screen
  - Providing various sensor data based on OBD
  - Recording and managing the real-time location information
- Information related to the driver's driving habits
  - Statistical data about the average mileage, numbers of sudden starting and braking, idle speed time, number of excessive RPM, congestion time, and excessive-speed time

## 3 Big-Data Analysis Model Using Vehicle Information

### 3.1 Construction of a Model for Vehicle-Big-Data Analysis

The construction of a model for analyzing vehicle big data can largely be divided into the construction of vehicle data collection environments and of a vehicle information center where all the data are to be collected and analyzed.

### 1) Vehicle Data Collection Environments

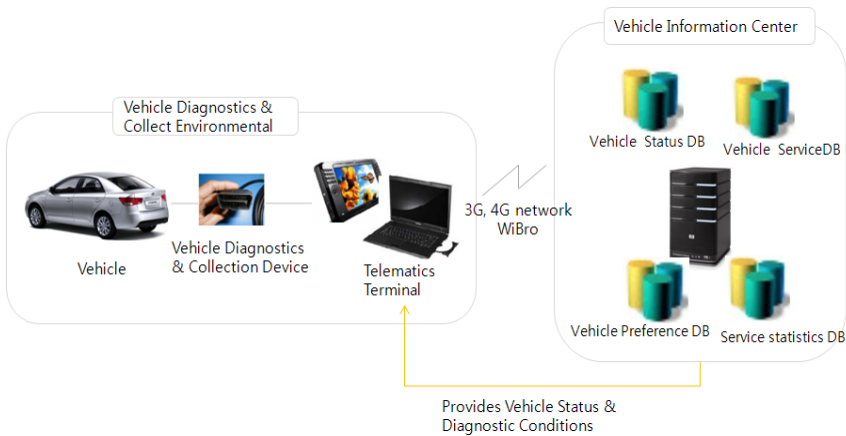
The environments consist of the OBD-II interface that collects data from the vehicle ECU, and of the smart appliances and smart appliance applications by which the user confirms the collected data and transmits them to the collection center.

### 2) Vehicle Information Center

The center carries out the functions of saving and management after analyzing the data transmitted from the vehicle diagnostic and collection equipment, for the user to receive his/her own information and to be provided with driving information and convenience service. The types of received data are divided into the 1<sup>st</sup> data (original data) and the 2<sup>nd</sup> data (original-data fabrications). The 2<sup>nd</sup> data are transmitted to the collection center after being converted into the vehicle information collection environments.

### 3) Schematic Diagram of the Big-Data Analysis System for Vehicle Use

The system to be constructed will collect the navigation information from a driver's vehicle, will conduct vehicle diagnosis, will provide information to the driver, and will transmit the collected data to the vehicle integrated information center for utilization.



**Fig. 1.** Schematic Diagram of the Big-Data Analysis System for Vehicle Use

### 4) Big-Data Analytical Factor Decision and Analysis

The vehicle navigation information analysis system selects the items to be managed regularly and to be provided to the driver for driving safety and convenience. In addition, the navigation status, driving tendency of each driver, etc. are analyzed and subdivided for use as basic data for supplying a suitable type of service to each driver.

In this paper, such data are largely divided into two kinds in the aspects of safety and vehicle management, as follows:

- Vehicle navigation and status management
  - Vehicle status: Selecting the items directly related to the navigation, like managing the vehicle error details, managing the real-time error occurrence and others, and checking the speed, RPM, disability code, and sensor status

- Vehicle navigation: Selecting the items necessary for the navigation, including the navigation time, average speed, travel distance, and movement path
- Driving tendency and statistical management
  - Driving tendency: Selecting the items related to the driver’s driving tendency, like the sudden acceleration, sudden braking, and number of navigations
  - Navigation statistics: Providing the driver with the details of the vehicle navigation by day, week, month, and year

## 4 Utilization of Big-Data Analysis Results Using Vehicle Information

If the vehicle navigation and status data is received, analyzed, and provided to the driver and/or to the users who need them, not only the driver’s own vehicle navigation management but also convenient service in association with various services will most likely be available. In addition, if the driver’s location and navigation speed data is provided, the service that will enable the other drivers who are travelling around such area to easily check the traffic situation will be available.

Fig. 4 shows the implicated correlation of various convenience service connections through the big-data analysis of the analysis items.

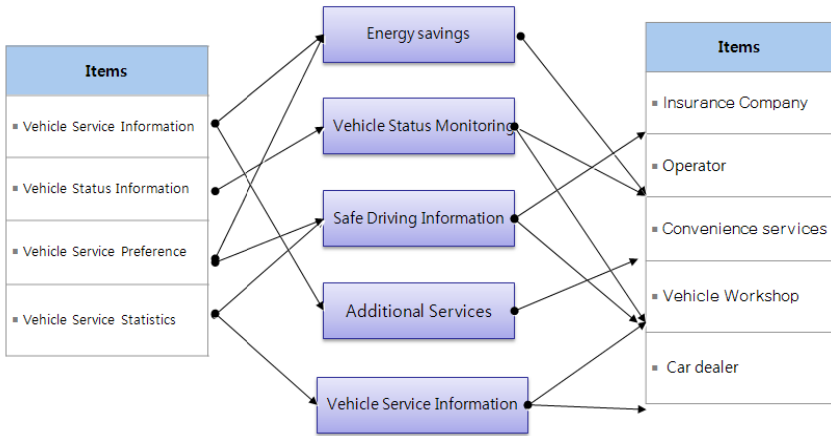


Fig. 2. Utilization of Big-Data Analysis Results Using Vehicle Information

## 5 Conclusion

The recent fusion of the automobile industry and information technology (IT) has introduced new technologies to satisfy the drivers’ demands related to being able to drive more conveniently and safely, and vehicles are being transformed into intelligent-type vehicles equipped with various sensors and network equipment. In addition, the environments are rapidly changing, allowing people to use their smartphones to check and

enjoy various types of services and convenience functions using vehicle navigation, status, and location information, among others.

In this paper, a model for providing and developing the services currently needed by drivers is proposed based on the monitoring and analysis of drivers' navigation details, driving habits, navigation status, etc. using the collected vehicle status and information. The results of this study can be applied to more effectively collect a large quantity of driver information so that drivers can save on costs, and can be considered a cornerstone for companies' establishment of a platform where they can develop optimal services and continuously manage their customers by investigating the driver's tendencies, etc.

## References

1. Jung, K.-H.: Smart Car and Future Social Changes. KT Economic Business Research Institute
2. E/E Diagnostic Test Modes, SAE standard J (1979)
3. Jang, S.-T.: A Development of Real-Time Monitoring System at Vehicle Based on OBD-II. Korea University of Technology and Education (August 2009)



# A Model for Analyzing the Effectiveness of Smart Mobile Communication Quality Measurement

Bong-Hwa Hong<sup>1</sup>, Hae-Jong Joo<sup>2,\*</sup>, and Sang-Soo Kim<sup>3</sup>

<sup>1</sup> Dept. of Information and Telecommunication, Kyung Hee Cyber University, 1 Hoegi-Dong, dongdaemun-Gu, Seoul

<sup>2</sup> Dept. of LINC, Dongguk University, #710, 82-1 Pil-dong 2-ga, Jung-gu, Seoul, 100-272, Korea

<sup>3</sup> Contents Vision Corp., #613, 82-1 Pil-dong 2-ga, Jung-gu, Seoul, 100-272, Korea  
bhhong@khcu.ac.kr, hjjoo@dongguk.edu, cqsky@paran.com

**Abstract.** The recent developments with regard to smart media require objective analyses of the effectiveness of the quality evaluations of the 4G and super-high-speed Internet; as such, it is necessary to establish a standardized model for analyzing the effectiveness of such quality evaluations. The studies related to the performance analysis of the information technology and system have focused on the identification of the correlation among various factors basically related to the performance of such technology and system. Such studies are ultimately relevant to the IS Success Model that can be used to successfully achieve the intended targets of the information system. This paper proposes an alternative list of performance indicators for the analysis of the effectiveness of the quality evaluation of the information technology and system, as well as a reference model for the procedure, definition, etc. for deducing the key performance indicators through the types and cases of common performance indicators applicable to the performance analysis from the past relevant studies. In addition, the effectiveness analysis model is reestablished and applied to systematically conduct effectiveness analysis of the quality evaluation business of smart media based on the literature and preceding case searches. The model of the evaluation system from the analysis perspective is as follows: plan (1<sup>st</sup>)→ do (progress: 2<sup>nd</sup>)→ see (result: 3<sup>rd</sup>).

**Keywords:** Quality Measurement Model, Effective Analysis Model, Smart Mobile Communication.

## 1 Introduction

This paper applies the types and cases of common performance indicators applicable to performance analysis from the past relevant studies to the procedure, definition,

---

\* Corresponding author.

etc. for deducing an alternative list of performance indicators for quality evaluation effectiveness analysis, and the key performance indicators. The studies related to the performance analysis of the information technology and system have focused on the identification of the basic correlations among various factors related to performance. Such studies are ultimately relevant to the IS Success Model, which can be used to successfully achieve the intended targets of the information system[1].

This paper aims to promote quality evaluation effectiveness analysis by quality evaluation business type (4G, super-high-speed Internet) by determining the reliability and validity of the deduced effects, by promoting smart mobile communication quality evaluation effectiveness analysis in a systematic and scientific way. Accordingly, this paper maintains the consistency and objectivity of the smart mobile communication business through the deduction of the common results, provides a foundation for establishing a standard model for the effectiveness measurement, and paves the way for the establishment of a generalized development stage system (degree of normality) through the effectiveness analysis of the quality evaluation business. To achieve this, this paper explains the related study “A Study on the Performance Analysis Model of the Information System” in chapter 2, proposes a definition and systematic diagram of the analysis of the effectiveness of the smart mobile communications quality evaluation, and presents the model indicator selections in chapter 3. Lastly, a conclusion is drawn.

## **2 Related Studies**

### **2.1 Initial Model of Delone and Mclean (1992), and Its Limits**

The factors affecting the success of the information system consist of the performance factors in six areas: system quality, information quality, system usage, user’s satisfaction level, individual performance, and organizational performance[1]. The performance measurement of the information system is applied using various performance measurement indicators with various properties, such as the properties of the system itself, the information property generated from the system, the users’ properties, and the organization’s properties.

The information system success factors consist of the relations affecting the stages quality--> use--> effects, as shown in <Table 1>; the factors affecting the quality of the information system consist of the system and information quality; the use of the system is subdivided into the information usage and the degree of satisfaction of the user; and the final effects are divided into the individual performance and the organizational performance affected by the individual performance.

As the initial model of Delone and Mclean (1992) presented in <Table 1> considers the influence factors of the information system itself, it exposes several limits. First, it omits the variable pertaining to the organizational situation where the information system is used; second, the “use” of the information system can be replaced with the “usefulness” of the system rather than simply its usability; third, despite the fact that the information system function itself contains the service factors that satisfy an organization’s IT demands, such factors are not considered in the

performance measurements of most models. Therefore, given that the role of the information department increases with time, the “service quality” area is appropriately added to the role of the information system department.

Delone and Mclean (2002) partially modified their models to accommodate the criticisms of other studies after 1992. “Service quality” was added to “system quality” and “information quality,” “use intention” was additionally introduced to the “system use” area, and the two were integrated into “pure effects” while considering that the benefits of separating the “individual influence” from the “organizational influence” were actually insignificant. The “pure effects” are characterized by the model that includes the “system use” and the reflux character re-affecting the “users’ satisfaction.”

**Table 1.** Six performance factors of Delone and Mclean

Stage	Quality ➡	Use ➡	Effects
<b>Area</b>	<b>System Quality</b>	<b>Information Usage</b>	<b>Individual Performance</b>
<b>Performance Factors</b>	<ul style="list-style-type: none"> <li>. Data precision</li> <li>. Data recency</li> <li>. Data contents</li> <li>. Use availability</li> <li>. Humane factors</li> <li>. Approach convenience</li> <li>. Reality of the requirements</li> <li>. Usefulness of the function</li> <li>. System precision</li> <li>. System flexibility</li> <li>. System accuracy</li> <li>. System integration</li> <li>. System efficiency</li> <li>. Resource availability</li> <li>. Response time</li> </ul>	<ul style="list-style-type: none"> <li>. Use amount/time</li> <li>. No. of questions</li> <li>. Connecting time, use functions</li> <li>. Used record volume</li> <li>. Use frequency</li> <li>. Volume of output report</li> <li>. Regular use volume</li> <li>. User type</li> <li>. Direct/indirect user</li> <li>. User/non-user</li> <li>. 1<sup>st</sup>/2<sup>nd</sup> user</li> <li>. Use characteristics</li> <li>. Use for intended purpose</li> <li>. Proper use</li> <li>. Used information types</li> <li>. User level</li> <li>. Repeated use, voluntary use</li> <li>. Motivation to use</li> </ul>	<ul style="list-style-type: none"> <li>. Interpreting the level of information</li> <li>. Learning level</li> <li>. Precise interpretation</li> <li>. Information recognition</li> <li>. Information recall</li> <li>. Confirming problems</li> <li>. Effectiveness of decision making</li> <li>. Quality of decision making</li> <li>. Improved decision making</li> <li>. Precision and time of decision making</li> <li>. Certainty and participation of decision making</li> <li>. Improvement of individual productivity</li> <li>. Change of decision making</li> <li>. Inducing management activity</li> <li>. Project outcome</li> <li>. Plan quality</li> <li>. Individual influence degree</li> <li>. Individual evaluation of the information system</li> </ul>
<b>Area</b>	<b>Information Quality</b>	<b>Users’ Satisfaction Level</b>	<b>Organizational Performance</b>
<b>Performance Factors</b>	<ul style="list-style-type: none"> <li>. Importance, suitability, and availability</li> <li>. Usability and interpretation possibility</li> <li>. Simplicity, form, and contents</li> <li>. Accuracy</li> <li>. Precision level</li> <li>. Sufficiency</li> <li>. Completeness</li> <li>. Reliability</li> <li>. Recency</li> <li>. Proper timing</li> <li>. Peculiarity</li> <li>. Comparability</li> <li>. Degree of quantification</li> <li>. Removal of distortion</li> </ul>	<ul style="list-style-type: none"> <li>. Satisfaction in a particular field</li> <li>. Overall satisfaction level</li> <li>. Single-item measurement</li> <li>. Multiple-item measurement</li> <li>. Information satisfaction level</li> <li>. Difference between the demanded and output information</li> <li>. Software satisfaction degree</li> <li>. Decision making satisfaction level</li> </ul>	<ul style="list-style-type: none"> <li>. Portfolio of application program</li> <li>. Scope of application program</li> <li>. No. of key application programs</li> <li>. Reduction of operational cost, decreasing no. of staffs</li> <li>. Increase of total productivity</li> <li>. Increase of revenue and sales</li> <li>. Increase of market share and profit</li> <li>. ROI and ROA</li> <li>. Net profit/costs</li> <li>. Cost/benefit ratio</li> <li>. Share price</li> <li>. Increase of work amount and product quality</li> <li>. Contribution level to the object achievement</li> <li>. Effectiveness of service</li> </ul>

**2.2 Model of Myers, Kappelman, and Prybutok (1997)**

Myers, Kappelman, and Prybutok (1997) reconstructed the evaluation model by adopting the contingency theoretical approach that included the service quality, added the work groups, and considered the organization and/or external environment at the same time by expanding the model of Delone and Mclean (1992) and the model of Pitt, Watson, and Kavan (1995).

They insisted that the “service quality” was considered in the information system proposed by them in that more IT demands should be satisfied, and they added the work group as an intermediate medium between the individuals and organization considering the organizational environment where the team role was emphasized[1].

### 2.3 Model of Parker and Benson(1998)

Parker and Benson (1998) presented a method that evaluated the economic influence on the companies affected by the information technology and system through the concept of 「information economics」 [1]. 「Information economics」 is a method that is based on the value chain of the company by expanding the profits to the concept of value, divides the value into the business area actually creating the values and the information technology area supporting such, and selects the information technology that is suitable for attaining the vision and implementing the strategy of the company through the connection and reconstruction of the costs and values created among such business and information technology areas. In addition, 「information economics」 is not limited to calculating the benefits in monetary terms, unlike the existing cost benefit analyses, but investigates the correlation in the economic aspect, considering the generated performance as a “value.”

## 3 A Model for Analyzing the Effectiveness of the Smart Mobile Communication Quality Evaluation

### 3.1 Conceptual Definition

**Table 2.** Model Concepts of Quality Evaluation Effectiveness Analysis

		Operant Definition
<pre> graph TD     Plan((Plan)) -- Efficiency --&gt; Do((Do (Progress)))     Do -- Practicability --&gt; Effectiveness((Effectiveness))     Effectiveness -- Effectiveness --&gt; Plan             </pre>	<b>Plan (1<sup>st</sup>)</b>	Formulating plans for funding, manpower, promotion, etc. for the establishment of a quality evaluation business, deducing the problems to be addressed and validating the reliabilities by evaluating their suitability to the overall business progress ( <b>Effectiveness Measurement</b> )
	<b>Do (2<sup>nd</sup>)</b>	Evaluating the direct activities, including the responses of the businesspeople/users/ government, investment, etc., and their influences on the results of the quality evaluation business ( <b>Practicability Measurement</b> )
	<b>See (3<sup>rd</sup>)</b>	Evaluating the quality improvement, satisfaction level, cost reduction, etc. that occur as the results of the response level to and activities of the businesspeople /users/government in relation to the quality evaluation business ( <b>Effectiveness Measurement</b> )

The effectiveness analysis model is reestablished and applied based on the literature and preceding case searches to systematically analyze the effectiveness of the quality

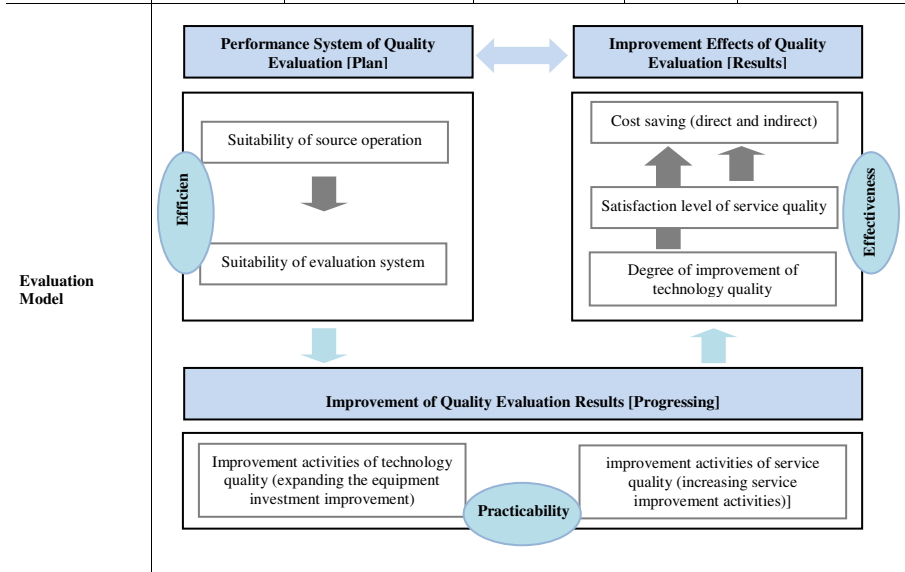
evaluation of the smart mobile communications (4G and LTE) businesses. The evaluation system is as follows (from the analysis perspective): plan (1<sup>st</sup>)→ do (progress: 2<sup>nd</sup>)→ see (result: 3<sup>rd</sup>)[2-5].

### 3.2 Systematic Diagram of the Models

The system was established to separately enable effectiveness measurement by deducing the evaluation areas, evaluation purposes, and subject institutions by business promotion stage; was designed to enable the effectiveness measurement of the overall process of the business promotion; and consists of a virtuous circulation structure that can be continuously improved by applying the results of each stage to the next business plan.

**Table 3.** Systematic Diagram of the Quality Evaluation Effectiveness Analysis Model

Business Stage	Plan (1 <sup>st</sup> )	Do (2 <sup>nd</sup> )	See (3 <sup>rd</sup> )		
<b>Evaluation Area</b>	Business operation	Improvement activities	Operational effects		
<b>Evaluation Purpose</b>	Efficiency	Practicability	Effectiveness		
<b>Creation Stage of Evaluation Effectiveness</b>	Input	Improvement (process)	Effects (outcome)		
			Quality	Satisfaction level	Cost saving
<b>Subject Institution</b>	Government (policy institution)	(Businesspeople, users, government)	Businesspeople	Users (clients)	Businesspeople, users, government



### 3.3 Selection of Quality Evaluation Effectiveness Analysis Indicators

To deduce the logical and reasonable indicators for effectiveness analysis, a literary search was conducted based on the preceding studies on the similar cases; the indicators were selected considering the evaluation areas and factors by business stage; and an indicator pool was completed based on such selection. After deciding the selection standard for five indicators using the SMART technique[6,7] from the indicator pool consisting of the preceding study cases, the initial draft for the indicators was prepared, and finally, the indicators were confirmed through the verification of the advisory committee consisting of experts in the relevant fields.

The selection of the optimized indicators cannot be achieved in a short time, but this is possible through continuous development. As such, it is necessary to have a procedure for filtering out the suitable indicators from the alternative indicator list composed by various studies, to measure the performance appropriate to the strategy by applying the changing strategy. A business like that involving quality evaluation considers the selection of indicators and the system among the indicators important because it has many ripple effects through the 2<sup>nd</sup> and 3<sup>rd</sup> stages because the effects occur directly. To satisfy such conditions, the standard for the indicator selection was used and applied with the SMART technique. The term “SMART” applies the specific level (Specific), measurability (Measurable), practicability (Action-oriented), relation (Relevant), and timeliness (proper Timing) as the indicator selection standards, and plays the role of showing which part is considered important in developing the indicators.

## 4 Conclusion

In this study, a common effectiveness measurement method was induced by quality evaluation business type (4G, super-high-speed Internet) by determining the reliability and validity of the deduced effects, for the promotion of smart mobile communication quality evaluation effectiveness analysis in a systematic and scientific way. Therefore, this paper maintains the consistency and objectivity of the smart mobile communication business through the deduction of the common results, provides a foundation for establishing a standard model for effectiveness measurement, and paves the way for the establishment of a generalized development stage system (degree of normality) through the effectiveness analysis of the quality evaluation business. In addition, in this study, models for evaluating all the effects on the overall quality evaluation business were developed and designed to provide a comprehensive view based on the procedures, so that the evaluations can include the effects occurring in the promotion of the quality evaluation business.

This paper defines the effects of quality evaluation and evaluates them by promotion stage, compares such effects by stage to deduce the excellent or inadequate stages, and proposes that the results thereof be applied by the government in its future business promotions.

## References

1. NIA V-RIR-09133, Analysis of effectiveness in 3G and high-speed internet quality test (December 2009)
2. Takahashi, A., Hands, D., Barriac, V.: Standardization Activities in the ITU for a QoE Assessment of IPTV. *IEEE Communications Magazine* 46(2), 78–84 (2008)
3. ITU-R Rec. BT.500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R (December 2002)
4. ITU-T Rec. P.910, Subjective video quality assessment methods for multimedia applications, ITU-T (September 1999)
5. Shumeli, R., Hadar, O., Huber, R., Maltz, M., Huber, M.: Effects of an Encoding Scheme on Perceived Video Quality Transmitted Over Lossy Internet Protocol Networks. *IEEE Tr. Broadcasting* 54(3), 628–640 (2008)
6. Kim, H., Choi, S.: A Study on a QoS/QoE Correlation Model for QoE Evaluation on IPTV Service. In: *IEEE ICACT 2010* (February 2010)
7. Dai, Q., Lehnert, R.: Impact of Packet Loss on the Perceived Video Quality. In: *IEEE INTERNET 2010* (September 2010)
8. ITU-T J.143 User Requirements for Objective Perceptual Video Quality Measurements in Digital Cable Television Series J: Transmission of Television, Sound Programme and Other Multimedia Signals Measurement of the Quality of Service (May 2000)
9. Yamagishi, K., Hayashi, T.: Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services. In: *IEEE ICC 2008* (2008)
10. Hurricane WAN Emulation & Network Simulation, PacketStorm Communications, Inc., <http://www.packetstorm.com/>

# Characteristics Analysis and Library Development for Common Lamps by Using PSPICE Modeling

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2,\*</sup>, Ho-Bin Song<sup>3</sup>, and Ok-Hwan Kim<sup>1</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Budaе-dong, Cheonan-si, Chungnam, 330-717, Korea  
{yckim59, owkim}@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College, 316, Daehak Road,  
Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

<sup>3</sup> Contents Vision CO. #613, 82-1, Dongguk University Chungmuro Media Center,  
Pildong 2-ga Jung-go, Seoul, 100-272, Korea  
songhobin@daum.net

**Abstract.** Recently, common lamps have been used in a variety of fields, including simple illumination science, and their domain of use is increasingly widening. In this study, we configure a library through modeling and verify its accuracy through simulations for widely used and representative lamps such as CCFL, fluorescent lamps, and HID lamps. On the basis of our experiments, we also perform a lamp simulation using PSPICE, which allows us to take advantage of the lamp library easily.

**Keywords:** CCFL, HID lamp, Fluorescent lamp, PSPICE.

## 1 Introduction

In the modern world, the common lamp is one of the most important and indispensable items and is used in various fields in addition to simple lighting.

However, the development and the application of various circuits using lamps require simulations that are currently performed with an alternative and simplified model, which is essentially a serial connection of loads such as resistors and inductors. Such an alternative simulation methodology can provide rough characteristics but has clear limitation in accurately verifying the functionality. If more detailed and precise lamp modeling is available, an accurate circuit analysis using lamps can be carried out accordingly.

To address such challenges, in this work, we have modeled widely used lamps and built a library. We have used PSPICE, a general-purpose and easy-to-use simulation tool. We have modeled a CCFL, a fluorescent lamp, and HID lamp, which have become popular recently and are now used widely, and applied them to our simulation setup, which has led to good functional characteristics as a result. With the output from this study, we expect that a user can obtain more accurate lamp simulation results with considerably less effort.[1]

---

\* Corresponding author.



## 2 Characteristics of Lamps and Modeling

### 2.1 CCFL Modeling

A cold cathode fluorescent lamp (CCFL) is suitable for miniaturization because of its simple structure. By changing the fluorescent material type and mixture ratio, we can obtain an arbitrary fluorescent color. Further, the CCFL has longer lifetime expectancy than a hot cathode fluorescent lamp and requires only a simple lighting circuit. Further, the CCFL shows a negative resistance characteristic, requires a ballaster to limit the negative current, and has resistor-like characteristics under a stable condition.[2]

Basically, the CCFL has four important parameters: firing potential for lighting, sustained voltage, frequency, and pipe current. The brightness of the CCFL depends not on the operating frequency but on the lamp current.

The effective voltage and current characteristics of the CCFL in this study have negative impedance at the sensitization level, and a lamp can represent two sensitization characteristics, as expressed in equation (1).

$$V_{\text{rms}} = 60.966 + 110.45 \cdot e^{-1.9404 \times I_{\text{rms}}} - 48.578 \times e^{-60.182 \times I_{\text{rms}}} \quad (1)$$

where the first term denotes the default value of the voltage-current curve, the second term shows the negative impedance nature, and the last term indicates the positive impedance nature.

The parameters specified in equation (1) can be derived by computing the least square root and the equivalent impedance to a resistor-like lamp at high frequency, as expressed in equation (2).

$$R_{\text{LAMP}} = \frac{V_{\text{rms}}}{I_{\text{rms}}} \quad (2)$$

From the above, we can compute the momentary voltage of a lamp  $V_t$  by using the momentary current of a lamp  $I_t$ , as shown in equation (3).

$$V_t = \frac{60.966 + 110.45 \cdot e^{-1.9404 \times I_{\text{rms}}} - 48.578 \times e^{-60.182 \times I_{\text{rms}}}}{I_{\text{rms}}} \times I_t \quad (3)$$

On the basis of equation (3), we modeled the CCFL. The momentary voltage of a lamp  $V_t$  can be described as a function of the momentary current  $I_t$  and the effective current  $I_{\text{rms}}$ . The effective value of the lamp current in a PSPICE model can be computed using an RC integration circuit, where the current source  $I_s$  is defined as the square of the momentary current of a lamp  $I_t$ , and the output voltage  $V_A$  can be obtained by integrating over the period  $T$ , as in equation (4).

$$V_A = \int_0^T I_t^2 \frac{dt}{T} = I_{\text{rms}}^2 \quad (4)$$

Therefore, the square root of  $V_A$  is the effective value of the current  $I_t$ . By using such an RC integration circuit to compute the effective current of a lamp, the lamp model equation in equation (3) can be transformed into the PSPICE model in Fig. 1.

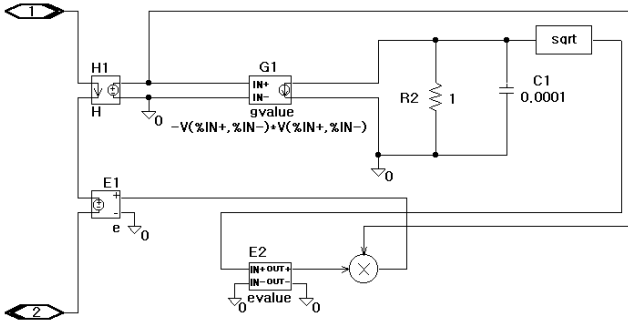


Fig. 1. PSPICE model for CCFL

## 2.2 Fluorescent Light Modeling

As a type of a low-pressure gas discharge lamp, fluorescent light can be best described by the flow of current through the carrier gas. Gas discharge or arc plasma occurs in a lamp tube. The discharge length consists of several regions between the positive column and the anode drop, which play an important role in the overall lamp behavior. Therefore, we model fluorescent light by focusing mainly on these two regions.

The Cassie equation where conductance  $G$  is a dependent variable is expressed in equation (5), which describes the arc behavior in the case of a high current.[3]

$$G = \frac{v \cdot i}{E_0^2} - \theta \frac{dG}{dt} \tag{5}$$

where  $E_0$  is the arc voltage,  $\theta$  is the arc time ratio,  $v$  is the lamp voltage, and  $i$  is the lamp current.

The Mayr equation is given as equation (6); it describes the arc behavior in the case of a low current.

$$G = \frac{i^2}{P_0} - \theta \frac{dG}{dt} \tag{6}$$

where  $P_0$  is the power loss.

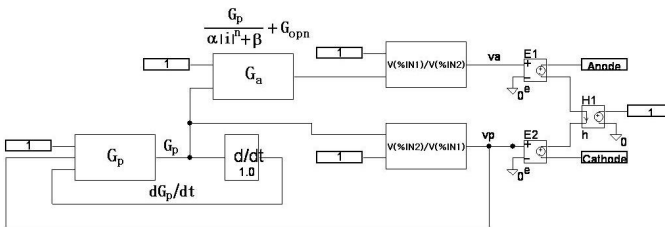


Fig. 2. PSPICE model for fluorescent light

The Mayr and Cassie equations are effective in the cases of 0 or low current and high voltage ranges, respectively. The positive column can be simulated using equation (5) and (6).

Fig. 3 shows the anode voltage and the lamp current of a fluorescent light.

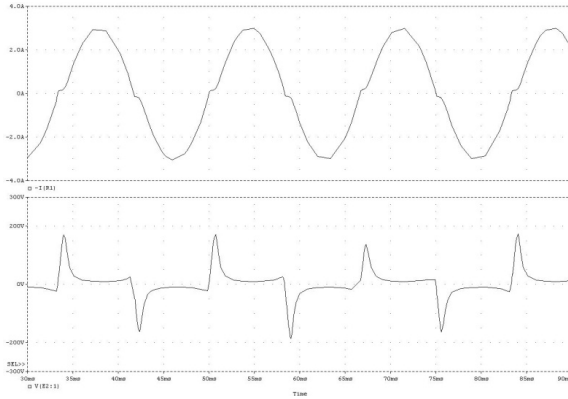


Fig. 3. PSPICE model for fluorescent light

### 2.3 HID Lamp Modeling

Our PSPICE model for a high-intensity discharge lamp (HID) can be built using physical properties, lamp parameters, and universal constants.[4]

The increase amount in arc heating  $dQ$  can be described as follows:

$$dQ = (P_{in} - P_{out})dt \tag{7}$$

where  $P_{in}$  is the input power to a lamp,  $P_{out}$  is the output power from the lamp, and  $dt$  is the time difference.

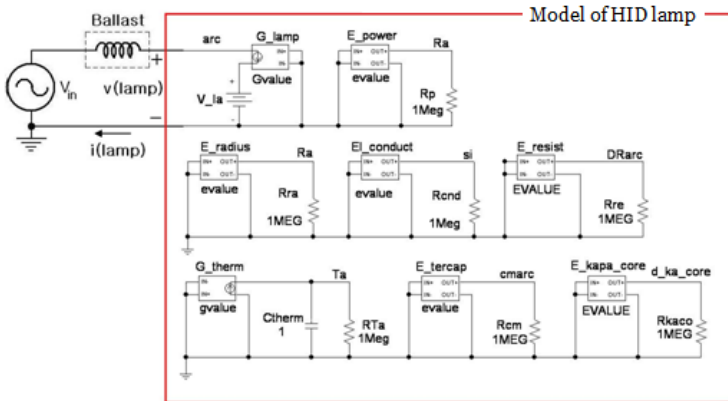
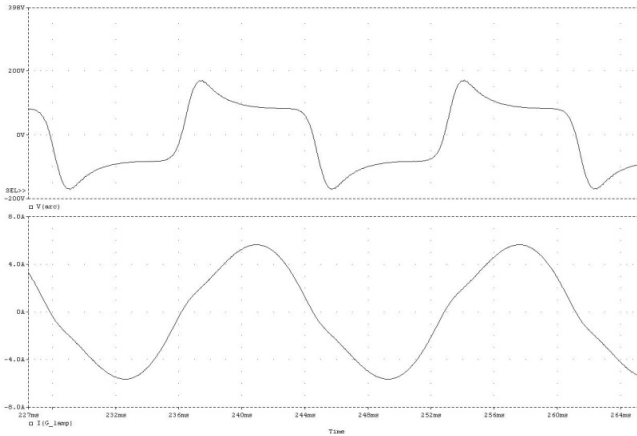


Fig. 4. PSPICE model for HID lamp

After applying interaction equations to equation (7), we can obtain the differential equation given in equation (8).

$$\frac{dT}{dt} = \frac{\frac{V_{lamp} i_{lamp}}{L_g} - 2\pi R_{el}[\varepsilon(T) + k(T - T_{amb})]}{\pi [c_{arc} \rho_{arc} R_{el}^2 + k_{Tred} c_{per} \rho_{per} (R_{tube}^2 - R_{el}^2)]} k\omega \quad (8)$$

Once we mathematically describe equation (8) in PSPICE, we can build the HID lamp model shown in Fig. 4. Fig. 5 respectively show the waveforms of the HID lamp obtained from our PSPICE simulation model at 60 Hz.



**Fig. 5.** Voltage and current curve for the HID lamp at 60[Hz]

### 3 Conclusion

In this study, we built an easy-to-use platform with PSPICE-based modeling and library for CCFLs, fluorescent lamps, and HID lamp. Further, we performed simulations to verify the correctness of our models and library. We hope that our library can allow more accurate simulations for lamps using only the basic parameters.

### References

1. Zaitso, T., Shigehisa, T., Shoyama, M., Ninomiya, T.: Piezoelectric Transformer Converter with PWM Control. In: Proc. IEEE APEC 1996, pp. 279–283 (March 1996)
2. Hwang, L.H., Yoo, J.H.: A study on the modeling of piezoelectric transformer for CCFL using a PSPICE. In: 7th International Conference on Power Electronics, ICPE 2007, pp. 122–126 (October 2007)
3. Perdigao, M., Saraiva, E.S.: MATLAB-SIMULINK fluorescent-lamp models. In: UPEC 2004. 39th International, vol. 2, pp. 855–859 (September 2004)
4. Paul, K.C., Takemura, T., Hiramoto, T., Yoshioka, M., Igarashi, T.: Three-Dimensional Modeling of an HID Lamp Operated by a Direct Current. In: ICOPS 2005, p. 191 (June 2005)

# A Study on Brake Torque for Traction Motors by Using the Electric Brake

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2\*</sup>, and Ho-Bin Song<sup>3</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Budaе-dong, Cheonan-si, Chungnam, 330-717, Korea  
yckim59@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College, 316, Daehak Road,  
Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

<sup>3</sup> Contents Vision CO. #613, 82-1, Dongguk University Chungmuro Media Center,  
Pildong 2-ga Jung-go, Seoul, 100-272, Korea  
songhobin@daum.net

**Abstract.** In this paper, a scaled model propulsion system was designed and tested for an electrical brake until stopping of the vehicles. A brake torque control method at the moment of vehicle stop was proposed accordingly. The test results for the electric brake were drawn while controlling the motor at low speed until the stop of the electric locomotives. Also, the speed detection method was used by implementing an observer that estimates the position and speed of rotor by using a resolver. Power converter was constructed as a converter-inverter system. Further, an improved brake method that uses only an electric brake till motor stop is proposed by comparing those in the blending brake that uses an air brake while reducing brake torque at vehicle stop.

**Keywords:** Electric brake, Traction motor, Brake torque, Converter-inverter.

## 1 Introduction

It has been reported that brake technology can pursue the effective energy use by expanding the use of a regenerative brake. Brake force is secured with technology development that expands a regenerative brake not only in the electric brake till stopping the motor but also in the high speed region. Also, brake force was reported to be increased and ride comfort is excellent by the new technology compared with the conventional.[1][2]

A pure electric brake possibly makes the mechanical brake device reduced by reducing air brake use. Moreover, it is superior in the maintenance and environmental aspects by reducing wear and tear of the brake shoe and lining as well as consequently generating dust.[3][4][5]

---

\* Corresponding author.

In this paper, a scaled model propulsion system was designed and tested for a electrical brake until stopping of the vehicles. A brake torque control method at the moment of vehicle stop was proposed accordingly. The test results for the electric brake were drawn while controlling the motor at low speed until the stop of the electric locomotives. Also, the speed detection method was used by implementing an observer that estimates the position and speed of rotor by using a resolver. Power converter was constructed as a converter-inverter system. Further, an improved brake method that uses only an electric brake till motor stop is proposed by comparing those in the blending brake that uses a air brake while reducing brake torque at vehicle stop.

## 2 Electric Brake Method Till the Stop of Electric Locomotive

The brake of an electric locomotive uses both an air brake and an electric brake. M car uses the air brake as a supplement for the shortage of brake force according to the electric brake specification. Since electric braking force is less in high speed region, speed reduction is executed with a blending brake which uses an air brake along with an electric brake, while at constant torque, speed reduction is executed with only an electric brake.[6] A vehicle is stopped by changing the electric brake into the air brake at the speed of around 6~7[km/h]. Stopping is executed by using the method as shown in Fig. 1(b). The proposed ideal brake process when electric brake is used to stop the vehicle is presented in Fig. 1(a).

