Exploring Dark Networks: From the Surface Web to the Dark Web

Hsinchun Chen, Ph.D., University of Arizona

Panel 3: Multi-Level, High-Dimensional Evolving and Emerging Networks (ML-HD-EEN)

October 12, 2017

“Leveraging Advances in Social Network Thinking for National Security: A Workshop”

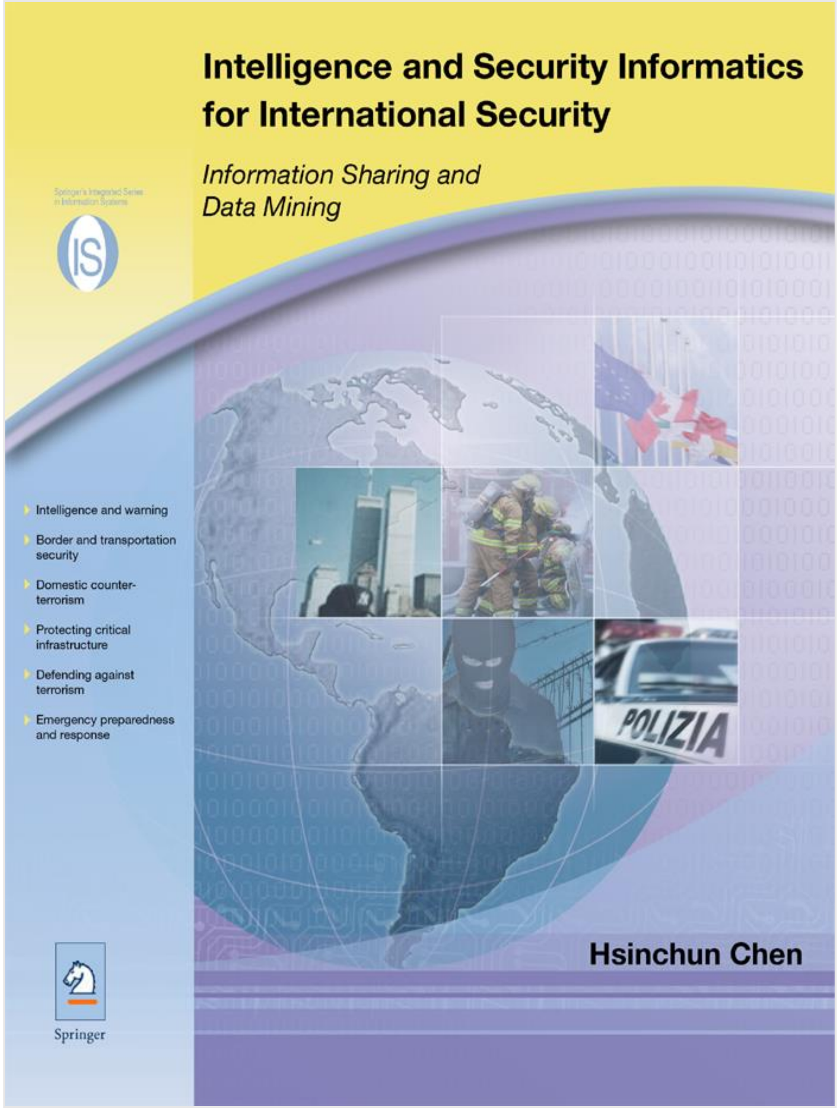
Acknowledgements: NSF, DHS, DOJ, DOD

My Background

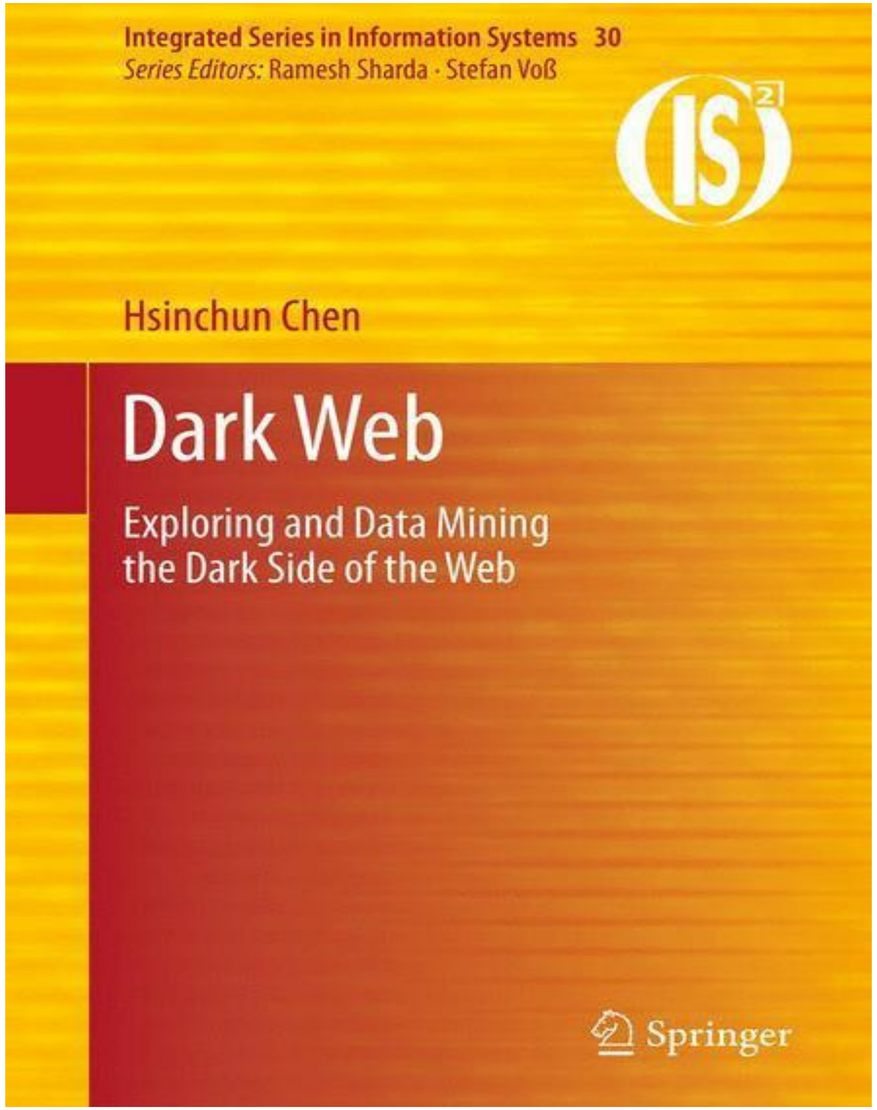
- Academic: Data/computational scientist; data/text/web mining, visualization; applied AI for security and health analytics
- Applications: Security analytics; dark networks
- Projects: COPLINK (1997-2009, gang/narcotic networks); Dark Web (2001-present, extremist/terrorist networks); Hacker Web (2009-present)

- SBE collaborators: M. Sageman, R. Breiger, T. Holt
- Agency collaborators: TPD, PPD, FBI, CIA, NSA, DHS (NSF, DOD)

