

# Cybersecurity Big Data Research for Hacker Community: A Topic and Language Modeling Approach



PI: Dr. Hsinchun Chen, University of Arizona; Co-PI: Dr. Weifeng Li, University of Georgia

## Challenges:

- The technical difficulties in hacker community collection
- The massive volume of the data
- The heterogeneity and covert nature of the data elements and their subtle linkages
- The need to comprehend subcultural nature of terms and concepts embedded in the hacking community across multiple foreign languages

## Solution:

- A large and comprehensive testbed of significant international online hacker communities
- An *adversarial deep representation* approach to learning the language-invariant representations from content in English and non-English hacker communities
- A *nonparametric supervised topic modeling* method for examining customer reviews of hacker assets
- A *scalable dynamic topic modeling* technique designed for incorporating expert knowledge of hacker communities

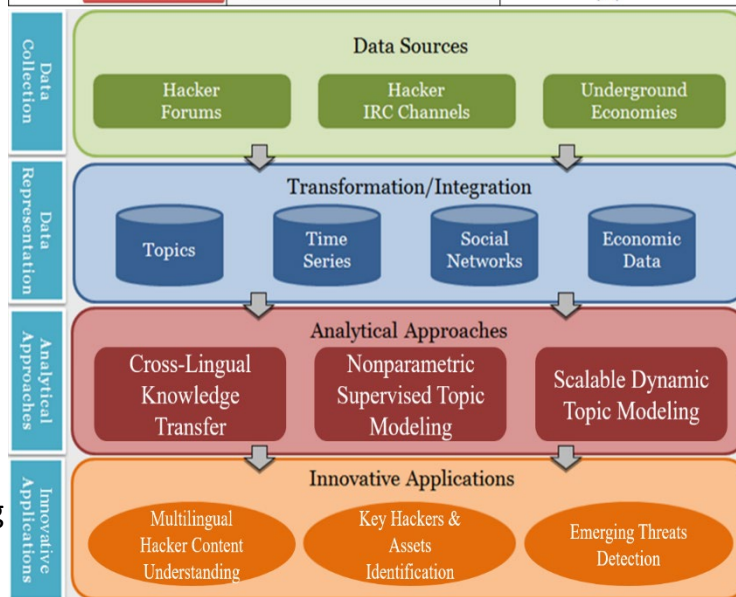
The screenshot shows a forum post for a C++/Delphi Cryptor v0.1. The post includes a title, a brief description, a README section with features and dependencies, and a list of attached files. The features listed include: CryptoGear encryption cipher, stub installation and registry startup, error message box display, PE assembly cloner, icon changer, stub size number, small stub size, and stub compression. Dependencies include Borland Delphi 7 IDE for GUI and Dev-C++ 5.11 IDE for Console and Stub. The post is tested on Windows 8 Professional x86.

## Scientific Impact:

- Understanding the human behaviors behind malicious online acts such as cybercrime and cyberterrorism is an important objective in gaining a better understanding of the cyber threat landscape.
- Hacker communities are of particular interest as they allow hackers to share malicious assets such as hacking tools, malware source code, and hacking tutorials with one another.

## Broader Impact:

- Benefiting several stakeholder communities: Intelligence and Security Informatics (ISI), NSF Scholarship-for-Service (SFS), National Cyber-Forensics & Training Alliance (NCFTA), and The Society for the Policing of Cyberspace (POLCYB).
- Integration into education: AZSecure SFS, NSA designated Center of Academic Education in Cyber Defense (CAE-CD) courses at UA, UA's online MS in Cybersecurity program, UGA's undergraduate area of emphasis in Information Security



NSF Award#1936370

PI: Dr. Hsinchun Chen ([hsinchun@email.arizona.edu](mailto:hsinchun@email.arizona.edu))

Co-PI: Dr. Weifeng Li ([weifeng.li@uga.edu](mailto:weifeng.li@uga.edu))