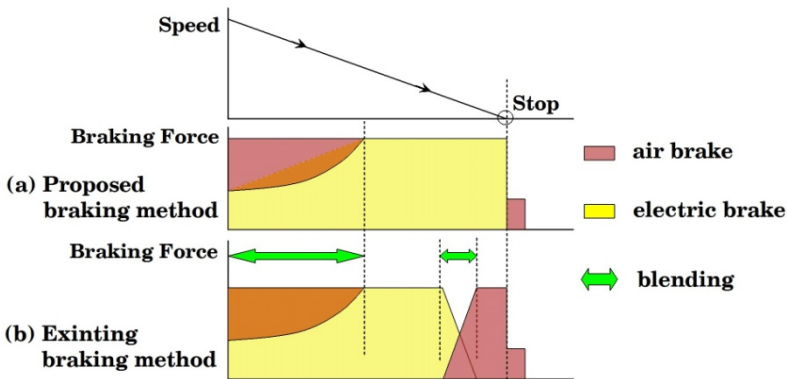


Fig. 1. Mode of electric brake until stop

For the torque reduction method in Fig. 2, section (1) is a zone wherein the vehicle is decelerated as brake torque  $T$ , section (2) is a reduction starting zone of torque at  $\omega_0$ , and section (3) is a control zone of brake torque.[7]

In section (3) of Fig. 2, when brake torque is made proportional to the vehicle speed, brake torque becomes equation (1).

$$T = k\omega \tag{1}$$

Under deceleration condition, there exists brake torque for section (1) of Fig. 2 and  $\omega_0$  which is decided by proportional factor  $k$  in equation (1). Besides, since the

condition for torque becomes zero in the torque control by Equation (1) is speed zero, it satisfies the condition for the stop moment.

Rotational speed and brake torque in the brake section can be expressed as equation (2).

$$\frac{d\omega}{dt} = -\frac{T}{J} \tag{2}$$

From equation (1) and equation (2), the rotational speed in section (3) becomes equation (3).

$$\omega = \omega_0 e^{-\frac{k}{J}t} \tag{3}$$

Reduction starting speed of brake torque  $\omega_0$  and deceleration time constant are decided by proportional factor  $k$ . When this proportional factor is applied to actual vehicles, it is a factor that needs to be controlled by considering ride comfort, etc.

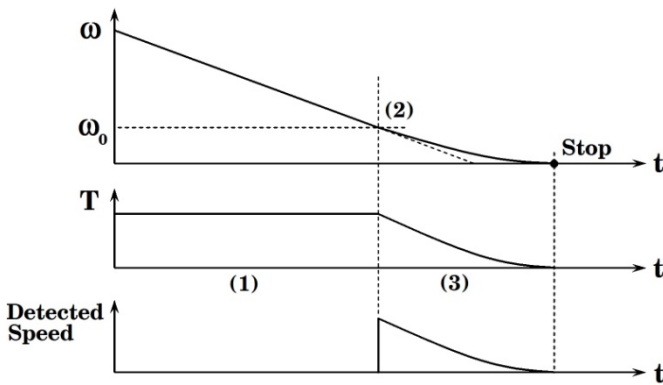


Fig. 2. Torque control at the moment of vehicle stop

### 3 Experimental of Motor Brake and Its Driving

When a vehicle is stopped by an electric brake, speed measurement at low speed and torque control are key functional elements. Especially, the precision in the speed measurement is extremely important to stop the motor smoothly. Therefore, position and speed of rotor were measured using a resolver in this study. By generating brake torque that is proportional to speed at low speed rather than generating suspension torque, travelling prevention after stop the rotor was achieved.

Fig. 3 is the measured wave form that stops rotor with torque proportional to the speed by detecting the speed that starts torque reduction when regenerative brake was executed during forward-reverse operation of motor.

In Fig. 3,  $i_q^*$  is the set value of current to the torque.  $\omega$  the magnified waveform estimated in the speed detector as low speed region. The recording time was limited as 17.5[rpm] for the forward-reverse direction for instrumentation. The third waveform in the Figure 3 is the detected speed waveform to control torque that is

proportional to the speed. The measurement was performed with the same magnification as  $\omega$ .  $i_a$  is the measurement of line current of motor.

The test equipment was adjusted as torque was varied depending on steps in the startup and brakes. The control was made to be switched at the point where torque and set torque becomes same by Equation (1) at the moment of stop while the regenerative brake was executed. A close examination of  $i_q^*$  in Fig. 3 gives an idea that current set value doesn't follow step changes at the motor stop. It is clear that torque control was switched by control program. In addition, by controlling torque without having off on the power converter even after the motor stops, suspension torque proportional to the speed could be maintained and stable movement could be observed. In Fig. 3, it can be seen that conversion of speed-torque control is executed at near 10[rpm].

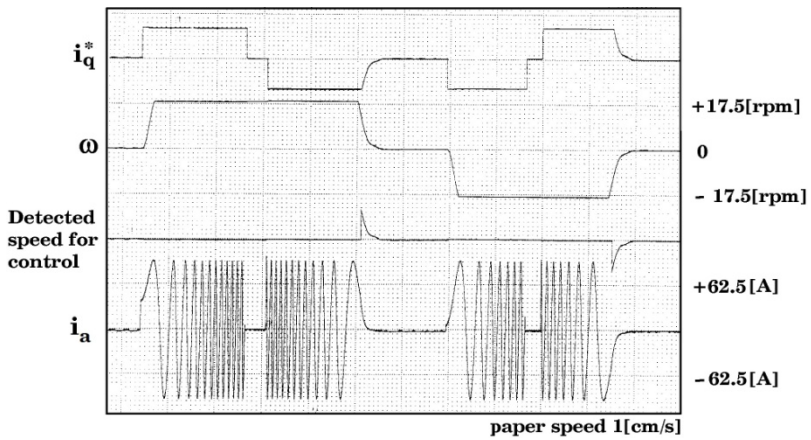


Fig. 3. Brake waveform of motor

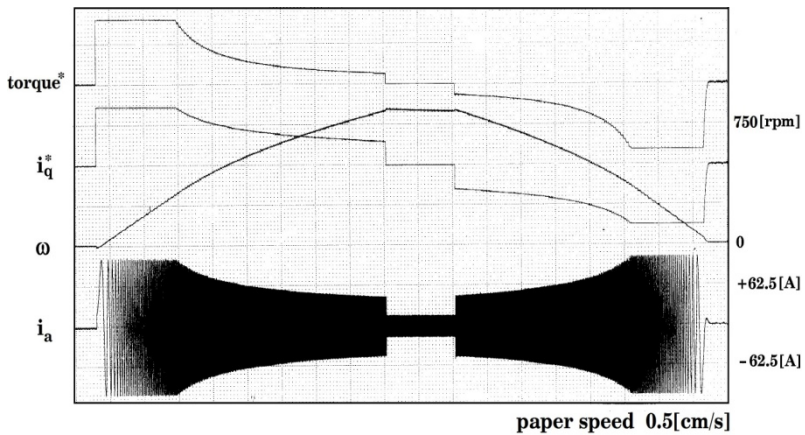


Fig. 4. Drive experiment for the inertial load



The vehicle operation was performed by a drive motor and stop it by a regenerative brake. In Fig. 4, the speed which startup the characteristic drive was set at 360[rpm] and torque was set zero at near 800[rpm]. After starting the regenerative brake, the motor was stopped with torque that is proportional to the speed. Torque was set as step changes same as previous experiments as shown in Fig. 4. The smooth driving at stop of motor was thereby observed.

## 4 Conclusion

A scaled model propulsion system was designed for the electric brake until stopping the vehicle. With experimental results, a control method of brake torque at the moment of motor stop was proposed. With test equipment, drive and brake tests were repeated in this study. Since the control method of brake force that was proportional to the speed at the moment of motor stop was implemented in the test, soft drive effect could be achieved at the stop only with the electric brake.

Besides, by carrying out an air brake movement at the same time of reducing brake torque at the stop, environmental perspective and energy use aspect and performance of vehicle could be maximized with the improved brake method that uses only an electric brake until vehicle stop as compared with those in the blending brake. The additional benefits of reducing vehicle weight and cost reduction for maintenance as well as ride comfort, energy efficiency and noise reduction could be expected.

## References

1. Kim, Y.-C., Song, H.-B., Cho, M.-T., Lee, C.-S., Kim, O.-H., Park, S.-Y.: A Study on the Improved Stability of Inverter through History Management of Semiconductor Elements for Power Supply. In: Kim, T.-h., Ramos, C., Kim, H.-k., Kiumi, A., Mohammed, S., Ślęzak, D. (eds.) ASEA/DRBC 2012. CCIST, vol. 340, pp. 155–162. Springer, Heidelberg (2012), doi:10.1007/978-3-642-35267-6\_20
2. Kraiem, H., Flah, A., Hamed, M.B., Sbita, L.: High Performances Induction Motor Drive Based on Fuzzy Logic Control. *IJCA* 5(1), 1–12 (2012)
3. Kim, Y.-C., Song, H.-B., Cho, M.-T., Moon, S.-H.: A Study on Direct Vector Control System for Induction Motor Speed Control. In (Jong Hyuk) Park, J.J., Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) *EMC Technology and Service. Lecture Notes in Electrical Engineering*, vol. 181, pp. 599–612. Springer, Heidelberg (2012), doi:10.1007/978-94-007-5076-0\_73
4. Ewald, F., Fuchs, M.A.S.: Analyses and Designs Related to Renewable Energy Systems. In: LLC 2011, pp. 557–687. Springer Science+Business Media (2011), doi:10.1007/978-1-4419-7979-7\_12
5. Aravindh Kumar, B., Saranya, G., Selvakumar, R., Swetha Shree, R., Saranya, M., Sumesh, E.P.: Fault detection in Induction motor using WPT and Multiple SVM. *IJCA* 3(2), 9–20 (2010)
6. Treerong, J.: Fault Detection of Electric Motors Based on Frequency and Time-Frequency Analysis using Extended DFT. *IJCA* 4(1), 49–58 (2010)
7. Pal, S., Tripathy, N.S.: Remote Position Control System of Stepper Motor Using DTMF Technology. *IJCA* 4(2), 35–42 (2010)

# FOLI Technique Algorithm for Real-Time Efficient Image Processing

Seongsoo Cho<sup>1</sup>, Kwang Chul Son<sup>2</sup>, Jongsup Lee<sup>3</sup>, and Seung Hyun Lee<sup>2</sup>

<sup>1</sup>Department of Electronic Engineering, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea

<sup>2</sup>Graduate school of Information Contents, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea

<sup>3</sup>Department of LINC, Dongguk University, #710, 82-1 Pil-dong 2-ga, Jung-gu, Seoul 100-272, Korea

{css, kcson, shlee}@kw.ac.kr, jsleearmy@dongguk.edu

**Abstract.** We proposed an image enhancement algorithm with less computational complexity than the existing one. In the proposed blocks for algorithm, the decimation filter provides the reduction of computational complexity using spatial reduction methods. The contrast enhancement block employs a probability density value (PDV) in order to control the excessive change of brightness. Cumulative distribution function (CDF) is estimated by using first order linear interpolation (FOLI) for real time processing. The colour enhancement block carries out saturation enhancement using saturation extension by maximum saturation enhancement (MSE) methods. It has been implemented in FPGA (field programmable gate array), and we have made an evaluation board for experiment. As compared to the conventional algorithms, the proposed algorithm provides better performance and lower computational complexity.

**Keywords:** FOLI, Human visual system, Image Processing, Real-time, FPGA.

## 1 Introduction

Driven by advances in image sensor technology and digital image processing technology, digital image devices are everywhere these days in user-friendly mobile phones and digital cameras for everybody to use. In particular, explosive demand for mobile phones and portable digital cameras as well as high expectations for high-quality output image are evolving technologies at a speed of light, which in turn motivate research to search for effective solutions to improve color, brightness as part of photo correction [1-4]. Therefore, many image enhancement techniques have been explained in some literatures [4-8] and the hardware cost is more for real time processing. The existing colour enhancement algorithms use spatial gamut mapping methods which are based on three one-dimensional look-up tables for Red, Blue and Green component adjustment. However, one-dimensional look-up table cannot correct display gamut exactly. There are many colour gamut mapping algorithms to

reduce any differences that might exist among the sets of colours obtainable on difference media or display devices [8-10]. Therefore, we propose an image enhancement technique to cover such shortcomings and increase the performance. The proposed algorithm carries out not only the contrast enhancement using histogram but also the colour enhancement using saturation extension. Simulation was performed using C to verify the proposed algorithm and VHDL was used for design implementation.

## 2 Conventional Algorithm

Virgil E. Vichers and Silverman proposed the plateau histogram equalization which is a modification of histogram equalization. Firstly, proper threshold value is selected which is represented by 'T'. As shown in equation (1), if the value of  $P(k)$  is greater than  $T$ , it is forced to equal  $T$ , otherwise it is unchanged, where  $P(k)$  is the histogram of an image [11].

$$P_T(k) = \begin{cases} P(k), & P(k) \leq T \\ T, & P(k) > T \end{cases} \quad (1)$$

where  $k$  represents the gray level of an image,  $0 \leq k \leq 255$ . An image enhancement is made by  $P_T(k)$  as shown in equation (2), (3).

$$F_T(k) = \sum_{j=0}^k P_T(j), \quad 0 \leq k \leq 255,$$

$$D_T(k) = \left\lfloor \frac{255 \cdot F_T(k)}{F_T(255)} \right\rfloor \quad (2)$$

where  $F_T(k)$  is cumulative histogram of an image,  $D_T(k)$  is the value of  $k$  after enhancement. The previous enhancement algorithm is histogram projection, while the value of subscript  $T$  is 1, and the histogram equalization where the subscript  $T$  equals to  $P_{max}(k)$ .

## 3 The Proposed Algorithm

In the proposed method for the contrast enhancement based on a CDF calculation by using a linear approximation, the dark part and the bright part can be utilised to enhance the visual quality of an image. The procedure for the proposed method is performed as follows, i) decimated image generation processing, ii) the CDF calculation by a linear approximation, iii) decision processing and applying  $IE_{factor}$  (Image Enhancement Factor). The flow chart of the proposed algorithm is shown in Fig. 1.

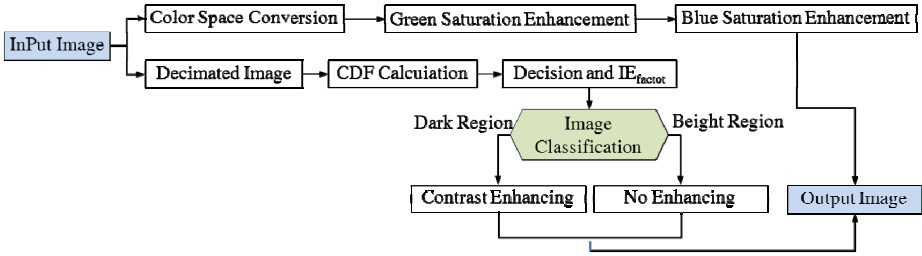


Fig. 1. Flow chart of the proposed algorithm

### 3.1 CDF Calculation by the FOLI Method

The conventional histogram equalisation method requires the entire pixel counter and memory in order to obtain the histogram information of the gray level of the input image, which makes the high computational complexity.

In the proposed method, FOLI is applied to calculate CDF for reducing the calculation time as well as its computational complexity. The proposed CDF is derived from sub-region value (SRV) and its cumulative distribution value (CDV). SRV represents a sub-region of histogram to calculate CDF, which is defined as (3).

$$SRV_k = \frac{L}{SRC}k \tag{3}$$

for  $k = 0, 1, \dots, SRC$ , In the case of 8-bit image,  $L = 255$ , the SRC is the number of sub-regions. In this paper, we choose  $SRC=4$  empirically. The reason is to prevent consumption of too many memory areas to materialize the hardware and deterioration of process speed due to too much computing when it is classified by the values of test results and too many areas.

CDV in the sub-region is defined as (4) for SRV.

$$CDV(X_{SRV_k}) = \sum_{j=0}^{SRV} X_j \tag{4}$$

for  $SRV_k = \frac{L}{4}, \frac{L}{2}, \frac{3L}{4}, L, 1, \dots, L$  and  $CDV(X_{SRV_k})$  represent the total sum of input image  $X_{SRV_k}$ .

The proposed transfer function by FOLI with CDV is defined as (5).

$$Y_n = (aX_{SRV_k} - an - 1) \times CDV(X_{SRV_k}) + a(n - X_{SRV_k}) \times CDV(X_{SRV_k}) \tag{5}$$

for  $n = 0, 1, \dots, L-1$ , where  $a = L/(X_{SRV_{k+1}} - X_{SRV_k})$ .

### 3.2 Decision Algorithm and Image Enhancement factor ( $IE_{factor}$ )

The existing histogram smoothing method used in the contrast control often generates excessive change in brightness. The proposed algorithm employs probability density

function (PDF) in order to prevent any sudden change since PDF provides information about contrast characteristics, which is derived from CDV. CDF and PDF are defined as (6) [6].

$$\begin{aligned}
 CDF(x) &= \sum_i^L P_i u(x - x_i) \\
 PDF(x) &= \frac{d}{dx} CDF(x) = \sum_i^L P_i \delta(x - x_i)
 \end{aligned}
 \tag{6}$$

where  $u(x)$  represents a unit step function and  $\delta(x)$  denotes an impulse function.

Probability density value (PDV) can be calculated from using CDVs in the sub-region. There are estimated errors in (7) between PDF and PDV. However, the proposed method does not require any accurate PDF since the whole image is used only to determine whether it is bright or dark. So, we need to compensate the errors by image pixel difference (IPD) derived from PDV. IPD is defined in (8) for decision algorithm.

$$PDV(X_{SRV(k+1,k)}) = CDV(X_{SRV_{k+1}}) - CDV(X_{SRV_k})
 \tag{7}$$

$$\begin{aligned}
 IPD1 &= Mag \left[ PDV \left( X_{\frac{L}{4},0} \right) \right] + Mag \left[ PDV \left( X_{\frac{L}{2},\frac{L}{4}} \right) \right] \\
 IPD2 &= Mag \left[ PDV \left( X_{\frac{3L}{4},\frac{L}{2}} \right) \right] + Mag \left[ PDV \left( X_{L,\frac{3L}{4}} \right) \right] \\
 &\quad \text{if}(IPD1 > IPD2) \text{under} \\
 &\quad \text{else} \\
 &\quad \text{Bright Image} \\
 &\quad \text{end if}
 \end{aligned}
 \tag{8}$$

By (9), the coefficient of contrast enhancement, ICDV, is defined as follows.

$$ICDV(X_{SRV}) = \begin{cases} CDV(X_{SRV}) & \text{for no enhancing} \\ IE_{factor} & \text{for enhancing} \end{cases}
 \tag{9}$$

Then, the output function  $Y_n$  can be expressed in (10) by substituting CDV by ICDV in (4).

$$Y_n = (aX_{SRV_k} - an - 1) \times ICDV(X_{SRV_k}) + a(n - X_{SRV_k}) \times ICDV(X_{SRV_{k+1}})
 \tag{10}$$

where  $SRV_k$  has  $16, \frac{L}{4}, \frac{L}{2}, \frac{3L}{4}$  and  $L$  for dark image and has  $0, \frac{L}{4}, \frac{L}{2}, \frac{3L}{4}$  and 235 for bright image.

## 4 Experiment

In order to compare the computing amount of the proposed algorithm and the existing algorithm, the frequencies of comparator, adder and multiplexer were compared. Each

computing device is the basic computing device to be used for hardware materialization and the computing amount can be compared by grasping the frequency of use of each. The measurement of frequency of use is the estimation based on algorithm performance.

**Table 1.** Computational Cost for Various Algorithms

	Comparator	Register	Adder	Multiplexer
CS	70	33	10	10
HE	256	514	257	258
BiHE	256	515	258	258
Proposed	10	25	45	18

The computational cost in table 1 was calculated assuming that all the architectures have been designed 24-bit adder and word length. Table 1 shows the calculation cost of hardware by each algorithm, and we can see less calculation was performed than the conventional algorithm. Especially, Adder show 1/5 times, comparator 1/25 times less than that of HE and BiHE, enabling us to know that it gives us superior performance with a very little calculation. In other words, the fact that outcome equivalent or better is obtained with smaller computing can be considered that it is easier in real time materialization than the existing algorithm.

As shown in Fig. 2-(a) is the original image which has large portion of blue and green colours. The processing of the blue and green colours in the image is not vivid enough. Thus it is required to be adjusted and the result using real-time colour gamut mapping method is shown in Fig 2-(b). By using the proposed approach, after applying the blue stretch and green scaling factors to the blue and green colours of the original image on each pixel adaptively based on its hue information, vivid video image can be obtained as shown in Fig 2-(c) and 3-(d). Fig. 2-(c) shows the results of performing only saturation enhancement algorithm while Fig. 2-(d) shows the results of simulating the entire enhancement for both contrast and saturation. We can see that the brightness and clearness increased more than those of the original images.



**Fig. 2.** Simulation results of conventional and proposed method (a) Original image, (b) Gamut mapping method, (c) Proposed method (without scalable contrast enhancement), and (d). proposed method with scalable contrast and emotion processing.

## 5 Conclusion

The proposed image enhancement method has low computational complexity for real time processing. The spatial reduction of input image was used to reduce the calculation amount. In order to prevent the excessive variation of the image brightness, the contrast enhancement algorithm was applied by separating and processing the dark region from the images. To make further improvement of the images saturations without the artifacts of colour reconstruction, the saturation enhancement algorithm was used. As an experimental result, the evaluation board was made and it was found that the proposed processor generated clearer and more natural image than original image. The simulation and experimental results show that the proposed algorithm provides better performance and low computational complexity. The reduced computational complexity and an enhanced image with improved clearness were found as a result of the proposed algorithm.

## References

1. Meylan, L., Alleysson, D., Sustrunk, S.: Model of retinal local adaptation for the tone mapping of color filter array images. *Journal of Optical Society of America A* 24(9), 2807–2816 (2007)
2. Choi, D.H., Jang, I.H., Kim, N.C.: Color Image Enhancement Based on an Improved Image Formation Model. *Jurnal of the IEEK SP* 4(6), 65–84 (2006)
3. Ke, C.: Adaptive Smoothing Via Contextual and Local Discontinuities. *IEEE Transactions on Pattern analysis and machine intelligence* 27(10), 1552–1567 (2005)
4. Park, Y.K., Kim, J.K.: A New Methodology of Illumination Estimation/ Normalization for Robust Face Recognition. In: *IEEE International Conference on Image Processing*, pp. 149–152 (2007)
5. Laih, C.S., Tu, F.K., Tai, W.C.: Global image enhancement using local information. *Electronics Letters* 30, 122–123 (1994)
6. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Transactions on Image Processing* 9, 889–896 (2000)
7. Kim, Y.T.: Contrast Enhancement Using Brightness Preserving Bi-Histogram Equalization. *IEEE Trans. on Consumer Electronics* 43(1), 1–8 (1997)
8. Cho, S., Shrestha, B., Joo, H.J., Hong, B.: Improvement of Retinex Algorithm for Backlight Image Efficiency. *Computer Science and Convergence* 114, 579–587 (2012)
9. Lee, C.S., Park, Y.W., Cho, S.J., Ha, Y.H.: Gamut Mapping Algorithm Using Lightness Mapping and Multiple Anchor Points for Linear Tone and Maximum Chroma Reproduction. *Journal of Image Science and Technology* 45(3), 209–223 (2001)
10. Chen, H.S., Kotera, H.: Three- dimensional Gamut Mapping Method Based on the Concept of Image Dependence. *Journal of Image Science and Technology* 46(1), 44–52 (2002)
11. Wang, B.-J., Liu, S.-Q., Li, Q., Zhou, H.-X.: A real-time contrast enhancement algorithm for infrared images based on plateau histogram. *Infrared Physics & Technology* 48(1), 77–82 (2006)

# Stabilization Inverse Optimal Control of Nonlinear Systems with Structural Uncertainty

Jong-Yong Lee<sup>1</sup> and Seongsoo Cho<sup>2</sup>

<sup>1</sup> Department of Culture, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea

<sup>2</sup> Department of Electronic Engineering, Kwangwoon University, 20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea  
{jyonglee, css}@kw.ac.kr

**Abstract.** Stabilization inverse optimal control for nonlinear systems with structural uncertainty is considered in this work. A theorem for the globally asymptotic stability based on the control Lyapunov function is presented from which a less conservative condition for the inverse optimal control is derived. The result is used to design an inverse optimal controller for a class of nonlinear systems, that improves and extends the existing results. The class of nonlinear system considered is also enlarged. The simulation results show the effectiveness of the method.

**Keywords:** Nonlinear Control, Control Lyapunov Function, Structural Uncertainty, Stability.

## 1 Introduction

Stability, optimization and uncertainty of the initial nonlinear control theory was actively discussed from Kokotovic and Arcak [1]. The core result addressed in this paper is the concept of control Lyapunov function (CLF) that relies on negative feedback and control. CLF was introduced by Artstein [2] and Sontag [3], and plays an important role in the stability theory.

In consequence, many researchers ranged over the effect of CLF and carried out important research outcomes [4-11]. Many researchers have chosen optimal stabilization as a theme of their thesis for the reason that optimal control is a tool designed to ensure the stability margin and robustness. The development of design methods for solving the inverse problem of optimal stabilization was proposed by Freeman and Kokotovic [4, 11]. In this paper, the inverse optimal control of nonlinear systems with structural uncertainty is considered. This study leads that global asymptotic stability in use of CLF to the theorem. Through this proof, less conservative condition is derived which can induce an inverse optimal control. The effectiveness of the proposed method is verified through simulation.



## 2 System Demonstration

Consider the nonlinear systems with uncertainty as follows.

$$\dot{x} = f(x) + \Delta f(x) + g(x)u \tag{1}$$

In this context,  $x \in R^n, u \in R^m$  are status of the system and the input.  $f(0) = 0$ , mapping  $f: R^n \rightarrow R^n$  and  $g: R^n \rightarrow R^n \times R^m$  are assumed smooth,  $\Delta f(0) = 0$  which mapping  $\Delta f: R^n \rightarrow R^n$  is property and structural uncertainty and can be expressed as equation(2).

$$\Delta f(x) = e(x)\delta(x) \tag{2}$$

Where  $e: R^n \rightarrow R^{n \times m}$  is a matrix given smooth function,  $\delta: R^n \rightarrow R^m$  is unknown vector-valued function. Also when presumed that  $\delta(x)$  is  $N(0) = 0$  with a given smooth function  $N: R^n \rightarrow R^+$ . Refer to the following,

$$\Gamma = \{ \delta(x): \|\delta(x)\| \leq N(x) \} \tag{3}$$

Where  $\|\cdot\|$  is indicates Euclidean norm. According to the entire  $x \in R^n$ ,  $\delta(x)$  or  $\Delta f(x)$  are acceptable when  $\delta(x)$  satisfies with respect to the equation (3).

Presume  $V: R^n \rightarrow R^+$  is a continuous function. If each  $x$  in  $x \neq 0$  is  $V(0) = 0$  and  $V(x) > 0$ ,  $V$  stated as positive definite and also in condition of  $V(x) \rightarrow \infty$  when  $\|x\| \rightarrow \infty$ ,  $V$  stated as proper.

In regard to  $V$  over the  $f$  Lie, a derived function is stated as  $L_f V(x) = \frac{\partial V}{\partial x} f(x)$ .

Definition 1: When each  $x \neq 0$  satisfies on equation (4), smooth and unique positive definite function  $V: R^n \rightarrow R^+$  is the CLF of system equation (1)[7].

$$inf_u \{ L_f V(x) + L_g V(x)u \} < -\|L_e V(x)\|N(x) \tag{4}$$

In condition,  $V(x)$  in equation (1) of CLF signifies equation (5).

$$L_g V(x) = 0, x \neq 0 \Rightarrow L_f V(x) < -\|L_e V(x)\|N(x) \tag{5}$$

Allusion 1: As from equation (5), a contented subset of  $L_g V(x) = 0$  is important in control. When  $L_g V(x) = 0$ , on certain  $x$ , if because  $L_f V(x) > 0$ ,  $V(x)$  is not Lyapunov function, of course nor CLF.

Definition 2: If towards certain  $\epsilon > 0$ , presuming  $\sigma > 0$  exists that fulfills  $\|x\| < \sigma$  which is  $x \neq 0$ ,  $V(x)$  is system of equation (1) satisfies small control property. In this case certain  $u$  that holds  $\|u\| < \epsilon$  exists to satisfy inequality (5).

Definition 3: Suppose  $k: R^n \rightarrow R$  has function of  $k(0) = 0$ . When fall apart from the origin as smooth and continuous at every point of  $R^n$ ,  $u(k) = k(x)$  on  $R^n$  represents as almost smooth[7].

### 3 The Primary Result

In this chapter, the core consequences of the study are presented. System equation(1) presents the almost smooth condition on feedback and sufficient conditions for global asymptotic stabilization.

#### 3.1 Global Stability Controller

Theorem 1: $V(x)$ is the system (1) of CLF and assume that it satisfies the small control property. At this point, in matter of acceptable  $\Delta f(x)$ , almost smooth state feedback control on system (1) subsists to allow asymptotic stability on band.

#### 3.2 Inverse Optimal Controller

In this part, the satisfaction of the property are discussed on feedback system of the system (1) which is the inverse optimal controller design.

1) The closed-loop system is the global asymptotic stable at the equilibrium point o  $x = 0$ f.

2) Input  $u$  is all about  $x$  that minimizes the cost function equation (16) regarding to  $\int(x) \geq 0$ and  $R(x) > 0$ .

$$J(u, x, x_0) = \sup_{\delta x \in \Gamma} \int_0^\infty (\int(x) + u^T R(x)u)dt \tag{6}$$

When approaching the inverse controller, the stabilizing controller  $u = k(x)$  is designer first, and then discovering  $\int(x) \geq 0$  and  $R(x) > 0$  for feedback  $u = k(x)$  to minimize the equation(6).

This method is determined by the designer that  $\int(x) \geq 0$  and  $R(x) > 0$  are selected inductively by feedback to stabilize rather than deductively, and this is called 'Inverse'.

Allusion 4: Sometimes it is difficult to discover a CLF for the system  $\dot{x} = f(x) + g(x)u$ .

In this paper,  $f(x)$  is decomposed to  $f(x) = f_1(x) + \Delta f_1(x)$ , and assume  $f_1(x)$  has the property that makes easier to get the CLF. It satisfies Theorem 1 and Theorem 2 about the CLF inverse optimal controller of  $\dot{x} = f(x) + g(x)u$ .

### 4 Example Cases

In order to validate the contents proposed in Chapter3, this chapter explains two example cases.

Example 1: Consider the following system of non-linear know the value of  $\theta$ .

$$\begin{aligned} \dot{x}_1 &= -x_1 + \frac{\theta}{2}x_2^2 + \theta(-2x_2 + x_3)x_3 \\ \dot{x}_2 &= -x_2 + x_3 \\ \dot{x}_3 &= -x_2 + x_3 + u \end{aligned} \tag{7}$$

Consider  $x = (x_1, x_2, x_3)^T$ ,

$$f(x) = \begin{pmatrix} -x_1 + \frac{\theta}{2}x_2^2 \\ -x_2 + x_3 \\ -x_2 + x_3 \end{pmatrix}, g(x) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \Delta f(x) = \begin{pmatrix} \theta(-2x_2 + x_3)x_3 \\ 0 \\ 0 \end{pmatrix}$$

Expressed as follows:

$$\begin{aligned} e(x) &= \begin{pmatrix} \theta x_3 & -2\theta x_3 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \delta(x) = \begin{pmatrix} x_3 \\ x_2 \end{pmatrix} \\ \Delta f(x) &= e(x)\delta(x)N(x) = \sqrt{x_2^2 + x_3^2} \end{aligned}$$

When  $V(x)$  has limit to the positive with fixed unique smooth function,

$$V(x) = \frac{1}{2} \left( x_1 + x_2 + \frac{\theta x_2^2}{2} \right)^2 + \frac{x_2^2}{2} + \frac{x_3^2}{2}$$

According to Theorem 2, the controller is same as the equation (8).

$$\begin{aligned} u &= -\frac{1}{2}B^{-1}(x)B^T(x) \\ &\begin{cases} -\gamma B^T(x) - B^T(x) \frac{\left( \frac{a(x) + \|c(x)\|N(x) + \sqrt{(a(x) + \|c(x)\|N(x))^2 + \|B(x)\|^4}}{\|B(x)\|^2} \right)}{\|B(x)\|^2} & ; B(x) \neq 0 \\ -\gamma B^T(x) & ; B(x) = 0 \end{cases} \end{aligned} \tag{8}$$

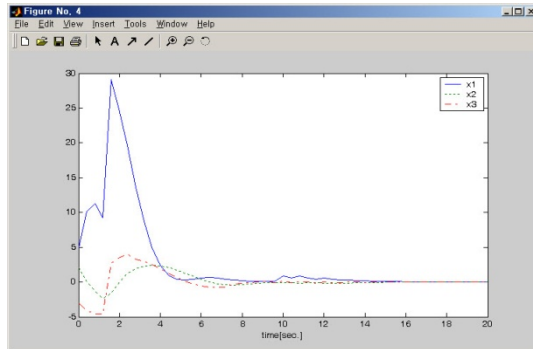


Fig. 1. The State of example 1

The closed-loop system without optimal limit of  $\theta$  which formed by system equation (23) and control equation (24) is global asymptotic stability of equilibrium point  $x = 0$  and  $J = V(x(0))$ .

To simulate, the initial states  $x(0) = (5, 2, -3)^T$ ,  $\theta = 1$ ,  $\gamma = 0.5$  were formed.

The simulation results are illustrated in Figure 1 and 2. As shown in the figure, it demonstrates that entire condition converges to 0.

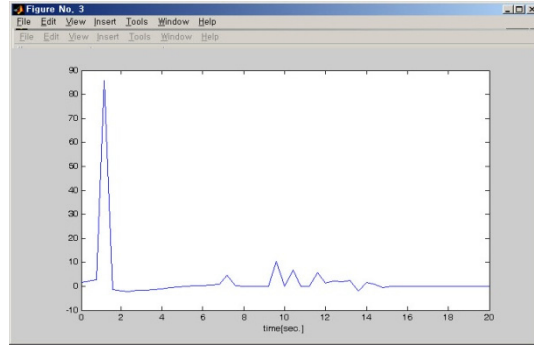


Fig. 2. The Control of example 1

For instance,  $V(x)$  is as follows.

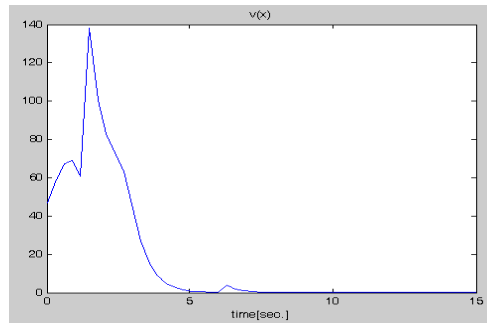


Fig. 3. The performance index of example 1

## 5 Conclusion

The inverse optimal control for nonlinear systems with structural uncertainty was discussed. The theory of global asymptotic stability was proposed on the basic imply on CLF function. In addition, the less conservative condition was derived in inverse optimal control. This method was used in nonlinear systems with structural uncertainty in order to design the optimal control station. The simulation results demonstrate that the proposed method is more effective. Through this method, the current system of nonlinear control techniques have improved.

## References

1. Kokotovic, P., Arcak, M.: Constructive nonlinear control. A historical perspective. *Automatica* 37(5), 637–662 (2001)
2. Artstein, Z.: Stabilization with relaxed controls. *Nonlinear Anal.Theory Methods Appl.* 7(11), 1163–1173 (1983)
3. Sontag, E.D.: A Lyapunov-like characterization of asymptotic controllability. *SIAM J. Control Optim.* 21(3), 462–471 (1983)
4. Freeman, R.A., Kokotovic, P.V.: Inverse optimality in robust stabilization. *SIAM J. Control Optim.* 34(4), 1365–1391 (1999)
5. Freeman, R.A., Primbs, J.A.: Control Lyapunov functions: new ideas from an old source. In: *Proc. IEEE Decision and Control Conf., Kobe, Japan*, pp. 3926–3931 (1996)
6. Mei, S.W., Shen, T.L., Sun, Y.Z., Lu, Q.: Passivation control of nonlinear systems with disturbances. *Control Theory and Appl.* 16(6), 797–806 (1999)
7. Sontag, E.D.: A universal construction of Artstein’s theorem on nonlinear stabilization. *Syst. Control Lett.* 13(2), 117–123 (1989)
8. Lin, Y., Sontag, E.D.: A universal formula for stabilization with bounded control. *Syst. Control Lett.* 16(5), 393–397 (1991)
9. Isidori, A.: *Nonlinear control systems*, vol. 2. Springer (1999)
10. Sepulchre, R., Jankovic, M., Kokotovic, P.V.: Constructive nonlinear control (September 27, 2011)
11. Freeman, R.A., Kokotovic, P.V.: Robust integral control for a class of uncertain nonlinear systems. In: *Proc. 34th IEEE Conf Decision and Control, New Orleans, LA*, pp. 2245–2250 (1995)

# Implementation of Wireless Electronic Acupuncture System

You-Sik Hong<sup>1</sup>, Bong-Hwa Hong<sup>2,\*</sup>, and Baek-Ki Kim<sup>3</sup>

<sup>1</sup>Dept. of Computer Science, Sangji University, Korea

<sup>2</sup>Dept. of Digital Media Engineering, Kyunghee Cyber University, Korea

<sup>3</sup>Dept. of Information & Telecommunication Eng.,

Gangneung-Wonju National University, Korea

h5674korea@gmail.com, bhhong@khcu.ac.kr, bkkim@gwnu.ac.kr

**Abstract.** In this paper we proposed diagnosis system in which we could use simultaneously pulse and tongue diagnosis data, and measurement data of oxygen saturation using technology of bluetooth. And also we developed adaptive wireless acupuncture system by using pulse diagnosis system to adjust strength and time of acupuncture and several acupuncture points of patients for whom intellectual fuzzy technology is applied. With this acupuncture system we can obtain optimal acupuncture time and use at any where and any time easily by input our physical condition to smart phone and in the web. We implemented smart wireless electronic acupuncture system to get acupuncture easily using intelligent diagnosis system.

**Keywords:** fuzzy rules, diagnosis, wireless, acupuncture.

## 1 Introduction

More than 60 percent of the electronic acu punctures are developed in the country using low frequency and the rest is developed using instantaneous electro stimulation. Existing low-frequency therapeutic apparatuses are simple frequency generator(16~32Hz) which attaches the electrodes to patient's diseased area. Patient cannot be treated effectively because it does not provide detailed frequency(three decimal places) but uncertain frequency. Furthermore, it can't find acupuncture points because it has no consideration for patient's sex, age, weight, illness, etc. And it causes problem that children and elderly people are bruised or wounded after getting electronic acupuncture because of inappropriate acupuncture time and strength.[1]. The pulse is considered an important factor in oriental medicine because observation of a person's pulse rate may reflect their health and illness. For example, if patient's heart stopped, it is very serious situation and this situation can be judged by pulse. Oriental doctors have considered pulse rates as important data in diagnosis. But the existing blood pressure pulse analyzer has some problem. It is uncertain whether the blood pressure pulse analyzing sensor is located precisely on the radial artery and it is difficult to

---

\* Corresponding author.

diagnose pulse exactly in different case of thick and thin forearm. Furthermore, the analogue type blood pressure pulse analyzer has problems with quantification of the blood pressure pulse. Although some people may have the same forearm length but the thickness of their blood vessel may differ. Therefore there is no set of data that is considered reliable enough to judge the accuracy of blood pressure pulse rates. Oriental doctors should not only judge the basic biological signals such as checking the pulse's size, strength, and speed, but should also consider the basic and quantitative analysis of the pulse in order to gain an accurate diagnosis. Also, the doctor should consider physical characteristics, such as the thickness of the skin and blood vessels, in order to reach an accurate conclusion. Therefore, measurement of the blood flow rate is a vital indicator in understanding the blood pressure rate and how the substances in the blood are transported.[2][3].