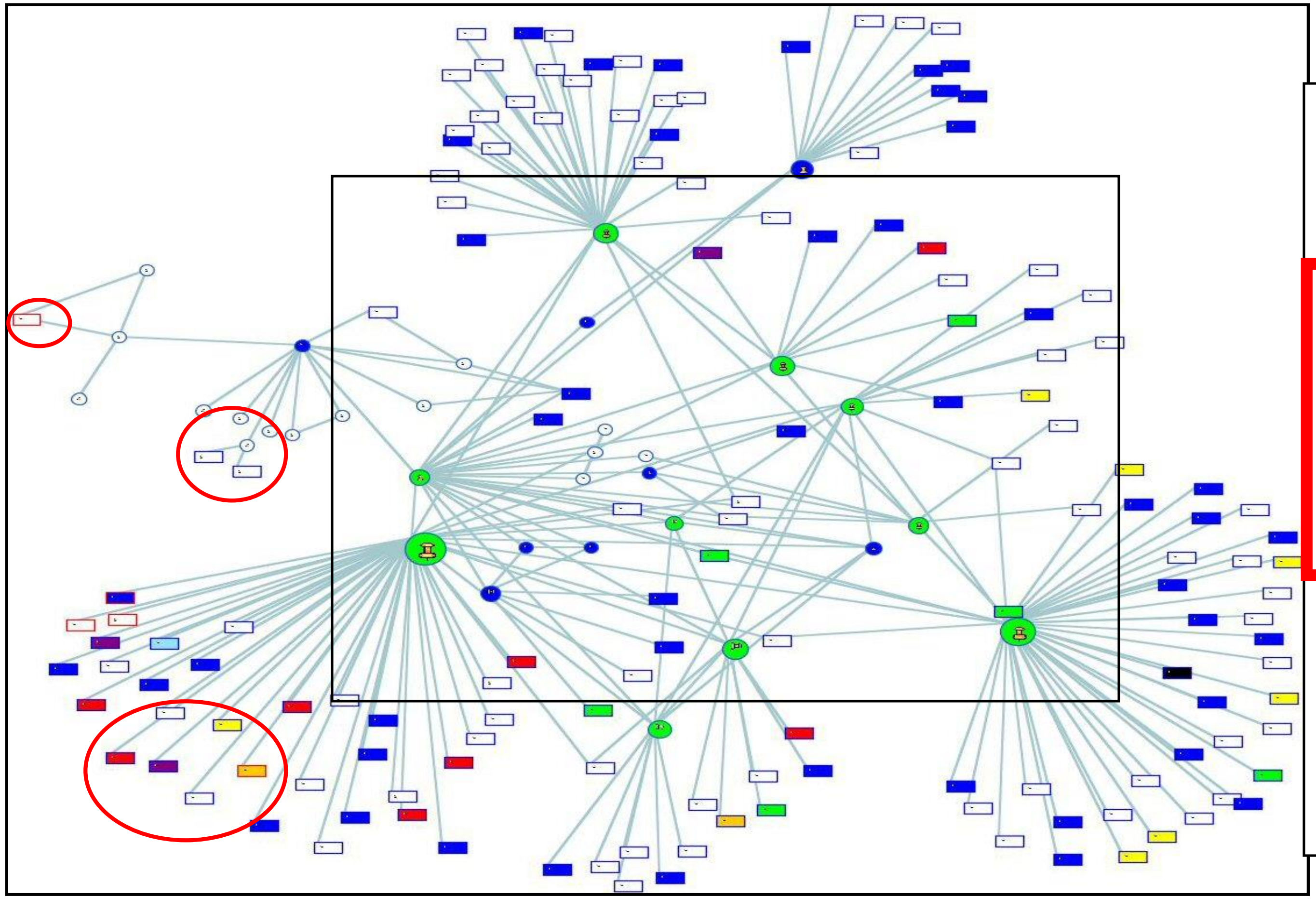
Springer, 2006



Springer, 2012



A Vehicle to Watch via its Networks?



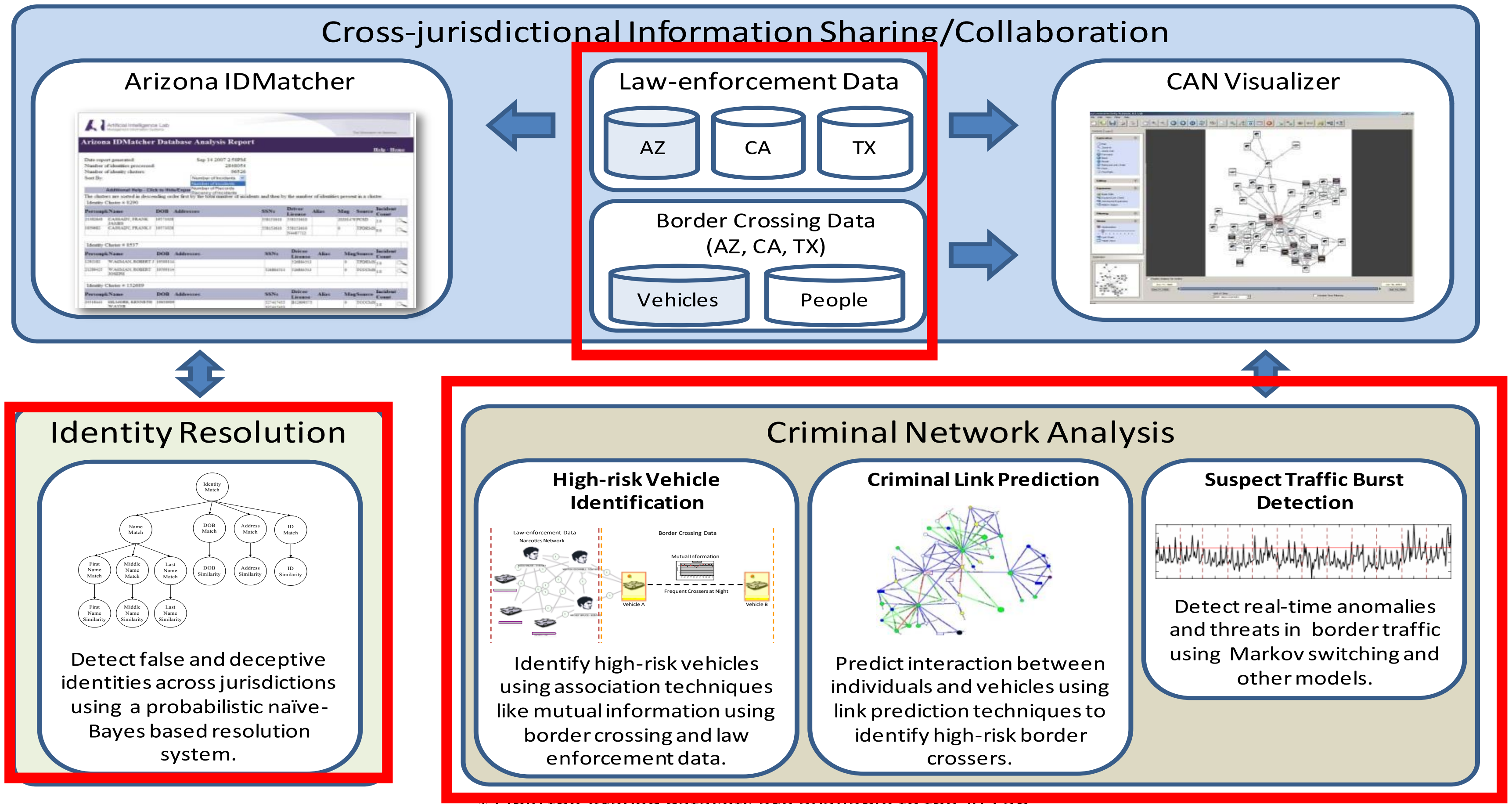
Shape Indicates Object Type
- circles are people
- rectangles are vehicles

Color Denotes Activity History
- Gang related
- Violent crimes
- Narcotics crimes
- Violent & Narcotics

Larger Size Indicates higher levels of activity

Border Crossing Plates are outlined in Red

COPLINK Identity Resolution and Criminal Network Analysis (DHS)

