CRII: SaTC: Identifying Emerging Threats in the Online Hacker Community for Proactive Cyber Threat Intelligence: A Diachronic Graph Convolutional Autoencoder Framework

Sagar Samtani, Ph.D. Assistant Professor, University of South Florida

Project URL: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1850362&HistoricalAwards=false

Project Overview: Background and Significance

- Cybersecurity experts have appraised the total cost of hacktivism, espionage, cyberwarfare, and other hacking at $450B annually.
- Many organizations aim to develop timely, relevant, and actionable intelligence (i.e., cyber threat intelligence) about emerging threats and key threat actors to enable effective cybersecurity decisions.
- The Dark Web is an emerging and viable CTI data source as it motivates millions of hackers from major geo-political regions (e.g., US, China, Russia, Middle East) to share malicious tools and knowledge.

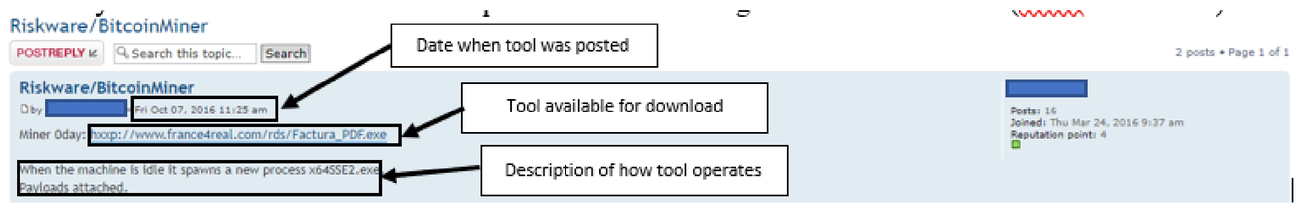


Figure 1. Example of a hacker providing a Bitcoin Miner 0-day exploit for free download

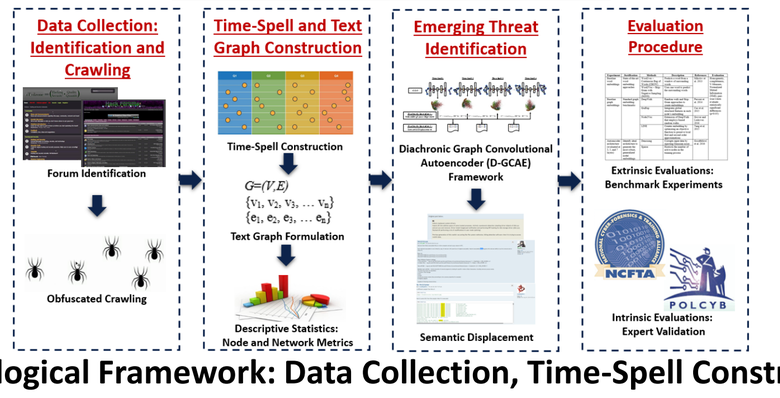
Key Challenges:

- Forum posts are unstructured, un-sanitized text.
- Hackers rapidly evolve their skills; thus, they develop new malware and augment existing exploits with novel functions.
- Unclear semantics of hacker terminology and how they shift over time.
- Existing CTI analytics are ill-equipped for these unique characteristics.
- Text analytics in hacker forum literature require significant extensions to generate valuable CTI.

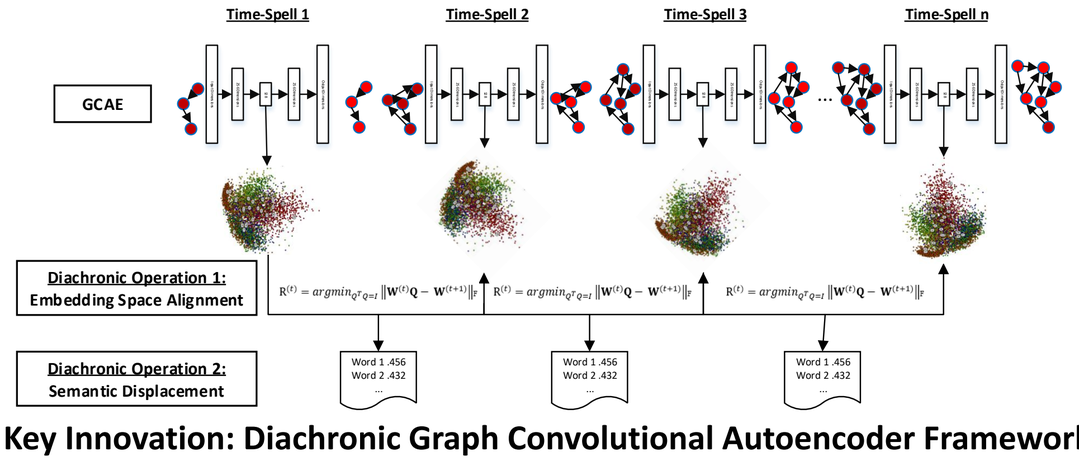
Scientific Impacts:

- This CRII SaTC project proposes a novel CTI framework designed to collect and identify emerging threats from multi-million record hacker forums.
- At the core of this framework, a novel computational algorithm called the Diachronic Graph Convolutional Autoencoder (D-GCAE).
- Draws upon and extends state-of-the-art in text graphs, diachronic linguistics, and unsupervised deep learning methodologies for emerging proactive CTI applications.

Proposed Solution



Methodological Framework: Data Collection, Time-Spell Construction, Emerging Threat Identification, and Internal/External Evaluations



Key Innovation: Diachronic Graph Convolutional Autoencoder Framework Novel Graph of Words for Hacker Content, Identifying Emerging Threats

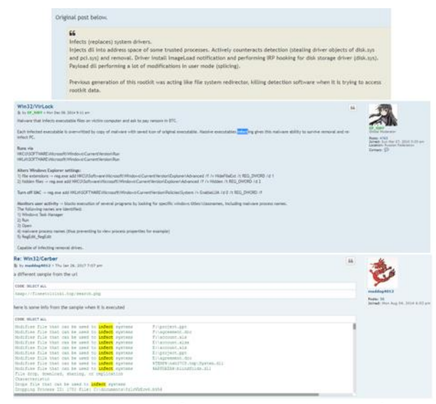
Broader Impact: Impact on Society

- Dissemination and integration of research to two international information sharing entities → 800+ partners across academia, industry, and government.

Broader Impact: Education and Outreach

- Integrating selected results into USF MS in Cyber program (#1 amongst veterans)
- Dissemination and outreach to Florida SUS via Cyber Florida (Florida Center for Cybersecurity)

Broader Impacts: Potential Impacts from Selected Results



- Integration of emerging threat functionalities into modern security software (e.g., SIEM).