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

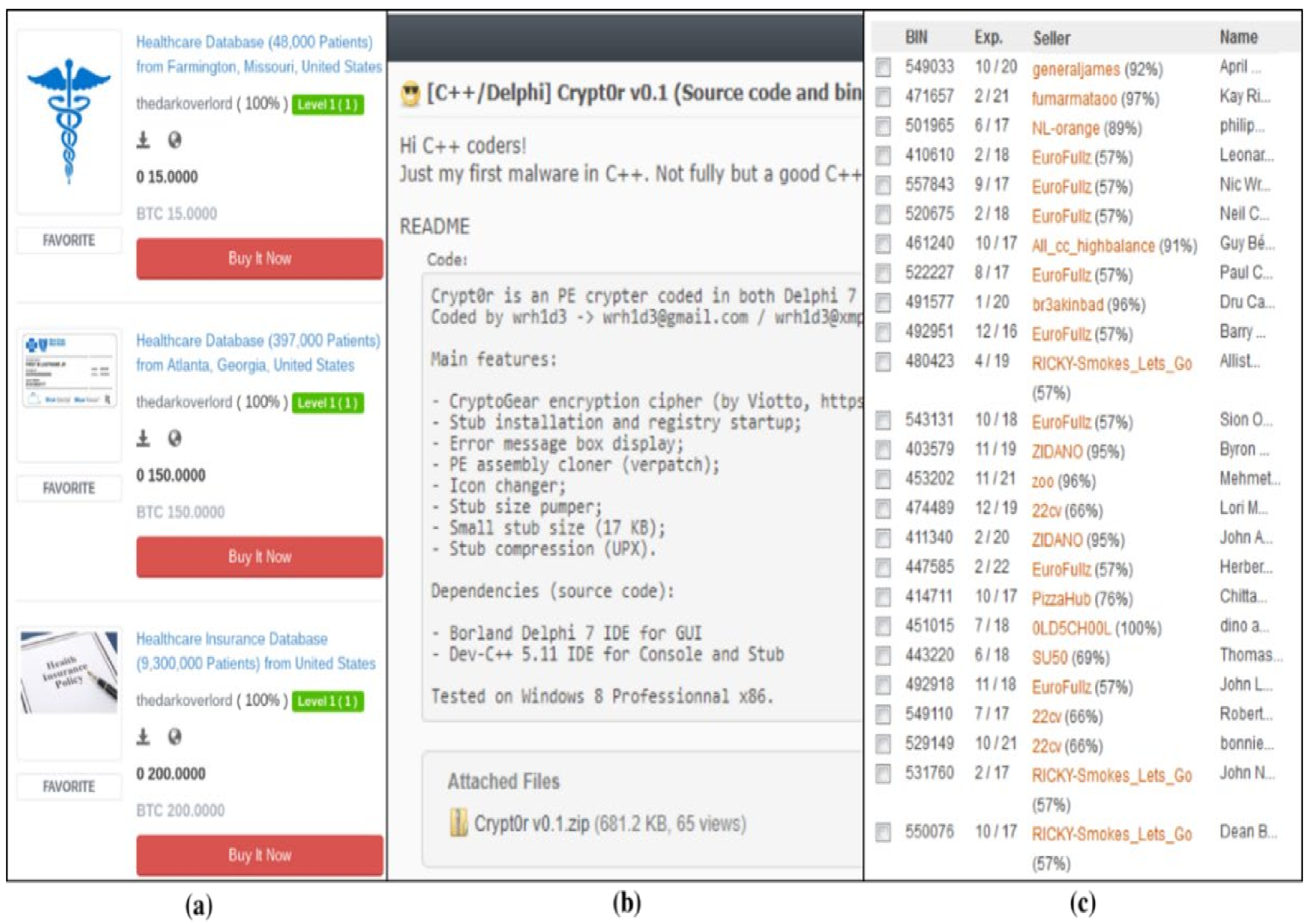
PI: Dr. Hsinchun Chen, Regents' Professor, ACM/IEEE Fellow, U. of Arizona (UA), AI Lab Director

Co-PI: Dr. Weifeng Li, U. of Georgia (UGA)

https://eller.arizona.edu/departments-research/centers-labs/artificial-intelligence/research/big-data

There is a growing perspective among both cybersecurity researchers and professionals that 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, which commonly exist as forums, Internet-Relay-Chat (IRC) channels, and underground economies, 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. While such communities make excellent candidates for research, the technical difficulties in data collection and analytics: the massive volume of data collection, the heterogeneity and covert nature of the data elements and their often not so obvious linkages, and the ability to comprehend common hacker terms and concepts embedded in communities across multiple regions present non-trivial challenges to researchers.