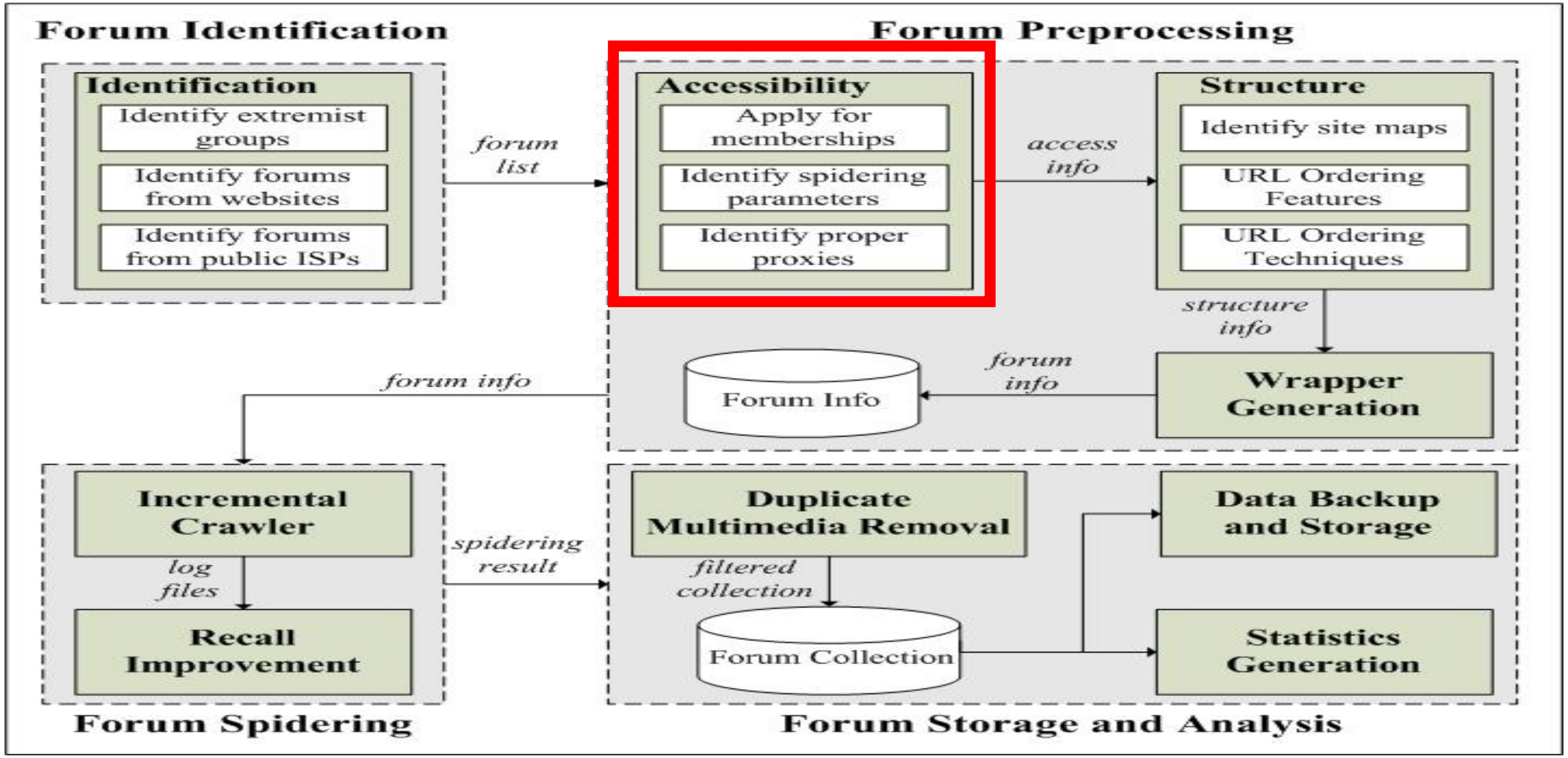


* Only the grayed datasets are available to the AI Lab

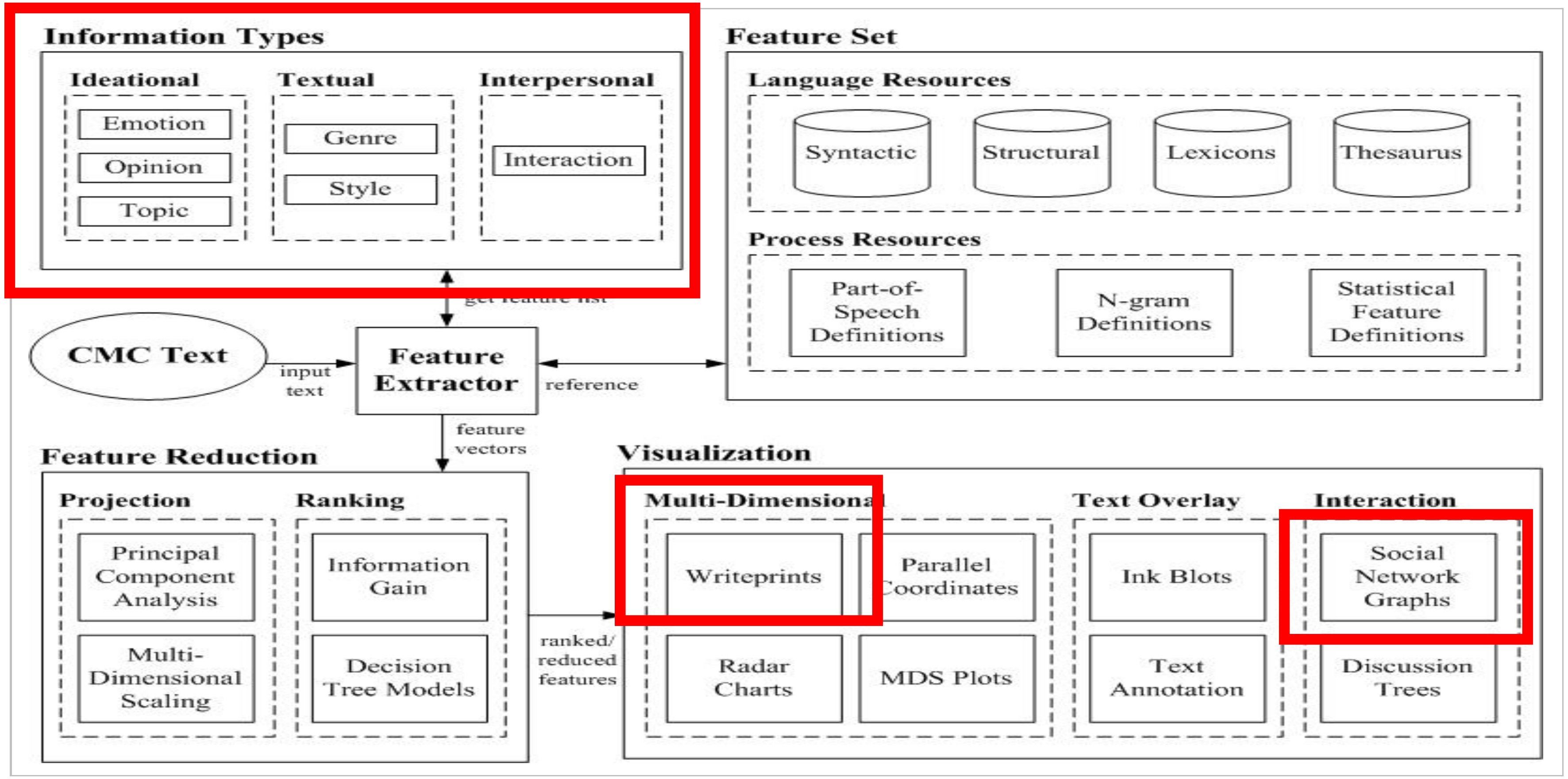
Dark Web Overview

- Dark Web: Terrorists' and cyber criminals' use of the Internet
- Collection: Web sites, forums, blogs, YouTube, etc.
- 20 TBs in size, with close to 10B pages/files/messages (the entire LOC collection: 15 TBs)

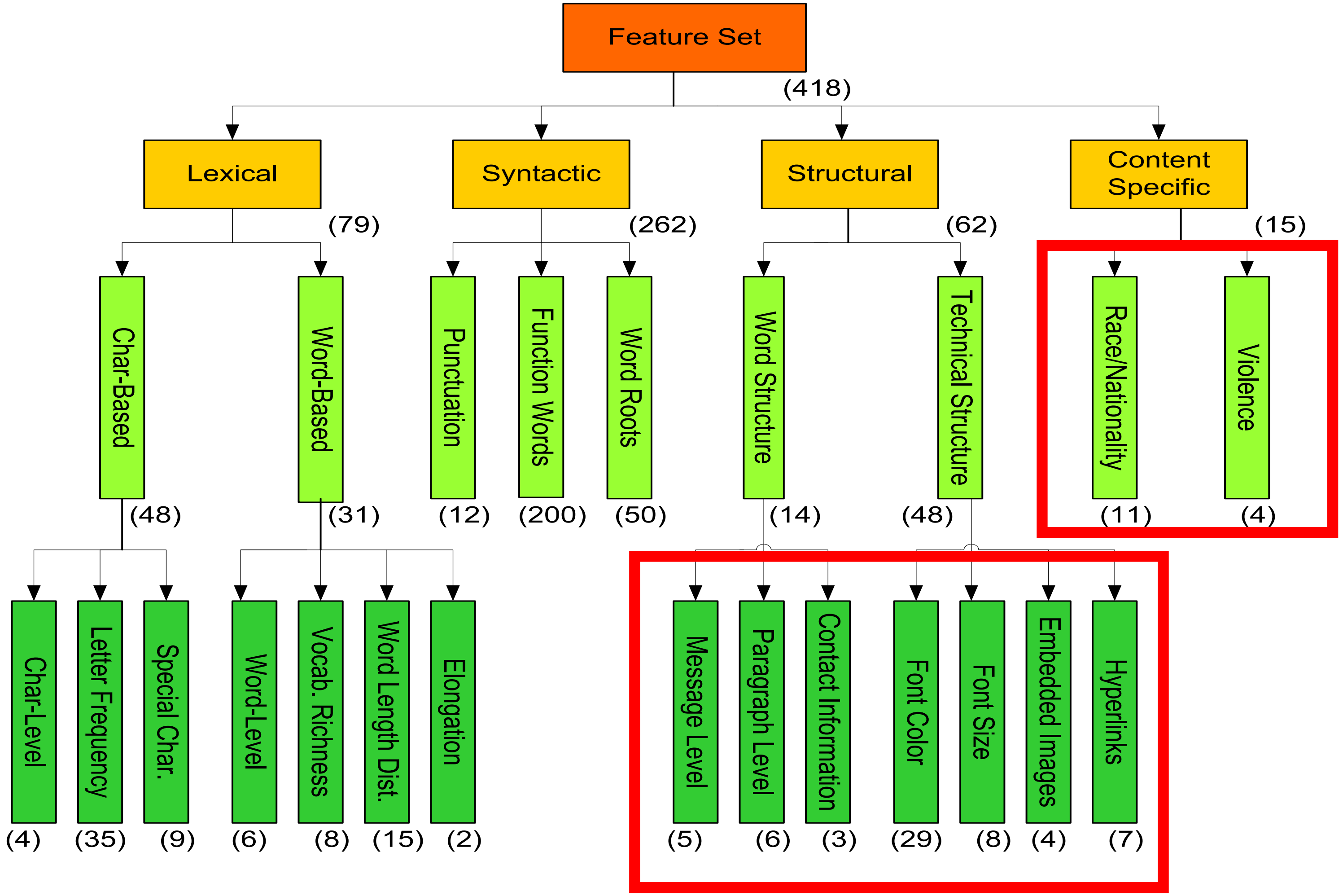
Dark Web Forum Crawler System: Probing the Hidden Web (Proxy, TOR)