The method of exiting diagnosis has problem which cannot diagnose the old and the infirm exactly because it does not take into consideration the condition of patient's gender, age, skin. To solve this problem, we analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We proposed the algorithm that diagnoses patient optimally considering patient's condition using intelligent fuzzy technique.[4][5]. In this paper we developed adaptive wireless acupuncture system by using pulse diagnosis system to adjust strength and time of acupuncture and several acupuncture points of patients for whom intellectual fuzzy technology is applied. And also we proposed diagnosis system in which we could use simultaneously pulse and tongue diagnosis data, and measurement data of oxygen saturation using technology of bluetooth. The composition of this paper is as follows. Section 2 is about Intelligent pulse diagnosis algorithm, section 3 is about wireless electronic acupuncture system, section 4 is about the simulation of wireless electronic acupuncture, and finally section 5 concludes.

## 2 Intelligent Pulse Diagnosis Algorithm

The intelligent pulse diagnosis system composed of three parts. The first part is composed of the sensor to detect the conductance which corresponds with injured part of human body, and reference signal generator to moderate the signal generated from patients is included. The second part is composed of DSP (Digital Signal Processor) board in which the signals are measured and to do a sort using fuzzy algorithm.[5] The last part is composed of computer system that displays the signal from DSP board to the monitor, and analysis software to diagnose the patients. Fig.1 shows the whole diagnosis algorithm for electronic acupuncture.

Pulse is beat-wave pattern of chest wall and great arteries according to heartbeat. The main purpose of pulse is observation of cardiomotility and blood movement. Recently study using physical characteristics shows that pulse wave pattern can change according to condition of blood vessel and blood circulation. The pulse wave pattern can be obtained by second differentiation of digital plethysmogram using physical specific status such as uncertain inflection point. In this paper, we classified a patient's physical condition into three categories, dangerous, ordinary, normal condition adapting pulse diagnosis algorithm using acceleration pulse wave pattern.

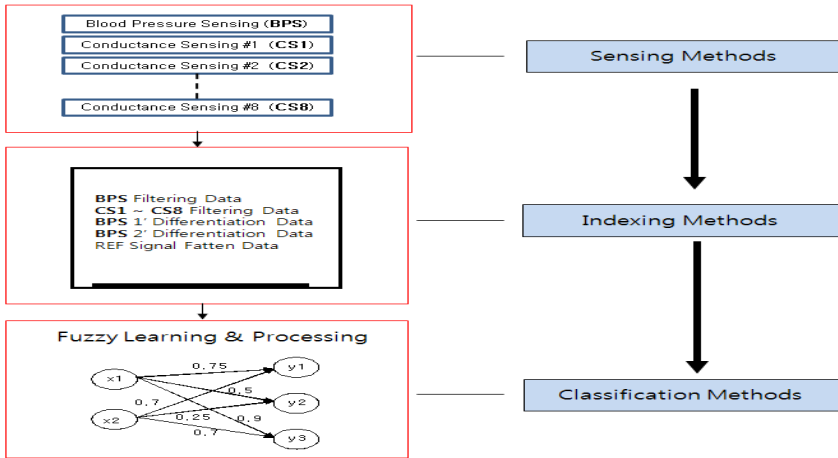


Fig. 1. Diagnosis algorithm for electronic acupuncture

Combination function of trust value: 1 and 2 type of fuzzy creation rule reduced type of 5 and 6 can come to the same node and conclusion through different inference path to infer fuzzy. In this node the same conclusion reach two of more different trust value. In this case combination function of trust value is used to recalculate trust value of conclusion.

$$\beta_c = \beta_{comb}(\beta_c, \beta_c^{old})$$

$$= \max(\beta_c, \beta_c^{old})$$

Here  $\beta_c^{old}$  is trust value of conclusion reached through inference path already,  $\beta_c$  is trust value of other conclusion reached through another inference path. If the 4 patients' (a, b, c, d) illness condition is end stages the value is displayed as 0.8-1.0 shown in the left, in case of middle stage the value is 0.4-0.7 and in case of first stage the value is displayed as 0.1-0.3. The value in the middle is displayed patient's physical condition. For example, if the patient's height is 150cm and weight is lower than 45kg the value is displayed as 0.1-0.3, in case of the height 151cm-170cm and the weight 46kg-70kg the value is displayed as 0.4-0.7, and in case of the height 171cm-200cm and weight 71kg-130kg the value is displayed as 0.8-1.0. In fig.1 the process to calculate fuzzy correction factor according to patient's physical condition is shown.

### 3 Wireless Electronic Acupuncture System

Wireless electronic acupuncture system with built in multi pad which can find out the condition of the patients automatically and treat the patients simultaneously. The system includes the function that it can treat the patients with acupuncture adjusted voltage, current, frequency oscillation automatically according to their physical



conditions. To perform the function, the system has function to sense and treat acupuncture simultaneously, and required logical and statistical data processing technique using fuzzy and exact analysis.[6]

Fig.2 shows whole diagram of the system. Shown in the left of the fig.2 we can see the 5-pad installed underneath the palm. Installing the 5 circle pad underneath the palm we can exchange the signal, and then adaptive acupuncture treatment can be done. At this point, measurement of the signal use the wireless type instead of cable type. Because the wireless type has advantage of convenience to get acupuncture, reduction of noise by use cable connected computer system and prevention of electric shock according to abrupt high-tension electricity.[7][8][9]

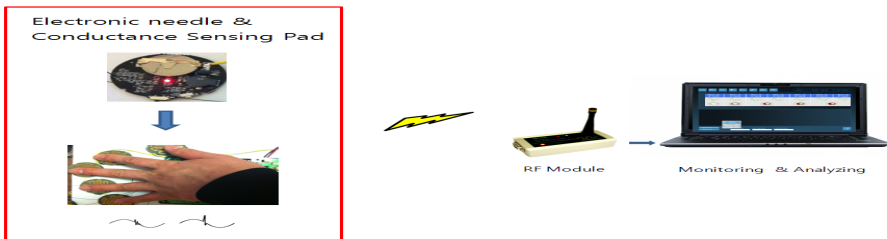


Fig. 2. Full diagram of the system

Information extraction of the human body to treat acupuncture is important not only data from body but also ages, sex, height and weight of patients. To do this, we make control variables using fuzzy algorithm before treatment of acupuncture.

Fig.3 shows Circuit of the Input signal AMP and Acupuncture signal. The part of sensing pad and contact point of the fingertip made of stripe array type to distribute contact point area evenly after plating with gold to reduce electric resistance.

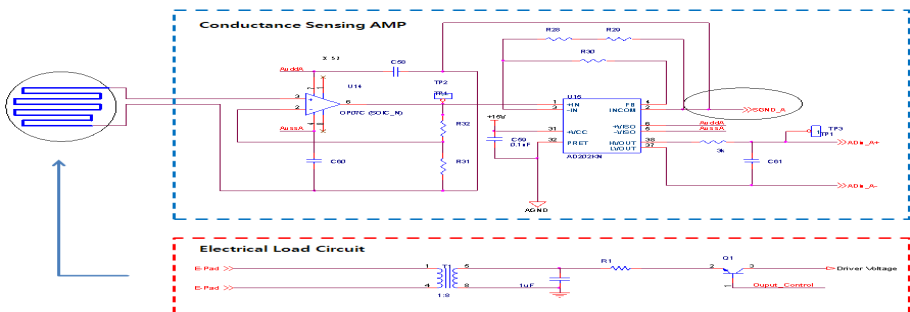


Fig. 3. Circuit of the Input signal AMP & Acupuncture signal

Fig.4 shows circuit of main processor and RF communication part. We made the compact circuit using 3V button battery, and provided expandability to measure another body part later.

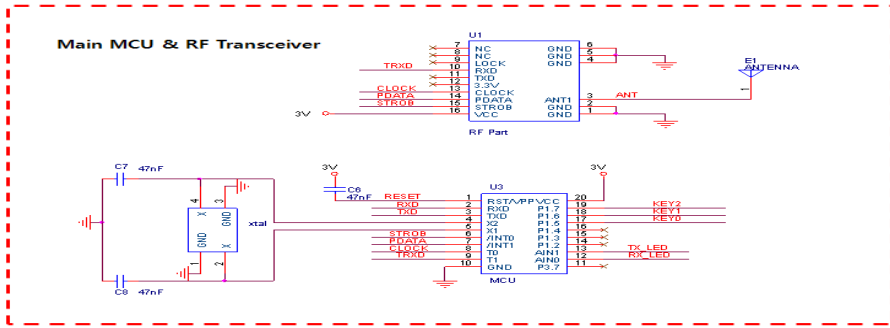


Fig. 4. Circuit of main processor and RF communication part

### 4 Simulation

In this paper, we proposed the optimal algorithm which could judge the remote medical diagnosis using fuzzy logic and fuzzy inference rules, and we simulated the process to calculate the optimal acupuncture time of body condition of patients. We produced the wireless communication part to transmit condition of patients' pulse, skin conductance and oxygen saturation data to user's terminal or remote medical terminal, and to receive the control signal from user's terminal or remote medical terminal.

To do this, we made the sensing pad, the circuit of AMP and acupuncture signal, wireless communication module and charging circuit for storage battery. And also we proposed the software including algorithm of analysis and control using fuzzy technique. Existing acupuncture system using DSP has complex structure, uses up a lot of electricity and it's big and expensive. But the adaptive wireless acupuncture system proposed in this paper is simple, inexpensive and safe. Fig.5 shows simulation of the glove type electronic acupuncture.

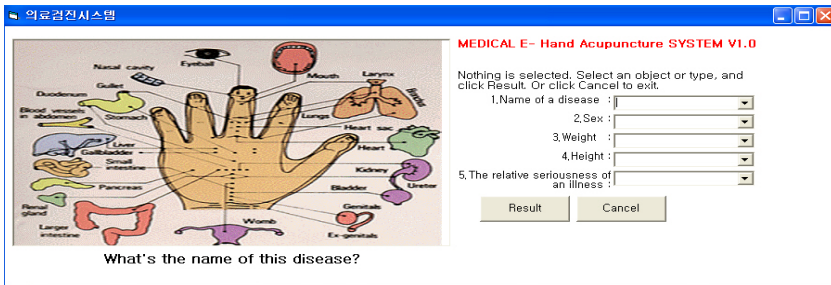


Fig. 5. Simulation of the glove type electronic acupuncture

To implement wireless system we used the way of RF data modem for wireless communication using Narrowband FSK. The feature of this way is robust to noise and it can transmit data easily by simple communication protocol. And this system is

adapt to design multi type data communication system and can be designed by low power, one 3V battery, in case of short distance. We considered not only resistance measurement but capacitive component to reduce error according to several condition of human body. To do this, we applied the pulse wave DC50V~200V, 500uA~1,500uA, intermittent stimulation of 5Hz~5KHz to main pad and fingertip and measured the voltage peak and phase frequency.

We used 470MHz band frequency and designed the system to change 21 physical frequency. And logical address of a channel corresponding to each adaptive acupuncture was assigned using polling technique and then called. The system supports half duplex communication. This way is suitable for the system because the system require low data and uses low speed communication relatively. The output power of wireless signal using button type battery is 1mW, and it is adequate to transmit data without noise. The speed of transmission is 1200~9600bps, wireless encoding uses way of Bi-phase Manchester code, communication between notebook computer and wireless modem uses RS232C.

Fig.6 shows wireless acupuncture system Android-based. In this system remote information is transmitted and received by process as follows.

- ① Measure bio information using sensor equipment for health care(pulse sensor, blood pressure/sugar sensor, ECG sensor, infrared thermometry sensor).
- ② Process measured bio information in the Hmote2420.
- ③ Transmit the information to Android platform through bluetooth using H-Andro210.

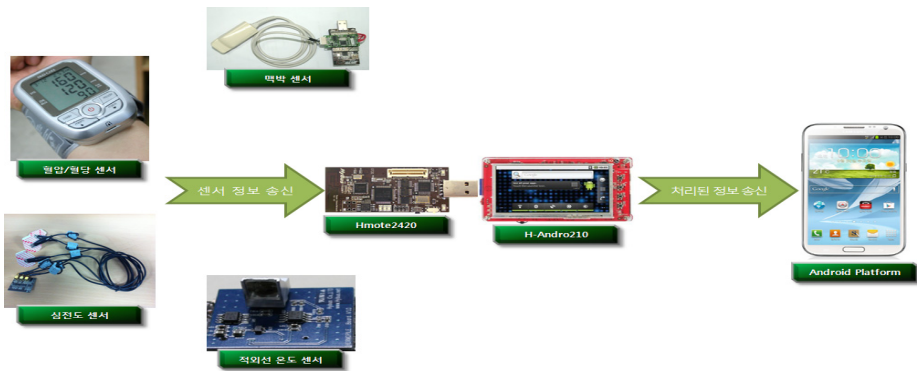


Fig. 6. Wireless acupuncture system

Fig.7 shows data the transmitter and the receiver using RF communication. For remote medical treatment, the transmitter acquire data from 4 sensors, and then transmit the data to receiver using RF communication.

Fig.8 shows transmit/receive system for ubiquitous network. It is made of MSP240CPU and CC2420 RF chip.



Fig. 7. Data transmitter and receiver using RF communication

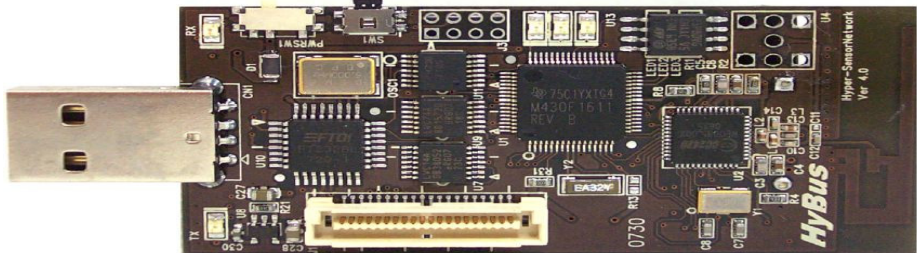


Fig. 8. Transmit/receive system for ubiquitous network

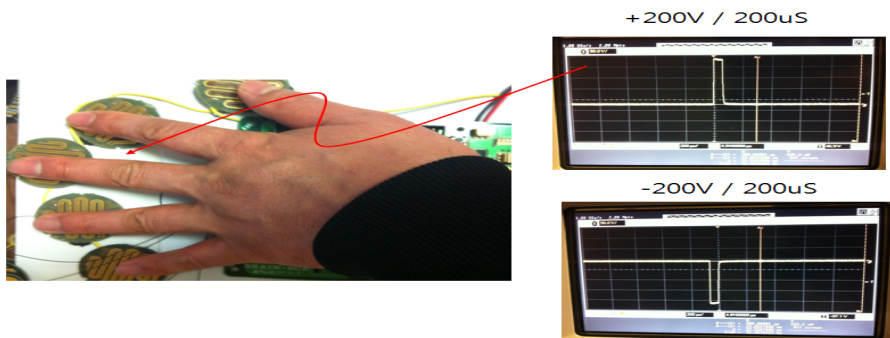


Fig. 9. Scene of electro stimulation to fingertips

Fig.9 shows scene of electro stimulation to fingertips using pads. To obtain signal, we send a reference signal to palm, and then decide body condition of patients on the basis of data obtained from pre-investigation using sensing pads and MCU attached to fingertips. At the same time signal processing is completed, electric stimulation signal generated by fuzzy algorithm is transmitted to sensing pads.

## 5 Conclusion

Existing acupuncture system using DSP has complex structure, uses up a lot of electricity and it's big and expensive. To solve this problems, we presented intelligent pulse diagnosis algorithm and wireless electronic acupuncture system.

Using proposed algorithm which judge the remote medical diagnosis based on fuzzy logic and fuzzy inference rules, we can calculate the optimal acupuncture time of body condition of patients. We made the sensing pad, the circuit of AMP and acupuncture signal, wireless communication module and charging circuit for storage battery. The intelligent wireless acupuncture system proposed in this paper is simple, inexpensive and safe compared with conventional acupuncture systems.

## References

- [1] Hong, Y.S., Kim, H.K., Kim, B.K.: Implementation of Adaptive Electronic Acupuncture System using Intelligent Diagnosis System. *International Journal of Control and Automation* 5(3), 141-152 (2012)
- [2] Lee, Y.J., Lee, J., Lee, H.J., Yoo, H.H., Choi, E.J., Kim, J.Y.: Study on the characteristics of blood vessel pulse area using ultrasonic. *Korea Institute of Oriental Medicine Researches* 13(3), 111-119 (2007)
- [3] Shaltis, P.A., Reisner, A.T., Asada, H.H.: Cuffless Blood Pressure Monitoring Using Hydrostatic Pressure Changes. *IEEE Trans. Biomed. Eng.* 55, 1775-1777 (2008)
- [4] Sik, H.Y., Kug, P.C.: Proc. of the Sixth International Fuzzy System Association, IFSA, pp. 461-464 (1995)
- [5] Baruah, H.K.: The Theory of Fuzzy Sets: Beliefs and Realities. *IJEIC* 2(2), 1-22 (2011)
- [6] Kim, S.W., Choi, Y.G., Lee, H.S., Park, D.H., Hwang, D.G., Lee, S.S., Kim, G.W., Lee, S.G., Lee, S.J.: Improvement of Pulse Diagnostic Apparatus with Array Sensor of Magnetic Tunneling Junctions. *J. Appl. Phys.* 99, R908-R910 (2006)
- [7] Lee, S.S., Ahn, M.C., Ahn, S.H.: A New Measurement Method of a Radial Pulse Wave Using Multiple Hall Array Devices. *J. Magnet.* 14, 132-136 (2009)
- [8] Haykin, S.: *Modem Wireless communication*. Prentice-Hall (2003)
- [9] Swami, A., Hong, Y.: *Wireless Sensor Networks: Signal processing and communications*. Wiley (2007)

# Network Based Intelligent Agent for Ubiquitous Environments

Dong W. Kim<sup>1</sup>, Ho-Dong Lee<sup>2</sup>, Sung-Wook Park<sup>3</sup>, and Jong-Wook Park<sup>3</sup>

<sup>1</sup> Dept. of Digital Electronics,  
Inha Technical College, Incheon, South Korea  
dwnkim@inhatec.ac.kr

<sup>2</sup> Mitos Company

<sup>3</sup> Dept. of Electronics,  
University of Incheon, Incheon, South Korea

**Abstract.** The network based intelligent agent (NBIA) system consists of intelligent software components. The framework of an NBIA system warrants reliable and stable network connection with real-time response between software components. Intelligent software components work on the resources distributed on the network, and the NBIA uses the results from the resources. In this way, the NBIA can provide high-quality intelligent service to users. Thus, a TCP/IP-based interface is designed to integrate software components into an NBIA system. This interface uses the server-client architecture to solve the bottleneck problem with the transfer rate of a wireless LAN.

**Keywords:** Network based intelligent agent, Ubiquitous environment.

## 1 Introduction

High-quality intelligent services need more system resources and a variety of environments. Also, these intelligent services work on a variety of operating systems or OS's (Windows, Linux, etc.), as well as on diverse hardware. Therefore, integrating these services on the network is a very difficult task.

TCP/IP is a traditional method of integrating components via a network. It is reliable and stable, and it is very easy to develop its components. Various architectures and integration schemes have been proposed [1-5]. With the development of components-based software engineering, a middleware such as CORBA is used to integrate components by virtue of its abstract network layer. The developer can use unlimited resources via the network without considering different OS's and diverse hardware. Furthermore, the use of middleware increases software reusability. Numerous architectures that use middleware have been reported [6-10].

The network based intelligent agent (NBIA) system consists of intelligent software components. Thus, the framework of an NBIA system warrants reliable and stable network connection with real-time response between software components. Intelligent software components work on the resources distributed on the network, and the NBIA uses the results from the resources. In this way, the NBIA can provide high-quality

intelligent service to users. Thus, a TCP/IP-based interface is designed to integrate software components into an NBIA system.

In the following, communications interface design based on TCP/IP, will be introduced. This scheme uses the server-client architecture to solve the bottleneck problem with the transfer rate of a wireless LAN.

## 2 Communications Interface Design Based on TCP/IP

Communications TCP/IP(Transmission Control Protocol/Internet Protocol) is a well known and the most popular communications protocol for Internet connection. In this section, communications interfaces based on TCP/IP are designed for an NBIA.

### 2.1 TCP/IP Protocol

TCP/IP consists of an IP(Internet protocol), which is an Internet protocol that uses the packet communication method, and a TCP(transmission control protocol).

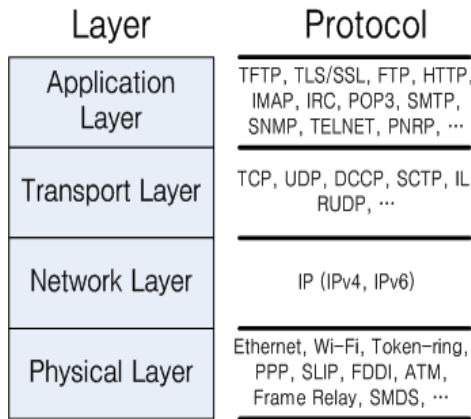


Fig. 1. Internet protocol stack

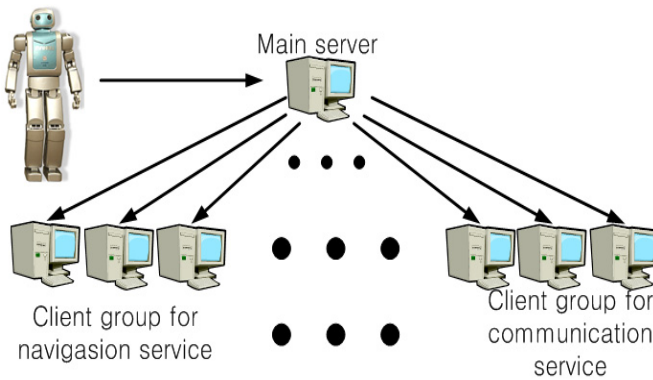
An IP does not guarantee packet forwarding, sending, and receiving. Also, the order of the packets sent and received may differ (unreliable datagram service). The TCP protocol operates over the IP, and the TCP warrants the transfer of the data packet and the order of the data packet. HTTP, FTP, and SMTP, including a large number of IP-based application protocols, run on TCP, which is why we call the protocol TCP/IP.

Figure 1 shows the TCP/IP protocol's stack and related protocols. The hierarchical structure is also shown.

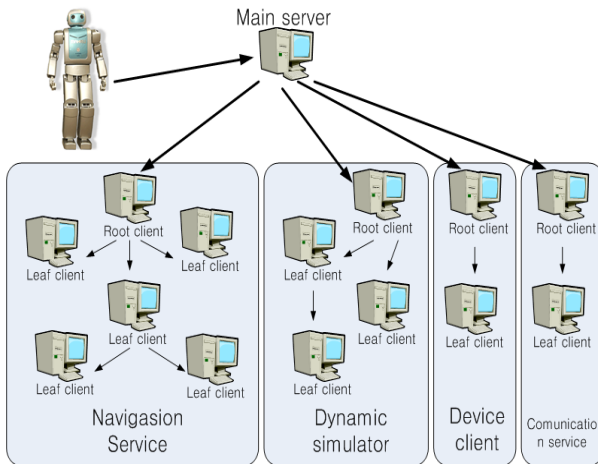
### 2.2 Development of a Network Core Using TCP/IP Protocol

Environment data such as video streams and audio streams from an NBIA are sent to each resource via the main server. Thus, if number of service components increases, then the data throughput of the main server is increased. This situation is shown in Figure 2. Of course there is the multicast protocol, but it is not suitable for sending large amounts of data reliably. In addition, it is difficult to control each resource based on their role, and it is difficult to adjust each resource based on its usage because it has no structural design. Also, the programming complexity of a main server is increased when the main server manages the resources.

To avoid these situations, the main server needs some methods that can classify each service component according to its role.



**Fig. 2.** A case of clients receiving information from the main server



**Fig. 3.** An example of an NBIA system architecture that uses the tree structure



In this section, a structural design such as the tree structure is used to solve the aforementioned problems. Each resource is formed as a branch of a tree structure. Only one client, the root of the tree structure, is connected to the main server. Thus, the root client receives the data transmitted from the main server, and then transmits the received data to other clients. In this way, the burden of the main server can be reduced.

Figure 3 shows an example of an NBIA system with the proposed tree structure. The system in Figure 3 consists of four service groups, an NBIA and a main server. For each service group that consists of multiple clients, only the root client is connected to the main server. Other clients are connected to the root client of the service group. Therefore, the main server manages only four root clients. Furthermore, each client in the service group is formed within the tree structure. The client's location and mission are shown clearly and conceptually in Figure 3. In addition, the main server can control the service group by controlling the root client.

### 2.2.1 Network Core

To develop the network framework described in Sec.2.2, a network core that can be configured into the network with a tree structure was designed. Also, the network core serves as a server and a client simultaneously. In Figure 4, the network core that was designed and implemented is shown. By default, the network core performs the roles of a server and a client. It can also, relay data from an input to multiple outputs. At this point, the kinds of data being entered are unlimited. With these structures, clients were configured with a tree structure on the network. Figure 5 shows an extension of a network core. An output of a network core can be an input of other network cores, and the direction of the data movement is bi-directional. Figure 6 shows an entire system configuration that uses this network core.

As shown in Figure 6, four network cores are connected to the main server. They are an input from an NBIA, a PDA, a client, and a service group that consists of four clients. At this point, the main server transmits an input from the NBIA to three clients, and the three clients that comprise the service group receive data from a root client that is connected to the main server. Thus, the root client that receives data from the main server is responsible for realizing the role of the main server with respect to other clients in the service group. Therefore, using the network core that serves as a server and a client simultaneously, each client will be able to organize itself into a tree structure. This reduces the load on the main server.

### 2.2.2 Basic Function of a Network Core

Table 1 describes the default functions of the proposed network core. A network core has four parts: the sender, receiver, parser, and sweeper.

Figure 7 shows a block diagram of the network core. As shown in Figure 7, the sender transmits data packets to other network cores. The receiver receives the data packet from other network cores. The sweeper assembles the packets from the receiver to the data. The parser is responsible for processing the data. A network core has these four basic functions and management roles.

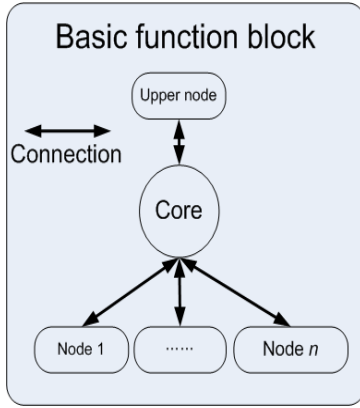


Fig. 4. Basic concept of a network core

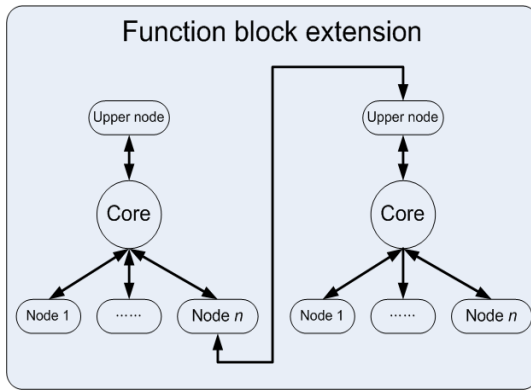


Fig. 5. Extension of a network core

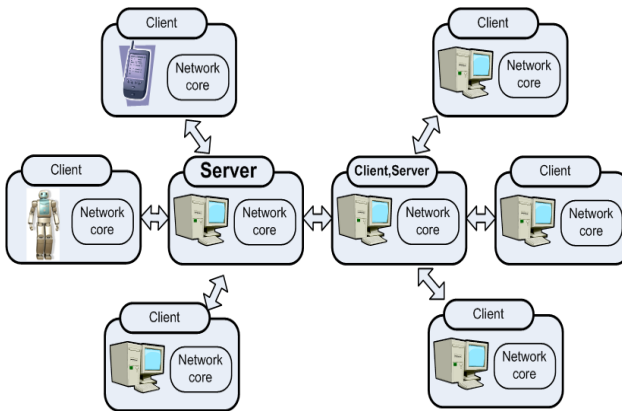
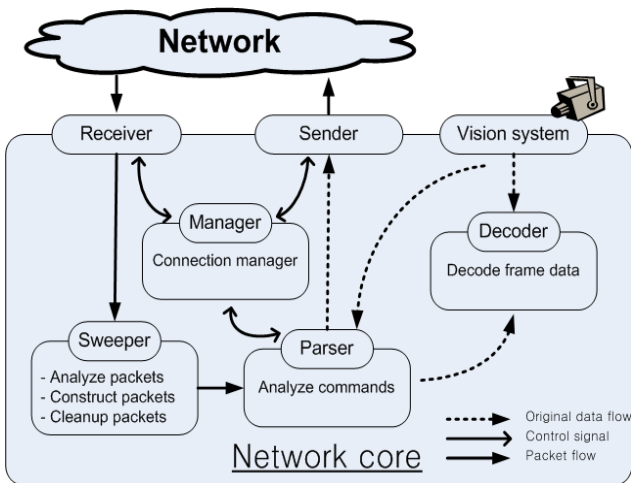


Fig. 6. A system structure that uses a network core

**Table 1.** Default functions of a network core

Functions of a network core
One-input node
$n$ -output node
Server functions
Client functions
Transfer of any data
Bulk data transfer
Core-to-Core communication
Decompression of MPEG4 streams
Compression and decompression of JPEG images
There are no limitations in the tree depth
DLL implemented as easily as possible for use with other applications



**Fig. 7.** Block diagram of a network core

### 3 Performance Analysis and Discussion

Table 2 shows the network transmission delay of a network core. The transfer rate is measured between cores through a wired LAN, a wireless LAN, and a local machine, respectively. It shows also quick network core data transfer. Note that during the data transfer, the network core partitions and assembles the data.

There is little difference in the delay depending on the size of the transmitted data, but the delay in a wireless LAN increases remarkably. Therefore, data compression is necessary to reduce the size of the data when the data is being sent over a wireless LAN.

**Table 2.** Transfer rate using a network core

Size of transferred data	Send/Receive ACK	Local machine	Internal machine	Wireless LAN
<b>Command only (28 bytes)</b>	<b>Send</b>	0.194 ms	0.053ms	0.038ms
	<b>Receive ACK</b>	2.152ms	2.161ms	2.838ms
<b>1000 bytes data with command</b>	<b>Send</b>	0.132ms	0.058ms	0.041ms
	<b>Receive ACK</b>	2.243ms	2.750ms	3.376ms
<b>4000 bytes data with command</b>	<b>Send</b>	0.232ms	0.101ms	0.071ms
	<b>Receive ACK</b>	2.218ms	3.209ms	11.166ms
<b>8000 bytes data with command</b>	<b>Send</b>	0.639ms	0.134ms	18.764ms
	<b>Receive ACK</b>	1.863ms	3.399ms	76.023ms

Using the proposed network core, the NBIA system can be configured quickly and easily, and the addition of new service components becomes very easy and fast.

The network complexity increases rapidly, however, with the addition of new service components and increased connectivity between service components. This phenomenon will occur when the service components and the network link increase. For example, if all the components make links to all the components, the network complexity will increase by  $N^2$ . This increased network complexity makes it difficult to understand the connections between the components. In addition, if a component caused the problem, it becomes hard to find the components and to recover the system. Moreover, the naming scheme that uses a fixed IP and a fixed port number based on the number enables a service component to execute on a specific resource and a specific environment. It will not be able to take full advantage of the distributed resources. To solve these problems, another trial will be needed in future.

## References

1. Schuehler, D.V., Moscola, J., Lockwood, J.: Architecture for a hardware based, TCP/IP content scanning system. High Performance Interconnects, 89–94 (2003)
2. Tang, W., Cherkasova, L., Russell, L., Mutka, M.W.: Customized library of modules for STREAMS-based TCP/IP implementation to support content-aware request processing for Web applications. In: Advanced Issues of E-Commerce and Web-Based Information Systems, WECWIS 2001, pp. 202–211 (2001)

3. Hansen, J.S., Riech, T., Andersen, B., Jul, E.: Dynamic adaptation of network connections in mobile environments. *IEEE Internet Computing* 2(1), 39–48 (1998)
4. Lee, K.B., Schneeman, R.D.: Internet-based distributed measurement and control applications. *IEEE Instrumentation & Measurement Magazine* 2(2), 23–27 (1999)
5. Liao, R.-K., Ji, Y.-F., Li, H.: Optimized Design and Implementation of TCP/IP Software Architecture Based on Embedded System. *Machine Learning and Cybernetics*, 590–594 (2006)
6. Wu, X.: A CORBA-based architecture for integrating distributed and heterogeneous databases. In: *Engineering of Complex Computer Systems, ICECCS 1999*, pp. 143–152 (1999)
7. Dhoutaut, D., Laiymani, D.: A CORBA-based architecture for parallel applications: experimentations with the WZ matrix factorization. In: *2001 Proceedings of the Cluster Computing and the Grid*, pp. 646–651 (2001)
8. Li, B., Lu, Z., Xiao, W., Li, R., Zhang, W., Sarem, M.: An architecture for multidatabase systems based on CORBA and XML. *Database and Expert Systems Applications*, 32–37 (2001)
9. Curto, B., Garcia, F.J., Moreno, V., Gonzalez, J., Moreno, A.: An experience of a CORBA based architecture for computer integrated manufacturing. *Emerging Technologies and Factory Automation* 2, 765–769 (2001)
10. Nishiki, K., Yoshida, K., Oota, M., Ooba, M.: Integrated management architecture based on CORBA. In: *Network Operations and Management Symposium*, pp. 3–15 (2000)

# A Study on the Clustering Scheme for Node Mobility in Mobile Ad-hoc Network

Hyun-Jong Cha<sup>1,\*</sup>, Jin-Mook Kim<sup>2,\*\*</sup>, and Hwnag-Bin Ryou<sup>3</sup>

<sup>1</sup> Defense Acquisition Program, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701

<sup>2</sup> Division of Information Technology Education, Sunmoon University,  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708

<sup>3</sup> Department of Computer Science, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701  
{chj826, ryou}@kw.ac.kr, calf0425@sunmoon.ac.kr

**Abstract.** A mobile ad-hoc network is an autonomous collection of wireless mobile nodes that organizes a temporary network without any network infrastructure. Due to node mobility, it is a challenging task to maintain the network topology. In this paper, we propose a stable clustering algorithm that uses node mobility for cluster formation. In the proposed algorithm, the node mobility is measured by counting the time of nodes entering into leaving from its transmission range. The node having the lowest mobility is selected as a cluster head. For topology maintenance with reduced control overhead, the cluster head adaptively controls the broadcasting period of hello message to the measured node mobility. Through computer simulations, it is verified that the proposed algorithm outperforms previous clustering algorithms in terms of control overhead, the rate of node mobility changes and the number of cluster head changes.

**Keywords:** Mobile ad-hoc network, Cluster, Node Mobility.

## 1 Introduction

In an ad-hoc network (MANET), a temporary communication network is set up with mobile nodes alone without the need for an existing network infrastructure. As such, it is particularly useful for places where a network infrastructure can't be installed. MANETs are applicable in diverse areas such as battlefields, emergency situations, education, and conferences.[1]

However, in a MANET, frequent delivery of control messages and reliable delivery of information are difficult due to changes in the network topology caused by frequent node movement, limited available bandwidth and electric energy.

To overcome these limitations, a clustering scheme can be used, which manages mobile nodes by partitioning them into groups, resulting in the following benefits:

---

\* First author.

\*\* Corresponding author.

efficient use of network resources, ease of mobility management, and reduction in control message overhead. Currently research into applying this clustering scheme is actively underway[2, 3, 4].

This paper addresses the problems of frequent changes to the topology caused by node movement in a 1-hop wireless ad-hoc network. Specifically, a cluster formation scheme is proposed, one that determines node mobility and elects the most stable node as the cluster head, as well as a clustering scheme which actively adjusts the HELLO message period (HP) according to the mobility of each cluster.

## **2 Related Work**

### **2.1 Clustering Scheme**

The most representative clustering schemes for a MANET are LID (Lowest ID Clustering) in which the node with the smallest ID is elected as the cluster head and CDS (Connected Domain Set) which is based on the domain set[5, 6].

However, in these clustering schemes 1-hop sized clusters are formed, and as such there is overhead from frequent cluster reformation. Although there is the 3hBAC (3-hop Between Adjacent Cluster heads) scheme in which 3-hop sized clusters are formed, it has the limitation of having to always form fixed 3-hop sized clusters. Existing 1-hop cluster head election schemes for MANETs include LID (Lowest ID) and HD (Highest Degree)[5, 7].

LID is referred to as an address-based clustering scheme as unique node IDs are used to elect the cluster head. Under this scheme, each of the nodes periodically broadcasts its ID in order to elect the cluster head, and the node with the lowest ID is elected as the cluster head[5].

In HD, the cluster head is elected by taking into account node connectivity, and so it is referred to as a connectivity-based clustering scheme. Each node broadcasts the information of its neighboring nodes at the same interval, and the node with the most number of neighboring nodes (density) becomes the cluster head. If thereafter it is the case that the density of a different node that joins the cluster is greater than the density of the current cluster head, it is chosen as the new cluster head[7].

### **2.2 Measure the Mobility of the Nodes**

The MP-AOMDV protocol is based on the AOMDV protocol. It is a routing technique that selects a more stable route by selecting, among several nodes, one with less movement as an intermediate node using GPS information. As the route can be selected based on the type of data to transmit based on multiple routes, it has the advantage of increasing the transmission rate in the network. However, as movement is predicted using GPS signals, they must be processed, and it doesn't work indoors[8, 9].

During route construction, each of the nodes calculates its forward coordinates using location information obtained via GPS, and infers the next location information. The time it takes to break away from the transmission range of two nodes is

calculated and the smallest value is found among the routes to the destination and set as the route effective time, and many routes are searched.

Afterward, in the route selection stage, the way the route is selected is different according to the type of data. For streaming data that has to be transmitted on a continual basis, the route with the greatest MET (transmission effective time) is selected. For the type of data that are small in size and must be transmitted quickly, the shortest route is selected based on the hopcount value. In the route maintenance stage, the route is reconstructed before getting broken (based on the route effective time), thereby allowing for continuous transmission.

### 3 Stable Clustering Scheme That Takes Mobility into Account

#### 3.1 Cluster Formation

**Broadcasting HELLO Messages.** In MP-AOMDV, mobility is measured at the time of data transmission request through a RREP. The constituent factors of a RREP are brought as they are. That is, separately from a RREP that occurs at the time of the transmission request, a HELLO message is periodically broadcast to neighboring nodes. The HELLO message contains the current location, the next location, and the transmission range, as with the contents added to a RREP packet in MP-AOMDV.