This project aims to address two research goals: (1) to advance current capabilities for large-scale identification, collection, and analysis of hacker communities, and (2) to make contributions to cybersecurity research by developing new big data and text analytics techniques that could enable cybersecurity researchers to conduct analyses on hacker community content and within other related domains.



Healthcare Database (48,000 Patients) from Farmington, Missouri, United States

thedarkoverlord (100%) Level 1 (1)

0 15.0000

BTC 15.0000

FAVORITE

Buy It Now

Healthcare Database (397,000 Patients) from Atlanta, Georgia, United States

thedarkoverlord (100%) Level 1 (1)

0 150.0000

BTC 150.0000

FAVORITE

Buy It Now

Healthcare Insurance Database (9,300,000 Patients) from United States

thedarkoverlord (100%) Level 1 (1)

0 200.0000

BTC 200.0000

FAVORITE

Buy It Now

(a)

[C++/Delphi] Crypt0r v0.1 (Source code and bin

Hi C++ coders!
Just my first malware in C++. Not fully but a good C++

README

Code:

```
Crypt0r is an PE crypter coded in both Delphi 7
Coded by wrh1d3 -> wrh1d3@gmail.com / wrh1d3@xmp

Main features:

- CryptoGear encryption cipher (by Viotto, https
- Stub installation and registry startup;
- Error message box display;
- PE assembly cloner (verpatch);
- Icon changer;
- Stub size pumper;
- Small stub size (17 KB);
- Stub compression (UPX).

Dependencies (source code):

- Borland Delphi 7 IDE for GUI
- Dev-C++ 5.11 IDE for Console and Stub

Tested on Windows 8 Professionnal x86.
```

Attached Files

Crypt0r v0.1.zip (681.2 KB, 65 views)

(b)

	BIN	Exp.	Seller	Name
□	549033	10 / 20	generaljames (92%)	April ...
□	471657	2 / 21	fumarmataoo (97%)	Kay Ri...
□	501965	6 / 17	NL-orange (89%)	philip...
□	410610	2 / 18	EuroFullz (57%)	Leonar...
□	557843	9 / 17	EuroFullz (57%)	Nic Wr...
□	520675	2 / 18	EuroFullz (57%)	Neil C...
□	461240	10 / 17	All_cc_highbalance (91%)	Guy Bé...
□	522227	8 / 17	EuroFullz (57%)	Paul C...
□	491577	1 / 20	br3akinbad (96%)	Dru Ca...
□	492951	12 / 16	EuroFullz (57%)	Barry ...
□	480423	4 / 19	RICKY-Smokes_Lets_Go (57%)	Allist...
□	543131	10 / 18	EuroFullz (57%)	Sion O...
□	403579	11 / 19	ZIDANO (95%)	Byron ...
□	453202	11 / 21	zoo (96%)	Mehmet...
□	474489	12 / 19	22cv (66%)	Lori M...
□	411340	2 / 20	ZIDANO (95%)	John A...
□	447585	2 / 22	EuroFullz (57%)	Herber...
□	414711	10 / 17	PizzaHub (76%)	Chitta...
□	451015	7 / 18	0LD5CH00L (100%)	dino a...
□	443220	6 / 18	SU50 (69%)	Thomas...
□	492918	11 / 18	EuroFullz (57%)	John L...
□	549110	7 / 17	22cv (66%)	Robert...
□	529149	10 / 21	22cv (66%)	bonnie...
□	531760	2 / 17	RICKY-Smokes_Lets_Go (57%)	John N...
□	550076	10 / 17	RICKY-Smokes_Lets_Go (57%)	Dean B...

(c)

Figure 1. Potential Threats in Selected Hacker Community Content: (a) stolen health data on a DNM; (b) crypter, a technology for ransomware on forums; (c) credit/debit cards and SSNs for sale in a carding shop.