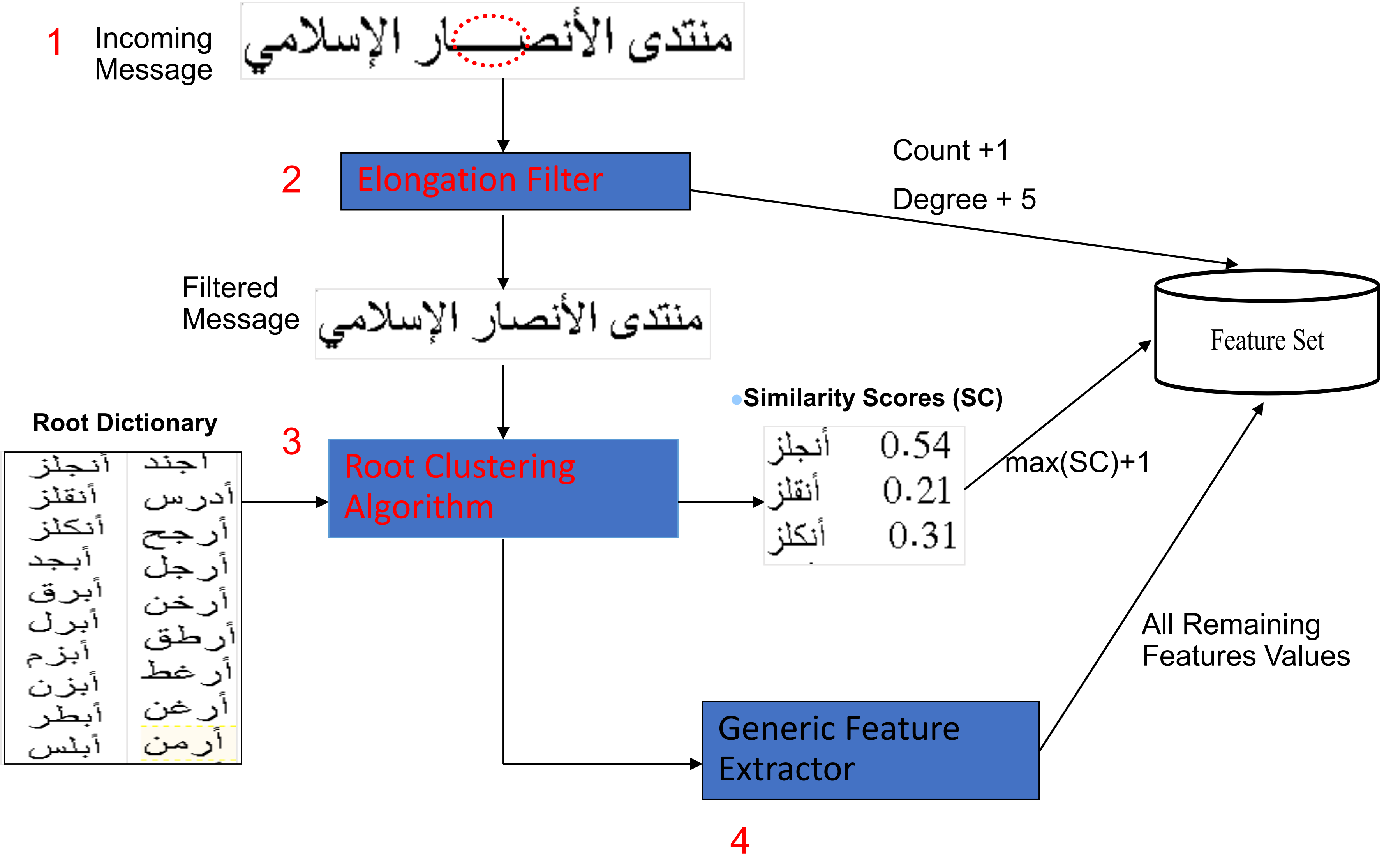
CyberGate for Social Media Analytics: Ideational, Textual and Interpersonal Information



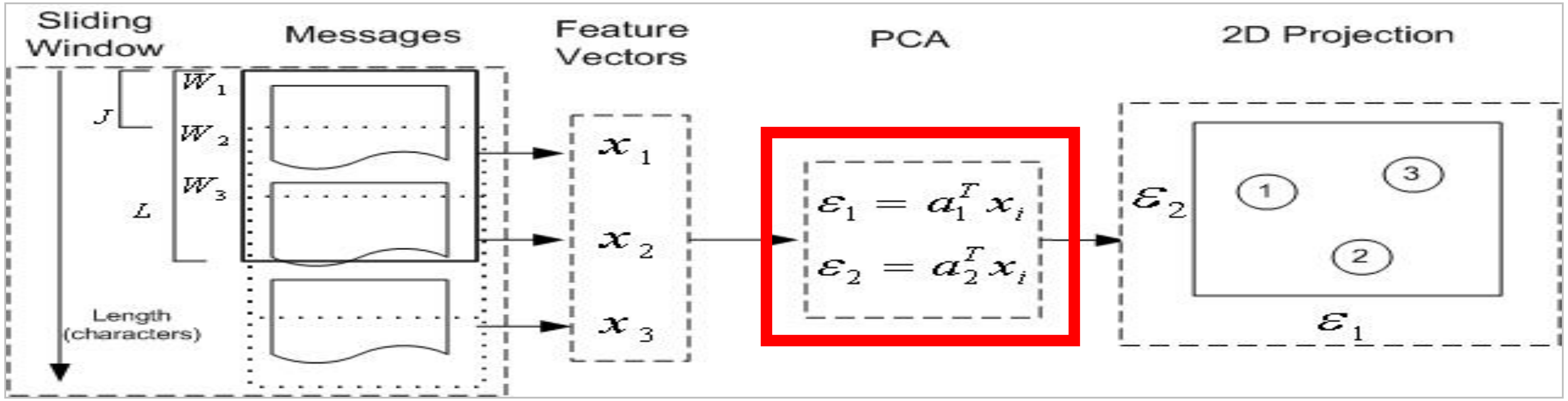
Arabic Writeprint Feature Set: Online Authorship Analysis