**Determining Node Mobility.** Each node calculates the time before the end of the transmission of the two nodes by calculating the RET (Route Effective Time) in MP-ADMDV through the HELLO message received from the area. The calculated RET is stored in the routing table; specifically the ID of the node and the RET value are stored.

Electing the cluster heads. In the cluster head election stage, cluster heads are elected based on the following procedure.

- ① All nodes broadcast a HELLO message at the initially set interval.
- ② Each node's own RET value is compared to the RET values of all other neighboring nodes to determine whether it is the cluster head or a member node.
  - If the node's own RET value < max. RET value of neighboring nodes then the node = cluster member
  - If the node's own RET value > max. RET value of neighboring nodes then the node = cluster head
  - If the node's own RET value = max. RET value of neighboring nodes then the node with the lowest ID = cluster head

### 4 Experiment

The following tests was done in order to check the performance of the clustering scheme proposed in this paper, which tested the number of times HELLO messages are broadcast as nodes move. In addition, the proposed scheme was compared with existing schemes LID and LIDAR.



Figure 1 shows the total number of HELLO messages sent by all nodes according to the node movement speed. As shown, as in LID cluster mobility is not reflected but a fixed HELLO message interval is used, the number of times a HELLO message is sent is constant. In contrast for LIDAR and the proposed scheme, as the HELLO message interval is adjusted according to mobility within the cluster, it changes as the movement speed of nodes changes.

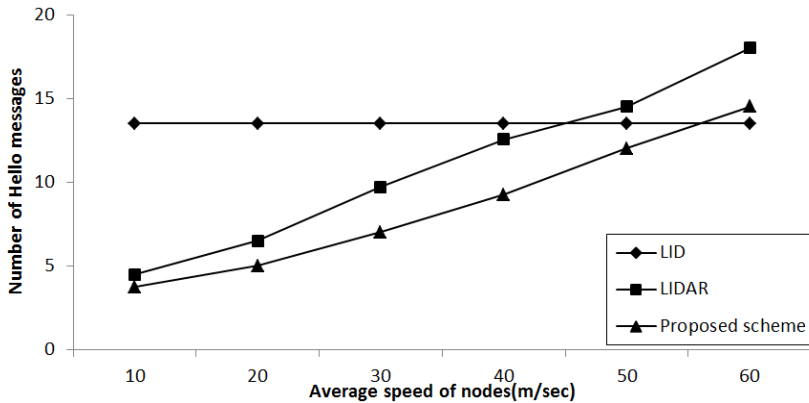


Fig. 1. Total number of HELLO messages as to maximum velocity of node

## 5 Conclusions

This paper proposed a scheme to form stable clusters in a mobile ad-hoc network by using node mobility. With existing clustering schemes, overhead occurs when clusters are reformed because node mobility is not taken into account, and because the size of clusters is small as well as fixed.

In this paper, node mobility is measured and clusters are formed based on cluster heads with low mobility, and as a result overhead from reforming clusters can be reduced, as well as routing overhead internal to the clusters.

For future work, a way of calculating mobility when the movement speed or direction of nodes arbitrarily changes in a mobile ad-hoc network needs to be studied, as well as optimal cluster formation based on this.

## References

1. Hong, X., Xu, K., Gerla, M.: Scalable routing protocols for mobile ad hoc networks. *IEEE Network* 16(4), 11–21 (2002)
2. Chatterjee, M., Sas, S.K., Turqut, D.: An on-demand weighted clustering algorithm (WCA) for Ad Hoc Networks. In: *Proc. of IEEE GLOBECOM*, pp. 1697–1701 (2000)
3. Jiang, M., Li, J., Tay, Y.C.: Cluster Based Routing Protocol (CBRP). Internet Draft draft-ietf-manet-cbrp-spec-01.txt (1999)

4. Wei, D., Chan, H.A.: A Survey on Cluster Schemes in Ad Hoc Wireless Networks. In: Proc. of IEEE Mobility Conference, pp. 1–8 (2005)
5. Ephermides, A., Wieselthier, J.E., Baker, D.J.: A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling. Proc. IEEE 75, 56–73 (1987)
6. Wu, J., Li, H.L.: On Calculating Connected Dominating Set for Efficient Routing in Ad Hoc Wireless Networks. In: Proc. of the 3rd International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications, pp. 7–14 (1999)
7. Yu, Y.J., Chong, P.H.J.: A survey of clustering schemes for mobile Ad Hoc Networks. IEEE Communications Surveys 7(1), 32–48 (2005)
8. Cha, H.-J., Han, I.-S., Ryou, H.-B.: QoS routing mechanism using mobility prediction of node in ad-hoc network. In: ACM MOBIWAC 2008, New York, pp. 53–60 (2008)
9. Perkins, C., Royer, E.: Ad-hoc on-demand distance vector routing. In: IEEE WMCSA 1999, pp. 90–100 (1999)

# Design of User Access Authentication and Authorization System for VoIP Service

Ho-Kyung Yang<sup>1</sup>, Jin-Mook Kim<sup>2\*</sup>, and Hwang-Bin Ryou<sup>3</sup>

<sup>1</sup> Defense Acquisition Program, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701

<sup>2</sup> Division of Information Technology Education, Sunmoon University,  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708

<sup>3</sup> Department of Computer Science, Kwangwoon University,  
#447-1, Wolgye-dong, Nowon-gu, Seoul, Korea, 139701  
{porori2000, ryou}@kw.ac.kr  
cal.f0425@sunmoon.ac.kr

**Abstract.** VoIP, which is used to deliver voice data on the Internet, is being welcomed as a means of replacing the PSTN. In VoIP, voice data are converted to Internet protocol data packets so that they can be delivered in an ordinary IP network. Thus, compared to ordinary telephone networks, it is low cost and highly extensible.

As VoIP services gradually gain traction, issues are coming to the fore, specifically security vulnerabilities and lowered service quality. To mitigate this, in this paper an authentication system is designed which an AA (Attribute Authority) server has added to VoIP in order to increase security and discriminate user access.

**Keywords:** VoIP, Access Control, Authentication System, Authorization System.

## 1 Introduction

When it comes to multimedia techniques, as networking techniques advance, the link with the Internet - which connects the entire world - is accelerating. Demands on services such as video conference and VoIP (Voice over Internet Protocol), which use the same IP (Internet Protocol) network to deliver multimedia data, including audio and video data, are quickly increasing.

Although VoIP efficiently provides voice communication between terminals, for it gain greater use, a variety of services is needed. Examples of additional services include various types, including call transfer, call forwarding when busy or when there is no response, call reservation, call waiting and call filtering. As a signaling protocol of VoIP for users to register services that they want at any time using a simple way, SIP and H.323 in particular are getting the attention.[1]

---

\* Corresponding author.

Although increase in the number of users is expected for VoIP, there can be various problems with the packet network from a security standpoint in that anyone can access it as it is an open network. While a PSTN can be attacked only by physically accessing it, when it comes to a VoIP even remote attackers can easily alter signaling messages or wiretap voice packets. Standardization of the SIP began at the IETF by considering extensibility, component reuse and interoperability as key criteria.

In addition, RFC 3261 recommends the use of a stable security model as a security technique for SIP. SIP provides secure messaging services using digest user authentication, TLS, and S/MIME. Media security is implemented by using SRTP (Secure RTP), which is currently being drafted.

Although using a stable security model can secure security, there is a disadvantage that the quality is drastically reduced for users, making it inconvenient for use.

In this paper a system is designed that addresses security problems caused by the increase in the use of VoIP services and for providing discriminate services according to user access privileges. This paper is organized as follows: Chapter 1 gives the introduction; Chapter 2 is on related research; Chapter 3 describes the proposed technique and system; Chapter 4 implements the system and analyzes its the performance; and Chapter 5 gives the conclusions.

## **2 Related Research**

### **2.1 VoIP**

VoIP is a service that uses the packet network originally designed for data communications for Internet telephony. It is a communication service that converts voice data to Internet protocol data packets so that calls can be made over the ordinary telephone network. Compared to the traditional telephone network service, it is low cost, supports multiple users simultaneously over the cable, and is highly extensible. Some of the protocols used are SIP and H.323 [2][3][4].

### **2.2 Attribute Certificate**

The attribute certificate refers to a type of certificate that plays a special role according to the particular environment rather than the certificate for personal identification as information protection services of various purposes increase in e-commerce. This type of certificate is used only for a specific goal and has a shorter lifespan than certificates used for personal identification. It can be used along with personal identification certificates. It has diverse applications in many fields such as network access control, billing according to access to contents, and web page access control.[4][5][6]

## **3 Proposed Technique**

The following are the prerequisite for the proposed technique. The AA server and KMS server goes through authentication beforehand and know each other's public key

values. The user generates a public key and a private key based on the PKI authentication technique, registers the public key with the KMS server and requests for a certificate to be issued. The KMS server includes the public key value of the ADD server when issuing the certificate.

### 3.1 User Registration Process

This is the process of registering the user before using the service. The register server issues a user certificate and the location server stores this. The register server and the location server are physically at the same location. The user registration algorithm, in which the user is registered with the register server and the certificate issued, is as follows.

### 3.2 Service Operation Structure

Communication using SIP involves going through a call connection process, during which various pieces of information may be leaked such as sender/receiver information, encryption technique and method of communication. Therefore a secure call setup is needed. An authentication server and a KMS server are added based on the SIP call setup in the existing VoIP environment for the authentication process.

The servers authenticate each other beforehand and share their public key values. In the call setup stage the sender first sends a hello message and its certificate to the proxy server, which checks the certificate and sends a response message that messages have to be encrypted.

The user sends an INVITE message by obtaining the public key of the proxy server from the response message for encryption. Before sending the message, its public key is generated based on that public key.

The proxy server sends to the AA server the INVITE message and a public key certificate that includes a random number R and a hash value (R).

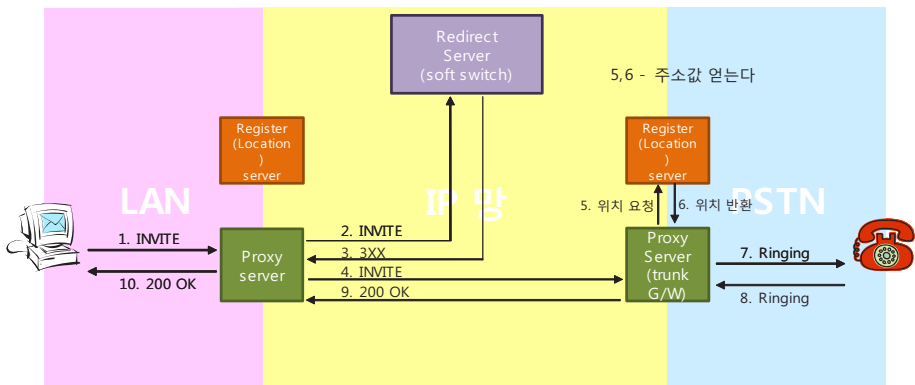


Fig. 1. SIP protocol session setup process

The user is identified using that certificate and the AA server sends the attribute certificate and the contents received from the proxy server. The SMS server receives that information, reviews the contents of the user certificate and the attribute certificate and sends the other party's address value and certificate.

The proxy server encrypts using the public key obtained from the other party's certificate and sends it. The proxy server on the receive side does authentication of the sender at SMS. Also, the sender's attribute certificate is verified at the AA server.

When this process is complete the proxy server sends a message to PSTN, and the telephone network sends the message using bell sounds.

If the process is successfully complete, a response message of "200 OK" is sent to indicate the call has been connected. The sender sends "ACK" to indicate that the message has been received successfully. This completes the call connection.

When secure call setup is complete, data transmission begins with the RTP protocol.

## 4 Implementation and Performance Analysis

### 4.1 Test Environment

For implementation of the proposed system, the RedHat 9.0, gcc V3.2.2 compiler of Linux Kernel V2.4.20.8smp was used.

The library was implemented based on text using a header file called "sip.h", which is freely provided on the Internet. TLS functionality was implemented using the openssl library.

### 4.2 Performance Analysis

This section describes the results of testing each of the following systems based on a test using a VoIP system based on the implemented SIP protocol: an ordinary VoIP system, a TLS-applied system, and the proposed system.

The tests were conducted based on the INVITE command which makes the CALL. It was assumed that account registration has been done.

**Comparison of Security Aspects of Each System.** Figure 2 compares the response times for different number of INVITEs in the VoIP system. For a typical system, the average response time is 5-6ms for 20 INVITE requests.

As shown in the figure, although the proposed system had a longer response time than a typical unsecure system, it was shorter than a TLS system.

**Comparison of Advantages and Disadvantages of Each.** While the ordinary VoIP system has fast response speed and low load on the system, its level of security is poor, and as a result systems with TLS added have become almost a de facto standard. But while these systems have excellent level of security, as a TLS session has to be set up every time a session is set up for each server, response time is slow and there is a lot of load on the system compared to ordinary systems.

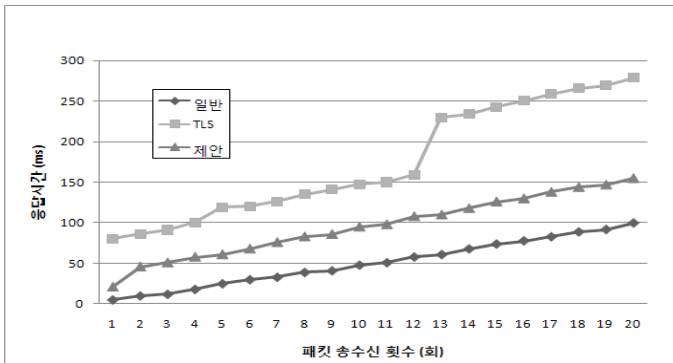


Fig. 2. The Response time with the number of INVITE

For the proposed system, however, an adequate level of security is provided while having fewer loads on the system than TLS-based systems.

As shown in Figure 4.5, the proposed system allows for user-specific access control, billing can be easily set, and various additional services can be provided for each user. A disadvantage is that a stage is needed for setting attributes for each user.

## 5 Conclusions

VoIP services which deliver voice data on the Internet are being welcomed as a means for replacing the PSTN.

As VoIP services gain more traction, problems started to appear in terms of QoS and security. In this paper an authentication system is designed which is made secure and provides differentiated services according to user access. It does this by adding an AA server to the VoIP sessions setup stage. For future work, ways to increase QoS would need to be studied.

## References

1. RFC 2617, HTTP Authentication: Basic and Digest Access Authentication, IETF (1999)
2. RFC 2402, IP Authentication Header, IETF IPsec WG (1998)
3. RFC 2246, The TLS Protocol Version 1.0, IETF TLS WG (1999)
4. 임채훈, "VoIP 시스템에서의 보안기술", (주)퓨처시스템자료
5. RFC 3261, SIP: Session Initiation Protocol (June 2002)
6. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. *J. Mol. Biol.* 147, 195–197 (1981), Session Initiation Protocol (sip) Working Group, <http://www.ietf.org/html.charters/sip-charter.html>

# Approach of Secure Authentication System for Hybrid Cloud Service

Jin-Mook Kim\* and Jeong-Kyung Moon\*\*

Division of Information Technology Education, Sunmoon University  
#100, Galsan-ri, Tangjeong-myeon, Asan-si, ChungNam, Korea, 336708  
{ca1f0425, moonjk}@sunmoon.ac.kr

**Abstract.** Desire for cloud service is very increasing recently. But users are using cloud service actually very few. Because a private cloud service is very expensive and have a many restrict condition. On the other hand, public cloud service, if you can use it than you have to pay a usage fee when you needed external resources that have been published. However, public cloud service is a fear that the privacy of personal information or leaked, put to save the information that requires security. Therefore, the demand for hybrid cloud service has occurred. But, in order to provide a hybrid cloud service, is often difficult to be applied as an authentication system that was used in the cloud service existing. Therefore, I would like to propose a secure authentication system for the hybrid cloud service in this paper. We will secure authentication system for the hybrid cloud service that is provided security, availability, applicability. Our proposed system has a very good capability in suggested condition.

**Keywords:** Hybrid cloud service, Authentication system, Secure authentication, MITM.

## 1 Introduction

Public cloud service is a consideration economical to individuals and businesses that is intended for use with virtualization and (Virtualization) Technology Company that provides a cloud service professional is to build the system resources of software or hardware. It is a service that allows you to provide a service to pay. However, in providing the service, it has a problem that the user associated with the storage of information privacy and security issues. Therefore, public cloud user exists only about 22% currently.

The Private cloud service, if the operation is to allow only inside the company, was also set to live with the problem and cost security, ease of use, companies of about 20 percent is used in the country.

Thus, services that allow you to selectively use only advantage of Private cloud service and Public cloud service is Hybrid cloud service.

---

\* First author.

\*\* Corresponding author.



However, there is no service provider that provides (inter-cloud services) hybrid cloud services in the country.

Inter - To provide the (Inter-cloud) service cloud, between network and cloud service providers, database, security service providers, issues of cost and service models that can agree on is each other, ease of use, security real problem is because too many to solve the problem.

In order to provide a Hybrid cloud service, and accurate understanding of the security threat elements that occur in the public cloud service and private cloud service always, based on this, so that it can provide security services suitable for hybrid cloud services. There is a need to most important security service. an authentication service for the individual objects that make up the cloud services and user authentication.

2-Factor authentication services are provided in the previous studies in this context is proposed. However, this alone, there is a difficulty to provide authentication services that are suitable for cloud computing a rapidly changing environment. Therefore, we would like to propose an authentication system of the new cloud services with improved 2-Factor authentication system proposed in the existing studies.

We can then provide authentication differentiated services for Public cloud service using the 2-Factor existing authentication systems, the Private cloud service, and the authentication system that applies RADIUS (Remote Authentication Dial In User Service) I proposed. New authentication system which we proposed is able to improve the safety, availability and adaptability than conventional systems.

The paper is organized as follows. A related study, we describe cloud services, 2-Factor Authentication service, for RADIUS Section 2. In the third chapter, I'm writing about the new authentication system we provide to see the security threats that can occur in the cloud service existing. In the fourth chapter, we investigated on the basis of new cloud security authentication system that has been proposed, availability, and applicability. It is the conclusion in Section 5 at the end.

## **2 Related Works**

In this paper, we describe the related work of three. We hope we have demonstrated for the cloud services that you have proposed to exist. Describes the 2-Factor Authentication service of the second, at the end, it was written about RADIUS for mutual authentication in a wireless network environment of exist.

### **2.1 Cloud Services**

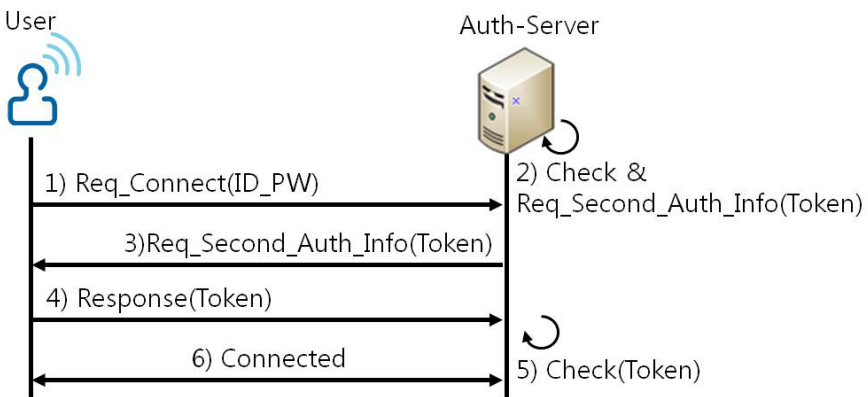
A cloud service is a service so that you can easily use resources of software and hardware of existing, so that you can share with the virtualization technology. Conventionally, by using a cloud service, it is possible to save the purchase costs of resources, use resources effectively to manage and what was used to buy resources directly or software required hardware it has the advantage of being able to.

Cloud services, can be divided into Hybrid cloud service Private cloud service, and Public cloud service, depending on the range to provide resources hardware or software. First Private cloud service is a system that is designed for a company to provide cloud services individually. There is a drawback, which can provide the strongest security, but the implementation cost is very high. Public cloud service is a system that can provide a cloud service published in the second. However, Public cloud service outflow of information of people who use the service, to provide security services have the drawback is difficult. Finally, Hybrid cloud service is a system that can be selectively applied to the advantages of Public cloud service and Private cloud service described above.

### 2.2 Two-Factor Authentication Service

The Authentication service, it is what service requestor have created the identity of their own to the system in advance, to examine the validity of the credentials of their own each time you request the service. The authentication service representative some ID-PW method, Token method, SSO, and PKI. It should be stored in the authentication server before the special user identification information each time the user requests a service, to check service utilization whether the request is valid, these, ID-PW, Token, Certificate it is a service to check and is correct. However, the authentication existing services, there is a point that is insufficient for use in the cloud service.

I think our authentication service in a variety of authentication system described above, suitable for the cloud with 2-Factor authentication service. 2-Factor authentication service is a method of mixing two or more ID-PW existing method, Token scheme, USIM method or MTM scheme can provide authentication services. Figure 1 shows an example of a method that can provide a 2-Factor authentication service.



**Fig. 1.** Example of 2-Factor authentication service (used ID-PW and Token)

As shown in Figure 1, after requesting authentication information sequentially two. It was confirmed by the authentication server, and designed to provide a service requested by the user. First, let's look at the example checks the PW and ID, to check additional Token that had been decided in advance with the authentication server.

### 2.3 RADIUS (Remote Authentication Dial In User Service)

In a wireless network environment, RADIUS is the authentication service to be able to provide authentication services to and from the service provider and the user with a token. The authentication server can generate a token, not only the authentication for the user identification, which was requested by issuing the Token to verify the identity information between the service provider and the user in advance it is intended to be able to provide authentication services for the service. Figure 2 shows for the RADIUS.

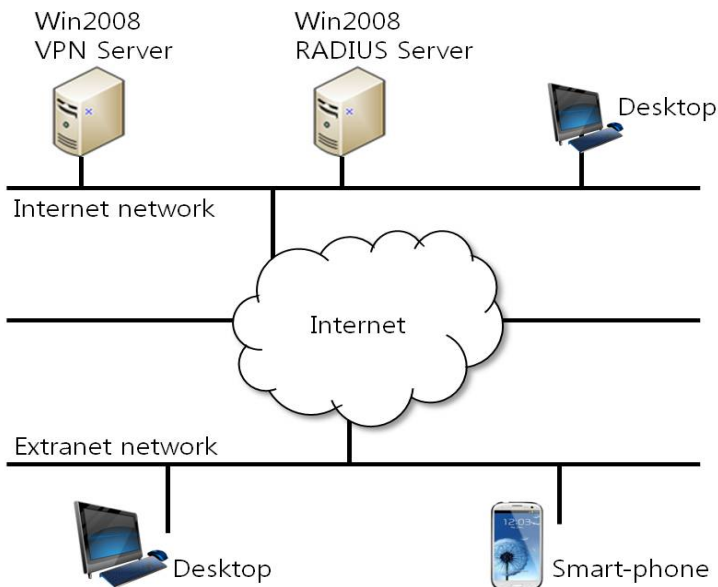


Fig. 2. RADIUS

As shown in Figure 2, smart-phone existing outside network or in order to request a service to connect to a desktop that is present in the internal network via the wireless devices and network environment phone and via the RADIUS server first save pre-authentication information currently is confirmed the authentication. And It connect wireless devices on the external network and desktop internal network through a VPN service.

### 3 Proposal System

#### 3.1 Threats of Hybrid Cloud Service

In order to provide the appropriate authentication system for hybrid cloud service, the correct understanding of the potential security threats that occur in existing cloud services is essential. Threat was expected to occur in hybrid cloud services are as the following list.

- (1) External network, middle browser attacks and man-in-the-middle attack is capable of generating between authentication systems that exist in the internal network. To destroy the availability of authentication system that exists in
- (2) Internal network, the threat of distributed denial of service attack and denial of service attack is large. The wireless device that is connected to
- (3) External network, the movement of the position occurs so frequently, device authentication of the device is difficult. I have a vulnerability to attack by script
- (4) Internal attacker Just the user authentication system
- (5) Existing provide authentication services to the internal network, but the protection of information and authentication of the public cloud services. You have the vulnerability to external network.

As list shown above, also cloud service, vulnerabilities can occur authentication service has on an existing network not only exist, special vulnerabilities characteristic of cloud services is caused by its existence to.

#### 3.2 Proposal System

Authentication system suitable for hybrid cloud services environment, our proposal system has a structure in which five independent components. Two most important components of the five authentication system provided are the RADIUS server and a Hash Machine. Component of the two processes the authentication of service request and status of the user. I have shown in Figure 3.

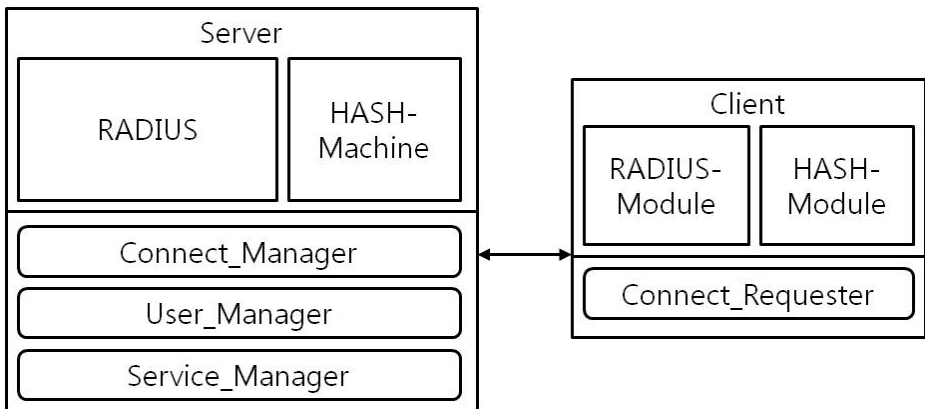


Fig. 3. Architecture of proposal system

Third component is the connection manager. When a user requests a cloud service, which manages the session - connection status - of the authentication service to complete the works. PW is the ID of the connection request user administrator; the user's fourth component is a component for creating users for the required information additional authentication, storage, and or management.

Finally, it is the creation of licensing authority information and services required for the authentication of the user service requirement, storage, service administrator to manage.

### 3.3 Procedures of Our Proposal System

Authentication system for hybrid cloud services which we proposed in the paper performs the processing procedure is divided into two stages increases. I showed in figure 4.

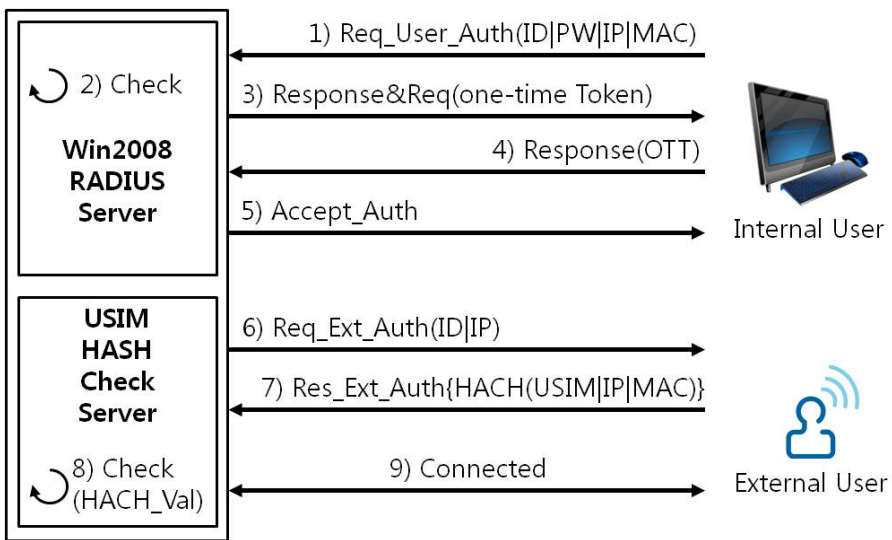


Fig. 4. Procedure of proposal system

The first process, even in the internal or external users, from step 1 to step 1, we describes the process of processing the service request in a wired environment. This was able to perform user authentication using the Token for temporary use with the ID-PW information and provide services.

- (1) Users to send the {ID-PW|IP|MAC} to the authentication server for request to user authentication.
- (2) RADIUS server checks the ID-PW value sent by the user.
- (3) User authentication request is valid, the authentication server, requesting to send a temporary token that responds that the user authentication request is valid, promised in advance.

- (4) The user generates and transmits a token to be used temporarily in the authentication server.  
 Make sure the temporary token sent by the user is correct.
- (5) If it is correct, authentication server performs user authentication.

The second process of treatment is from step 6 to step 9. In this procedures, we shows the process when it requests a service using a wireless device. In order to check the user service requirement whether valid.

- (6) The server is required to verify the IP address and the user ID. USIM stored in the wireless device to the service request, IP.
- (7) The user responds to the value that is treated by the hash function that is promised in advance MAC.
- (8) Authentication server to verify the hash value that the user has transmitted. If the service request, user is legitimate.
- (9) System provides the appropriate services to the user.

## **4 Appraisalment of Proposed System**

### **4.1 Appraisalment of Security**

Authentication system of hybrid cloud proposed is a safe middle browser attacks and man-in-the-middle attacks. we in this paper. Between the authentication server and the user, by a malicious user receives the token temporarily or ID-PW, the user authentication system that we provide we, temporary token also perform man-in-the-middle attack since it is not possible to know the value of the secret used to encrypt the value of the token and method as promised in advance to generate, it is not possible to perform intermediate browser or attack middle attacks.

### **4.2 Appraisalment of Availability**

Our proposed hybrid cloud services to provide authentication system, that allocated before the firewall have advance against of DoS(denial of service attack) or DDoS(distributed denial of service attacks). And it cannot harm to our system placed in the DMZ area by the availability.

### **4.3 Appraisalment of Applicability**

In this paper, we propose a hybrid authentication system for cloud service that consist to private cloud service and public cloud service by complementary relationship. And we configure the network environment for hybrid cloud service in the DMZ area. By doing so, we offer cloud services to existing authentication systems, while maintaining the same applicability can be improved.

## 5 Conclusion

Virtualization Technology requires a computer system with an existing one cloud service to connect a lot of resources to improve the ability to provide services to users, is a technique that can be expected. But a private cloud service to connect with many dedicated resources to provide users high cost because it has the disadvantage that occurs. In contrast, public cloud service is installed on the external network to the existing hardware resources by connecting technology to provide services to public safety has a problem because it is.

In this paper, Hybrid cloud service to receive the most attention in the future was expected. Hybrid cloud service, but for the certification system and related services are until now not sleep.

We the above environmental constraints Hybrid cloud service was proposed for the authentication system. Our proposed Hybrid cloud service authentication system for 2-Factor authentication service with existing RADIUS service who presented in ways designed to take advantage of.

Our proposed Hybrid cloud service authentication systems for security, availability, in terms of applicability as compared to existing methods are expected to be excellent. In this paper, a hybrid cloud services, but restrictive for all matters were not considered. In future studies, in order to provide hybrid cloud services, which may occur for a variety of constraints further review of the existing authentication system to accept the lack of information about the in-depth research studies continue to perform better plans.

## References

1. Kim, T.H., Kim, I.H., Min, C.W., Yeom, Y.I.: Security Technical Trend of Cloud Computing. *Computer Science Magazine* 30(1), 30–38 (2012)
2. Jansen, W., Grance, T.: *Tuidelines on Security and Privacy in Public Cloud Computing* (2011)
3. Mannan, M., Kim, B.H., Ganjali, A., Lie, D.: Unicorn: Two-factor Attestation for data Security. In: *Proc. Of the 18th ACM Conference on Computer and Communications Security* (2011)
4. Bang, Y.H., Jeong, S.J., Hwang, S.M.: Security Requirement Development Tools of Mobile Cloud System. *Information and Communications Magazine* 28(10), 19–29 (2011)
5. Hubbard, D., Sutton, M.: *Top Threats to Cloud Computing V1.0*. Cloud Security Alliance (March 2010)
6. Liao, I.-E., Lee, C.C., Hwang, M.S.: A password authentication scheme over insecure networks. *J. Comput. System Sci.* 72(4), 727–740 (2006)

# Intelligent Inference System for Smart Electronic Acupuncture

You-Sik Hong<sup>1</sup>, Chang-Hoon Choi<sup>2\*</sup> and Baek-Ki Kim<sup>3</sup>

<sup>1</sup>Dept. of Computer Science, Sangji University, Korea

<sup>2</sup>School of Computer Information, Kyungpook National University, Korea

<sup>3</sup>Dept. of Information & Telecommunication Eng.,

Gangneung-Wonju National University, Korea

h5674korea@gmail.com, hoon@knu.ac.kr, bkkim@gwnu.ac.kr

**Abstract.** In this paper, we proposed the system that diagnoses a patient optimally considering the patient's condition using intelligent fuzzy technique. We designed the system to respond to the various patterns and to sense the situation which potential difference is changed according to the patient's painful part simultaneously. It contains the function that a patient can search the exact point of electronic acupuncture and check on optimal strength and time of electronic acupuncture considering the patient's body conditions. The system includes the hardware to provide protection function for safety and to support the multimode function of electronic acupuncture through change of control mode.

**Keywords:** inference, fuzzy rules, acupuncture, diagnose.

## 1 Introduction

Oriental doctors have considered pulse rates as important data in diagnosis. But the existing blood pressure pulse analyzer has some problem. It is difficult to standardize the pulse exactly because thickness of their blood vessels is different even if the thickness of two person's forearm is equal. And it is uncertain whether the blood pressure pulse analyzing sensor is located precisely on the radial artery. The analogue type blood pressure pulse analyzer has problems with quantification of the blood pressure pulse.[1] Therefore there is no set of data that is considered reliable enough to judge the accuracy of blood pressure pulse rates. In order to gain an accurate diagnosis, oriental doctors consider the patient's pulse by the basic biological signals such as checking the pulse's size, strength, and speed, and also the basic and quantitative analysis of the pulse. And the doctor should consider physical characteristics, such as the thickness of the skin and blood vessels, in order to reach an accurate conclusion.[2],[3] But the method of exiting diagnosis has problem which cannot diagnose the old and the infirm exactly because it does not take into consideration the condition of patient's gender, age, skin. Most of the conventional

---

\* Corresponding author.



electronic acupuncture systems are made by using low frequency and the rest are made by using momentary electro-stimulation. They can't treat the patients effectively because it uses uncertain and vague frequency. Furthermore, it can't find acupuncture points because it has no consideration for patient's sex, age, weight, illness, etc. And it causes problem that children and elderly people are bruised or wounded after getting electronic acupuncture because of inappropriate acupuncture time and strength. In this paper, to solve these problems, we proposed the algorithm that diagnoses patient optimally considering patient's condition using intelligent fuzzy technique. We analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We classified the patients by their body, illness and age, and calculated the exact time of electronic acupuncture suitable for patient's physical condition using fuzzy logic and inference. The composition of this paper is as follows. Section 2 is about disease inference algorithm, section 3 is about the simulation of the system, and finally section 4 concludes.

## 2 Disease Inference Algorithm

If human is taken with a disease, the electric resistance of the diseased part is higher than around part. The inherent current of human body does not flow well in the diseased part due to high electric resistance. Small current flows in the diseased part, as a result, absolute current of cell is decreased. In other words, if inflammatory reaction, various disease and cancer occur in the human body, then pain, part fever, edema and seizure are appeared. If human body is injured and got an infection in skin and bodily tissue, muscles are contracted to protect him. For instance, the reason of occurrence of pimple, atopic dermatitis and lentigo is that the electric resistance of these parts is high, and the parts become to status of nonconductor because of oxygen deprivation. So skin disorder appears, and gets an infection in the skin or skin tissue is dead. In these cases, if the patient gets electronic acupuncture in the diseased part for 1 minute, his blood circulation is promoted by penetrating blocked aeremia and supplying bioelectricity. The fine current from electronic acupuncture(13~500  $\mu\text{A}$ ) strengthen the ATP five times, activate the tissue cell and go on smoothly metabolism. In the majority of cases, the intensity of blood pressure which shows the dynamics of blood flow from heart is measured by sensor pad attached to the heart. In oriental medicine case, it is measured by sensors attached to the arteries in the wrist.

DSP board presented in this paper for implementation of intelligent pulse diagnosis system is designed to respond to the various patterns and to sense the situation which potential difference is changed according to patient's painful part simultaneously.[4]

In this paper, if result H is not 100 percent in spite of evidence E in the rule; IF E THEN H, we can express this rule with conditional probability  $P(H|E)$  using methodology of probability. And this conditional probability can be found using Bayesian theory.[5],[6]

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^n P(E|H_j)P(H_j)}$$

Although this Bayesian theory is very clear theoretically, many problems can be occurred in case of application of the real issues. First, to know the conditional probability  $P(H_i|E)$ , we have to know  $P(H_i)$  and conditional probability  $P(E|H_j)$ . For example, let us suppose that  $E$  is symptom of patient’s body and  $H_i$  is disease, to know the probability  $P(H_i|E)$  of a certain disease  $H_i$ , a prior probability of each disease  $P(H_j)$  and probability  $P(E|H_j)$  have to be given. But there are frequent occasions that the data of these cases is not enough. Second, in the equation at above, each disease  $H_i$  is has to mutually exclusive:  $P(H_i \cap H_j) = 0$ , but this assumption cannot be satisfied because three types of disease can be occurred in any case of patients simultaneously. Suppose the probability of disease  $H$  is 0.7 when three symptoms  $E_1, E_2, E_3$  are all true. This conditional probability is summarized as follows.  $P(H|E_1 \cap E_2 \cap E_3) = 0.7$ . Hear 0.7 is subjective value of probability. Be that as it may, it does not mean that the probability of not occurrence of disease  $H$  is 0.3 when the three symptoms are all true as below.