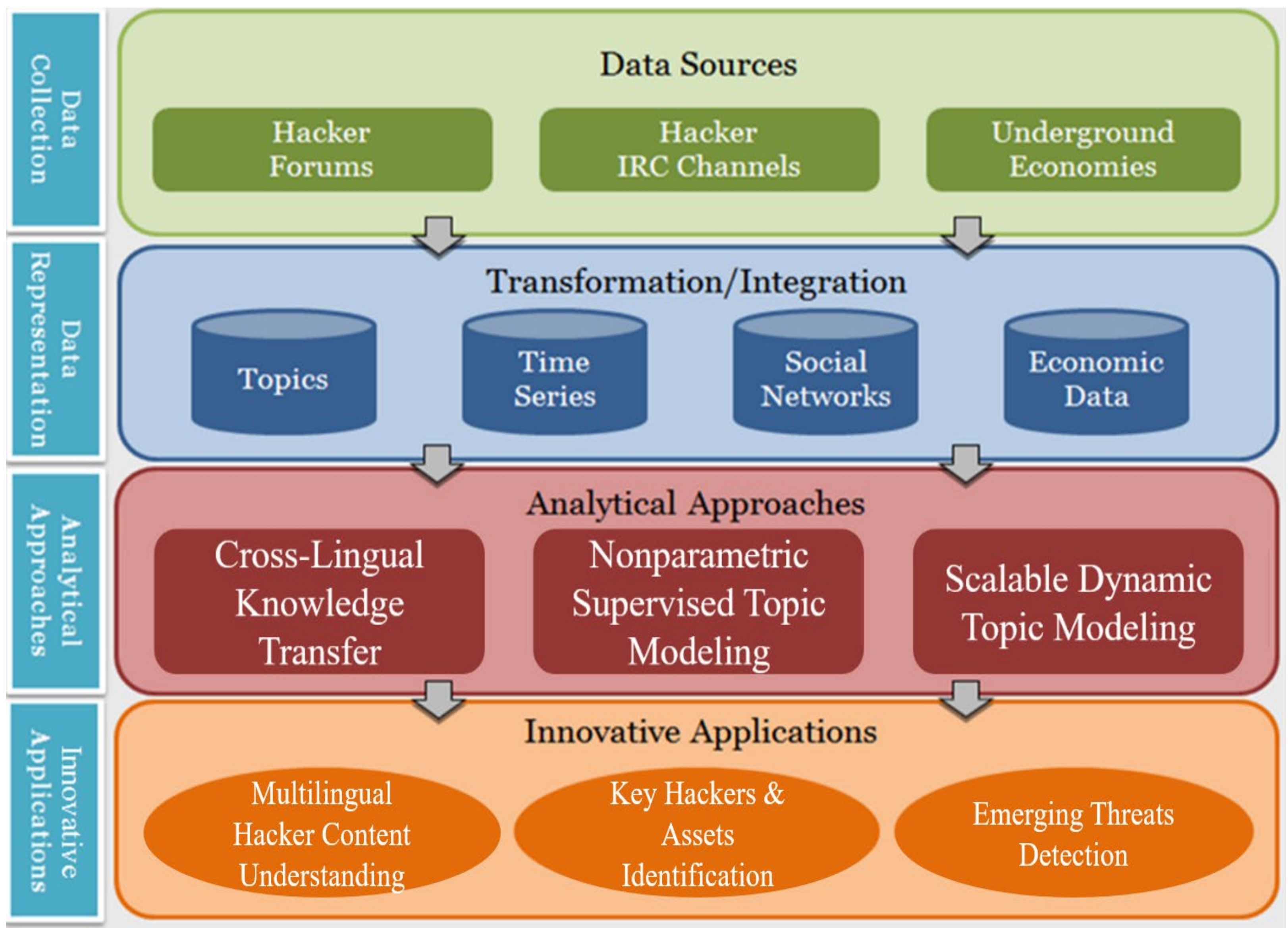


Figure 2. Proposed Framework of Scientific Impacts

Intellectual Merit:

Developing a large, comprehensive and longitudinal testbed of significant international online hacker communities:

- Web crawlers equipped with counter anti-crawling mechanisms and IRC listeners capable of connecting to hidden services

Understanding multilingual hacker community content to facilitate multilingual analytics:

- An adversarial deep representation approach to learning the language-invariant representations from content in English and non-English hacker communities

Identifying the key hacker assets such as malware and stolen data dumps:

- A nonparametric supervised topic modeling method for examining customer reviews of hacker assets to infer their quality

Modeling temporal attributes of hacker community content and finding evidence of potential emerging threats:

- A scalable dynamic topic modeling technique designed for incorporating expert knowledge of hacker communities and converging efficiently

Societal Impact:

- National Cyber-Forensics & Training Alliance (NCFTA) and The Society for the Policing of Cyberspace (POLCYB), serving as primary domain advisory and end user feedback groups, as well as liaisons to disseminate knowledge to others in the cybersecurity community
- Planning for extensive collaboration with the greater ISI communities and NSF SFS communities to evaluate and provide feedback on our proposed analytical approaches

Integration into Education:

- Significant research roles for master's and Ph.D. students in the UA's cybersecurity-based AZSecure Scholarship-for-Service (SFS) program
- NSA designated Center of Academic Education in Cyber Defense (CAE-CD) courses at UA (150+ students)
- UA's online MS in Cybersecurity program
- UGA's undergraduate area of emphasis in Information Security (~40 students)

Dissemination:

- Cybersecurity conferences and academic outlets such as IEEE ISI, Women in Cybersecurity (WiCyS) (800+ participants), and NSF SaTC PI meeting (400 participants)
- NCFTA's and POLCYB's Annual Conferences each draw 1,000+ attendees from law enforcement, government, and industry
- NSF-funded AZSecure DIBBs-ISI repository designed to share dozens of multi-million record security datasets and tools