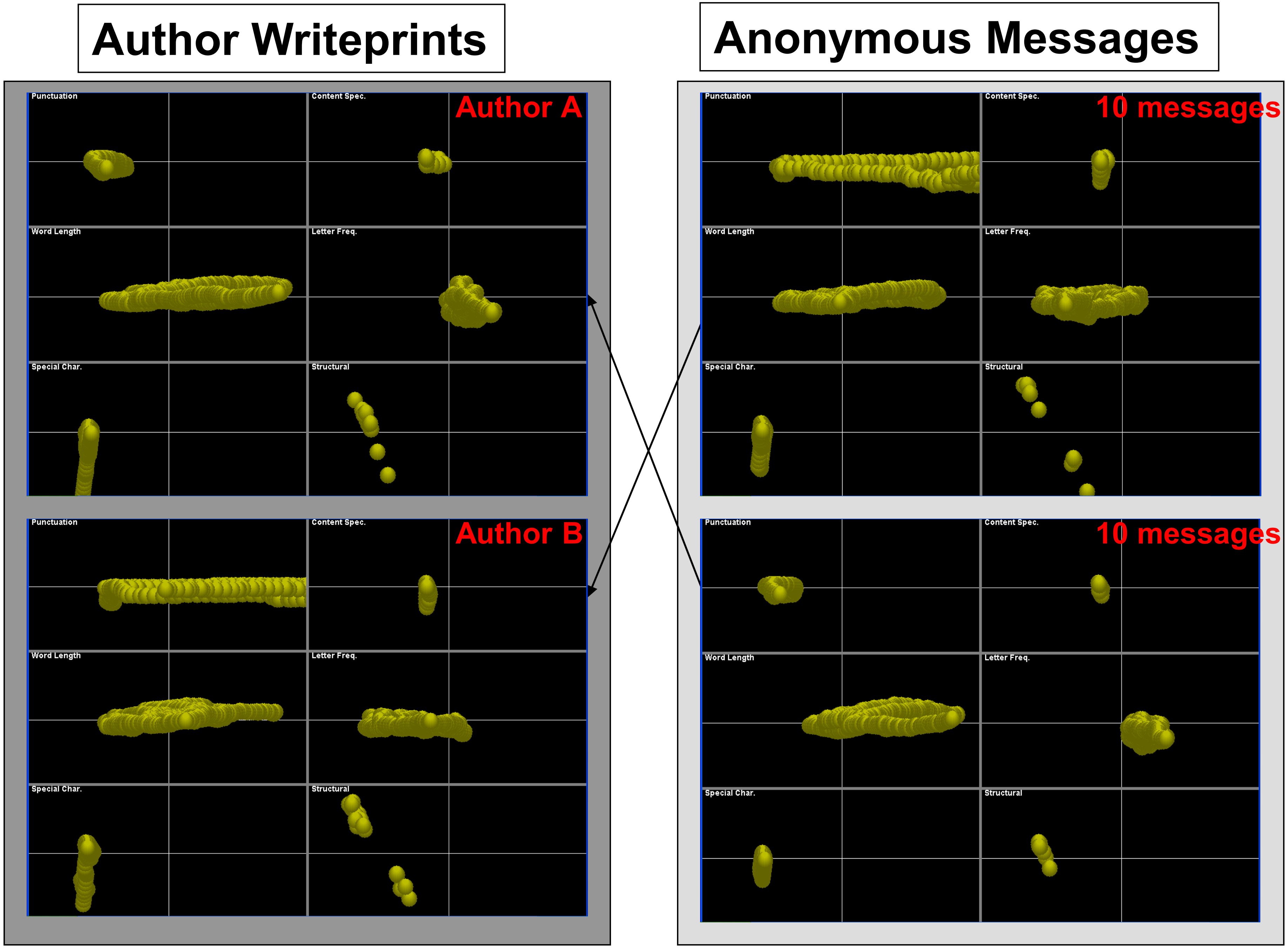


Arabic Feature Extraction Component



CyberGate System Design: Writeprints

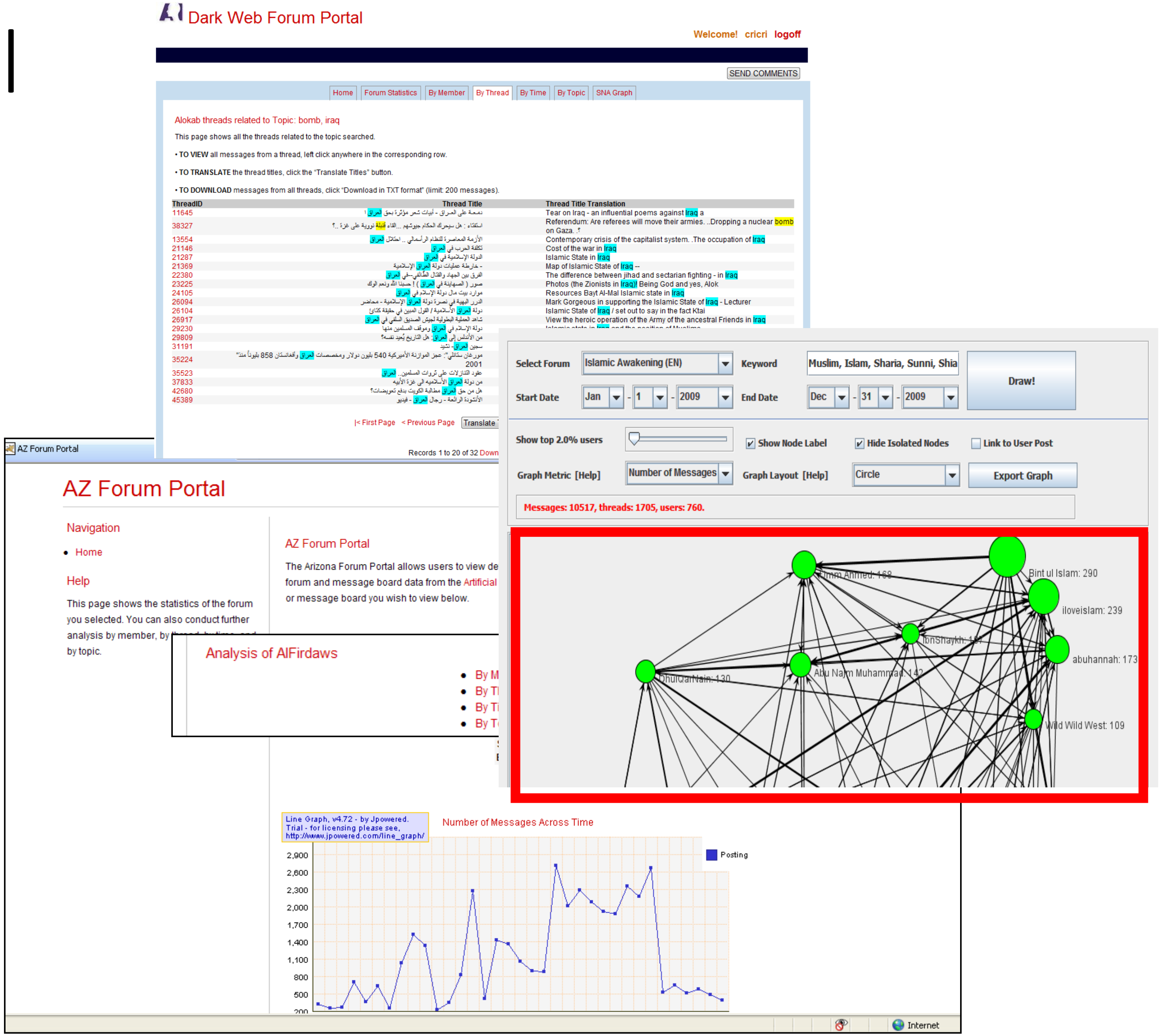


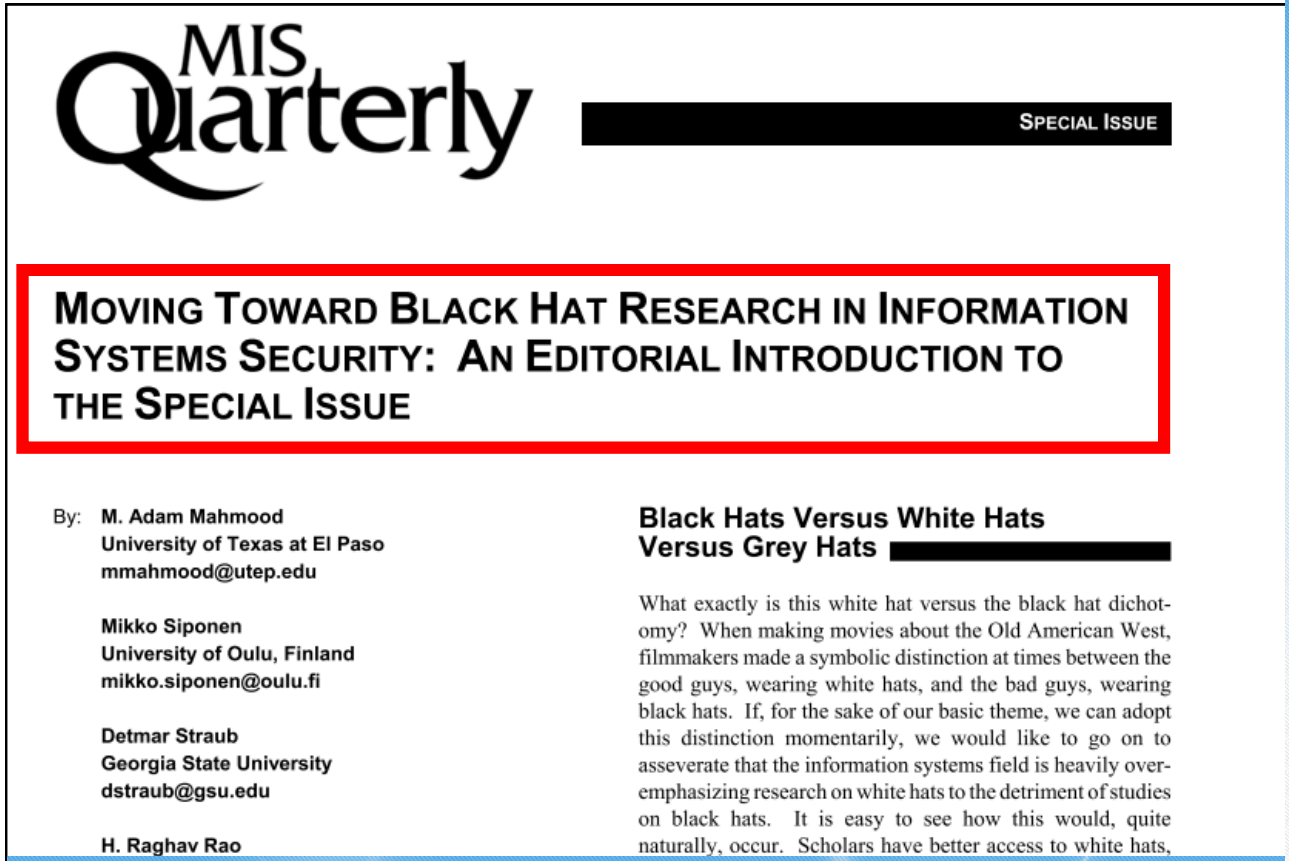


Author Writeprints
Anonymous Messages
Author A
Author B
10 messages
10 messages
Punctuation
Content Spec.
Word Length
Letter Freq.
Special Char.
Structural

AZ Forum Portal

- 13M messages (340K members) across 29 major Jihadi forums in English, Arabic, French, German and Russian (VBulletin)
- Linking members over time




MIS Quarterly

SPECIAL ISSUE

MOVING TOWARD BLACK HAT RESEARCH IN INFORMATION SYSTEMS SECURITY: AN EDITORIAL INTRODUCTION TO THE SPECIAL ISSUE

By: M. Adam Mahmood
University of Texas at El Paso
mmahmood@utep.edu

Mikko Siponen
University of Oulu, Finland
mikko.siponen@oulu.fi

Detmar Straub
Georgia State University
dstraub@gsu.edu

H. Raghav Rao

Black Hats Versus White Hats Versus Grey Hats

What exactly is this white hat versus the black hat dichotomy? When making movies about the Old American West, filmmakers made a symbolic distinction at times between the good guys, wearing white hats, and the bad guys, wearing black hats. If, for the sake of our basic theme, we can adopt this distinction momentarily, we would like to go on to asseverate that the information systems field is heavily over-emphasizing research on white hats to the detriment of studies on black hats. It is easy to see how this would, quite naturally, occur. Scholars have better access to white hats,


National Science Foundation
WHERE DISCOVERIES BEGIN

QUICK LINKS