$P(H|E_1 \cap E_2 \cap E_3) = 0.3$ . The probability 0.3 is calculated based on the axiom of probability,  $P(H|E) + P(\bar{H}|E) = 1$ . Therefore, this means that 0.7 is not the value of probability. The reason why the above axiom is not true is that 0.7 is sure of disease  $H$ , but 0.3 does not mean the disease  $H$ . This means that trust and distrust are treated separately. 0.3 is just a unknown part. The meaning of ‘Do not know’ or ‘ignore’ is different from ‘refute’. According to the axiom of probability, probability of disease  $H$  is assured as 0.7 and not disease as 0.3. If trust and distrust are coexisted together, it would be rather to lower the strength of trust. Namely the trust must be reduced as 0.4. If the probability is remained as 0.7, the rest probability 0.3 is not considered as distrust but unimaginable of unknown area. Fig.1 shows correction factor using fuzzy rules.[7]

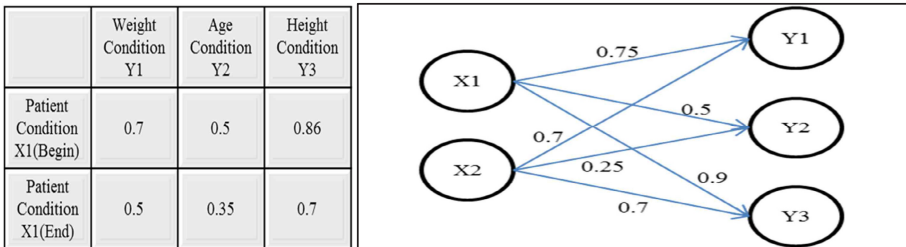


Fig. 1. Correction factor using fuzzy rules

Degree of belief represents degree of confirmation and it is expressed as difference of belief and disbelief as below.

$$CF(H,E) = MB(H,E) - MD(H,E)$$

Here CF means degree of belief about hypothesis H when the evidence E is given, MB means measure of increased belief about H caused by E and MD means measure of increased disbelief about H caused by E. CF means net increase of trust caused by given evidence.  $CF > 0$  means  $MB - MD > 0$  and a given evidence increases trust of hypothesis,  $CF = 1$  means that the hypothesis is proved clearly by evidence. There is two cases when  $CF = 0$ , in case of  $MB = MD = 0$  nothing can be trusted, in case of  $MB = MD > 0$  trust is offset by distrust.  $CF < 0$  means increase of trust that hypothesis is negation.  $CF = 0.7$  means trust 70 percent greater than distrust. The difference between MB and MD is important, not the each value

### 3 Acupuncture System Using Intelligence

In this paper, we use the intelligent algorithm for pulse diagnosis as follows. In this paper, it tried to solve these problems using intelligent fuzzy rules.

$$e = R - Y$$

$$Ce = e_2 - e_1$$

Where, Y: optimum pulse feeling judgment

R: Criteria Input  
 e: Error  
 Ce: Error Displacement  
 e<sub>2</sub>: Current Error

In this paper, in order to solve this kind of problem, it uses compositional inference while using the fuzzy rule. Fuzzy compositional rule of inference is applied to come up with a calibrating constant in order to derive an accurate result (considering the patient's physical condition) in analyzing the blood pressure pulse. In existing method, an oriental doctor infers one pulse wave out of 28 pulse wave and diagnoses the patient. It is difficult to know whether pulse detection sensor is located in the radial artery exactly or not by using existing pulse checker. And in the case of different body type and the thickness of a forearm, it is difficult to take pulse exactly. And also It is difficult to standardize the pulse with analog pulse checker. For example, even if the thickness of two person's forearm is equal, it is difficult to standardize pulse exactly because thickness of their blood vessels is different.

In this paper, we used TMS320VC33(TI) as main DSP of electronic acupuncture system, HY29F040 as flash ROM to store OP code, and 512K×32 capacity as Main operating RAM. The signal measured in sensor passes isolation amplifier(AD202) through primary filtering after being amplified at AMP stage. This is necessary and to prevent the fatal electrical accident occurred in body. The system was designed to control the changes of power supply such as voltage range of 10uV~1V and current range of 100uA~10mA using D/A and FET to experiment various patterns continuously base on the point of view that an electrical conductivity is varied with the body characteristics and patient's affected part. The system includes the hardware to provide protection function for safety and to support the multimode function of electronic acupuncture through change of control mode.

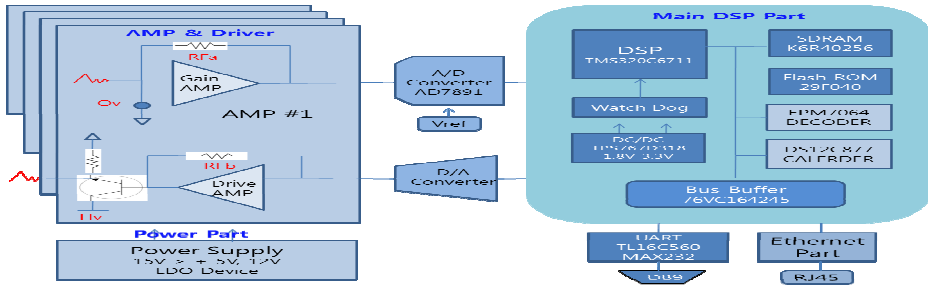
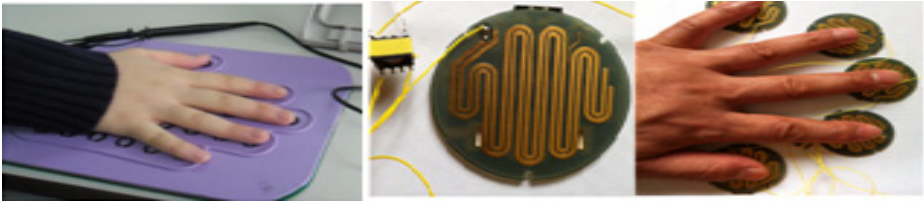


Fig. 2. Circuit diagram of acupuncture system

Fig.2 shows circuits of the electronic acupuncture system. If random order signal of selected function is entered control part, the signal is transformed and performs the output order corresponding to the signal data. Although pulse wave occurred in the human body is not fast in comparison with brainwave, the system requires fast A/D converter and DSP board with floating point arithmetic operation for real-time analysis, secondary differentiation and fuzzy relearning according to shape change of new pattern. But D/A converter does not need to be fast or accurate in comparison with A/D converter. For personal security this part also requires isolation of power supply, for this requirement, serial type D/A with 4 signals (CS, SCL, SDI, SDO) and photo coupler to isolate signal are added. The system was designed to control all data created in DSP board through RS232C or USB port and to accept result data in real-time. Dedicated serial controller was built in the system to support RS232C, communication speed is in the range of 1,200BPS to 115,000BPS. Also the system support USB2.0.

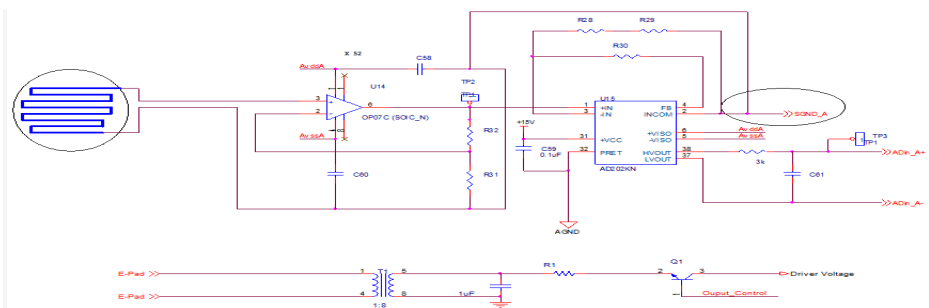
#### 4 E-Acupuncture Pad with Built-In Multi-Active

What is a multi-pad with a built-active JEUNJACHIM (Electronic-Acupuncture) depending on a patient's current body status? Based on this information, the patient meets the voltage and current self-oscillation. The frequency with the ability to automatically advanced procedure is called JEUNJACHIM. In order to perform these functions simultaneously with the sensing of JEUNJACHIM, one is required to possess the ability to perform surgery, derive accurate analysis from fuzzy logic and process statistical data. Electrical resistance of the body including long-term resistance, internal resistance and the surface can be divided into exposed skin. Resistance of the body when the DC voltage is based on the pure resistive component can be considered only based on the basis, when the impedance of the AC voltage should be considered. That body electrical conductors if you think skin, blood, muscles and other body each part of the voltage and current for the resistive component and capacity components are separated by impedance and its size, the electrical conduction path, the contact voltage, the contact area, and energizing time, is applied differently depending on the frequency may occur.



**Fig. 3.** Multipad with a built in electronic acupuncture

Figure 3 illustrates the basic theory of Electronic-Acupuncture. In addition, these changes in a person's age, gender, humidity, temperature, weight and fat accumulation is based on the changes. The requirements when considering the electrical resistance of human skin in general is based on the amount of approximately  $2500\Omega$ . However, the same voltage and current is applied even if the amount of contact area and pain change in resistance over time are different. In the electrical resistance of human body tissues, regardless of the DC and AC power is almost constantly appear if time longer JUAL heat due to temperature rise of tissue resistance is slightly reduced. When the electricity in the human body typically conduct a minimum of power to feel the flow of the AC voltage is  $1\text{mA} \sim 2\text{mA}$  for men. In contrast, direct the flow of power is smaller than the stimulus at least five double-road sensing current flow caused by the voltage applied, even though I do not feel the flow of electricity. Thus, in the treatment of **JEUNJACHIM**, electricity is AC rather than DC voltage with the voltage of the aneurysm and the frequency and voltage, over current change as a real hand acupuncture procedures, a small battery that has the same effect as a treatment is likely to be seen. In the experiment, according to AC current that can safely come off as self a man  $16\text{mA}$  ( $60\text{Hz}$ ) women  $10.5\text{mA}$  ( $60\text{Hz}$ ) is about the human body can withstand DC current is approximately  $74\text{mA}$  men for women is approximately  $50\text{mA}$ . But it also including a person's body size and weight may appear slightly different depending on the requirements. In this paper, the voltage between  $15\text{V} \sim 50\text{V}$  AC voltage to the change of  $5\text{Hz} \sim 1.2\text{KHz}$  and current  $500\text{uA} \sim 1500\text{uA}$  given in the current experiments were carried out. Figure 4 illustrates the electronic acupuncture circuit.



**Fig. 4.** Electronic acupuncture circuit

## 5 Conclusion

In this paper, we proposed the system that diagnoses a patient optimally considering the patient's condition using intelligent fuzzy technique. We analyzed the fine distinction considering thickness of skin and blood vessels and pulse, weather big or small, strong or weak and fast or slow. We classified the patients by their body, illness and age, and calculated the exact time of electronic acupuncture suitable for the patient's physical condition using fuzzy logic and inference.

We designed the system to respond to the various patterns and to sense the situation which potential difference is changed according to the patient's painful part simultaneously. It contains the function that a patient can search the exact point of electronic acupuncture and check on optimal strength and time of electronic acupuncture considering the patient's body conditions. The patients can be treated with optimally, and acupuncture time can be shortened and strength of acupuncture can be moderated considering their condition. This system is useful for remote medical examination and treatment.

**Acknowledgement.** This research was supported by Kyungpook National University Research Fund, 2012.

## References

- [1] Haider, A.W., Larson, M.G., Franklin, S.S., Levy, D.: Systolic Pressure, Diastolic Blood Pressure, and Pulse Pressure as Predictors of Risk for Congestive Heart Failure in the Framingham Heart Study. *Ann. Inter. Medi.* 138, 10-17 (2006)
- [2] Shaltis, P.A., Reiser, A.T., Asada, H.H.: Cuffless Blood Pressure Monitoring Using Hydrostatic Pressure Changes. *IEEE Trans. Biomed. Eng.* 55, 1775-1777 (2008)
- [3] Lee, Y.J., Lee, J., Lee, H.J., Yoo, H.H., Choi, E.J., Kim, J.Y.: Study on the characteristics of blood vessel pulse area using ultrasonic. *Korea Institute of Oriental Medicine Researches* 13(3), 111-119 (2007)
- [4] Hong, Y.S., Kim, H.K., Kim, B.K.: Electronic Acupuncture system with built-in Multipad using intelligence. In: *Proc. Of The 2012 Advanced Information Technology and Sensor Application*, p. 162 (2012)
- [5] Sik, H.Y., Kug, P.C.: *Proc. of the Sixth International Fuzzy System Association, IFSA*, pp. 461-464 (1995)
- [6] Garg, M.L., Ahson, S.I., Gupta, D.V.: A Fuzzy Petri-nets for Knowledge Representation and Reasoning. *Information Processing Letters* 39, 165-171 (1992)
- [7] Hong, Y.S., Kim, H.K., Kim, B.K.: Implementation of Adaptive Electronic Acupuncture System using Intelligent Diagnosis System. *International Journal of Control and Automation* 5(3), 141-152 (2012)
- [8] Leung, K.S., Lam, W.: Fuzzy Concepts in Expert Systems. *IEEE Computer*, 43-56 (September 1988), [10] Looney, G.C., Alfize, A.: Logical Controls via Boolean Rule Matrix Transformation. *IEEE Trans. on SMC* 17(6), 1077-1082 (November/December 1987)

- [9] Looney, G.C.: Fuzzy Petri Nets for Rule- based Decision Making. IEEE Trans. on SMC 18(1) (January/February 1988), [11] O'Rourke, M.F., Kelly, R.P., Avolio, A.P.: The Arterial Pulse, 1st edn. Lea & Febiger, Philadelphia (1992)
- [10] Gong, Y., Chen, H., Pu, J., Lian, Y., Chen, S.: Quantitative investigation on normal pathological tongue shape and correlation analysis between hypertension and syndrome. China Journal of Traditional Chinese Medicine and Pharmacy (2005)

# Electric Braking Control System to Secure Braking Force in the Wide Speed Range of Traction Motor

Young-Choon Kim<sup>1</sup>, Moon-Taek Cho<sup>2,\*</sup>, and Ok-Hwan Kim<sup>1</sup>

<sup>1</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Buda-dong, Cheonan-si, Chungnam, 330-717, Korea  
{yckim59, owkim}@kongju.ac.kr

<sup>2</sup> Dept. of Electrical & Electric Engineering, Daewon University College,  
316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

**Abstract.** In this paper, a vehicle stopping method using an electric brake until a traction motor is stopped is studied. At the moment of vehicle stop, electric brake is changed to control mode wherein torque is reduced at a low speed. Gradient is controlled by estimating the load torque of motor thereby traction motor is not rotated after stop. In addition, coasting operation and brake test were performed from normal-opposite operation and start using a small-scale model comprising the inertial load equipment and the power converter. Further, traction motor was made to be equipped with a suspension torque. Pure electric braking that makes traction motor stopped by an air brake at the time of stop was also implemented. Constant torque range and constant power range were expanded during braking so that braking force was secured with the electric brakes even in high speed region. Therefore, vehicle reduction effect could be expected by reducing parts related with an air brake which is not used frequently by using a pure electric brake in the M car in wide speed region. Further, maintenance of brake system could be reduced, Besides ride comfort of passenger in the electric rail car, energy efficiency improvement, and noise reduction effect could be additionally expected.

**Keywords:** Electric brake, Traction motor, Constant torque.

## 1 Introduction

Brake technology in high-speed region is executed in parallel with air brake that supplements electric braking force. Still, air brake use frequency has to be minimized to improve performance of vehicles. Since air brake in the electric rail car is basically the equipment which applies mechanical friction, the noise and dust generated in the process of brake are caused to reduce the performance of vehicle.[1],[2]

In the braking system of electric rail car, braking is extended in the high-speed region, thus operating range is expanded provided that traction motor secures insulation that withstands overvoltage to secure electric braking power in a high-speed region.

---

\* Corresponding author.

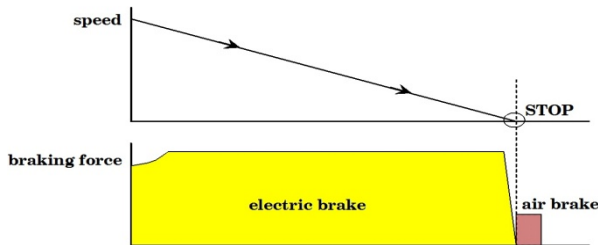


Therefore, in this study, the control method that secures brake force of electric rail car in a wide speed range where traction motor is driven with an assumption that a device which absorb regenerative power is installed in electric rail cars. By using electric brake until securing braking force and stop the vehicles at high-speed region, air brake which was essential due to shortage of existing electric brake needs not be used.[2],[3]

The possibility of securing brake force having a constant power region was tested in the braking test for the small-scaled test system. Also, pure electric brake for a wide speed range of traction motor was realized by using an auxiliary power converter which sends regenerative power equivalent to regenerative motion of main power converter for drive and terminal voltage increases due to expansion of constant power drive to the DC bus.

## 2 Method of Electric Brake

The electric brake method suggested in this study is presented in fig. 1. In high-speed region, drive range was expanded and drive at constant torque control region was expanded to secure electric brake force to minimize deceleration changes as much as possible. At stop mode, suspension torque was given to prevent driving. After stopping electric rail cars, it is made to be stopped by air brake.



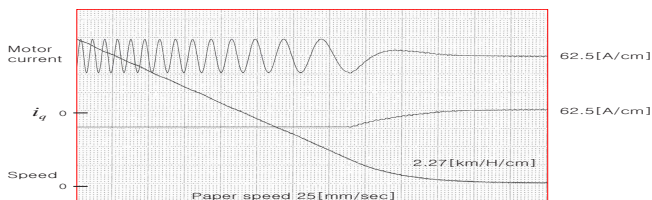
**Fig. 1.** Electric brake method

In this study, a drive motor is comprised of permanent magnet type synchronous motor (PMSM) in the device comprised of motor for inertial mass and load. Drive motor was stopped by electrical brake till stop by vector control was used. Electric braking was executed until stopping the drive motor. Stop control was executed by position detection of rotor by resolver and the mode of assuming the speed and load torque was used.

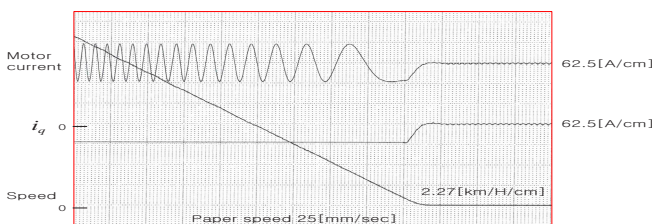
A brake test was conducted by the proposed method after accelerating drive motor and then driving it at constant speed. Test equipment was a small-scale model of direct drive system. When rotational speed is 824[rpm], the speed of the electric rail car is equivalent to 120[km/h].

The gain of velocity feedback and time constant of filter are related with the time which reduces braking torque at the moment of the vehicle stop. During this time, since the deceleration ratio is changed, it considerably affects on the ride comfort. Therefore, it is desirable to increase feedback of speed. Fig. 2 and Fig. 3 show the

measurement results of stopping motion according to velocity feedback and time constant of filter. With increase in the time constant of filter, the oscillation of current was reduced and stable drive was observed. Also, the larger the velocity feedback, the shorter was the duration of torque reduction, i.e., deceleration change duration became shortened.



**Fig. 2.** Velocity feedback: 50, filter time constant: 0.0384[s]



**Fig. 3.** Velocity feedback: 250, filter time constant: 0.0384[s]

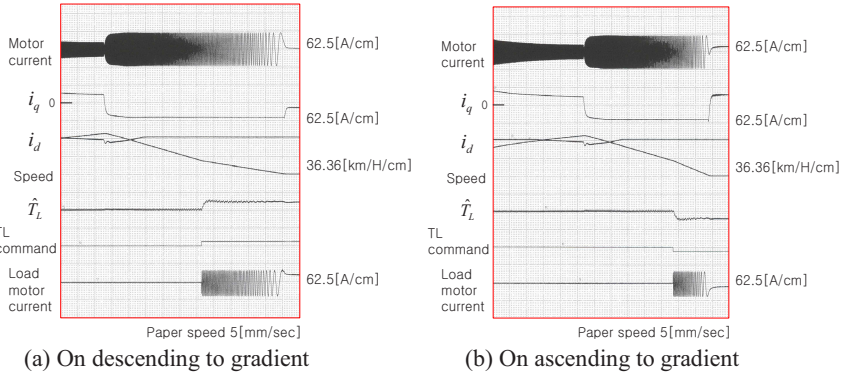
In case a load torque is existed, gradient of transfer at stop control acts as load torque. Fig. 4 shows the measurement result of brake after acceleration till 80[km/h]. It shows the brake condition under positive(+) torque and negative (-) torque of a loaded motor. These are corresponding to up motion and down-motion of electric rail cars on the gradient.

In Fig. 4, load torque is assumed even after a vehicle stops. The suspension torque is generated under a gradient condition. Therefore, a pure electric brake is executed to remove the electric brake after stopping by the air brake. This kind of vehicle can be environment-friendly by reducing noise and dust and can improve performance of electric rail car.

### 3 Securing the Braking Force and Test Result

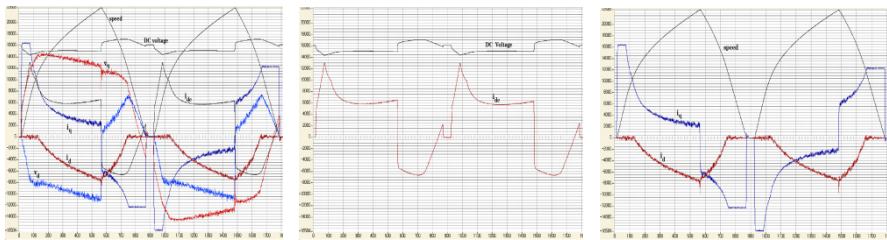
Because voltage of the inverter can be secured with the brake of motor and therefore brake of constant torque in total interval can be achieved, motor should be driven by limiting voltage to be saturated under any voltage conditions. During brake, increases in the brake force for the trolley voltage, limit in constant power driving by flux driving, and magnifying method of brake pattern by inserting series resistance were proposed.

Fig. 5 shows the case of characteristics driving by constant power driving. Driving speed was 65[km/h] with the braking characteristics. It shows driving from the maximum speed 120[km/h]. Voltage variation characteristics of transfer also affects on the braking characteristics. The higher the voltage variation, the higher is the brake force.



**Fig. 4.** Response of load torque estimator (velocity feedback): 150, filter time constant: 0.084[s] (Gain of estimator  $k_i = 58.6, k_p = 62$  of estimator)

Fig. 5 shows the case of characteristics driving by constant power driving. Driving speed was 65[km/h] with the braking characteristics. It shows driving from the maximum speed 120[km/h]. Voltage variation characteristics of transfer also affects on the braking characteristics. The higher the voltage variation, the higher is the brake force.



Condition: ratio: 0x3500, Characteristics drive conversion point: 0x23c0 65[km/H], transfer conductance: 0.1[Ω]

**Fig. 5.** Transfer drop and constant power brake

Since resistance is made short according to speed, brake force is secured in high-speed region. On the contrary, when speed becomes reduced, resistance becomes short which improves efficiency. The series resistance expands constant drive region, thus it is possible to secure brake force for the wide speed range.

Constant torque brake is regarded as ideal for wide speed range. Fig. 6 shows the vector diagram wherein an armature resistance of motor is ignored. Here, point A indicates the starting point of constant torque brake using a resistance drop, while point B indicates the case where all the resistances were short. Resistance drop can be estimated using Eq. (1).

$$Ri_q = k\omega\phi - \sqrt{v_{\max}^2 - (\omega Li_q)^2} \tag{1}$$

At point B where all the series resistance becomes short, the resistance drop should be 0. Therefore, it is when Eq. (1) becomes  $\omega$  that is 0 and this is the maximum point of constant torque driving. Meanwhile, the regenerative power of motor becomes Eq. (2).

$$k\omega\phi i_q = (Ri_q + \sqrt{v_{\max}^2 - (\omega Li_q)^2}) i_q \tag{2}$$

Within the parentheses bracket in Eq. (2), the first term is the resistance drop and the second term is the component which is in-phase with induced electromotive force in inverter voltage. Since resistance drop has a constant current, it brings voltage drop in proportion with resistance, thus it has a constant size regardless of speed.

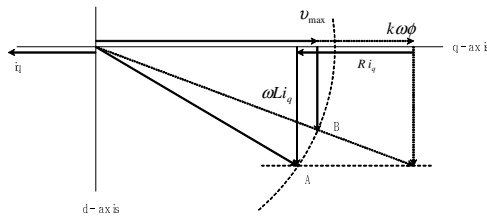


Fig. 6. Vector diagram during brake

## 4 Conclusion

The electric brake is a method used to stop the vehicle by changing torque at low speed to control the mode at the moment of vehicle stop. The load torque for the gradient was assumed so as vehicle location is not rotated after stopping.

The braking test was carried out with a small-scale model comprised of a inertial load system and a power converter. The model was tested from starting till coasting operation and braking test including normal-opposite operation.

Suspension torque after stop was implemented in the model and a pure electric brake that stops the model with air brake. Therefore, if there is no problem like emergency shutdown when there is a fault in the electrical system, the air brake can be omitted in the M car. Besides, it is possible to reduce weight of the vehicles.

Pure electric brake was also realized in the system. During braking the car, the brake force could be secured only with electric brakes even in high speed region by expanding constant torque region and constant power region.

## References

1. Kim, Y.-C., Song, H.-B., Cho, M.-T., Lee, C.-S., Kim, O.-H., Park, S.-Y.: A Study on the Improved Stability of Inverter through History Management of Semiconductor Elements for Power Supply. In: Kim, T.-h., Ramos, C., Kim, H.-k., Kiumi, A., Mohammed, S., Ślęzak, D. (eds.) ASEA/DRBC 2012. CCIS, vol. 340, pp. 155–162. Springer, Heidelberg (2012)
2. Kim, Y.-C., Song, H.-B., Cho, M.-T., Moon, S.-H.: A study on direct vector control system for induction motor speed control. In (Jong Hyuk) Park, J.J., Jeong, Y.-S., Park, S.O., Chen, H.-C. (eds.) EMC Technology and Service. LNEE, vol. 181, pp. 599–612. Springer, Heidelberg (2012)
3. Kovudhikulrungsri, L., Koseki, T.: Speed Estimation in Low-Speed Range for an Induction Motor to Realize Pure Electric Brake. I.E.E. Japan Joint Technical Meeting on Transportation & Electric Railways and Linear Drive, TER-00-38 LD-00-65, 19–24 (July 2000)

# Optimized Design of Charger for Electric Vehicles with Enabled Efficient CCCV Mode Movement

Ji-Yong Chun<sup>1</sup>, Young-Choon Kim<sup>2\*</sup>, and Moon-Taek Cho<sup>3</sup>

<sup>1</sup> Div. of Automotive Engineering,  
Ajou Motor College  
jychun@motor.ac.kr

<sup>2</sup> Div. of Mechanical and Automotive Engineering College of Engineering,  
Kongju National Univ. 275 Budae-dong, Cheonan-si, Chungnam, 330-717, Korea  
yckim59@kongju.ac.kr

<sup>3</sup> Dept. of Electrical & Electric Engineering,  
Daewon University College,  
316, Daehak Road, Jechen-si, Chungbuk Province, 390-702, Korea  
mtcho@mail.daewon.ac.kr

**Abstract.** This paper presents a charger technology for the electric vehicles that enables charge and discharge not only for low voltage and high voltage but also for any battery type by using a high performance DSP. The proposed charger was made to function as generalized fast and low battery by using PWM buck converter that runs with CCCV(Constant Current Constant Voltage) mode. Besides, by designing the controller as fixed-type and varied suiting to the load characteristics, constant output was ensured even during power trip. Also, by controlling the battery type, charge, and discharge, a/s becomes easy. This battery would be possibly implemented not only in the Off Board Charger of the electric vehicles but also in the On Board Charger of EREV(Electric Range Extender Vehicle) in future.

**Keywords:** Electric vehicle, DSP, PWM, CCCV, EREV.

## 1 Introduction

A battery charge/discharge system proposed by this study has a computation part to automatically control itself by perceiving mode change during charging and discharging the battery by using DSP Controller MC56F8345. Also, a module was constructed so that charging and discharging are quickly executed.[1],[2]

For that, a system was made to adjust charging current of various steps according to the load by supplying voltage and current suiting to the imposed load by adjusting the time ratio of PWM(Pulse Width Modulation) during control even if battery voltage is different from Li-ion and lead storage type. In addition, battery for the

---

\* Corresponding author.

electric vehicle was made as plug-in system in case the battery voltage is a slow charger-type. The capacity of battery was set up as 2[kVA] level considering that DC output voltage is 310[V] when AC voltage is input. At the same time, in case of fast chargers, capacity was designed as 10[kVA]. Therefore, input current into the battery 1[C-rate] was set as 3[A] based on the current input into the battery around 35[A] so that charge and discharge mode of minimum ten steps could be switched.

## 2 Design of EV Charger

Since there is one MOSFET switching device inside power converter, and microprocessor controller source is required, a SMPS(Switching Mode Power Supply) that supplies multi-source required in each circuit like gate circuit of controller and power converter by RCC(Ring Choke Converter) as in Fig. 1 by dc source from battery was constructed.

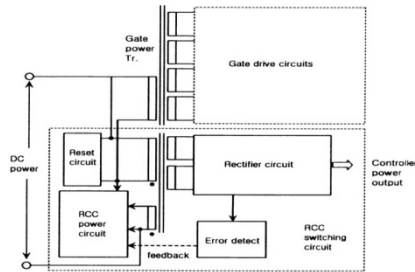


Fig. 1. MOSFET Gate and Power Circuit

For the driving source, ac source 380/200[V] was commutated for fast charger. Maximum output voltage de 310[V] was also set during diode commutation of 220[V] at second side.

Insulation element photocoupler was used so that signal from output port of microprocessor transmits signal to high power circuit at drive circuit terminal. Circuits were insulated by OP-AMP to reduce the effect of noise at controller terminal to control voltage and current. DC offset was made to be adjusted to convert power to digital value. Also, the dc voltage input is an important factor in PWM modulation when output voltage is controlled in DC/DC converter. In addition, since dc voltage is high voltage-type, control circuit is necessarily required. Therefore, a hall element was used to electrically insulate the dc signal.

To filter peak voltage by carrying out single phase full wave of ac voltage  $100\sqrt{2}$  at source side, 20 numbers of electrolytic condensers having capacity  $2200[\mu\text{F}]/63[\text{V}]$  were used. Since load is battery and it is working as voltage source at output side, capacitor was not needed. It was arranged by programming so that output side can function as CC(Constant Current) mode.

Meanwhile, iron core was used as an inductor. Also, a current constant-type reactor capacity 1[mH] was used to stably maintain current 15[A] needed by load.

### 3 Controller Design

Buck chopper was constructed as 10[kW] level, and power part was constructed to drive as continuous mode under load higher than 10[%] at switching frequency 15[kHz]. Output voltage was set as ripple less than 1[%].

The switching frequency was set as 15[kHz] considering temperature restriction by switching loss at IGBT. Requirement for transient response in power converter was not included since the expected change of load condition was huge.

The voltage controller that includes feed forward compensator is same as Eq. (1). The control diagram is presented in Fig. 2.  $k_f$  is the feed forward compensation gain and optimum value was chosen by frequency response characteristics of output voltage for the load current.

$$d(t) = k_p(v_{ref}(t) - v_o(t)) + k_i \int (v_{ref}(t) - v_o(t)) dt - k_f \frac{di_o(t)}{dt} \quad (1)$$

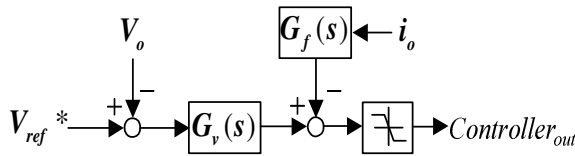


Fig. 2. System Control Block Diagram Using Feedforward Compensator

### 4 Simulation

The algorithm using feed forward compensator proposed in this study was implemented in the PWM buck chopper. To examine the overall control characteristics of proposed algorithm, a control block was constructed by using MATLAB/Simulink as shown in Fig. 3.

The overall control system which was used in the simulation was constructed as shown in Fig. 3. System parameter for the output voltage was set as 300[V] when input voltage was 400[V] as in Table 1. As a test condition, loads were input in the sequence of 10[%](16.36[Ω])→100[%](1.636[Ω])→10[%](16.36[Ω]).

When the feed forward compensator of load current was implemented from the simulation results as in Fig. 4 and Fig. 5, the transient response and steady-state characteristics were significantly improved. Therefore, the characteristics equivalent to the dynamic characteristics required by the system proposed in this paper could be obtained.



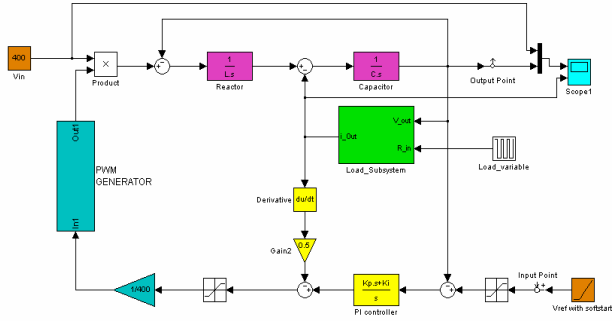


Fig. 3. SIMULINK Block Diagram of PWM Buck Chopper

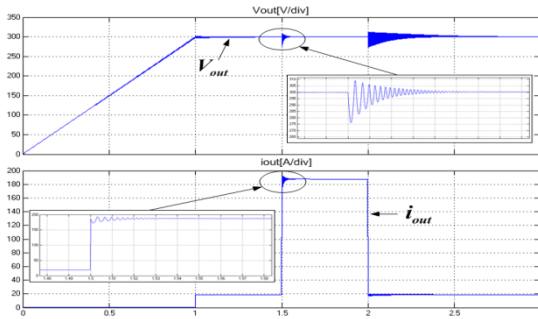


Fig. 4. With PI Controller

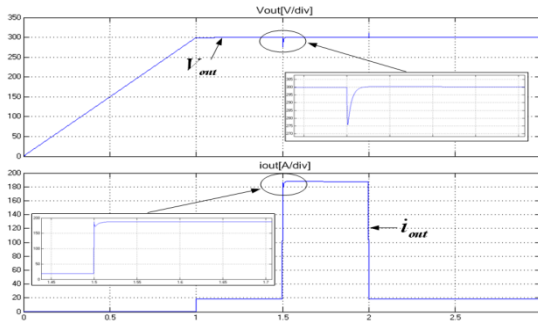


Fig. 5. With Load Current Feedforward Compensator

## 5 Test Results

Fig. 6(a) shows a current waveform between IGBT emitter and base during PWM modulation(upper figure) and the voltage wave form which was input during battery charging(lower figure). Fig. 6(b) shows the voltage(upper figure) and current wave form(lower figure) which were input during battery charging.

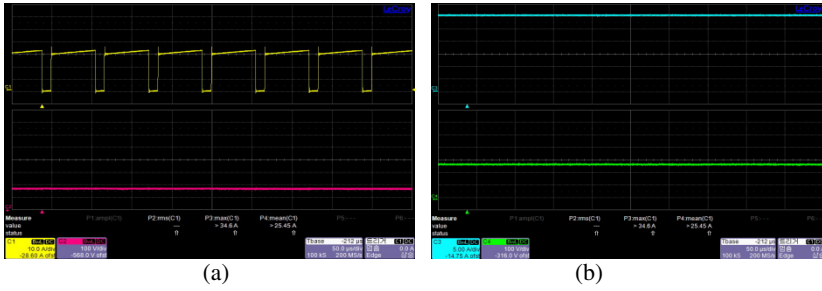


Fig. 6. (a) PWM Modulation Waveform and (b) Output Voltage and Current Waveform

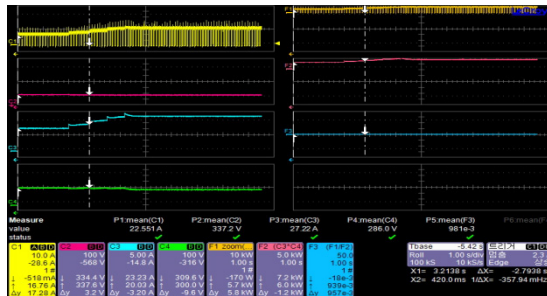


Fig. 7. Input Voltage, Output Voltage, and Current Waveform in Load Test / Efficiency Measurement

The load test was performed to check the stability when load was changed in the sequence of no load→light load→ heavy load→ full load under the same condition as shown in Fig. 7. The efficiency measurement shows that the efficiency of more than 95[%] was confirmed. As can be seen from Fig. 7, the efficiency characteristics of 98.1[%] was obtained.

## 6 Conclusion

The present paper presents a charger technology for the electric vehicles that enables charge and discharge not only at lower voltage and high voltage range but also irrespective of battery type by using a high performance DSP.

The charger technology proposed in this study could be equipped with generalized fast and low function by comprehensively implementing high voltage semiconductor device MOSFET for large power as a low voltage type and IGBT for high voltage type. The generalization of battery was realized by using PWM buck converter that runs with CCCV(Constant Current Constant Voltage). In addition, by designing the controller fixed type and variable type suiting to the load characteristics, continuous output was ensured even during power trip. A/S was also convenient by controlling the battery type, charge, and discharge. Further, it was applied in the power supply unit linking with battery that enabled the synchronization of power converter and drive power design technology for the interface.

## References

1. Maharjan, L., Yamagishi, T., Akagi, H.: Active-Power Control of Individual Converter Cells for a Battery Energy Storage System Based on a Multilevel Cascade PWM Converter. *IEEE Transactions on Power Electronics* 27, 1099–1107 (2012)
2. Zhang, Z., Xu, H., Shi, L., Li, D., Han, Y.: A unit power factor DC fast charger for electric vehicle charging station. In: 2012 7th International Power Electronics and Motion Control Conference (IPEMC), vol. 1, pp. 411–415 (2012)

# A Virtual Cluster Scheme Technology for Efficient Wireless Sensor Networks

Seongsso Cho<sup>1</sup>, Bhanu Shrestha<sup>1</sup>, Young-gi Kim<sup>2</sup>, and Bong-Hwa Hong<sup>2</sup>

<sup>1</sup> Department of Electronic Engineering, Kwangwoon University,  
20 Kwangwoon-ro, Nowon-gu, Seoul 139-701, Korea

<sup>2</sup> Department of Information Communication, Kyunghee Cyber University, Korea  
{css,bnu}@kw.ac.kr, smileknock@gmail.com, bhhong@khcu.ac.kr

**Abstract.** Once nodes are deployed in the wireless sensor network, as it is impossible to replace batteries, the amount of transmittable information depends on how to use the limited energy efficiently for longer network life. Virtual Cluster Routing (VCR) builds the efficient virtual cluster considering node compactness, selects the closest node and manages the routing table to reduce overhead significantly by referencing routing table information within the virtual cluster without communicating with other clusters via the head, and has a lower transmission delay and higher survival time than the routing scheme.

**Keywords:** Sensor Network, MANET, LEACH, Sensor Node, Sink Node.

## 1 Introduction

The Wireless Sensor Network (WSN) consists of small sensor nodes including microcontrollers, wireless transmitters, and sensing modules. And node information is transmitted to the data collecting node, sink node mostly via multi-hop communication. At first, for the military purpose, the WSN was developed to monitor and reconnoiter areas inaccessible to people by deploying many sensor nodes. However, its usage expanded to environmental monitoring, building risk analysis, patient monitoring, medical service and others to collect various information [1]. In the sensor network, modeling the Mobile Ad-Hoc Network (MANET) environment, without fixed bases such as Access Points (AP), relatively many sensor nodes are deployed in the wide sensor field to create various dynamic topologies and have nodes autonomous and independent from each other [2-3]. What should be considered here is how to use the limited energy resource of the sensor node efficiently. The sensor node is powered by the battery, which cannot be replaced or recharged due to its operational environment. When assessing the WSN's performance, energy efficiency, accuracy of detected data and service quality are considered. Among them, the key item is energy efficiency. As the sensor consumes energy over time, energy efficiency decides how long the sensor can operate before energy runs dry [4]. In the sensor network, the existing routing scheme can be divided into plane-based one and hierarchy-based one depending on the node configuration. While nodes exchange

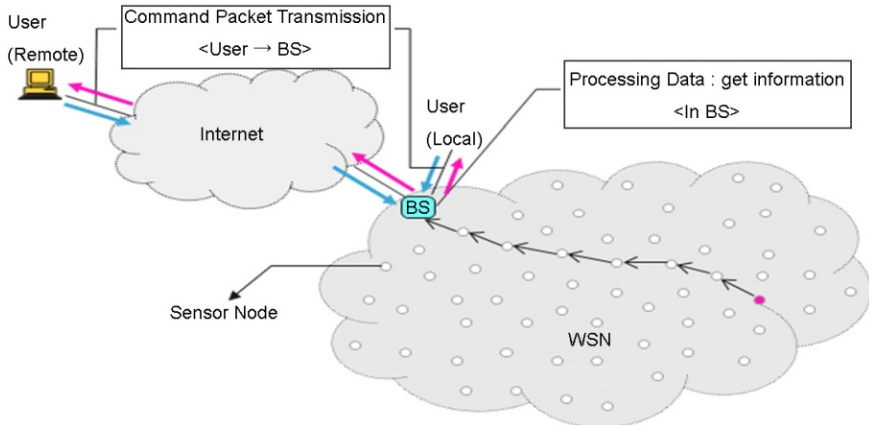
information with each other at the same level in the plane-based routing scheme, many clusters are built to create the node hierarchy for data transmission in the hierarchy-based routing scheme [5-6].

