CAREER: Securing Cyberspace: Gaining Deep Insights into the Online Underground Ecosystem

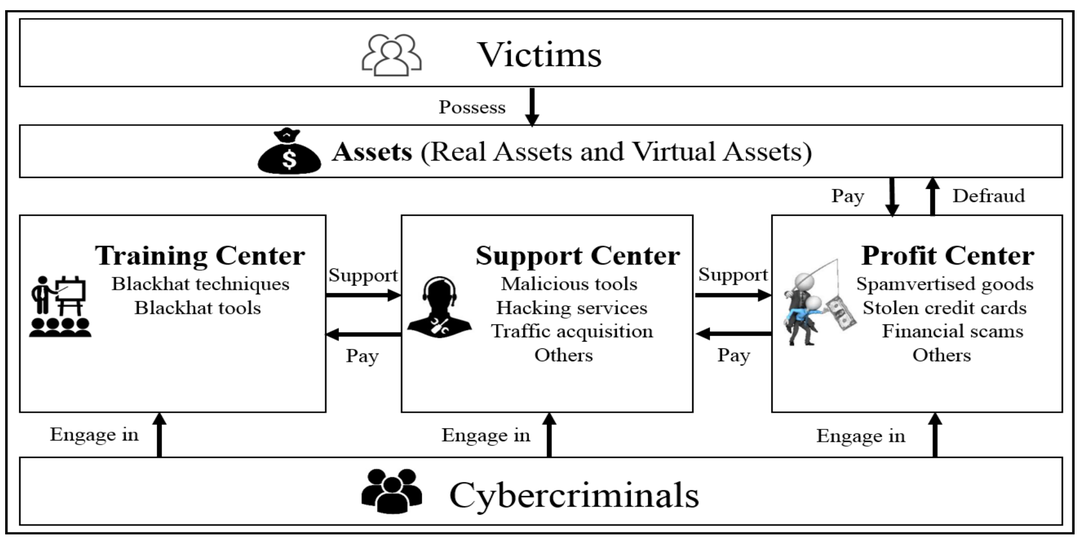
Award ID#: CNS-1940859(1845138)

PI: Yanfang (Fanny) Ye
Leonard Case Jr. Associate Professor
Department of Computer and Data Sciences (CDS)
Case School of Engineering (CSE)
Case Western Reserve University (CWRU)
yanfang.ye@case.edu

Project Description

Cybercrime is increasingly enabled by an online underground ecosystem, within which are anonymous forums and dark web platforms for cybercriminals to exchange knowledge and trade in illicit products and services. **The goal of this project** is to design and develop an integrated computational framework for in-depth investigation of the online underground ecosystem and thus to help secure cyberspace by producing data-driven interventions of cybercrimes.

Overall Structure of Online Underground Ecosystem

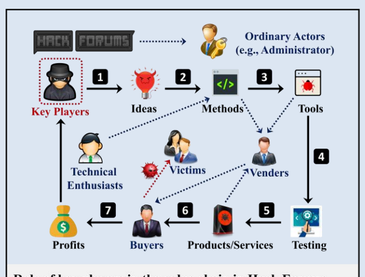


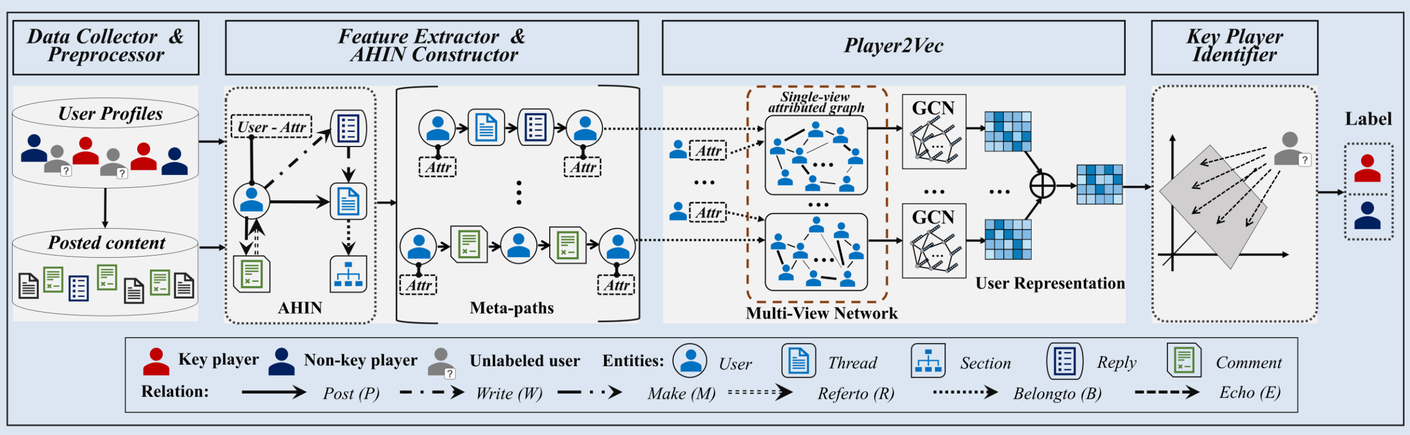
This proposed project seeks for ***data-driven intervention*** to address the key challenges of online underground ecosystem analysis, which focuses on the up-to-date crimeware and crimeware-as-a-service (CaaS). The research is three-fold:

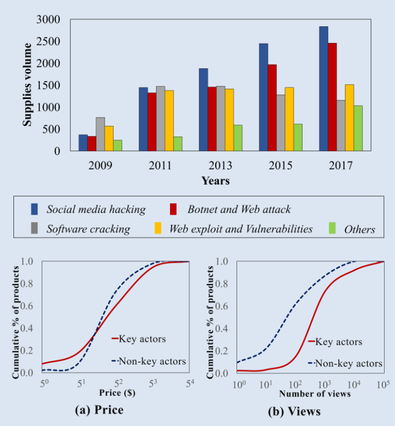
1. Design methods and develop scalable techniques to automate the analysis of online underground markets;
2. Develop novel frameworks for cross-market user identification and profiling;
3. Design innovative models and propose algorithms for structural analysis of cybercriminal social networks to gain deep insights into the organization and operation of cybercriminals in the ecosystem.

Solution (technical approach, key innovations, new contributions)

In order to combat the evolving cybercrimes, with the support of this project, we propose and develop an intelligent system (i.e., *iDetective*) to automate the analysis of underground forums for the identification of key players (i.e., users who play the vital role in the value chain). In *iDetective*, we first introduce an attributed heterogeneous information network (AHIN) for user representation and use a meta-path based approach to incorporate higher-level semantics to build the relatedness over users in underground forums; then we propose *Player2Vec* to efficiently learn node (i.e., user) representations in AHIN for key player identification. with the support of this project, the PI's work on adversarial machine learning to combat crimeware won the prestigious ***AICS 2019 Challenge Problem Winner*** and she also recently received the ***IJCAI 2019 Early Career Spotlights***.







The trend of different kinds of products or services advertised by identified key players in Hack Forums.

Key players vs. non-key players on price and views.

Selected Publications:

1. Yiming Zhang*, Yujie Fan*, **Yanfang Ye**, Chuan Shi, Liang Zhao. "Key Player Identification in Underground Forums over Attributed Heterogeneous Information Network Embedding Framework", International Conference on Information and Knowledge Management (CIKM), 2019. *(19.4% acceptance rate)*
2. Yiming Zhang*, Yujie Fan*, Wei Song, Shifu Hou*, **Yanfang Ye**, Xin Li, Liang Zhao, Chuan Shi, Jiabin Wang, Qi Xiong. "Your Style Your Identity: Leveraging Writing and Photography Styles for Drug Trafficker Identification in Darknet Markets over Attributed Heterogeneous Information Network", WWW, 2019. *(20% acceptance rate for short paper)*
3. **Yanfang Ye**, Shifu Hou*, Lingwei Chen*, Jingwei Lei, Wenqiang Wan, Jiabin Wang, Qi Xiong, Fudong Shao. "Out-of-sample Node Representation Learning for Heterogeneous Graph in Real-time Android Malware Detection", 28th International Joint Conference on Artificial Intelligence (IJCAI), 2019. *(17.9% acceptance rate)*
4. Deqiang Li, Qianmu Li, **Yanfang Ye**, Shouhuai Xu. "Enhancing Robustness of Deep Neural Networks Against Adversarial Malware Samples: Principles, Framework, and Application to AICS'2019 Challenge". The AAAI-19 Workshop on Artificial Intelligence for Cyber Security (AICS), 2019. **AICS 2019 Challenge Problem Winner**.
5. Qingzhe Li, Amir Alipour-Fanid, Martin Slawski, **Yanfang Ye**, Lingfei Wu, Kai Zeng, Liang Zhao. "Large-scale Cost-aware Classification Using Feature Computational Dependencies", IEEE TKDE, 2019.

Scientific Impacts and Broad Impacts to Society and other Domains

- **Scientific Impacts.** The proposed work could reach other fields such as social or economic sciences (e.g., the business model of CaaS, which includes actors, value chains and operation modes, can facilitate the research in economic science).
- **Societal Impacts.** This project will benefit to scientific communities and society as a whole by developing interventions into online crime to secure cyberspace for its users.
- **Impacts to other Domains.** The developed techniques can be used in other security domains, such as spammer detection and anti-fraud. The proposed methods can be applicable in different learning tasks in data mining and machine learning, such as multi-class classification and network embedding.

Integrating Research with Education

- **Curriculum Development Activities.** PI Ye has developing a new graduate-level course *CS591L Cyber Security and Big Data Analytics* and a new undergraduate-level course *CE349/444 Practicing Cybersecurity: Attacks and Countermeasures*.
- **Robust Outreach Efforts** to K-12, general public, undergraduate, graduate, minority, and women in cybersecurity. The establishment of a **cybersecurity lab** through this project will enhance education and workforce training in cybersecurity

PI Ye: STEM-Engineering Challenge Camp (All Female) & Summer Coding Camps.

"Innovation, Research, Education - for a Better World!"

Interested in meeting the PIs and our works? Attach post-it note below!