SEARCH

HOME FUNDING AWARDS DISCOVERIES NEWS PUBLICATIONS STATISTICS ABOUT NSF FASTLANE

Discoveries

Discoveries
Search Discoveries
About Discoveries

Discoveries by Research Area
Arctic & Antarctic
Astronomy & Space
Biology
Chemistry & Materials
Computing
Earth & Environment
Education
Engineering
Mathematics
Nanoscience
People & Society
Physics

Email Print Share

Discovery
When hackers talk, this research team listens

Online conversations help fill critical gap in cybersecurity knowledge about attackers' motivations, possible targets

NSF-supported researchers have shed new light on how hackers communities interact.
Credit and Larger Version

October 8, 2015

Hsinchun Chen leads a research project that explores the motivations of cyberattackers.
Credit and Larger Version


National Science Foundation
WHERE DISCOVERIES BEGIN

QUICK LINKS

SEARCH

HOME FUNDING AWARDS DISCOVERIES NEWS PUBLICATIONS STATISTICS ABOUT NSF FASTLANE

Funding

Find Funding
A-Z Index of Funding Opportunities

Email Print Share

Directorate for Computer & Information Science & Engineering

Secure and Trustworthy Cyberspace (SaTC)

CONTACTS

Name	Email	Phone	Room
Jeremy Epstein	jepstein@nsf.gov	(703) 292-8338	1175


National Science Foundation
WHERE DISCOVERIES BEGIN

QUICK LINKS

SEARCH

HOME FUNDING AWARDS DISCOVERIES NEWS PUBLICATIONS STATISTICS ABOUT NSF FASTLANE

Funding

Find Funding
A-Z Index of Funding Opportunities
Recent Funding Opportunities
Upcoming Due Dates

Division of Graduate Education

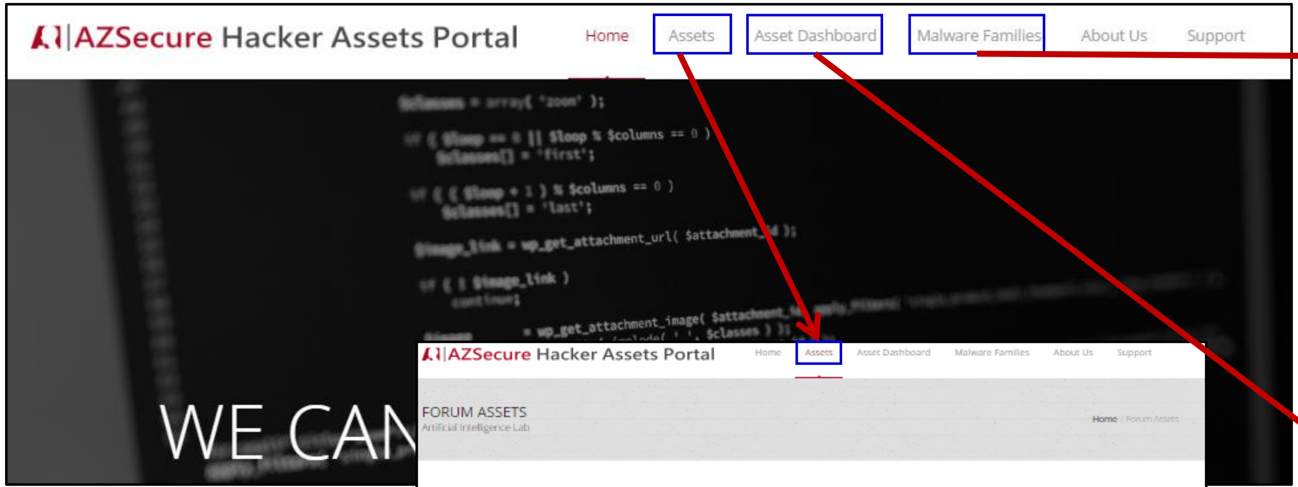
CyberCorps(R) Scholarship for Service (SFS)