If the plane-based routing scheme is applied to the large-scale sensor network, routing table management and increased routing messages increase routing overhead. To address this problem, as a hierarchy-based routing scheme, Low Energy Adaptive Clustering Hierarchy (LEACH) has been suggested [7-8]. In case of LEACH, due to inefficient clusters and increased dependence on cluster heads, routing overhead increases. In this study, to overcome these problems, combining the plane-based routing scheme with the hierarchy-based one, Virtual Cluster Routing (VCR) is suggested. VCR builds efficient VCR considering node compactness. The virtual cluster communicates with other clusters not only by heads but also by referencing the routing table within the virtual cluster to select the nearest node. Also, as only routing information within the cluster is maintained, overhead for routing table management can be reduced significantly.

## 2 Related Research

### 2.1 Overview of the Sensor Network Routing Protocol

Figure 1, shows the sensor network where each node reads information through its sensor [9].



**Fig. 1.** Wireless sensor network architecture

Depending on the network type, the protocol can be divided into the plane routing one and the cluster-based hierarchical routing one. In the plane routing protocol, the network is considered as one area, all nodes equally participate in routing. The cluster-based protocol partitions the network into several clusters and classifies nodes into the hierarchy by its role [4]. Data collected by lower nodes are transmitted to upper nodes, and then the upper nodes combine them to send to the BS. Well-known protocols are LEACH, LEACH-Centralized

(LEACH-C), and TEEN [10-11]. Also, based on network’s operating modes and target applications, it can be classified into proactive one and reactive one. In the proactive network, nodes in the field run only during their cycle to sense and collect data, and then send it to their upper nodes. For periodic data monitoring, LEACH and LEACH-C are suitable. On the other hand, in the reactive network, all nodes in the field sense data sequentially, react to a data change immediately, and then send changed data to their upper node directly. This is suitable for time-critical applications. The most well-known example is TEEN.

## 2.2 LEACH Protocol

The WSN using LEACH consists of multiple clusters. Each protocol, there are the cluster head (CH) which controls all sensor nodes in the cluster, fuses data from sensor nodes, and then send it to BS; and non-CH which collects and sends data to its CH. Especially, as the CH should fuse data from non-CHs, and then sends it to the remotely-located BS, much energy is consumed. So, to have all nodes play a role as CH evenly, whenever the round begins, the CH is selected from all nodes according to specified probability [10].

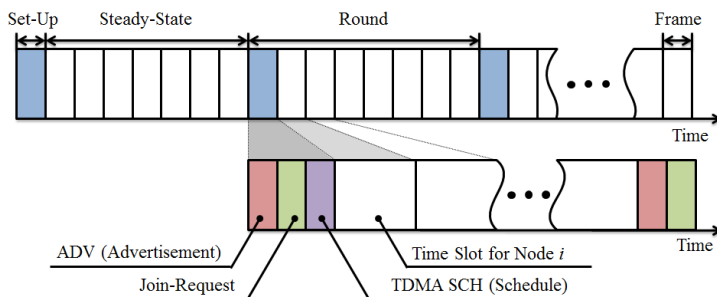


Fig. 2. Timeline showing operation of LEACH

As shown in Figure 2, operation and configuration of the LEACH protocol consist of rounds. In each round, by starting up with set-up, where the head is selected to form a cluster, and then the steady state where data is transmitted from non-CH to CH, and then CH to BS [9-10]. During the unassigned slot time, the node switches to the sleep mode to save energy. When one round ends, a new round begins, a new CH is selected, and then the above procedure is repeated. Even when previously-collected data is the same as the one sensed currently, data is transmitted to the CH. In other words, as unnecessary data is sent among member nodes, consuming energy. At the same time, as the CH is selected according to probability and the cluster is built based on the selected CH’s location, the cluster can be built in the unfavorable geologic structure.

## 2.3 Characteristics of the Sensor Network

In the sensor network, there are processors which process detectable and collected information, small sensor nodes which send such information, and sink nodes which collect

and send such information to the outside [12]. Different from existing networks, the sensor network is basically designed to automatically collect remote information, and is widely used in scientific, medical, military and commercial applications.

The sensor network differs significantly in its application, control and configuration. In the traditional network, Quality of Service (QoS) should be guaranteed, and the configuration, routing and mobility control of mobile nodes for high bandwidth use is important. However, in the sensor network, as many small sensors are running in the environment where people cannot access easily and power cannot be resupplied, energy control for sensor nodes is very important. Furthermore, compared to traditional wireless environments such as Ad Hoc, several hundreds to tens of thousands nodes are compacted to create a sensor network. Therefore, routing many nodes can create routing overhead. This issue should be addressed in the large-scale sensor network.

### 3 Virtual Cluster Routing (VCR)

In The virtual cluster exchange routing information only with its closest node to build a network. At this point, the sensor node in the virtual cluster can send data to different cluster nodes or sink nodes without going through the virtual cluster head. To build a virtual cluster, based on node compactness, virtual cluster heads are selected. And then, based on selected heads, multiple virtual clusters are built. For partition nodes not included in the virtual cluster, the virtual cluster is built. Also, to exchange data among virtual clusters, their level is set.

To select the virtual cluster head, all sensor nodes send the ADV message to themselves and nearby nodes. At this point, the ACK message for the ADV message is not sent. Based on the sum of ADV messages from surrounding nodes, adjacent node information is determined. If a node receives messages whose sum is bigger than the standard value based on the node compactness, that node is selected as the cluster head. If two of neighboring nodes are selected as the cluster head, the node with a bigger ADV message sum will be selected as the virtual cluster head.

To minimize the overhead in building the virtual cluster, the virtual cluster head is selected only once. At this point, the standard value for head selection is cut in half. More nodes can be involved in selecting the virtual node head. The formula for selecting the cluster head is shown in Formula (1).

$$\text{Head Selection Standard} = [(N * \pi R^2) / 2A] \quad (1)$$

When the cluster head is selected in the LEACH scheme, according to Formula (2), each node calculates its possibility to become the head cluster.  $C_i(t)$  is an indicator function. During  $r \bmod(N/k)$ , if the node was the cluster head, the indicator function's value is 0, and if not, it is 1. In other words, if a node was the head during  $r \bmod(N/k)$  even once, it cannot be selected as a head again.

$$P_i(t) = \begin{cases} \frac{k}{N - k(r \bmod \frac{N}{k})} & : C_i(t) = 1 \\ 0 & : C_i(t) = 0 \end{cases} \quad (2)$$

In Formula (2),  $i$  is node's indicator,  $t$  is time,  $N$  is number of nodes,  $k$  is number of clusters, and  $r$  is round. During a certain round, as the head is selected from nodes which have not been a head before, the number of rounds increases, resulting in a simple increase in  $P_i(t)$ . This pattern repeats at the cycle of  $N/k$  to have all nodes selected as a head node having equal probability.

## 4 Test and Performance Evaluation

### 4.1 Test Environment

In As shown in Table 1, network size, number of nodes, transmission range and number of heads were selected for various sensor network topologies to conduct tests.

**Table 1.** Test Environments

Topology	Network Size	No. of Nodes	Transmission Range	No. of Heads
A	500m*500m	100	100m	6
B	1000m*1000m	200	200m	12
C	2000m*2000m	300	300m	18
D	4000m*4000m	400	400m	25
E	5000m*5000m	500	500m	31

### 4.2 Survival Time Test

To evaluate the energy efficiency of VCR in the sensor network, the survival time was compared with the existing routing scheme. In this test, the initial energy of 100J was given to Topologies A and E, and then CBR data was transmitted by every 0.5 seconds. Table 2 shows network survival times. As shown here, Topology A has a long survival time in the order of AODV, DSDV, LEACH, and VCR.

**Table 2.** Comparison of Network Survival Times

Routing Scheme	Topology A	Topology E
AODV	1249	199
DSDV	1392	213
LEACH	1434	229
VCR	1521	253

VCR survived 22%, 9% and 6% longer than AODV, DSDV, and LEACH respectively. In the AODV scheme, as route searching messages increased rapidly whenever a new route was set, energy consumption increased proportionally to route searching messages. In the DSDV scheme, as 100 routing tables were maintained to send data, this led to overhead, resulting in high energy consumption. In the LEACH scheme, as the overhead caused by routing tables and routing messages was reduced, energy was less consumed than AODV and DSDV. However, as clusters were formed every round, energy was consumed more than the proposed VCR scheme. Also, in the



test with Topology E where there was a significant increase in the number of nodes, transmission range and network size, VCR's survival time was longer than AODV, DSDV and LEACH by 27%, 19% and 11% respectively.

## 5 Conclusion

In the sensor network, the plane-based routing scheme increases overhead due to routing table management and increased routing messages. To address this problem, the hierarchy-based routing scheme is suggested and has lower overhead by managing the routing table with multiple clusters. However, inefficient cluster formation and dependency on the cluster head selected every round leads to cluster overhead. In the VCR scheme, considering the compactness of sensor nodes existing in the network, the virtual cluster head is selected, resulting in higher efficiency. Also, as the virtual cluster is formed only in the highly compacted area, cluster overhead can be reduced. Moreover, as the virtual cluster level is set, data is sent only within the virtual cluster where the node belongs by referring to the routing table. This can reduce overhead caused by routing messages and routing tables. The test confirmed the proposed VCR scheme had the lower transmission delay in the large-scale sensor network with increased network size, transmission range and number of nodes.

## References

1. Ian, F., Akyildiz, S.W., Sankarsubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. *IEEE Communications Magazine*, 102–114 (August 2002)
2. Anna, H.: *Wireless Sensor Network Designs*. John Wiley & Sons, Ltd. (2003)
3. Paolo, S.: Topology control in wireless ad hoc and sensor networks. *ACM Computing Surveys (CSUR)* 37(2) (2005)
4. Heintzelman, W.B.: Application-specific protocol architectures for wireless networks. Ph.D. dissertation, Mass. Inst. Technol, Cambridge (2000)
5. Jiang, M., Li, J., Tay, Y.: Cluster Based Routing Protocol, IETF Draft (August 1999)
6. Chatterjee, M., Das, S.K., Turgut, D.: An On-demand Weighted Clustering Algorithm (WCA) for Ad-hoc Networks. In: *Global Telecommunications Conference (GLOBECOM 2000)*, vol. 3, pp. 1697–1701. IEEE (2000)
7. Luo, H., Liu, Y., Das, S.K.: Routing correlated data in wireless sensor networks: A survey. *IEEE Netw.* 21, 40–47 (2007)
8. Zhang, Y., Gulliver, T.A.: Quality of Service for Ad-hoc On-demand Distance Vector Routing. *IEEE Conference Publications* 3, 192–196 (2005)
9. Cho, S., Shrestha, B., La, K.-H., Hong, B., Lee, J.: An Energy-Efficient Cluster-Based Routing in Wireless Sensor Networks. In: Kim, T.-H., et al. (eds.) *FGCN 2011, Part I. CCIS*, vol. 265, pp. 15–22. Springer, Heidelberg (2011)
10. Koutsonikola, D., Das, S., Hu, Y.C., Stojmenovic, I.: Hierarchical Geographic Multicast Routing for Wireless Sensor Networks. *Wireless Netw.* 16, 449–466 (2010)
11. Dimokas, N., Katsaros, D., Manolopoulos, Y.: Energy-efficient distributed clustering in wireless sensor networks. *Journal of Parallel and Distributed Computing* 70, 371–383 (2010)

# An Intelligent e-Services Composition Platform for Ubiquitous Baby Care: The Case Study of Life and Commercial Support Services for Property Management

Chih-Kun Ke<sup>1</sup>, Yi-Jen Yeh<sup>2</sup>, and Chiao-Min Chang<sup>1</sup>

<sup>1</sup>Department of Information Management,  
National Taichung University of Science and Technology,  
40401 Taichung, Taiwan R.O.C.

<sup>2</sup>Department of Division for Mobile Internet Software Technology,  
Network Services & System Technology, Information and Communications  
Research Laboratories, Industrial Technology Research Institute,  
31040 Hsinchu, Taiwan R.O.C.  
{ckk, s1802b109}@nutc.edu.tw,  
EthanYeh@itri.org.tw,

**Abstract.** Homecare is a kind of property life service of the life and commercial support services for property management. How to supply an adaptive homecare service to community residents becomes an interest research topic of the building management and maintenance industry, particularly for enhancing community management quality. Besides, Information and Internet techniques have changed human lifestyle. Various mobile devices let user to enforce some service in the ubiquitous environment. Therefore, this work presents an intelligent e-services composition platform that community residents can get composite homecare e-services in the ubiquitous environment. An use case of baby care demonstrates the proposed platform that uses advanced techniques to create the resident profile. The resident profile records basic personal attributes and requirements. According to the resident profiles collected in the knowledge base, the resident fetches relevant e-services to compose the reasonable composite baby care e-services. Finally, the adaptive baby care information from the reasonable composite baby care e-services is recommended to residents to enjoy their lives in the ubiquitous environment. Hopefully, supplying the adaptive baby care information can help residents to have good life quality.

**Keywords:** property management, e-services composition, baby care, adaptive recommendation, ubiquitous.

## 1 Introduction

The life and commercial support services for property management are closely connected with the lives of residents, including property agent service and consultation, administration, property life service, commercial support, etc. Homecare is a kind of

property life service of the life and commercial support services for property management [1-3]. How to supply an adaptive homeware service to community residents becomes an interest research topic of the building management and maintenance industry, particularly for enhancing community management quality. Besides, Information and Internet techniques have changed human lifestyles. Service providers construct an innovative e-service platform to provide the various e-services over the Internet [3]. E-service is the modern trend for creating new usage patterns to attract and retain users. Various mobile devices let user to enforce some e-service in the ubiquitous environment [4]. However, in a service-oriented ubiquitous environment, quality of service (QoS) is an important significant requirement in evaluating a service [4], especially a composite e-service. Organization uses an agreement to coordinate the service providers and users [5]. Information retrieval techniques [6] are used to extract key terms from a user personal data. The extracted key terms form a profile to represent the information needs of users for acquiring the services [7]. Therefore, how to enforce the agreement in various e-services convergence is a service quality consideration.

This work presents an intelligent e-services composition platform that community residents can get composite e-services in the ubiquitous environment. A use case of baby care demonstrates the proposed platform that uses advanced techniques to create the resident profile. The resident profile records basic personal attributes and requirements. We use an agreement to coordinate the fetched e-services and compose the e-services in evaluating the composite service quality. The agreement is designed according to a resident profile [5]. The various contents of e-services are configured based on meeting the different service requirements in an agreement. Finally, the adaptive baby care information from the reasonable composite baby care e-services is recommended to residents to enjoy their lives in the ubiquitous environment. A baby care use case is shown the demonstration [4]. In the ubiquitous environment, resident family member uses tablets to get the baby care information, e. g., feeding status, schedule of baby vaccination, and real-time messages. Besides, resident family members can use the home media center, e.g., smart television, to have fun together in browsing the comprehensive baby care information, e. g., statistics of baby growth, photos, videos, and relevant baby messages.

The rest of this paper is organized as follows. Section 2 reviews pertinent literature on property management in Taiwan and e-service composition. Section 3 then introduces the proposed intelligent e-service composition platform. Section 4 demonstrate a prototype system. Conclusions are finally, drawn in Section 5, along with recommendations for future research.

## **2 Related Works**

### **2.1 Property Management in Taiwan**

In Taiwan, property management is categorized into building and environment usage management and maintenance, life and commercial support services, and property management. The category, life and commercial support services, includes area such

as property agent service and consultation, administration, property life service (community network, homecare and nurse services, delivery and logistics), life product and commercial support [1]. Although the definition of property management differs among countries, its main focus is to achieve three objectives, including increasing land value, enhancing safety of the work and life environment, and decreasing the cost of building maintenance and resource waste.

Generally, property management involves comprehensive community management, including building maintenance, environment construction, life functionality, and relevant service items to assist various activities associated with resident life. Traditional services, e.g., management of postal services, security management, equipment and device maintenance, public health and environmental management, are assigned to property management companies. In sum, property management involves human development of building including property service, dynamic line management, public equipment item, community construction, etc [2]. Through differences in composition and implementation, property management aims to maximize property value [3]. Recently, the Internet has facilitated human life. Creating innovative service is a new trend supporting resident life. For example, network home delivery, repair service, and traffic planning service can promote resident life quality and performance [1-3].

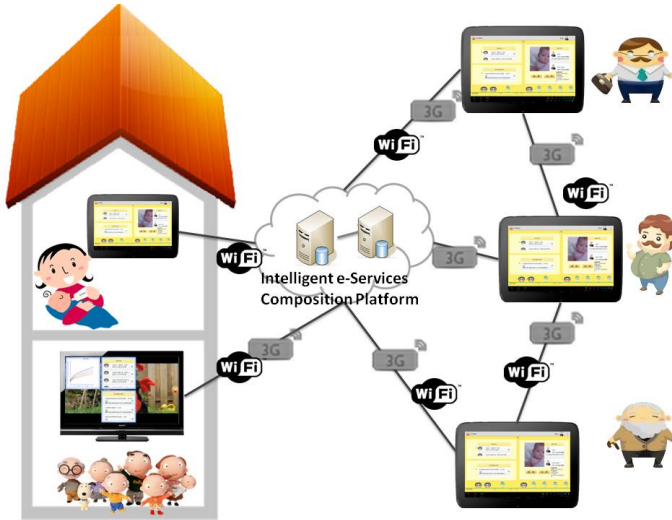
## 2.2 E-Services Composition

In service-oriented computing paradigm, e-services are self-aware, self-contained modules and perform functions that can range from answering simple requests to executing business processes. E-services can be described, published, located, discovered, programmed, and configured using XML-based technologies over a network [8]. The web services business process execution language (BPEL), web service description language (WSDL), universal description, discovery, and integration infrastructure (UDDI), simple object access protocol (SOAP) techniques and Internet standards, e.g., hypertext transfer protocol (HTTP), are enforced to reconstruct an integrated e-service system from non-network-based systems or deployed e-services over standard middleware platforms. Designing high performance, expandable, and reliable e-services has recently been a key topic [3, 8].

Although the present era is one of information overload, users remain starved of information. Users are dissatisfied with single types of information, and integrate numerous types of data to compose various information contents, e.g., comprehensive product catalog. Integrating different types of information platforms to provide total solution service is important. E-service may be a software component that can be integrated with different platforms. The integration of different platforms may cause communication problems. E-service can dynamically integrate different platforms with load balance and thus solve these communication problems. The loose coupling property makes the e-services to compose a comprehensive e-service with various functionalities. This enables different platforms to co-operate smoothly. The composite e-service is used to accomplish specific business tasks in enterprise application integration (EAI), e.g., the broker or auction mechanism over Internet is a kind of composite e-service in the ubiquitous environment [3, 7].

### 3 Intelligent e-Services Composition Platform

This section introduces the proposed intelligent e-services composition platform and its functionality of e-services composite processing in the ubiquitous environment, is shown in figure 1.



**Fig. 1.** An intelligent e-Services composition platform for ubiquitous baby care

The intelligent e-services composition platform uses information retrieval techniques to collect key terms from the personal, operational data and the context of the user operation. The extracted key terms form a profile to represent the information needs of users for acquiring the e-services. Besides, the agreement is constructed according to a user profile for quality of service (QoS) evaluating. User can fetch the desire e-services from the e-services pool. Then, the platform identifies the e-services' communicate interface to compose the collected e-services and gathers relevant contents of the composite e-services from the knowledge base. According to predefined agreement, content filtering process adaptive evaluates and filters out the irrelevant e-services' contents from the resident requirements. Finally, the proposal platform presents the evaluated contents in a service interface and adaptive recommend to the resident. The functionality is illustrated by an implementation in the tablets and the home media center. The proposed experimental e-service composition platform is built on a tablet which works on an Android-based operation system. In the ubiquitous environment, family member, e.g., father, mother, grandpa, grandma, uncle, or auntie, etc., uses tablets to get the baby care information, e. g., feeding status, schedule of baby vaccination, and real-time messages. Besides, family members can use the home media center, e.g., smart television, to have fun together in browsing the comprehensive baby care information, e. g., statistics of baby growth, photos, videos, and relevant baby messages. The Android SDK [9] function calls which explained the proposed platform working are shown in research [4].

## 4 Prototype System Demonstration

The development environment of smart devices includes a Microsoft Windows 7 32bit operating system. The development environment is Google Android SDK 4.0.3, Google Android Development Tools (ADT) version 20.0.3, Java Development Kit (JDK) version 1.7, Eclipse Classic version 4.2.1, SQLite, and SQLite Expert Professional version 3.3 for database construction. The smart device used is an Asus Eee Pad Transformer TF201, Google Android version 4.0.3. The operations of e-service composition process is shown in figure 2.

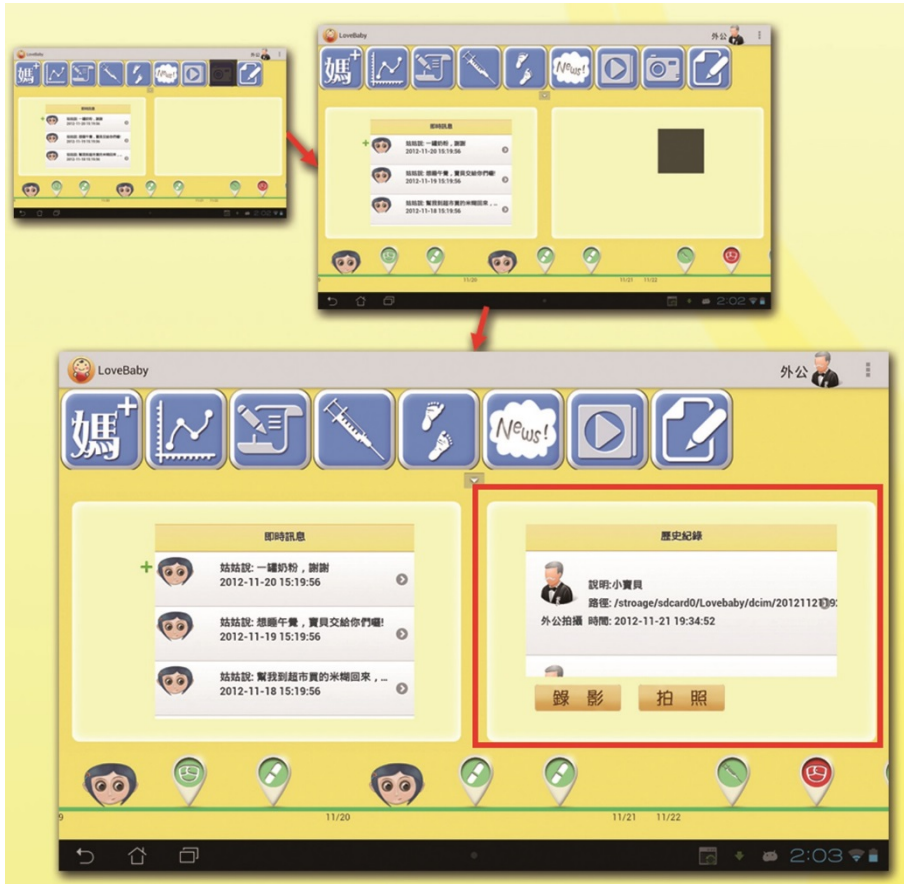


Fig. 2. The operation of e-service composition

The system initialization process is presented as three steps, including initialize, produce a tag (identification), and produce a block operations. The e-service fetching process is presented as three steps, including fetch, drag, and put operations. The e-service composition process is presented as four steps, including compose, get, filter, and close operations.

## 5 Conclusion

This work presents an intelligent e-services composition platform that community residents can get composite e-services in the ubiquitous environment. The use case of baby care illustrates the adaptive baby care information produced from the reasonable composite baby care e-services. Then, the baby care information is recommended to resident's family to enjoy their lives in the ubiquitous environment. Hopefully, supplying the adaptive baby care information can help residents to have good life quality. Future studies can pay more attention to design resident feedback mechanisms. Feedback can help the proposed e-service composition platform perform intelligent turning and learning to improve service quality incrementally. Furthermore, the recommendation technique is considered and combined with more intelligent methods, e.g., case-based reasoning, fuzzy inference methods, to increase its effect.

**Acknowledgement.** This research was supported in part by the Industrial Technology Research Institute and the National Science Council of Taiwan (Republic of China) with a NSC grant NSC 102-2410-H-025-017.

## References

1. Taiwan Institute of Property Management, <http://tipm.org.tw/>
2. Property Management Development Framework and Action Plan of Taiwan, <http://www.cepd.gov.tw/dn.aspx?uid=1240>
3. Ke, C.K., Su, J.Y., Chang, S.F.: A Novel Service Platform for Message Negotiation of E-services: Case Study of Life and Commercial Support Services for Property Management in Taiwan. *Journal of Convergence Information Technology* 7(7), 292–302 (2012)
4. Ke, C.K., Yeh, Y.J., Jen, C.Y., Tang, S.W.: Adaptive Content Recommendation by Mobile Apps Mash-up in the Ubiquitous Environment. In (Jong Hyuk) Park, J.J., Barolli, L., Xhafa, F., Jeong, H.-Y. (eds.) *Information Technology Convergence*. LNEE, vol. 253, pp. 567–574. Springer, Heidelberg (2013)
5. Cosmin Silaghi, G., Dan Șerban, L., Marius Litan, C.: A Framework for Building Intelligent SLA Negotiation Strategies under Time Constraints. In: Altmann, J., Rana, O.F. (eds.) *GECON 2010*. LNCS, vol. 6296, pp. 48–61. Springer, Heidelberg (2010)
6. Richrdo, B.Y., Berthier, R.N.: *Modern Information Retrieval*. The ACM Press, New York (1999)
7. Liu, D.R., Ke, C.K., Lee, J.Y., Lee, C.F.: Knowledge maps for composite e-services: A mining-based system platform coupling with recommendations. *Expert System with Applications* 34(1), 700–716 (2008)
8. Papazoglou, M.P.: *Web Services: Principles and Technology*. Prentice-Hall (2007)
9. Android SDK, <http://developer.android.com/sdk/index.html>
10. Kerlinger, F.N.: *Foundations of Behavioral Research*, 3rd edn. Harcourt Brace College Publishers, New York (1992)
11. Kaiser, H.F.: An index of factorial simplicity. *Psychometrika* 39(1), 31–36 (1974)
12. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quart.* 13, 319–339 (1989)

# An Authoring System of Creating Graphic Map for Item Search Based on Library OPACs

Hsuan-Pu Chang<sup>1,\*</sup>, Wei-Ting Chang<sup>2</sup>, and Shi-Pin Fong<sup>1</sup>

<sup>1</sup> Department of Information and Library Science,  
Tamkang University, Taiwan  
musicbubu@gmail.com

<sup>2</sup> Department of Computer Science and Information Engineering,  
Tamkang University, Taiwan  
tt90089@gmail.com

**Abstract.** In this paper we proposed an authoring system that allows librarians to be able to construct and maintain a graphic-based navigation system by themselves that provides the graphic searching function based on original On-line Public Access Catalog (OPAC) systems. On the other hand, patrons can use their mobile devices to display the created graphic map indicating the positions of required books. Through transforming a floor plan into spatial information connected to call numbers, which is equivalent to book addresses, the position accuracy is increased to a coordinate block divided by a shelf.

**Keywords:** OPAC, call number, graphic-based map, library, navigation.

## 1 Introduction

The majority of libraries today employ OPACs as portals for searching and browsing resources. More specifically speaking, OPAC [1][2][3] is an information retrieval system characterized by short bibliographic records, mainly of books, journals, and audiovisual materials available in a library. For these library materials, a corresponding call number [4] is given for every search item. These call numbers are assigned based on the Dewey Decimal Classification system [5], which the library uses to categorize its books. The conventional search using OPACs have no information about where a book is physically located in a library but leaving it up to patrons to figure out where it is. In a larger library it may not be obvious how materials are organized or categorized. Once you find the right section you may still have to go through many shelves to find what you are looking for. Bazlan, M. J., & Rasam, A. R. A in Malaysia [6], investigated the current types of problems users faced in library, more than 50% respondents have difficulty to find the location of desired book, and 80% respondents agreed that current search methods could be improved with mapping direction and graphic visualization.

---

\* Corresponding author.



The indoor navigation techniques have various applications including the use in a library. Ng, W. W. Y et al. [7] utilized RFID to carry out the implementation in a library, as well as proposed a solution of a restriction caused by the small width of book spin for tags. Hui Li & Xiangyang Gong [8] described an approach to integrate indoor map and outdoor map together using Google map APIs and discussed the navigation service in library based on the integrated map. Kazuki Watanabe et al.[9] presented a library navigation system named LiNS, using combination Web, sensor and smartphone technologies that also provides the function of finding the route from the current location to the point where the object the user needs is located. In addition, the “See Also” by which the user can get related and recommended information may develop a user’s new interest. Rong-Yuh Hwang [10] implemented the mobile navigation system with bluetooth technology and IrDA. Bill Rogers et al. [11] designed a 3D browsing interface for graphically navigating a large collection of documents, as in a library. Three-dimensional scene rendering technique allows the user to view the inside landscape of a library from different perspectives.

In spite of many solutions have been proposed to improve the searching experience, there still have been challenges need to be overcome in reality such as high cost and less flexibility. Almost all visualized navigation systems in libraries are customized because the building environment differs from each other, and it needs a significant charge to carry out the navigation functions covering all items held in a large library. On the other hand, the shelves and items both could be added and removed, that is to say the system must provide the editing functions to make it practical in reality. As a result, the main purposes of this authoring tool are 1) allowing librarians to be able to build and maintain a graphic-based navigation map by themselves, 2) extending the searching function based on the original OPAC system, 3) let patrons use their mobile devices to display the created graphic map in which the required book location is designated

## **2 System Implementation**

### **2.1 Authoring and Searching Flow**

As the Fig.1 shows the steps of constructing a graphic map by a librarian are 1) using the authoring system to drag and drop graphic objects for building the floor plan, 2) inserting call number information corresponding bookcase objects, 3) saving each floor’s map information to the spatial database. The steps of searching a required item by a patron are 1) using OPACs to search his required items, 2) passing the call numbers retrieved from OPACs to the navigation system, 3) displaying the graphic map with navigation signs that designate the physical places of the required items.

### **2.2 Authoring Functions**

To carry out the previous authoring and searching process, we implemented the following functions including graphic object editing, call number and spatial information storage. The detail introductions are as follows:

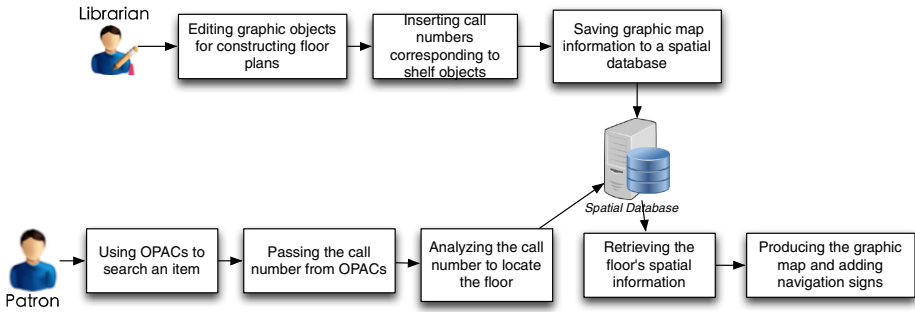


Fig. 1. Librarian authoring and patron searching operation flow

• **Graphic Object Editing**

The authoring system provides the drag and drop function that allows a librarian selecting a variety of objects to the locations where he want to place them. These objects are separated into two main categories; the bookcase objects represent various sizes and types of bookcases used to hold items in a library, other objects are irrelative for holding items but used to decorate a complete floor space including stairs, desks, pillars etc. The Fig. 2 demonstrates a floor plan example made by the authoring system.

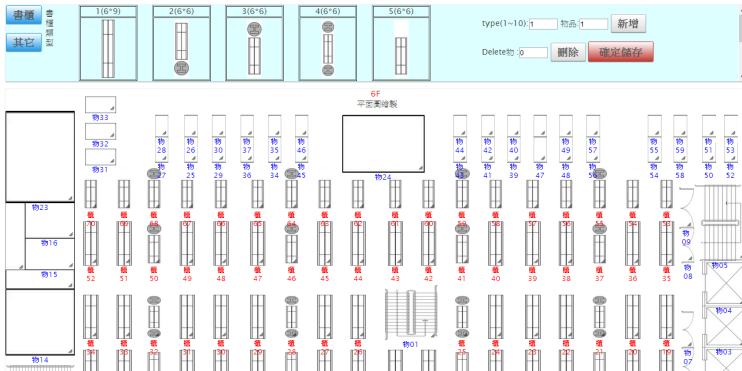
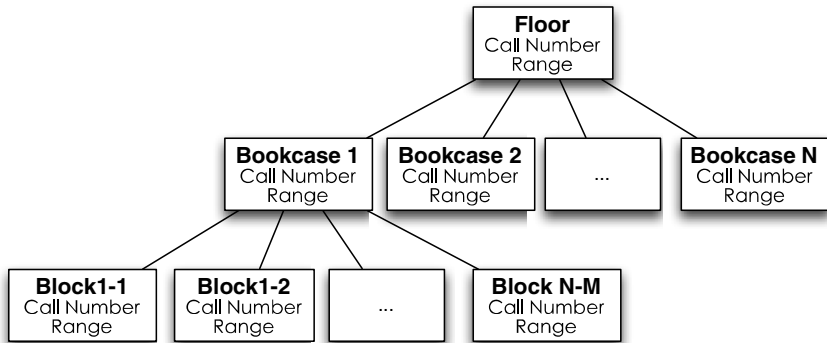


Fig. 2. An example of authoring a floor plan

• **Call Number and Spatial Information Storage**

The database stores not only the object properties and spatial parameters but also the call number information corresponding to different levels of library objects. As the Fig. 3 shows, there are three levels data are stored in database.



**Fig. 3.** Three levels of storing call number

- First level: Preserving the call number range within a floor.

At this level, a floor's basic spatial information is stored including how many bookcases are allocated and the call number range of each floor.

- Second level: Preserving the call number range of each bookcase located within the floor.

At this level, each bookcase's detail information in a floor is stored including its coordinates, width and depth, object type, and call number range.

- Third level: Preserving the call number range of each coordinate block divided from a shelf.

A shelf in a library is usually big enough to be divided into plenty blocks for management and searching. In order to increase the accuracy of designating the required book location, each shelf has a specific table to store every block's call number range within it. The divided blocks of a shelf are similar to x and y coordinates; x coordinate is a given number of blocks along the horizontal axis starting from the block (block 0) on the extreme left of a shelf, y coordinate is a given number of blocks along the horizontal axis starting from the lock (block 0) at the top of a shelf.

### 2.3 Graphic Map Presentation

The authoring system is a web-based tool that librarians don't have to install any program and the created map is also web-based content so that it can be displayed by a patron's mobile device. According to the input call number the target floor spatial information is retrieved and transformed as the Fig. 4 shows. The navigation graphic first designates the target bookcase with different color then further depicts the bookcase structure with coordinate blocks that indicates the accurate position of the target item on the shelf.

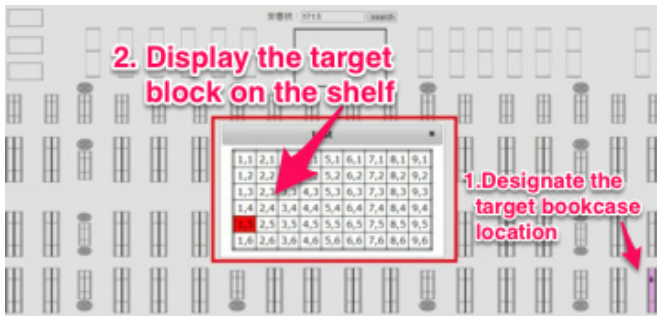


Fig. 4. The target item is indicated by graphic map

### 3 Conclusion

The main purpose of this research is allowing librarians to have the ability to construct and maintain a graphic navigation system for their library. The graphic navigation system does not rely on additional gadgets or advanced technologies but can seamlessly integrate OPACs that have been used in most libraries. The required item can be accurately indicated the position on a shelf. But it will cost a lot of time at first time to construct the database connecting spatial objects and call number information. How to reduce the burden and simplify the constructing works with batch processing is one of our future works.

### References

1. Thanuskodi, S.: Use of online public access catalogue at annamalai university library. *International Journal of Information Science* 2(6), 70–74 (2012)
2. Antelman, K., Lynema, E., Pace, A.K.: Toward a twenty-first century catalog. *Information Technology and Libraries* 25(3), 128–139 (2013)
3. Beccaria, M., Scott, D.: Fac-back-OPAC: An open source interface to your library system. *Computers in Libraries* 27(9), 6–8 (2007)
4. Gombrich, E.H., Gombrich, E.H., Gombrich, E.H., Gombrich, E.H.: *Art and illusion: A study in the psychology of pictorial representation*. Phaidon London (1977)
5. Wiegand, W.A.: The “amherst method”: The origins of the dewey decimal classification scheme. *Libraries & Culture* 33(2), 175–194 (1998)
6. Bazlan, M.J., Rasam, A.R.A.: Development of geographical based library information system(GeoLIS). Paper presented at the 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC), pp. 248–252 (2012), doi:10.1109/ICSGRC.2012.6287170
7. Ng, W.W.Y., Qiao, Y.-S., Lin, L., Ding, H.-L., Chan, P.P.K., Yeung, D.S.: Intelligent book positioning for library using RFID and book spine matching. Paper presented at the. In: Paper presented at the 2011 International Conference on Machine Learning and Cybernetics (ICMLC), pp. 465–470 (2011), doi:10.1109/ICMLC.2011.6016840
8. Li, H., Gong, X.: An approach to integrate outdoor and indoor maps for books navigation on the intelligent mobile device. Paper presented at the 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN), pp. 460–465 (2011), doi:10.1109/ICCSN.2011.6013872

9. Watanabe, K., Takahashi, T., Ando, T., Takahashi, K., Sasaki, Y., Funakoshi, T.: LiNS: A library navigation system using sensors and smartphones. In: 2010 International Conference on Paper presented at the Broadband, Wireless Computing, Communication and Applications (BWCCA), pp. 346–350 (2010), doi:10.1109/BWCCA.2010.94
10. Hwang, R.-Y.: The design and implementation of mobile navigation system for the digital libraries. In: Proceedings of the Sixth International Conference on Information Visualisation, pp. 65–69 (2002), doi:10.1109/IV.2002.1028757
11. Rogers, B., Cunningham, S.J., Holmes, G.: Navigating the virtual library: A 3D browsing interface for information retrieval. In: Paper presented at the Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems, pp. 467–471 (1994), doi:10.1109/ANZIIS.1994.397010

# Cloud-Based Traveling Video Editing System

Joseph C. Tsai and Neil Y. Yen

Foundation of Computer Science Lab., The University of Aizu, Japan  
jctsai@u-aizu.ac.jp

**Abstract.** As the development of media and network, more and more social media services are published in recent years. In this paper, we present a system based on cloud computing. Feature points searching algorithm and image-based rendering are used to generate a meaningful video. We will use feature points to identify the key frames from users' photos database. The Google map API is also utilized to generate the traveling trajectory by the GPS information from photos. A set of metadata is employed to make the generating video much colorful and accurate. Finally, we propose image-based rendering based on cloud computing to combine the key frames and videos into a new videos.

