Beyond Simple Searching: Open Source Exploitation of the Chinese Internet

James Mulvenon, Ph.D.

July 2007

Defense Group Inc.

Center for Intelligence Research and Analysis

Advancing U.S. Intelligence Through Innovative Research, Analysis, and Public Outreach

What is the Center for Intelligence Research and Analysis?

- CIRA's mission is to provide cutting-edge research and analysis on critical issues to the US Government
- CIRA successfully executes this mission using
  - Hunter-killer teams of cleared Chinese, Korean, Farsi, Arabic linguist-analysts and fieldwork consultants working on hard targets issues
    - CIRA's staff currently has 10 cleared Chinese linguist-analysts
  - Networks of high-value functional specialists (e.g., cultural anthropologists, cognitive neuroscientists, electrical engineers)
  - Digital network intelligence teams for Internet datamining and network topology mapping of foreign computer networks
  - Operations support capabilities, including classified document forensics and secure open source architectures

CIRA's Open Source Philosophy

- CIRA uses open source exploitation…
  - …to provide some or all of the answers to hard target questions
  - …to complement classified data for true all-source analysis
  - …to provide cueing for follow-on technical and human collection and analysis
- CIRA utilizes a multi-faceted research methodology and infrastructure
  - Advanced linguistic capabilities
  - Extensive fieldwork and field capabilities
  - Deep, anonymized Internet exploitation
  - Extensive collection of foreign books, newspapers and technical journals
  - Broad network of subject matter expertise here and abroad
  - Clearances, accredited SCIF, secure comms, secure open source architectures

Open Source and China: Understanding the Milieu

- The Chinese language itself presents the greatest barrier to open source exploitation
  - "China's first line of national defense"
  - "China's first layer of encryption"
- Reading Chinese vs. understanding Chinese
  - Linguists pose difficult dilemma
    - Native speakers are difficult to get cleared
    - Yet cultural nuance cannot be learned in a classroom
  - There is no substitute for immersion language training
    - Cuts against the grain of the current security mindset

UNCLASSIFIED//FOUO//DGI/CIRA
PROPRIETARY

Open Source and China: Understanding the Medium

- Chinese Internet is still a largely untapped gold mine
  - Soon Chinese language pages will outnumber all others
- But “Great Firewall” DMZ poses important barrier to unfettered searching
  - Increasingly large “dark nets” blocked to foreign IP addresses
  - Requires searching from within China - easier than you think!
- Even within China, native language search engines are censored for politically sensitive content
  - Google.com (Chinese) vs. Google.cn, Baidu, Others
  - Despite Baidu’s PR campaign, no evidence that domestic engines work better in Chinese
- Blogs and BBS have good material, but needle in a stack of needles
  - Local proprietary software poses technical challenges for collection

UNCLASSIFIED//FOUO//DGI/CIRA
PROPRIETARY

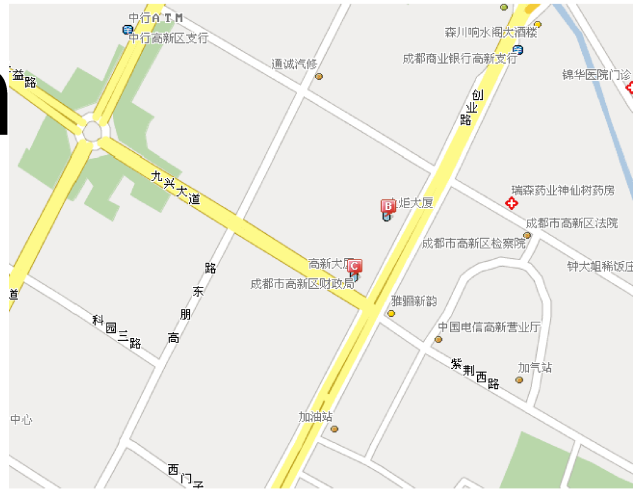
Open Source and China: Digging Below the Surface

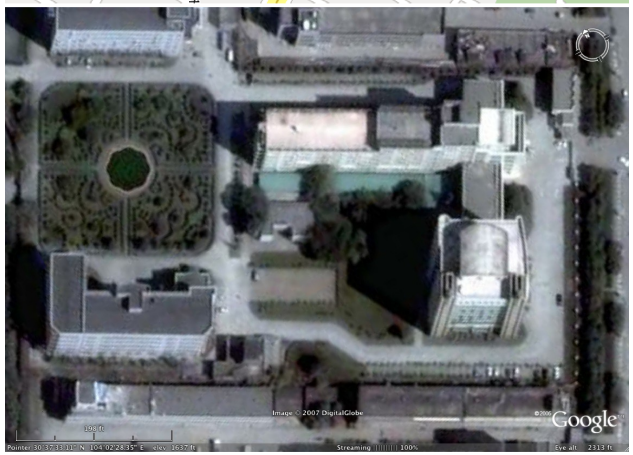
- Port 80 (HTTP) is not enough
  - Web browsing provides access to most of the desired content
  - But critical contextual details and additional information is available on the command line
- Simple and legal network topology mapping provides a fuller picture
  - Whois, nslookup, telnet, even portscanning

UNCLASSIFIED//FOUO//DGI/CIRA
PROPRIETARY

Open Source and China: Adding the Geospatial Dimension

- Location, location, location!
  - Accurate geolocation data is often the best way to unravel entity identities and relationships
- Relatively simply to find geolocation data for Chinese entities
- Street addresses often sufficient to find location in online GIS websites
  - Bracketing technique also works well
- Easy to map onto Google Earth for satellite imagery





UNCLASSIFIED//FOUO//DGI/CIRA
PROPRIETARY

Conclusions and Implications

- Possibilities of open source in China
  - China is “opaque” if one does not read Chinese
  - In fact, the Chinese system is much more transparent than commonly believed
  - Some subject lend themselves particularly well to open source exploitation
- Limitations of open source in China
  - Some subjects are completely opaque
  - Some subjects reveal only a macro-level discussion
    - Still tremendously valuable as cueing for follow-on efforts

UNCLASSIFIED//FOUO//DGI/CIRA
PROPRIETARY