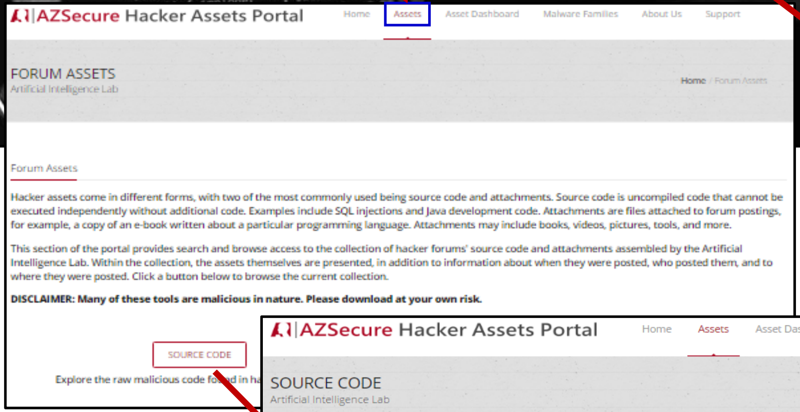
CONTACTS

Name	Email	Phone	Room
Dongwon Lee	dlee@nsf.gov	(703) 292-4679	
Victor P. Piotrowski	vpiotrow@nsf.gov	(703) 292-5141	
Paul Tymann	ptymann@nsf.gov	(703) 292-2260	

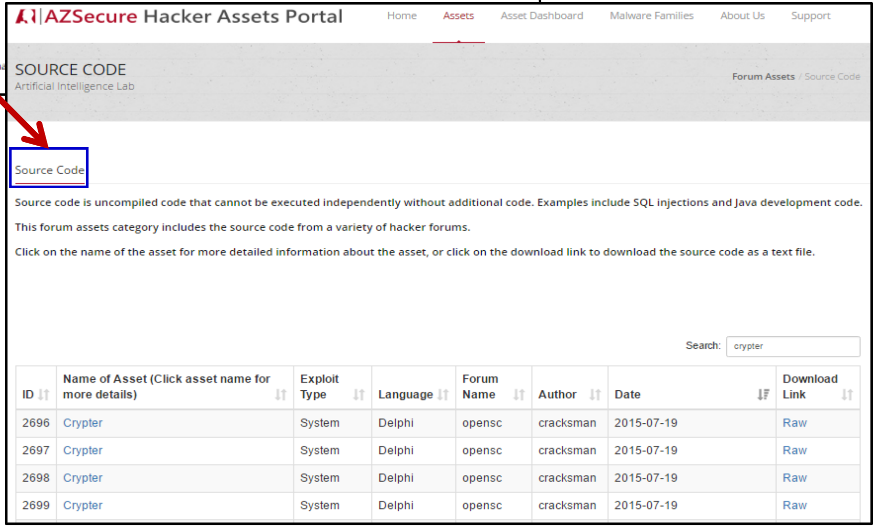
Hacker Assets Portal V2.0 – Overview

(a) Home page, linking to (b & c) **Assets**, (d) **Dashboard**, and (e) **Malware Families**:

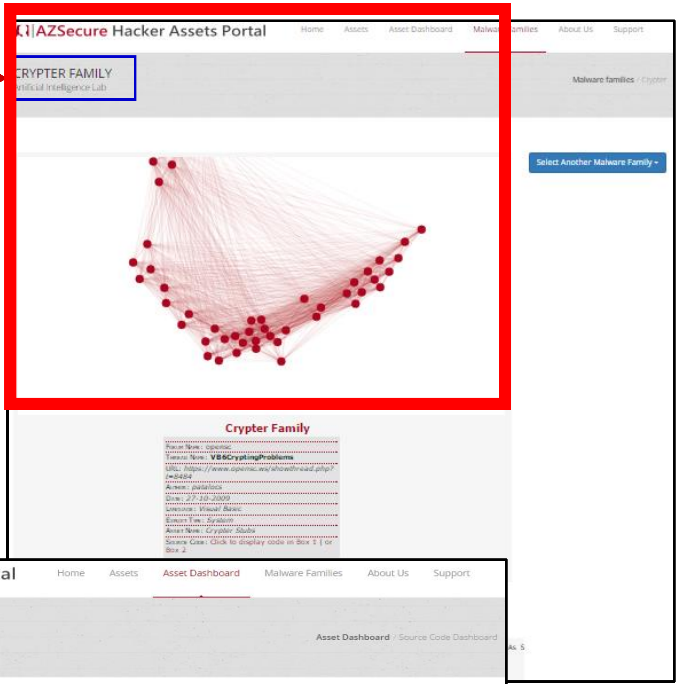




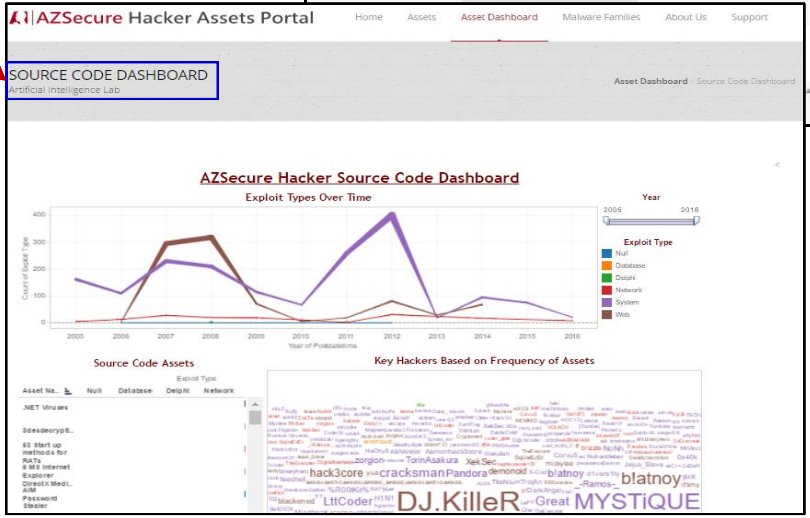
(b) **Assets** page, linking to Source Code and Attachments



(c) Source Code page; sortable by asset name, exploit type, date, etc.