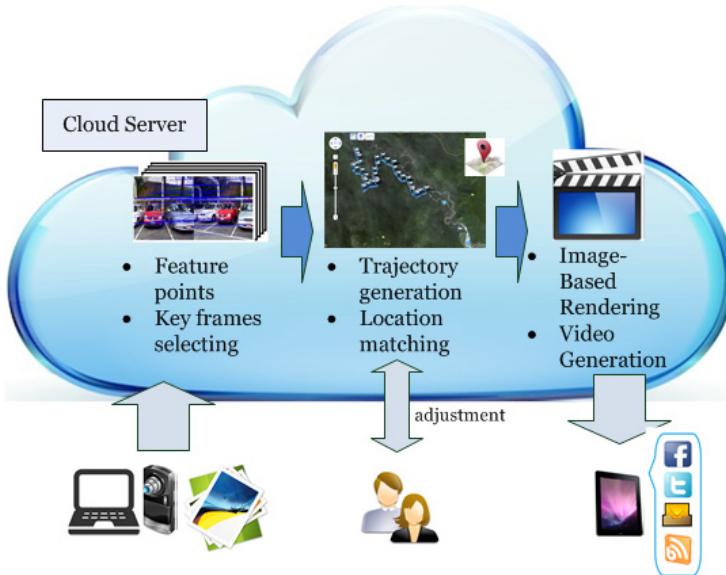
**Keywords:** Social Media, Video generating, Image-Based Rendering, Cloud computing.

## 1 Introduction

In the recent years, smart phones are become popular in our life. The cell-phone manufacturing companies focus on the development of the sub-devices, such as camera, GPS and mobile data network, of the smart phone. Therefore, more and more people will use smart phone to record their travel or memory. Besides, people also enjoy sharing their photos, videos or experience by Internet, such as blog and facebook. The traveling experience will be summarized as a diary or story. That is a wonderful way to record and share the memory. However, this kind of sharing gateway is a little hard to read by elders. They are not good at using or controlling computers. It will be a load if they need to use computers or network to receive their children's or grandchildren's traveling record. In addition, some bloggers or users sometimes feel tired in summarizing the photos or videos. Especially, if the period of the travel is over five days, it's more different to outline the experiences into articles. To avoid or solve this kind of situation, we proposed a system based on feature points searching algorithm and image-based rendering to generate videos automatically.

Image-based Rendering (IBR) is a good technology for combining images into a video and generating additional frames between two photos. It just cost a little time to look for feature points in each frame, most of all, it is real-time in display. The concept of IBR is that capture some geometric information from images, such as geometric proxies, epipole consistency and minimal angular deviation [1], and then we can generate the video with key frames and additional frames. View-dependent texture mapping (VDTM) [2, 10, 14] is one effective IBR method. It can establish a basic 3D

model by detecting the edge of the building in the photo, and then use the geometric constraints to optimization the model. Finally, the texture will be pasted to the model and generates new views. VDTM is simple, quick and only requires small numbers of photos to synthesis realistic 3D images. But the disadvantage is that it must be used in ordinary buildings or objects which edges are very obvious. Another method is light field/lumigraph [3,4]. Instead of using a plenoptic function [8], they propose a new 4D function to represent the flow of light. This method needs to take lots of photos to ensure that there are lights go through every direction. It must spend a lot of time in data processing, because of the large number of ray. But the result is better than VDTM in rendering of images. View Interpolation/View Morphing [5, 6] refers to the images information near the new viewpoints, and the camera position must on the baseline. (Baseline is a straight line connecting two cameras.) The main idea is using the geometric properties to determine pixels in new viewpoints from left or right photos. This method is easy and requires a small number of photos, but the scope of new vision is limited and cannot produce the correct picture if there are no reference images.



**Fig. 1.** The context of the proposed server

The Fig. 1 shows the concept of our system. At first, users have to update their photos or films to the system by smartphone or other smart devices. The system will compute the feature points of each frame and decide the key frames. Second, the travelling trajectory will be generated according to GPS information. Finally, the new video can be created by key frames and trajectory. In section 2, we will give a specified description to our scheme. We explain our method in section 3. Conclusion is in the section 4.

## 2 Overview of Our Scheme

In this paper, we proposed a cloud-based system for users updating their traveling photos or videos. It can refer to the information of the files to generate a new video automatically. The scheme can be divided into the following three steps, feature points generation, traveling trajectory generation and new video generation. At first, we have to choose the key frames from the input files. In order to select suitable frames for video generation, we use Affine-SIFT (ASIFT) [7] to extract the feature points from the frames. The feature points of each frame will be utilized to retrieval with the image database. It aims to sift the important frames from all files. We also use the information of feature points as the data set to compare with the scenes on the target location.

After the candidate key frames selecting, there are two steps processed in the same time. The travel trajectory will be generated according to the key frame. As the photos taking by smart phone or other smart devices, the information of GPS will be used to create the path. The Google map and the GPS data set are used in this step. In order to make the result more accurate, an interactive controlling system is designed in this step. Users can adjust the error information into a correct one. Finally, we would like to utilize Image-Based Rendering algorithm to connect all key frames. In additional, the feature points are used in this step to generate continue-like frames between each key frame. It can let the video looks similar to change the angle and make the video more smooth.

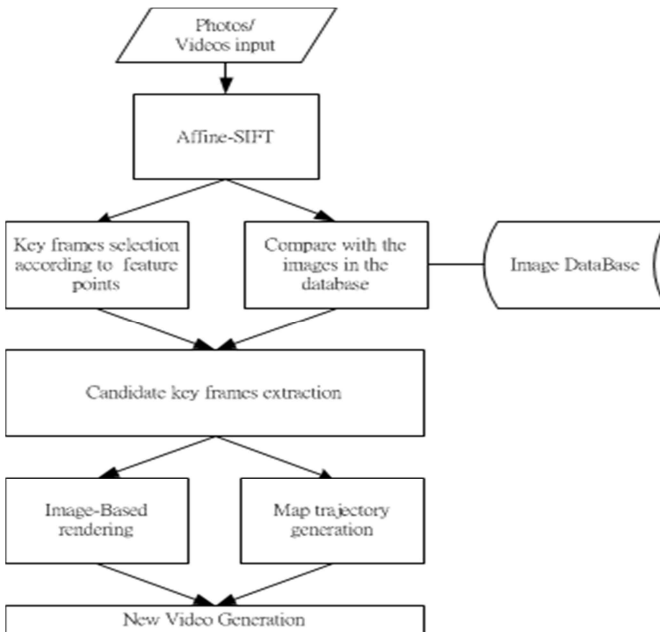


Fig. 2. Flow chart of our system



As Fig. 2 shows, we use ASIFT algorithm in the first step. After the feature points extract, the key frames will be selected by the information. The new frames and travel trajectory can be generated by referring to the key frames. Finally, the video will be created with the key frames, additional frames and travelling trajectory.

### 3 Automated Video Generation

In this section, we would like to introduce the methods of our proposed system. How to select the key frames from the data base and use current frames to generate an interpolation frame is a challenge issue in our system.

#### 3.1 Feature Points Generation

In feature points detecting, one of famous methods is Scale-invariant feature transform (SIFT) algorithm [9], SIFT algorithm can extract and describe local feature points from source images. Therefore, many methods will utilize SIFT algorithm to detect features between two images that are obtained in the same scene but shooting in difference view angles. SIFT algorithm is an effect and robust approach that using in feature matching. However, the frames in our experimental samples are not continual frames. SIFT algorithm will extract feature points with inaccuracy while the shooting angle is too large. Another situation, the structure is too smooth or the feature points between two images are too similar, will also lead to poor results. In this case, SIFT algorithm may not be able to find enough feature points in images.

In order to solve the above problems, J. Morel et al. have proposed Affine-SIFT (ASIFT, [7]) to extract more feature points from sets of images with large shooting angle. The original SFIT algorithm utilizes four parameters to calculate and record the zoom, translation and rotation of frames. Different to SIFT, the authors in [7] proposed two new parameters to replace the original four in SIFT. Because ASIFT algorithm increase some concepts that will become robust than the SIFT algorithm. But in processing time, ASIFT algorithm will spend more processing time than SIFT algorithm. According to our experiment that are not much different in processing time.

To check which frames are key frames, we will ignore the frame with the similar descriptions of feature points to other frames. Our strategy in key frames selecting is keep the frames with most feature points and other frames will be invisible. But the ignored frames are not deleted from the data base, we will use them as the reference in video generation.

#### 3.2 Travelling Trajectory Generation

After extracting all key frames, the map trajectory can be generated by the Global Positioning System (GPS) information. There are many kinds of on-line map APIs are released in the recent years. Such as Google Maps, Virtual Earth and Yahoo Maps are widely used in various applications. Several problems will be considered in choosing suitable API to process map data, (1) the map display, (2) image data set, and (3)

address location. Make comparison to the published Map APIs, we select Google Maps as our experimental tool.

There are some definitions in our system when we extract the GPS information from the photos,

**GPS Record:** There are four tuples  $(i, t, x, y)$  used in our experiment, where  $i$  is the file id from user's devices,  $t$  is the timestamp,  $x$  and  $y$  are the Euclidean coordinates.

**Trajectory:** A trajectory  $T$  is sequences of GPS information of input photos or key frames. It can be represent as  $T = G_1 \rightarrow G_2 \rightarrow \dots \rightarrow G_n$ , where  $G$  is the GPS information of the key frames. A trajectory describes the movement of a user. The original GPS records can be seen as the set of the trajectories of all the users.

We would like to let users receive the trajectory information automatically, therefore the Map API is used to generated the data.

### 3.3 The Video Generation

After getting the feature points of each frame, we can refer to the information to generate a new video. In addition, the trajectory is also generated in the above section. It can help check the scene changed. There are two sub-steps in this section, the first is image-based rendering and the second is scene changed. However, the performance of the rotation computing is not good enough, we apply it into cloud computing.

#### Image-Based Rendering

In this section, we will compute the rotation of each continual frame. The information of feature points is computed in section 3.1, we can use the feature points to generate a homography matrix. As the homography property, we know it includes rotation and translation information of the camera, so through the decomposition of homography matrix, the angle of rotation can be extracted.

$$K^{-1} * H * K = [r_1, r_2, T] \quad (1)$$

$K$  is the intrinsic parameter,  $T$  is the translation matrix, and  $r_1, r_2$  is the two row vectors of rotation matrix  $R$ . We can use camera calibration by the chessboard to get intrinsic parameters  $K$ , but it changes with different cameras. We assume that the camera setting is the same, because we always use the same device during one trip. Therefore  $K$  is a fixed value, and can be reused.

We know that when an object rotates an angle  $\beta$  along the  $Y$  axis of a Cartesian coordinate system, the rotation matrix has a stationary formatting. Therefore, normalizing the rotation matrix  $R$  and we can get the angle of rotation  $\beta$ . This angle is the camera rotation angle between two frames. On the other hand, we can determine the camera position by  $\sin\beta$ . If the destination image is on right side of the source image,  $\sin\beta$  is positive. Otherwise, it is negative.

After the rotation is computed, the information can be used to generate the additional frames. The feature points and angles are used as the parameters. We employ view warping in this paper. As feature points and angles are known, view warping algorithm can generate several smooth frames.

### Scene Changed

In this section, we can know the scene is changed by the information of GPS or feature points. Therefore, we just use image morphing to generate a sequence of images.

## 4 Conclusion and Future Works

We proposed a video generation system in this paper. Users can update their photos with GPS information on the system. Our system will refer to the information of the photos to generate the corresponding data, such as feature points, trajectory and video. Our future work will focus on the performance of the system.

## References

1. Buehler, C., Bosse, M., McMillan, L., Gortler, S.J., Cohen, M.F.: Unstructured Lumigraph Rendering. In: SIGGRAPH 2001 Conference Proceedings. ACM Siggraph Annual Conference Series, pp. 425–432 (2001)
2. Debevec, P., Taylor, C., Malik, J.: Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach. In: SIGGRAPH 1996, pp. 11–20 (1996)
3. Levoy, M., Hanrahan, P.: Light field rendering. In: SIGGRAPH 1996, pp. 31–42 (1996)
4. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: SIGGRAPH 1996, pp. 43–54 (1996)
5. Chen, S.E., Williams, L.: View Interpolation for Image Synthesis. In: Proc. SIGGRAPH 1993, Computer Graphics Proceedings, Anaheim, California, August 1-6. Annual Conference Series, pp. 279–288. ACM SIGGRAPH (1993)
6. Seitz, S., Dyer, C.: View Morphing. In: SIGGRAPH (1996)
7. Morel, J.M., Yu, G.: ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences* 2(2), 438–469 (2009)
8. Adelson, E.H., Bergen, J.R.: The plenoptic function and the elements of early vision. In: *Computational Models of Visual Processing*, pp. 3–20 (1991)
9. Brown, M., Lowe, D.G.: Recognising Panoramas. In: *Proceedings of the 9th International Conference on Computer Vision (ICCV 2003)*, Nice, France, pp. 1218–1225 (2003)
10. Blinn, J.F., Newell, M.E.: Texture and Reflection in Computer Generated Images. In: *CACM*, pp. 542–547 (1976)
11. Debevec, P.E., Yu, Y., Borshukov, G.D.: Efficient view-dependent image-based rendering with projective texture mapping. In: *Eurographics Rendering Workshop 1998*, pp. 105–116 (1998)

# Linked Data-Based Service Publication for Service Clustering

Incheon Paik, Wuhui Chen, Banage T.G.S. Kumara,  
Takazumi Tanaka, Zhenni Li, and Yuichi Yaguchi

The School of Computer Science and Engineering  
The University of Aizu, Japan  
paikic@u-aizu.ac.jp,  
{chenwuhui21,btgsk2000,lizhenni2012}@gmail.com,  
himura.eco.ttt.sw@nifty.com,  
yaguchi@u-aizu.ac.jp

**Abstract.** In this paper, we propose an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. First, we propose Linked social service which is published on the open web by following Linked data principles with social link, then, a spatial clustering algorithm is proposed to enable visualization for reducing the using thresholds. Finally, experiment is conducted to show the effectiveness of our proposed approach.

## 1 Introduction

Web services have been considered to have a tremendous impact on the web, as a potential solution for supporting a distributed service-based economy on a global scale. However, despite outstanding progress, uptake on a Web scale has been significantly less than initially anticipated. From investigation in several technological perspectives of Web services, the reasons can be mainly described by the following:

First, a lack of available and ubiquitous ontologies for service annotation results in higher using threshold for service provider in service publication stage. To better support service discovery, composition and execution, Semantic web services has been proposed as a key to maximize a higher level of automation by enriching services with semantic annotation and has already shown their benefits [1]. However, up until now, the impact of Semantic web services on the open web has been minimal due to lack of available and ubiquitous ontologies for service annotation in service publication stage. In current ontology engineering field, there is still a large deficiency of uniform and ubiquitous ontologies in many application domains. This is due to the fact that creating ontology requires many engineers to cooperate with each other.

Second, traditional approaches have not provided visualization of the clusters. In traditional algorithms there is not any method to get the measurement or clue to identify the density variation within cluster and cluster position relative to the other clusters on the space. Another issue of traditional algorithms is, in iterative steps these

algorithms consider about the similarity of limited number of services (e.g., similarities of cluster representatives like cluster centers of intermediate clusters). So if there are any false positive members in intermediate clusters, then it will affect to the cluster performance. Furthermore traditional clustering algorithms are failed to achieve higher noise isolation.

In order to address the aforementioned issues, we propose an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. To reduce user's usage thresholds with service consumer, we apply spatial clustering technique called the Associated Keyword Space (ASKS) [2] with projection from a 3D sphere to a 2D spherical surface for 2D visualization. To support the semantic service annotation, Linked social service is built on a web of data which is an outstanding body of knowledge (light weight ontologies and data expressed in their terms) that can help to significantly reduce the effort for creating semantic annotations for services.

The remainder of this paper is structured as follows: in Section 2 we propose Linked social service to connected distributed services with social link. In Section 3, we propose a hybrid method to measure similarity of terms based on the property of Linked data. Then in Section 4 a spatial clustering algorithm is proposed to enable visualization for reducing the using thresholds. And then the evaluations of effectiveness of our approach are done in Section 5. The final section gives the conclusion and future work.

## 2 Linked Social Service

The advent of the Web of Data together with social principles can constitute the final necessary ingredients that will ultimately lead to a widespread adoption of services on the Web. Firstly, the evolution of the Web of Data is highlighting the fact that light weight semantics yield significant benefits that justify the investment in annotating data and deploying the necessary machinery. This initiative is contributing to generating an outstanding body of knowledge (light weight ontologies and data expressed in their terms) that can help to significantly reduce the effort for creating semantic annotations for services. Secondly, the recent evolution around Linked data has shown that linking data over the Web can lead to large quantities of very useful data with a low cost. Rather than isolated data islands, connecting distributed structured data into a single data space can lead to reused data, discover data from relevant data and integrate data from large numbers of formerly unknown data sources. This new scenario provides suitable technologies and data, as well as the necessary economic and social interest for the wide application of services technologies on a Web scale.

Our previous work [3] proposed Linked social service to construct a global social service network based on linked data principle for better quality of service discovery and service composition. In the global social service network, services described in lightweight ontologies are interlinked to related services from different sources functionally across the Web and in turn external services may link to them functionally using social link. However, our previous work has ignored the calculation of service

similarity and service cluster with visualization for reducing the using thresholds. In this paper, we focus on service discovery by service cluster with visualization for reducing the using thresholds.

### 3 Interlinking Term with Similarity

Currently, very large data sets have been published as linked data. Linked data is technology that permits mechanical collection of data of web [4], and they make semantic link and network. These links can be used for link based mining and semantic data mining. Calculating similarity using linked data can consider relationship and words type which are semantic information.

In this section, we propose a novel method to calculate term similarity using linked data. Calculation of words similarity by link based approach need the linked data set containing many words data and links. We use DBpedia [5] to calculate words similarities. DBpedia is a project to convert Wikipedia contents to linked data. The English version DBpedia offers 3.77 million things data. We can access to DBpedia dataset at online using database language such as SQL.

#### 3.1 Similarity of Nodes

This subsection explains our proposed calculation method using linked data. This method has 3 parts. First part is number of paths in shortest path. When number of paths is low, similarity is high. Second part is similarity by property and third part is similarity by structure of nodes. Final similarity is addition of 3 parts shown in the following formula:

$$\begin{aligned} \text{Similarity}(A, B) = & L(A, B) + \text{SimP}(A, B) \\ & + \text{SimO}(A, B) \end{aligned}$$

In this formula,  $\text{Similarity}(A, B)$  is similarity between the word A and the word B.  $L(A, B)$  is number of paths at the shortest path.  $\text{SimP}(A, B)$  and  $\text{SimO}(A, B)$  will be explained in later sections 3.1 and 3.2.

DBpedia dataset have some types. We take Raw Infobox Properties (information in wikipedia infobox) in DBpedia dataset types. Link properties of linked Raw Infobox Properties data nodes by the given word data node are compared to link properties of linked Raw Infobox Properties data nodes by the other word data node. Words similarity is high if there are many same link properties. But too common properties are cut (for example name, text). Additionally when the word data node has many Raw Infobox Properties data nodes links, one link's importance is low.

They are Links of other data nodes which Apple or Banana has. For example, The Apple data node have links of Species, familia , unrankedDivisio, and phosphorusMg. Apple and Banana have Species property and PotassiumMg property concurrently. So they are similar. A more detail is as following formula:

$$\text{SimP}(A, B) = \frac{RIP(A \cap B)}{RIP(A) + RIP(B)}$$

In this formula,  $SimP(A,B)$  is similarity by property.  $RIP(A)$  is Raw Infobox Properties in the word A.

### 3.2 Similarity of Nodes

In this subsection, we focus on Ontology Infobox Types. This method compares data nodes having Ontology Infobox Types property between word nodes. The calculating method is shown as the following formula:

$$SimO(A, B) = \frac{OIT(A \cap B)}{OIT(A) + OIT(B)}$$

In this formula,  $SimO(A,B)$  is similarity using Ontology Infobox Types.  $OIT(A)$  is Ontology Infobox Types in the word B. For example, Apple and Banana have links of [Plant<http://umbel.org/umbel/rc/Plant>](http://umbel.org/umbel/rc/Plant) and [EukaryoticCell<http://umbel.org/umbel/rc/EukaryoticCell>](http://umbel.org/umbel/rc/EukaryoticCell) concurrently.

## 4 Service Clustering with Similarity

Current clustering approaches use traditional clustering algorithms such as agglomerative and k-means as the clustering algorithms. Traditional approaches have not provided visualization of the clusters. They show clusters with service groups on distance base. The conceptual clusters are mainly useful for machine. But visualization helps for human's manipulation of the service clusters and gives inspiration for a specific domain from visual feedback. In this research we apply spatial clustering technique called Spherical Associated Keyword Space (SASKS) which we proposed in our previous work [2]. SASKS algorithm is modified version of Associated Keyword Space (ASKS). ASKS is an extended multidimensional scaling algorithm [6] and able to represent services in 3-D space by using the service's similarity. Another advantage of ASKS is that, it can achieve higher noise isolation. This result to increase the precision of service clusters.

### 1) Distance Measure of SASKS

Let  $k$  denote the dimension of the space in which services are located. Distance between two services is given by  $D_{ij}$ ;

$$D_{ij} = -f(x_j^{(k)} - x_i^{(k)}) \quad (1)$$

Where  $x_i$  and  $x_j$  are locations of the service  $i$  and  $j$  respectively.  $f$  has a parameter  $a$  and is defined using (2).

$$f(x) = \begin{cases} |x|^2, & |x| < a \\ 2a|x| - a^2, & |x| \geq a \end{cases} \quad (2)$$

Where parameter  $a$  is density control parameter. Clustering efficiency and the calculation load are both strongly influenced by the parameter  $a$ . Figure. 1 shows nonlinear distance function  $f(x)$  used in ASKS.

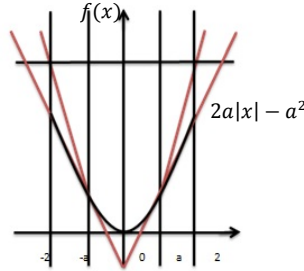


Fig. 1. Distance function  $f(x)$  used in SASKS

**2) Iterative Solution of Nonlinear Optimization**

The criterion function of SASKS is given by (3).

$$J(x_1, x_2, \dots, x_n) = \sum_i \sum_j \{-M_{ij} f(x_j^{(k)} - x_i^{(k)})\} \rightarrow \max \tag{3}$$

Where  $M_{ij}$  is the affinity value between services  $i$  and  $j$ . Here we use similarity score between services as the affinity value. The partial derivative of  $J$  with respect to  $x_i$  provides the formula for determining the values of  $x_i$  that maximize  $J$ :

The following (4) iterative computation converges to the solution  $x_i$ :  $ij = 1, 2, \dots, n$ ,  $k = 1, 2, 3, \dots, p$  and  $t = 1, 2, \dots$

$$x_i^{(k)}(t + 1) = \frac{\sum_{j=1}^n M_{ij} \{D(x_j^{(k)}(t) - x_i^{(k)}(t))(x_j^{(k)}(t))\}}{\sum_{j=1}^n M_{ij} D(x_j^{(k)}(t) - x_i^{(k)}(t))} \tag{4}$$

Following three constraints must be enforced at each step of the iterative computation for all service locations,  $x_i$  ( $i = 1, 2, \dots, n$ ).

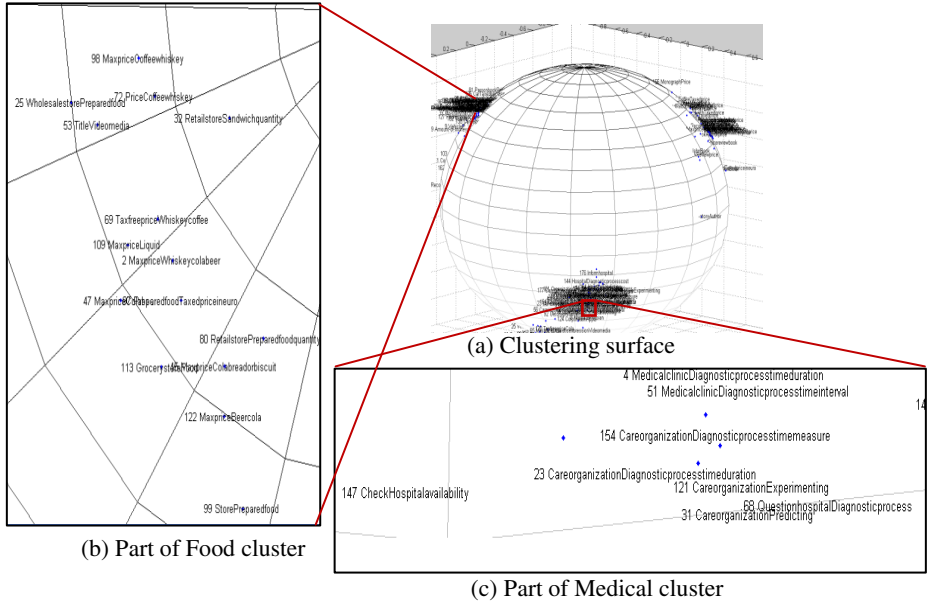
- make the original point as the center of gravity for the services;
- obtain covariance matrices such that dispersion in any direction creates the same value;
- Uniformalize the services radially from the origin.

Uniformalization is useful for clustering noisy data and otherwise tends to distribute the connections too evenly across the data.

**5 Evaluation**

This experiment was conducted on Microsoft Windows 7, Intel core i7-3770, 3.40 GHz and 4GB RAM. ASKS algorithm was implemented using MATLAB. WSDL documents were gathered from the real-world Web service providers and Web service repositories. We performed manual classification in order to categorize the Web service data set to compare the results. Book, Medical, Food, Film and Vehicle were the identified categorizes. As we mentioned, cluster efficiency strongly influenced by the density control parameter  $a$  in (2). We have done the experiments with  $a = 0.2$  and with 100 iterations.





**Fig. 2.** Result of spatial clustering and visualization

Figure. 2 shows result of spatial clustering approach. On the spherical surface the services are distributed according to their similarity. When analyzing the spherical surface, we observed five main regions where services are placed and we show that similar services in same domain are placed into one region. We observed clear separation of regions and density variation of services within the region that can be considered these regions as service clusters.

Highlighted areas in Fig. 2(a) show some similar services. Fig. 2(b) and Fig.2(c) show parts of Food and Medical clusters respectively. Highlighted area in Fig. 2(c) shows that more similar services are placed in same area within the cluster. For example CareorganizationExperimenting service is placed inside the cluster more closely to the careorganizationDiagnosticprocesstimeduration and careorganizationpredicting service than checkHospitalavailability service.

## 6 Conclusion

In order to reduce the using thresholds for both service provider and service consumer, we proposed an approach to publish services based on Linked data principles and discover services by service cluster with visualization for reducing the using thresholds. Linked social service was proposed to publish service on the open web by following Linked data principles with social link. Then, a spatial clustering algorithm

was proposed to enable visualization for reducing the using thresholds. Finally, experiment was conducted to show the effectiveness of our proposed approach. In the future work, by connecting isolated service islands or services repositories into service social network, we expect that our approach can make service requirement for service-based economy at global scale clear so that our approach can impel service providers to publish their services on the web as a piece of service social network and motivate service consumer to use services from global social service network.

## References

1. Pilioura, T., Tsalgatidou, A.: Unified Publication and Discovery of Semantic Web Services. *ACM Trans. Web* 3, 1–44 (2009)
2. Yaguchi, Y., Oka, R.: Spherical visualization of image data with clustering. In: 2012 4th International Conference on Awareness Science and Technology (iCAST), pp. 200–206. IEEE (2012)
3. Chen, W., Paik, I.: Improving Efficiency of Service Discovery Using Linked Data-based Service Publication. *Inf. Syst. Front.* (September 2012), doi:10.1007/s10796-012-9381-x
4. Bizer, C., Heath, T., Lee, T.B.: Linked Data: Principles and State of the Art. 17th Int'l World Wide Web Conf. (2008)
5. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A Nucleus for a Web of Open Data. In: Proc. 6th IEEE Int'l Semantic Web Conf., pp. 722–735 (November 2008)
6. Platzer, C., Rosenberg, F., Dustdar, S.: Web service clustering using multidimensional angles as proximity measures. *ACM Transactions on Internet Technology (TOIT)* 9(3), 1–26 (2009)

# The Practical Quality Model for Cloud Learning System

Anna Kang<sup>1</sup>, Leonard Barolli<sup>2</sup>, HwaYoung Jeong<sup>3,\*</sup>,  
JongHyuk Park<sup>4</sup>, and Hae-Gill Choi<sup>5</sup>

<sup>1</sup> Dongguk University, Korea

<sup>2</sup> Department of Information and Communication Engineering,  
Fukuoka Institute of Technology (FIT), 3-30-1 Wajiro-Higashi,  
Higashi-Ku, Fukuoka 811-0295, Japan

<sup>3</sup> Humanitas College of Kyunghee University, Hoegi-dong, Seoul, 130-701, Korea

<sup>4</sup> Department of Computer Science and Engineering,  
Seoul National University of Science and Technology

<sup>5</sup> Dept. of Information and Communication, Kyunghee Cyber University,  
Hoegi-dong, Seoul, 130-701, Korea

anakang37@gmail.com, barolli@fit.ac.jp, hyjeong@khu.ac.kr,  
parkjonghyuk1@hotmail.com, hgchoi@khcu.ac.kr

**Abstract.** Learning system has increase user needs with many and various request for more convenient and efficient learning performance. In order to provide this point to user, the develop method for learning system needs to change to more recent it and to apply the characteristics fitting on their system. However the research to identify the quality factors for cloud learning system is very lacked. In this paper, we propose the practical quality model for cloud learning system. For this purpose, we identify their criteria from the existing one, IS (information system) factors from Mclean and Dealon Model. By this proposal, we want to suggest basic criteria fitting on the important factor for cloud learning system.

**Keywords:** QoS model, cloud learning system, cloud computing, quality attributes, Mclean and Dealon model.

## 1 Introduction

Many information science researchers contend that service quality is very important variable. It affects the critical factor for system success. As competition in the information system service industry grows and the managers have to justify the cost of information systems, it is very critical that reliable instruments be developed to measure both service quality and system success [1]. In this manner, in 1992, DeLone and McLean (D&M) proposed the success model for IS (information system) as the dependent variable of the field. Their research of the literature resulted in a taxonomy of IS success consisting of six factors: System Quality, Information Quality, Use,

---

\* Corresponding author.

User Satisfaction, Individual Impact, and Organizational Impact. The model also addressed the various relationships among these success factors; but, many researchers announced to request that the model needed "further development and validation" [2]. They aimed their research for four conclusions as follows.

1. The researcher has a broad list of individual dependent variables to choose from IS.
2. Significant reductions in the number of different dependent variable measures are needed so that research results can be compared.
3. There are very few IS field study research that try to measure the influence of the IS effort on organizational performance.
4. IS success is a multidimensional construct and needs to be measured.

In the system development environment, various and high techniques are approving and proposing recently such as cloud computing, ubiquitous computing, context awareness and so on. Especially, cloud computing is to provide computing technology that is to support convenient, on-demand, network based service application such as service networks, service servers, storage, applications, and their interaction services between the user and service provider [3].

In this manner, learning system area needs to change develop method of their system for more efficient approach for users (learners, teachers and developers). However, quality model for cloud learning system lacked.

In this paper, we propose important attributes for quality of cloud learning system. For this purpose, we identify the criteria for cloud computing and consider the correlation between their factors. And we extract the criteria from the factors of DeLone and McLean (D&M) model to fit on learning characteristics.

## **2 DeLone and McLean (D&M) Model and Cloud Learning System**

### **2.1 DeLone and McLean (D&M) Model**

DeLone and McLean proposed D&M model to identified categories for informatio system success. Their suggestion was six factors of IS success: System Quality, Information Quality, Use, User Satisfaction, Individual Impact, and Organizational Impact. System Quality was measured to the technical level of communication, and Information Quality was equivalent to the semantic level of communication. Use related to "receipt of information." User Satisfaction and Individual Impact were associated with the "information's influence on the recipient." Organizational Impact means the "influence of the information on the system." Fig 1 shows the original DeLone and McLean model [1].

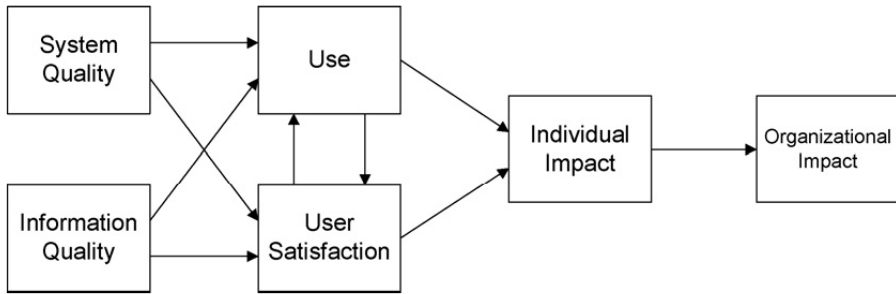


Fig. 1. DeLone and McLean's original model for IS success

Later, they updated their system model for more detail. The main changes concerned quality, and service quality, Intention to Use was included in the model. The final factors were information quality, systems quality, service quality, intention to use, use, user satisfaction and net benefits as shown in Fig 2 [4, 5].

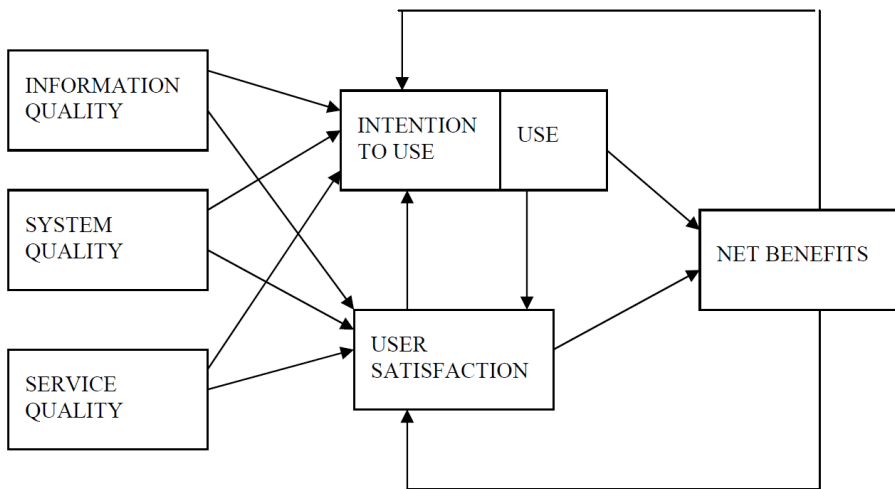


Fig. 2. Updated DeLone and McLean's model for IS success [4]

## 2.2 Cloud Learning System

Cloud computing technology has increasing by many enterprises and organizations. It basically involves a various independent technologies such as hardware devices, virtualizations, distributed processing, network system, web services, and so on [5].

E-learning system offers new possibilities for learning process and leads to dramatic changes in education practice. This changes affect not only the educational institutions but also the enterprise efficiencies [7]. Therefore education researchers proposed the develop method for learning system that is to apply cloud computing

technique. Mohd [6] presented cloud learning system structure with LCMS (Learning Content Management System) as shown in Fig 3. But it was very simple and was not enough to show the cloud learning structure.

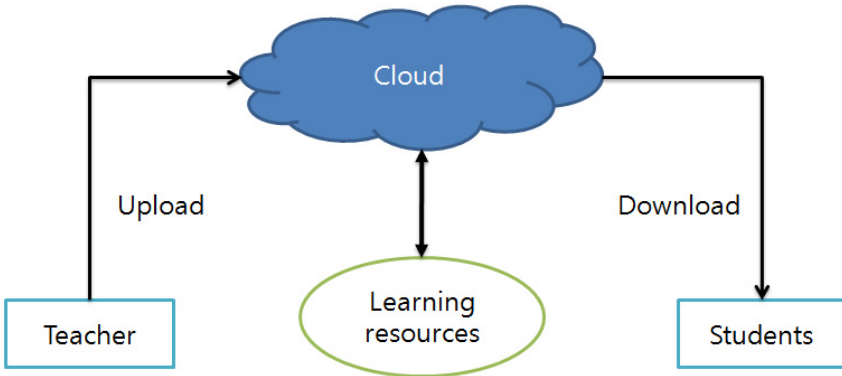


Fig. 3. Cloud learning system [6]

### 3 The Quality Attribute for Cloud Computing

This paper aims to identify the quality model for cloud learning system. For this process, we extract some criteria from exist research (Fig 4) and modify their system

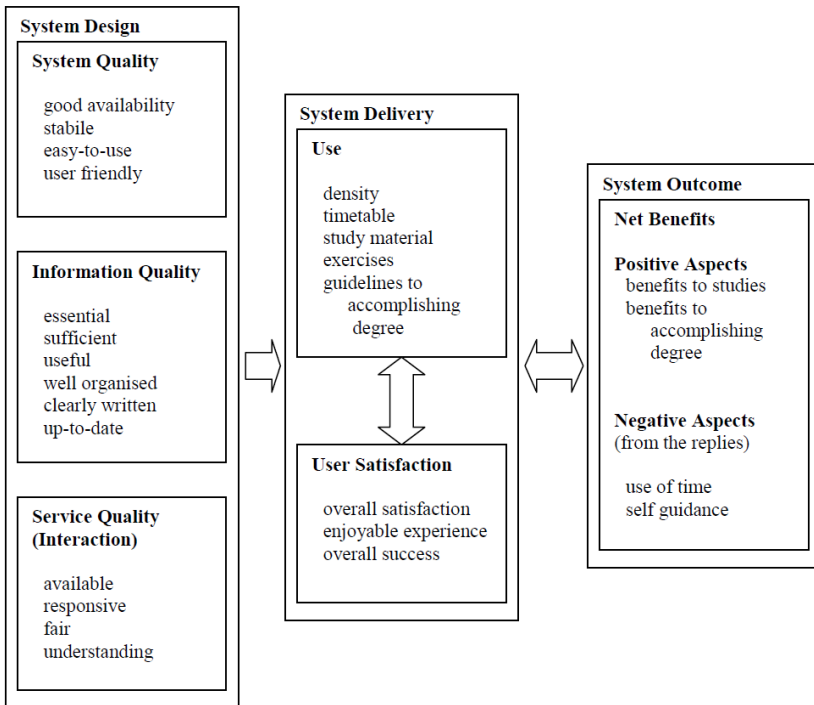


Fig. 4. The quality model for virtual learning system by Halonen et al [5]

structure. Halonen et al [5] presented virtual learning system in cloud computing as shown in Fig 4.

By their research, we propose cloud learning system architecture as shown in Fig 5. In this system, System Quality consist of 5 factors, User friendly, Easy to use, Changeability, Operability and Maintainability. Information Quality has Accuracy, Usefulness, Efficiency and Correctness. And finally, Service Quality has Usability, Availability, Adaptability, Functionality, Interoperability, Completeness, Reliability, and Suitability.

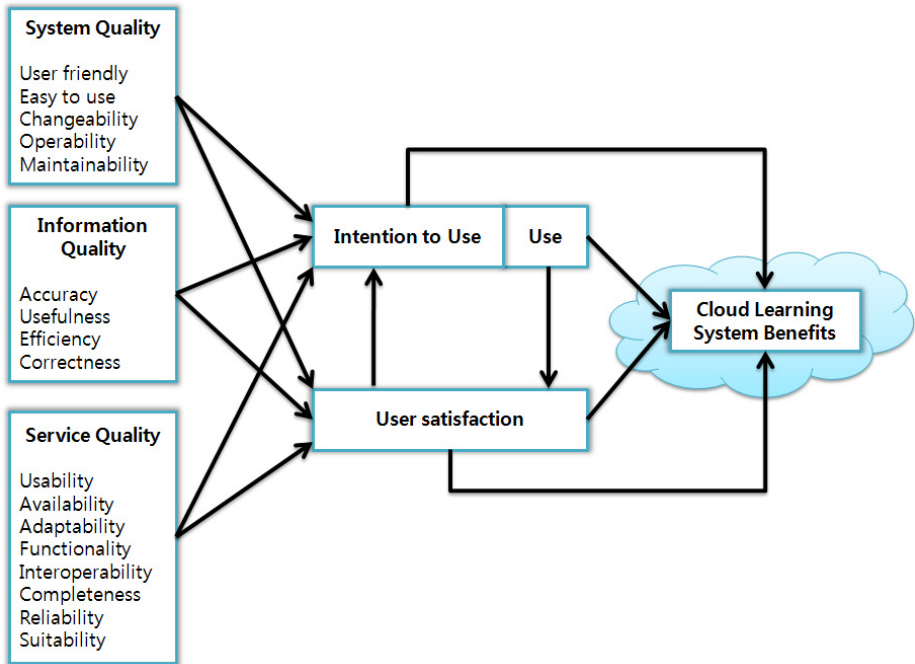


Fig. 5. The quality model for cloud learning system

## 4 Conclusion

In this paper we proposed quality model for cloud learning system. The qualities consist of 3 factors such as System Quality, Information Quality, and Service Quality. System Quality has 5 factors; User friendly, Easy to use, Changeability, Operability and Maintainability. Information Quality has 4 factors; Accuracy, Usefulness, Efficiency and Correctness. And finally, Service Quality has 8 factors; Usability, Availability, Adaptability, Functionality, Interoperability, Completeness, Reliability, and Suitability. All the criteria affect to user satisfaction and to cloud learning system benefits.

**Acknowledgment.** This research supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2013-H0301-13-4007) supervised by the NIPA (National IT Industry Promotion Agency).

## References

- [1] Petter, S., McLean, E.R.: A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information & Management* 46, 159–166 (2009)
- [2] Landrum, H., Prybutok, V.R.: A service quality and success model for the information service industry. *European Journal of Operational Research* 156, 628–642 (2004)
- [3] Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., Rajarajan, M.: A survey of intrusion detection techniques in Cloud. *Journal of Network and Computer Applications* 36, 42–57 (2013)
- [4] DeLone, W.H., McLean, E.R.: Measuring e-commerce success: applying the DeLone & McLean Information systems success mode. *International Journal of Electronic Commerce* 9(1), 31–47 (2004)
- [5] Halonen, R., Acton, T., Golden, W., Conboy, K.: DeLone & McLean success model as a descriptive tool in evaluating the use of a virtual learning environment. In: *International Conference on Organizational Learning, Knowledge and Capabilities* (2009)
- [6] Murah, M.Z.: Teaching and learning cloud computing. *Procedia - Social and Behavioral Sciences* 59, 157–163 (2012)
- [7] Jia, H., et al.: Design of a performance-oriented workplace e-learning system using ontology. *Expert Systems with Applications* (2010), doi:10.1016/j.eswa.2010.08.122



# Realizing the Right to Be Forgotten in an SNS Environment

Cheol Ho Sin, Nam A. Kim, Byeong Woo Go, Kim Seong Min,  
Jae Dong Lee, and Jong Hyuk Park\*

Department of Computer Science and Engineering  
Seoul National University of Science and Technology  
{sch1992, skadkdla, terrtn90, enh1535,  
jdlee731, jhpark1}@seoultech.ac.kr

**Abstract.** As the right to be forgotten and privacy protection are becoming more important in modern society, the security threats that exist on Social Networking Services (SNS) create critical problems. Information leaked from security accidents on SNS is provided to a third party and used for malicious intentions, generating additional damages. This study discusses the concept of the right to be forgotten and the security threats on SNS that invade the privacy of an owner of information. In addition, it proposes a model that has User Privacy Policy (UPP), Dual Watermarking Scheme, and FaceCloak functions that provide strong information controllability so that we can resolve security threats and realize the right to be forgotten.

**Keywords:** SNS, privacy protection, the right to be forgotten.

## 1 Introduction

Along with the development of internet technology, Social Networking Services (SNS) became places for people around the world to communicate. Even though a lot of people post their everyday lives and private information on SNS, threats exist because of vulnerable security. When there is a security accident on SNS, an individual's private information is exposed. When it is exposed to an attacker, information can be provided to a third party without the owner's intention, and it can be used for harmful purposes that generate additional damages. It also causes serious privacy problems when information is recovered fully to its original form by combination and analogy [1, 2].

This study recognizes that this problem is caused by the fundamental characteristics of digital memory and suggests a technical resolution. This study suggests a model that resolves security threats on SNS by realizing the right to be forgotten.

The constitution of this study is as follows. Chapter 2, Related Works, discusses the concept and the range of the right to be forgotten, as well as the concepts behind

---

\* Corresponding author.

SNS, the security threats to SNS, and the technology that is used to protect privacy on SNS. Chapter 3 discusses the requirements for realizing the right to be forgotten and how to realize this right in an actual SNS environment. Chapter 4 provides a conclusion.

## 2 Related Works

### 2.1 The Right to Be Forgotten

The right to be forgotten became a hot issue as the EU announced General Data Protection Regulation, which contains this right as major concept, on January 25th, 2012. The right to be forgotten is the right to have access to one's own personal information dispersed on the internet without delay and to be able to correct or delete them. It is a wider concept compared to the right to oblivion, which prevents people from having access to an individual's criminal records or court records. The right to be forgotten includes a range of information, including not only personal information but all information related to an individual [3, 4].

### 2.2 Social Network Services (SNS)

Boyd et al. defined SNS as web-based service with the following functions: to open a profile or open it partially, to make a contacts list, and to search for other people's contacts. Facebook, MySpace, and Twitter are representative examples of SNS. Users put their information in the profile section and open it to the public to let people know about them and to make personal relationships. Relationship lists provide link that connect to the profiles of other people. This relationship list is, again, shared with others. In addition, a user can leave a message on other users' profiles.

### 2.3 Security Threats on SNS

The following threats exist on SNS:

- **The right to control one's information is lost:** It is difficult for an owner of information to recognize that the information that he or she provided to SNS is shared by strangers through the SNS's sharing functions. Information posted on SNS can be easily saved or captured and redistributed to others. Moreover, information necessary for network device operation, such as connect time, IP address, and visit history, is provided and can be collected regardless of an owner's intention. For such information, it is hard for an owner to perform his or her right to delete or correct [2].
- **Approach from an unauthorized third party:** It is not only personal information provided by the owner of the information that can be collected and used by an unauthorized third party, but also individual information dispersed on SNS. This information can be recovered fully to original information by combining with or inferring to other information. This may include information that the owner does not want to provide to others [2].

- **Security weakness:** Widgets provided by unverified third parties have security weaknesses and can be a target of XSS attack. Some cases reported that malicious codes were installed in the profiles of attacked widgets. Some left links that connected to spam or phishing sites by inviting random people or by posting articles using software. In addition, an SNS Aggregator, which manages several SNS all together, is operated by only one account [2].

## 2.4 Privacy Protection Technology on SNS

There are some technologies for privacy protection on SNS:

- **User Privacy Policy (UPP):** UPP is a technology that allows an SNS user to set a privacy policy for each piece of information that he or she posts on SNS. The owner of the information can control the information individually or generally according to privacy settings. UPP consists of a policy element that includes OWNER, RECEIVER and ACCESS-RIGHT. The policy and each element consist of 1:m. OWNER is the owner of the information and the receiver is a subject who approaches the information. The access right is a privacy setting for the information that the receiver approaches [6].
- **Dual Watermarking Scheme:** The dual watermarking scheme is Digital Right Management (DRM) technology that records the ownership of a file and records the information from the original file uploaded on SNS on a multimedia file using a watermarking technique. If there's no watermark when the owner of the information uploads a file, it inserts an ownership and semi-fragile watermark. When another person re-uploads the file with a watermark, it comprehends the owner of information and decides whether to allow it to be posted or not according to its privacy policy. After this, it checks to see if the file has been changed. If it has been changed, it notifies the owner of the information so that they can decide whether to allow it to be posted or not [7].
- **FaceCloak:** FaceCloak shows an unknown third party a fake profile to protect the owner of the information's privacy on SNS. It is basically operated by three keys: the master key, the personal index key, and the access key. The master key and the personal index key are shared with the friends of the owner of the information, while the access key belongs only to the owner. The owner of the information encrypts and stores the actual information to the third server. Then the owner saves the index value using personal index key and provide fake information on facebook. Users sharing the owner's master key and personal index key search encrypted actual information by using personal index key and read decrypted actual information by using master key. On the other hand, to the third party that doesn't have master key and personal index key, fake profiles that are stored on facebook becomes visible. It should have the access key when the owner of the information to save or modify the new information to the third server [1].

### 3 How to Realize the Right to Be Forgotten in an SNS Environment

#### 3.1 Consideration When Realizing the Right to Be Forgotten

- **Right to control one’s own information:** The owner of the information should be able to know the distribution channel of his or her information and have easy access to the information for approaching, editing, and deleting. When the owner of the information has the right to control his or her own information, the right to be forgotten can also be realized naturally. Besides, SNS technology should provide the owner with easier and more intuitive methods for tracking and controlling his or her information.
- **Disallowance of unauthorized approach:** Many problems that occur on SNS, such as SNS spam, are generated from unauthorized approaches. To prevent such threats, SNS should be able to keep information safe from unauthorized approaches. A user without such authorization for the information should be kept from approaching, collecting, or using the information. In addition, SNS should provide alternatives for problems such as unauthorized copying.
- **Security:** SNS should provide a safe security system. All saved and transferred information should be processed using safe encoding algorithms and protected from script attacks, such as XSS. SNS servers should guarantee file integrity so that major files, such as source files, can be protected from malicious programs, such as viruses. It should also be able to maintain strict confidentiality and availability.

#### 3.2 Proposed Model

Figure 1 shows the model that realizes the right to be forgotten in an SNS environment.

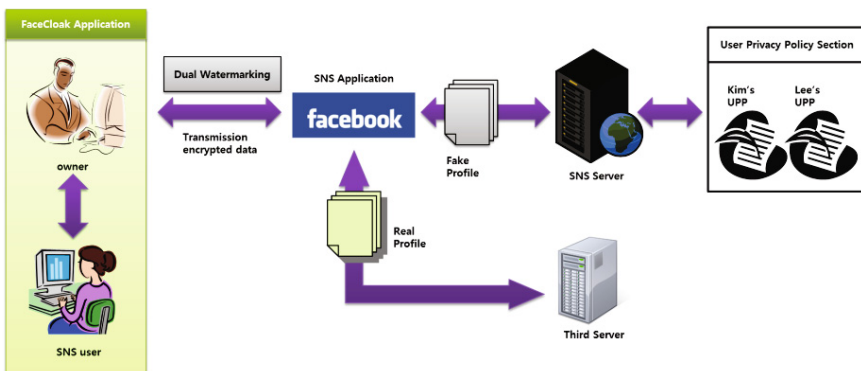


Fig. 1. Proposed model architecture

The owner is the owner of the information and the SNS user is the user of the SNS. The SNS application is an application that provides functions so that information can be posted and shown to users. The SNS server is where actual SNS data is saved and managed. FaceCloak saves a fake profile. A third server is where the actual profile is saved and managed through FaceCloak. The owner of the information can set the approachability and range of disclosure for each piece of information through UPP. When another user approaches the information, it understands the authorization of the user based on UPP and records the approach or sends a request for authorization to the owner of the information. This allows the owner of the information to understand how his or her information is collected, used, and distributed, which realizes the right to control one’s own information. The dual watermarking scheme protects the owner of the information from information distribution or distortion against the owner’s intention. Moreover, it provides an ownership search for a dual watermarked file, so that the owner can understand the information distribution process and easily approach, edit, and delete the information. In addition, tamper detecting using a semi-fragile watermark informs the owner of the information when the information is edited and asks for authorization. In this way, the owner can manage unwanted information editing or distortion. FaceCloak provides a fake profile to a third party to protect the owner of the information from the reckless collection of information. It also provides safe security, as it saves the encoded information on a third server.

### 3.3 Use Case Scenario

Figure 2 is a use case scenario for the suggested model.

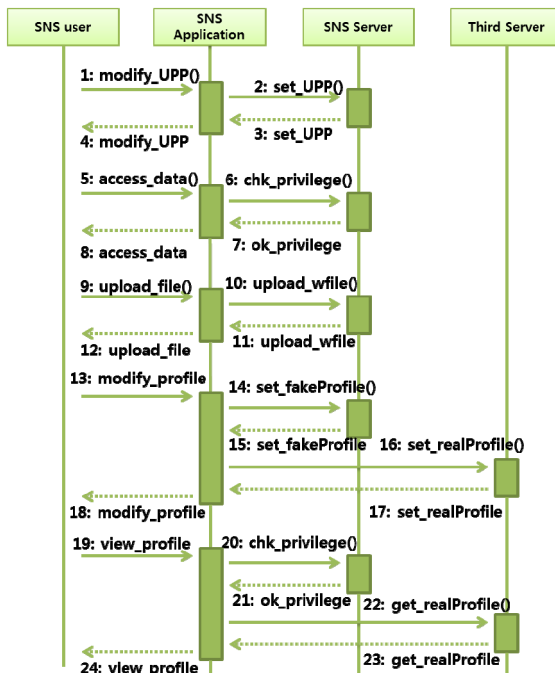


Fig. 2. Use Case Scenario

The process of this model is as follows:

- (1) SNS user → SNS application: modify\_UPP()  
The SNS user makes a request to the SNS application to change the detailed Privacy Settings on UPP.
- (2) SNS application → SNS Server: set\_UPP()  
The SNS application delivers to the SNS Server the UPP changes, saves them in the server, and returns the changed result to the SNS application and the SNS user.
- (3) SNS user → SNS application: access\_data()  
The SNS user makes a request to the SNS application to access the information.
- (4) SNS application → SNS Server: chk\_privilege()  
The SNS application delivers the request to access to the SNS server. The SNS server confirms access authority and returns the applicable items to the SNS application and the SNS user.
- (5) SNS user → SNS application: upload\_file()  
The SNS user makes a request to the SNS application for a file upload. The SNS application checks to see if it has dual watermarking. If not, it makes a watermark.
- (6) SNS application → SNS Server: upload\_wfile()  
The SNS application delivers the watermarked file to the SNS server. The SNS server then checks the owner of the information and the changed file uploading authority and uploads the file. The upload result is returned to the SNS user and the SNS application.
- (7) SNS user → SNS application: modify\_profile()  
The SNS user makes a request to change the profile by sending fake profile information and encoding the actual profile information and access key to the SNS application.
- (8) SNS application → SNS Server: set\_fakeProfile()  
The SNS application delivers and saves the fake profile information to the SNS server. The saved information is then returned to the SNS application.
- (9) SNS application → Third Server: set\_realProfile()  
The SNS application delivers and saves the access key and the actual profile information on the third server. The saved information is returned to the SNS application and the SNS user.
- (10) SNS user → SNS application: view\_profile()  
The SNS user makes a request to the SNS Application to view the profile.
- (11) SNS application → SNS Server: chk\_privilege()  
The SNS application checks with the SNS server to see if the SNS user has the authority to access. It returns a fake profile if the user does not have the authority. If the user has the authority, it receives the master key and the personal index key from the owner of the information.
- (12) SNS application → Third Server: get\_realProfile()  
The SNS application uses the personal index key to receive the coded profile from the third server and decodes it using the master key. The information is then delivered to the SNS user.

## 4 Conclusion

This study discusses security threats on SNS and suggests UPP, the dual watermarking scheme, and FaceCloak as alternatives to resolve these threats. It also suggests a model that realizes the right to be forgotten in an SNS environment. This model manages information access in detail through UPP and guarantees the user the right to control his or her own information by notifying the owner of the information of the distribution process of the information and the file and allowing control over the information, while also providing dual watermarking scheme technology. In addition, FaceCloak provides safe security by managing the exposure of information to a third party and by encoding information. Moreover, applying information expiration dates in this model can prevent the leakage of old information that the owner of the information has not managed for a long time. The proposed model allows the owner of the information to realize the right to be forgotten by providing the right to control his or her own information and by controlling unauthorized approaches to ensure security. However, many studies on the right to be forgotten say that the right cannot be realized with technological measures. Therefore, there should be social systems that can be connected to the proposed model. Above all, the owner of the information should recognize his or her rights as an owner as well as the risks arising from providing information.

**Acknowledgements.** This work was supported by the Korea Foundation for the Advancement of Science and Creativity (KOFAC) grant funded by the Korean government (MEST).

## References

1. Luo, W., Xie, Q., Hengartner, U.: FaceCloak: An Architecture for User Privacy on Social Networking Sites. In: 2009 International Conference on Computational Science and Engineering, p. 26, 29–30. IEEE Computer Society (2009)
2. Hogben, G.: ENISA Position Paper No.1 Security Issues and Recommendations for Online Social Networks. In: ENISA, pp. 8–16 (2007)
3. Castellano, P.S.: The Right to Be Forgotten under European Law a Constitutional Debate. *Lex Electronica* 16(1), 18 (Hiver/Winter 2012)
4. Druschel, P., Backes, M., Tirtea, R.: The right to be forgotten - between expectations and practice. In: ENISA, p. 3 (2011)
5. Danah, M., Boyd, N.B.: Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 210–230 (2008); 2008 International Communication Association, 211 (2008)
6. Aïmeur1, E., Gambs, S., Ho, A.: UPP: User Privacy Policy for Social Networking Sites. In: 2009 Fourth International Conference on Internet and Web Applications and Services, pp. 269–271. IEEE Computer Society (2009)
7. Zigomitos, A., Papageorgiou, A., Patsakis, C.: Social Network Content Management through Watermarking. In: 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 1383–1385. IEEE Computer Society (2012)

# Author Index

- Abdulkarim, Muhammad 1079  
Adnan, M.R.H. Mohd 627  
Afrianty, Iis 723  
Ahmad, Faudziah 989  
Ahmad, R.B. 661  
Ahmad, Wan Fatimah Wan 1079  
Airuddin, Ahmad 989  
Alam, Bashir 633  
Ali, Zaid G. 661  
An, Phan Thanh 179  
An, Zhulin 201  
Anh, Huynh Tuan 89  
Anh, Nguyen Phuong 95  
Anitha, R. 1201  
Anusha, Kannan 1219  
Apirakkan, Orapan 743
- Bae, Ji-Hye 597, 1241  
Bae, Nam-Jin 321, 327  
Baharudin, Baharum B. 217, 1145  
Bandara, Kasun 249, 459  
Bang, Hyo-Chan 385, 411  
Barolli, Leonard 369, 467, 1437  
Bi, Juan 709  
Bin, Xie 981  
Bing, Han 893  
Binh, Huynh Thi Thanh 333, 345  
Biswas, Siddhartha Sankar 633  
Bo, Wang 893
- Cai, Wandong 241  
Cai, Zhonghua 751  
Callaghan, Michael 655  
Cao, Xi-zheng 997  
Cha, Hyun-Jong 531, 1365
- Chang, Chiao-Min 1411  
Chang, Hsuan-Pu 1417  
Chang, Wei-Ting 1417  
Chao2, Xu 901  
Chen, Aiwang 241  
Chen3, Chen 901  
Chen, Chen-Hsiang 677  
Chen, Chih Chung 1265  
Chen, Chin-Fa 685  
Chen, Dongxiang 1151  
Chen, Guirong 241  
Chen, Hua-Ping 853, 1045  
Chen, Li-Wen 885  
Chen, Ping Kuo 1265  
Chen, Rui 783  
Chen, Wuhui 1429  
Chen, Xiang-Lan 853, 1045  
Chen, Yong 967  
Chen, Zhe 1109  
Cheng, Qiang 837  
Chipulis, Valeri 7  
Chitra, S. 1225  
Cho, Chang-Suk 115  
Cho, Kum Won 1185, 1195  
Cho, Kyungryong 321  
Cho, Moon-Taek 497, 503, 559, 565,  
609, 1325, 1331, 1393, 1399  
Cho, Seongsoo 509, 1337, 1343, 1405  
Cho, Yong-Yun 301, 309, 321, 327  
Choi, Byeong-Cheol 161  
Choi, Chang-Hoon 551, 1385  
Choi, Hae-Gill 1437  
Choi, Ikhwang 1185  
Choi, Mi-Jung 173



- Choi, Min 603  
 Chou, Chun-Mei 769  
 Chowdhury, Olly Roy 301, 309  
 Chun, Ji-Yong 565, 1399  
 Chun, Se-Hak 1251  
 Chung, Ji Ryang 353  
 Chung, Yeon-ho 459  
 Chung, Y.H. 249  
 Cullen, Mark 655
- Dai, Feng 843  
 Dang, Duc-Tai 685  
 Deivanai, Kathir 1235  
 Deng, Yubo 869  
 Desruelle, Heiko 13  
 Diao, Boyu 201  
 Dinh-Duc, Anh-Vu 763  
 Do, Luu-Ngoc 71  
 Do, Phung 339  
 Doja, M.N. 633  
 Dong, Li 877  
 Dong, Xin 1109  
 Dong Lee, Jae 467, 1443  
 Dongri, Yang 1023  
 Duan, Ping 1061  
 Duan, Xiaolian 967  
 Dung, Tran Nam 339
- Fan, Hong 649  
 Fan, Yachun 131  
 Feng, Shuguang 83, 147  
 Fong, Shi-Pin 1417
- Geyi, Liu 911  
 Ghazali, Rozaida 153  
 Ghorashi, Seyed Hamid 65  
 Gielen, Frank 13  
 Girish, G.P. 1129  
 Go, Byeong Woo 1443  
 Gong, Jian 1031  
 Gu, Peihua 837  
 Gu, Xingsheng 811  
 Guo, Ga-ifang 997  
 Guo, Wen 843
- Hai, Nguyen Ngoc 179  
 Hajmohammadi, Mohammad Sadegh 1053  
 Han, Seung-Wan 161  
 Haron, Habibollah 627, 723
- Hasbullah, Halabi 1137  
 Hassan, M. Fadzil 829  
 Hassim, Yana Mazwin Mohmad 153  
 Hassnawi, L.A. 661  
 Hayakawa, Tomokazu 225  
 He, Wei-ping 997  
 Hien, Vu Thanh 89  
 Hikita, Teruo 225  
 Ho, Diem N. 763  
 Hong, Bong-Hwa 489, 509, 515, 583, 1317, 1349, 1405  
 Hong, Gwang-Soo 1, 293  
 Hong, Soon Goo 261  
 Hong, Suck-Joo 483, 1311  
 Hong, You-Sik 515, 551, 1349, 1385  
 Hooi, Yew Kwang 829  
 Hou, Miao 863  
 Hsiao, Hsi-Chi 769  
 Huang, Chiung En 1265  
 Huang, Jia 709  
 Huang, Xiaoping 621  
 Huashun, Li 757  
 Hui, Liu 901  
 Huimin, Deng 877  
 Huo, Bei 937  
 Husni, Husniza 1259  
 Hwang, Junsik 453  
 Hwang, Soon-Hak 1251
- Ibrahim, A.L. 695  
 Ibrahim, Roliana 1053  
 Idrus, Rosnah 643
- Jeon, Wanho 1185  
 Jeon, Woongryul 423, 435  
 Jeong, Byungjin 445  
 Jeong, Chang-Sung 571  
 Jeong, Heon 315  
 Jeong, HwaYoung 1437  
 Jeong, Ji-Seong 1171  
 Jeong, Yeon-Kyu 1  
 Jeong, Young-Sik 369, 603  
 Jia, Yuyue 967  
 Jiang, Xiao-Wu 1045  
 Jiang, Yuyu 1039  
 Jianhua, Hao 919  
 Jin, DuSeok 1195  
 Jinlian, Du 957  
 Jo, Eun Byeol 209  
 Joo, Hae-Jong 483, 489, 1311, 1317

- Juncheng, Pan 877  
 Jung, Chai Young 1171  
 Jung, Hanmin 1179  
 Jung, Hyosook 453  
 Jung, In-Yong 571  
 Jung, Joon-Yong 1171  
 Jung, Low Tang 217, 1137  
 Jung, Seung Eun 1171  
 Jung, Sungmo 59  
 Jung, Taedong 141  
 Jung, Youngman 391, 435  
  
 Kadir, Mohammed R.A. 723  
 Kang, An Na 467, 1437  
 Kang, Anna 369  
 Kang, Fengju 1121  
 Kang, Jinsuk 77  
 Kang, Min-Jeoung 1171  
 Kang, Won Min 467  
 Katsumata, Masashi 233  
 Ke, Chih-Kun 1411  
 Khang, Mai Trong 95  
 Kim, Baek-Ki 515, 551, 1349, 1385  
 Kim, Bum-Soo 173  
 Kim, Byung-Gyu 1, 293, 1241  
 Kim, Deokho 125  
 Kim, Do-Hyeong 1171  
 Kim, Donghoon 279  
 Kim, Donghyun 59  
 Kim, Dong-Ju 185, 255  
 Kim, Dong W. 523, 1357  
 Kim, Eung-Kon 315, 805  
 Kim, Geun-Hye 1171  
 Kim, Hea-Suk 167  
 Kim, Hong-Geun 301, 309  
 Kim, Hye-Sun 475, 1303  
 Kim, Hyoung-Ro 475, 1303  
 Kim, Hyunduk 185, 255  
 Kim, Hyun-Woo 369  
 Kim, Jeeyeon 391  
 Kim, Jeong-Nyeo 161  
 Kim, Jeong-Uk 51  
 Kim, Jeongyeun 411, 417  
 Kim, Jin-Mook 531, 537, 543, 1365,  
 1371, 1377  
 Kim, Jin-Sung 1171  
 Kim, Jiye 423, 435  
 Kim, June Young 577  
 Kim, Juyoung 379  
 Kim, Kyungah 125  
  
 Kim, Minwoo 125  
 Kim, Nam A. 1443  
 Kim, Ok-Hwan 497, 559, 609, 1325,  
 1393  
 Kim, Sang-Soo 489, 1317  
 Kim, Seoksoo 59  
 Kim, Sukil 1171  
 Kim, Sung-Ki 293, 597, 1241  
 Kim, Sung Min 209  
 Kim, Sunwoo 445  
 Kim, Tae-Jung 1  
 Kim, Yilip 385, 429  
 Kim, Yonggon 445  
 Kim, Young-Choon 497, 503, 559, 565,  
 609, 1325, 1331, 1393, 1399  
 Kim, Young-gi 509, 1405  
 Kim, Young-Kuk 51  
 Kim, Younho 1185  
 Ko, Yeonghae 405, 591  
 Koh, Yoon Mi 273  
 Kruahong, Thongchai 643  
 Kumara, Banage T.G.S. 1429  
 Kun, Li 1089  
 Kuznetsov, Roman 7  
 Kwak, Jin 285  
 Kwak, Kyung-Hun 321  
 Kwon, Kyung Hee 273  
 Kwon, Ohmin 445  
  
 Lakkhawannakun, Phoemporn 775  
 Lakshmi, M. 1211  
 Lawal, Ibrahim A. 1137  
 Le, Hiep 339  
 Le, Thanh Manh 615  
 Leanjay, Nattayanee 643  
 Lee, Byunggil 379  
 Lee, Guo Haur 729  
 Lee, Hanwook 435  
 Lee, Heng Wei 643  
 Lee, Ho-Dong 523, 1357  
 Lee, Hyun-Woo 603  
 Lee, Jae-Joon 77  
 Lee, Jongsup 1337  
 Lee, Jong-Yong 1343  
 Lee, Ju Hwan 209  
 Lee, Kyu-Chul 1195  
 Lee, Minwoo 167  
 Lee, Samuel Sangkon 363  
 Lee, Sang-Heon 185, 255  
 Lee, Sangho 1171, 1179

- Lee, Sanghun 173  
 Lee, Seungbock 1171, 1179  
 Lee, Seung Hyun 1337  
 Lee, Seungwoo 1171, 1179  
 Lee, Wong Poh 799  
 Lee, Wonjae 603  
 Lee, Yongmyoung 141  
 Lee, Youngsook 391, 423  
 Lee, Yunho 475, 1303  
 Le-Huu, Khoi-Nguyen 763  
 Ler, Shir Ni 791  
 Li, Huang 1073  
 Li, Jia 1061  
 Li, Peipei 201  
 Li, Shoupeng 869  
 Li, Tang Jia 669  
 Li, Yongda 621  
 Li, Zhenni 1429  
 Liang, Hongtao 1121  
 Lifei, Xing 957  
 Lijun, Hua 1073  
 Lim, Chang-Gyun 315  
 Lim, Chen Kim 703, 729  
 Lim, Jae-Deok 161  
 Lim, Jaesung 77  
 Lim, Seung-Ho 577  
 Lin, Haizhuo 1067  
 Lin, Iuon-Chang 677, 685  
 Linh, Pham Thuy 677  
 Linkui, Qu 757  
 Liu, Dongjing 649  
 Liu, Fangbing 43  
 Liu, Guangqing 751  
 Liu, Xin 1031  
 Liu, Zhen 1039  
 Liu, Zhifeng 837  
 Liwen, Chen 919  
 Lu, Chun 1031  
 Lu, Zhang 1159  
 Luan, Hua 131  
 Lv, Haiyang 1061
- Ma, Yong-Fei 885  
 Madhusudhanan, B. 1225  
 Mahmuddin, M. 735  
 Md. Said, A. 695  
 Meng, Li 1007  
 Meng, Mei 951  
 Min, Kim Seong 1443  
 Mingyi, Wu 911
- Mohammed, Athraa Jasim 1259  
 Moon, Jeong-Kyung 543, 1377  
 Moon, Yang-Sae 167, 173  
 Mu'azu, Abubakar Aminu 695, 1137  
 Mukherjee, Saswati 1201  
 Mun, Jongho 391, 423
- Nam, Dukyun 1185  
 Nam, Yoonho 435  
 Nan, Li 943  
 Nasien, Dewi 723  
 Ngan, Nguyen Hoang 95  
 Ngo, Son Hong 345  
 Nguyen, Vu Thanh 89, 95, 339  
 Nhat, Vo Viet Minh 107  
 Ning, Wang 757  
 Niroopan, Pararajasingam 249, 459  
 Nisar, K. 695
- Ogier, Jean-Marc 799  
 Oh, Yeon-Jae 805  
 Osman, Mohd Azam 643, 799
- Paik, Incheon 1429  
 Palma, Freida 261  
 Park, Chan-seob 293  
 Park, Dae-Heon 301, 309  
 Park, Dong-Hwan 385, 411  
 Park, Dong-Su 483, 1311  
 Park, Jang-Woo 301, 309, 321, 327  
 Park, Jong Hyuk 467, 571, 603, 1437, 1443  
 Park, Jong-Wook 523, 1357  
 Park, Jun-Young 115  
 Park, Kidam 391  
 Park, Kiseo 583  
 Park, Mi-Jeong 315  
 Park, Namje 379, 385, 405, 411, 417, 429, 591  
 Park, Seongbin 453  
 Park, Sung-Wook 523, 1357  
 Park, Sung Yun 209  
 Park, Sun-Rae 1195  
 Peizhang, Xie 1165  
 Peng, Liu 1023  
 Peng, Wang 757  
 Phaudphut, Comdet 191  
 Piao, Yuxue 19

- Qian, Ping 1039  
 Qiao, Liu 929  
 Qiao, Ruixuan 43  
 Qin, Rong 1151  
 Qin, Zhiguang 709  
 Qing, Hu 1073  
 Qiu, Tie 43  
 Quan-bin, Tang 943  
 Qun, Dai 1287  
 Quoc, Nguyen Hong 107  
 Quoc, Nguyen Trong 95
- Rajyalakshmi, G.V. 1219  
 Razali, Radzuan 1079  
 Ro, Won Woo 125  
 Ruilan, Zhang 1115  
 Rujirakul, Kanokmon 191  
 Ru-tong, Li 1007  
 Ryou, Hwang-Bin 537, 1371  
 Ryou, Hwnag-Bin 531, 1365  
 Ryu, Nuri 185, 255
- Sabudin, Maziani 643  
 Sakthidharan, G.R. 1225  
 Sandhya, P. 1211  
 Saravanan, P. 1225  
 Selamat, Ali 1053  
 Seo, DongBum 571  
 Seo, Dongmin 1171, 1179  
 Seo, Jinwon 285  
 Seol, Ye-In 51  
 Seresangtakul, Pusadee 743, 775  
 Shafe, Afza 1079  
 Shah, Peer Azmat 695, 1137  
 Shahrudin, Muhammad Shahrizan 735  
 Shariff, Azmi M. 829  
 Shen, Chien-Hua 769  
 Shen, Jie 1109  
 Sheng, Yehua 1061  
 Shin, Chang-Sun 301, 309, 321, 327  
 Shin, Ilhoon 141  
 Shon, Myoung-Kyu 185  
 Shrestha, Bhanu 509, 1405  
 Shusheng, Gu 973  
 Sim, Alex Tze Hiang 65  
 Sin, Cheol Ho 1443  
 Singh, Manmeet Mahinderjit 669  
 Sirageldin, Abubakr 217  
 Sirisangtragul, Wanna 743  
 Sivamani, Saraswathi 321, 327
- Sohn, Myoung-Kyu 255  
 So-In, Chakchai 191  
 Son, Kwang Chul 1337  
 Son, Nguyen Hoang 107  
 Song, Ho-Bin 497, 503, 1325, 1331  
 Song, Wang 957  
 Song, YingHui 863  
 Soomlek, Chitsutha 191  
 Sun, Guojiang 783  
 Sun, Wen-Liang 885  
 Sung, Won-Kyung 1179
- Taheri, Shahrooz 65  
 Talib, Abdullah Zawawi 703, 729, 799  
 Tan, Kian Lam 703  
 Tanaka, Takazumi 1429  
 Tang, Shengjing 783  
 Tao, Xu 1089  
 Thai, Nguyen Quoc 95  
 Ting, Mei 1287  
 Tsai, Joseph C. 1423  
 Tung, Nguyen Xuan 333
- Udompongsuk, Kanokporn 191  
 Usha Devi, Gandhi 1219
- Van, Thanh The 615  
 Van Hoai, Tran 179  
 Vijayalakshmi, S. 1129  
 Vu, Thanh T. 763
- Waikham, Boonsup 191  
 Wang, Danghui 621  
 Wang, Fei 201  
 Wang, Huang 853, 1045  
 Wang, Huibin 1109  
 Wang, Jianhua 837  
 Wang, Jilong 1067  
 Wang, Lisong 819  
 Wang, Qing 751  
 Wang, Rong 241  
 Wang, Wei 997  
 Wang, Xuesong 131  
 Wang, Yu 967  
 Wei, Zhang 1089  
 Wenliang, Sun 919  
 Won, Dongho 391, 423, 435  
 Woo, Junhyun 141  
 Wu, Can 837

- Wu, Chia-Chun 685  
 Wu, Shizhong 869  
  
 Xia, Feng 43  
 Xiao, Chengwei 649  
 Xiao-Fan, Mu 929  
 Xiao-Jian, Ding 981  
 Xin, Xiong 1073  
 Xingpeng, Zhou 1165  
 Xiong, Yongping 869  
 Xu, Bingqing 101  
 Xu, Dongliang 863  
 Xu, Huijie 241  
 Xu, Yongjun 201  
 Xu, Yongxuan 853  
 Xu, Zhenhao 811  
  
 Yaguchi, Yuichi 1429  
 Yahya, Abid 661  
 Yahya, Khairun 799  
 Yamakami, Toshihiko 35  
 Yang, Ho-Kyung 537, 1371  
 Yang, Hyung-Jeong 71  
 Yang, Xin-rong 937  
 Yanhong, Song 911  
 Yeh, Yi-Jen 1411  
 Yen, Neil Y. 1423  
 Yeo, Young-Hyun 1241  
 Yi, Gangman 279, 353  
 Yi, Hyunyi 19  
 Yi, Jeong Hyun 19  
 Yi, Zhang 1287, 1295  
 Yibing4, Wu 901  
 Yin, Yu 757  
 Yin, Zhimin 27  
 Ying, Wang 1023  
 Yinghui, Song 877  
  
 Yong, Suet-Peng 1145  
 Yong, Xia 1007  
 Yoo, Kwan-Hee 1171  
 Yoo, Seung-Hyeok 315  
 Yoo, Sujin 453  
 Yoon, Hyunsoo 445  
 Yu, Chen 1017  
 Yu, Shang 885  
 Yu, Xiangzhan 27  
 Yujun, Zhu 951  
 Yusheng, Zou 911  
 Yusof, Yuhanis 717, 1259  
 Yussiff, Abdul-Lateef 1145  
  
 Zain, Azlan Mohd 627  
 Zainon, Wan Mohd Nazmee Wan 729,  
 791  
 Zeng, Yanyang 1121  
 Zhan, HongJun 1097, 1103  
 Zhanchang4, Yang 919  
 Zhang, Hongli 27, 863  
 Zhang, Huatang 967  
 Zhang, Junhua 1273  
 Zhang, Lichen 83, 101, 147  
 Zhang, Qian 819  
 Zhang, Shun 811  
 Zhang, Siyang 1061  
 Zhang, Tao 869  
 Zhang, Wei 1097, 1103  
 Zhang, Yang 1279  
 Zhekun, Li 911  
 Zhi-guang, Qin 929, 1007  
 Zhixin, Feng 1115  
 Zhongwei, Xu 951  
 Zhou, Mingquan 131  
 Zu-Feng, Wu 929