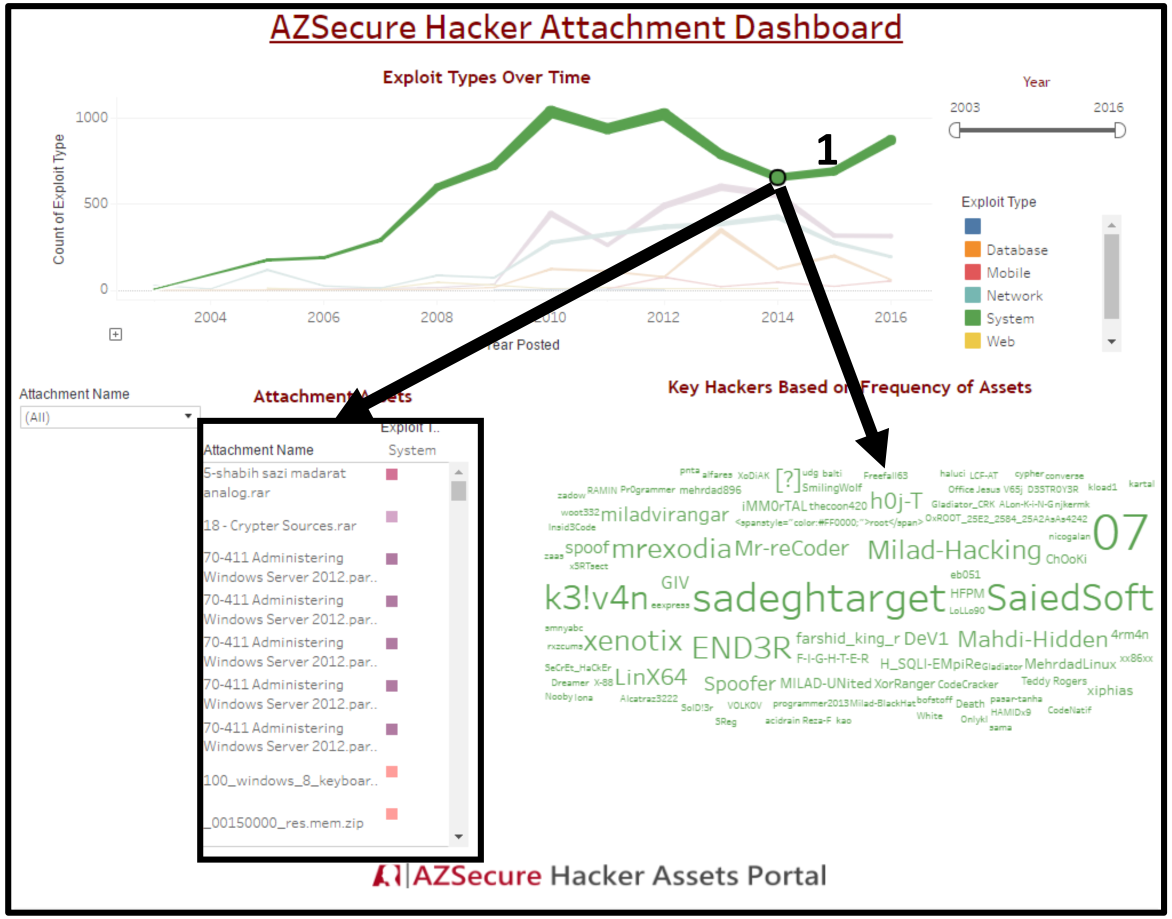


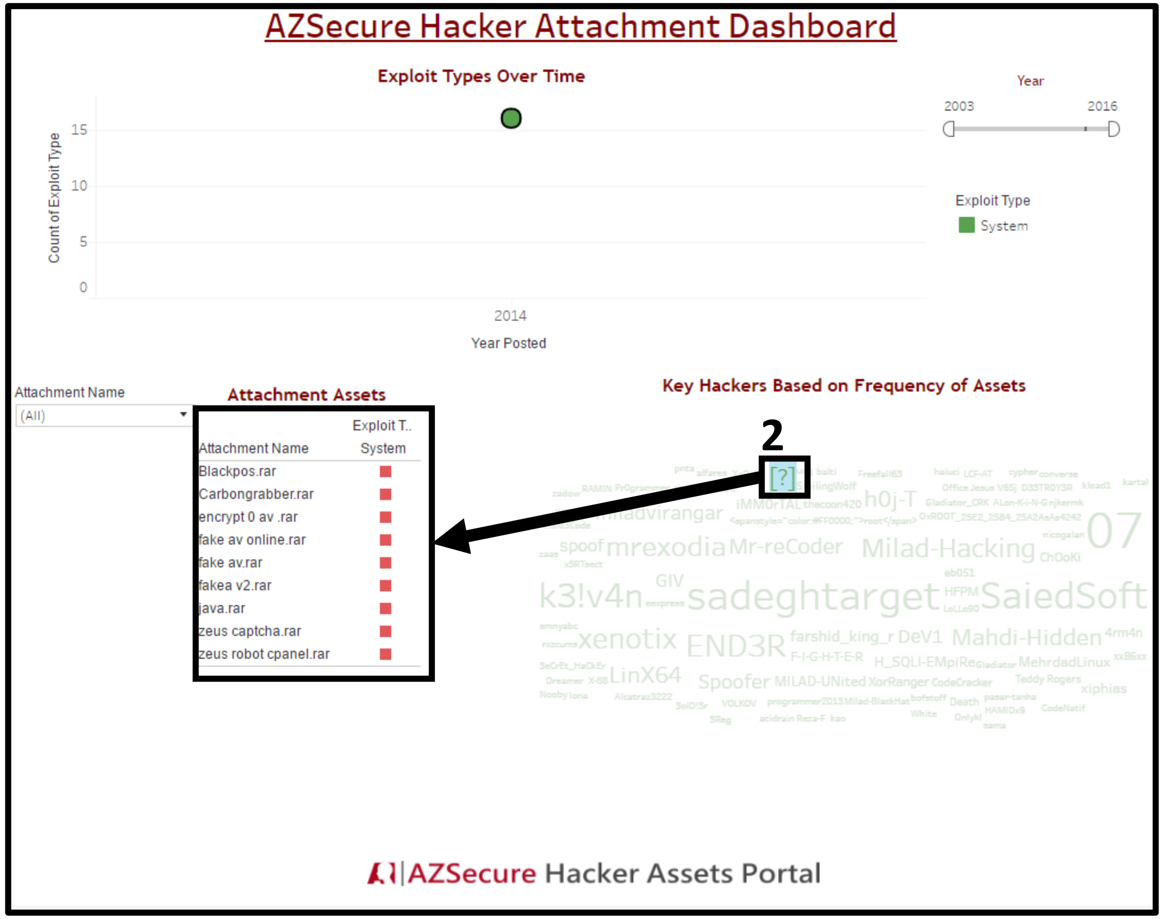
(e) **Malware Families**, for depicting relationships among assets over time (Crypter Family shown)



(d) **Dashboard** for drill-down analysis of hackers & assets over time

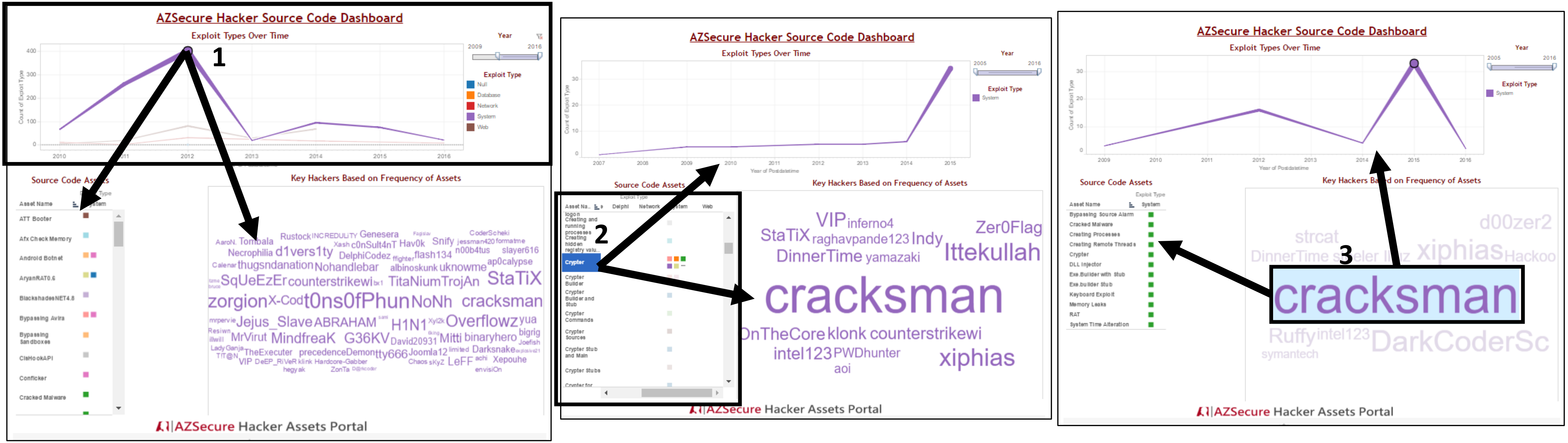
Cyber Threat Intelligence (CTI) Example – Bank Exploits





1. Filtering on 2014, when BlackPOS was posted, shows assets and threat actors at that time.
2. Filtering the actor who posted BlackPOS reveals that he posts other bank exploits (e.g., Zeus).
   - Provides intelligence on which hacker to monitor.

Cyber Threat Intelligence (CTI) Example – Crypters



1. Filtering on a specific time point (highest peak):
2. Filtering on a specific asset (crypters, a key technology for Ransomware)
3. Filtering a specific crypter author (Cracksman) shows the trends and types of assets he posted.

Selected Challenges for ML-HD-EEN in Dark Networks

- Identifying data Sources: availability (data stovepipes, data integration; RMS, RDBMS); web OSINT (<u>surface web, deep web, dark web; TOR, ICT</u>); data types (structured vs. unstructured; <u>multi-lingual, multi-media; source code, attachment, tutorial</u>), data biases (<u>noise, deception, adversarial; vigilante, honeypot, APTs</u>)

- Recognizing nodes: levels/dimensions (who/what/where/when/why/how); entity extraction and recognition (<u>identity resolution, web authorship analysis, writeprint</u>)

- Establishing links: linked by associations (labeled links, probabilistic links); linked by time/space (<u>same-time-same-place; border crossing, hotspot</u>); linked by conversations (<u>linguistic cues and styles; ICT, forums</u>)

- Analyzing network patterns: (many SNA techniques)

- Tracking changes over time: stream data collection & mining (update & alert; anomaly detection, concept drift; <u>emerging cyber threats and DarkNet Markets</u>

Selected Solutions & Directions for ML-HD EEN

- Comprehensive & timely OSINT data collection: from the surface web to the dark web; across level/dimension, over time

- Data integration and SNA extraction: AI assisted entity/relationship recognition/integration; across level/dimension, over time

- Methodological foundations: dark networks, hidden networks; noise, deception, adversarial intent

- Data analytics: advanced social media analytics, stream data mining, adversarial machine learning, BIG DATA analytics; across level/dimension, over time

For questions and comments

hchen@eller.Arizona.edu

http://ai.Arizona.edu