Intelligent Malware Detection Utilizing Novel File Relation-Based Features and Resilient Techniques for Adversarial Attacks

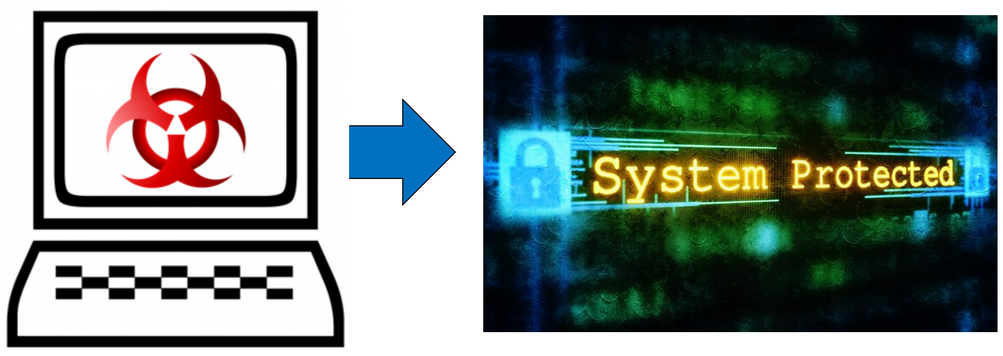
PI: Yanfang (Fanny) Ye
Leonard Case Jr. Associate Professor
Department of Computer and Data Sciences (CDS)
Case School of Engineering (CSE)
Case Western Reserve University (CWRU)
yanfang.ye@case.edu

Award ID#: CNS-1946327(1618629)

Project Description

Driven by the considerable economic benefits, both the sophistication and quantity of malware have significantly increased. To protect legitimate users from the evolutional malware attacks, we **aim to** develop more powerful and resilient techniques at both feature and model levels which are capable of protecting the users against increasingly sophisticated cyber attacks.

How secure is your computer?



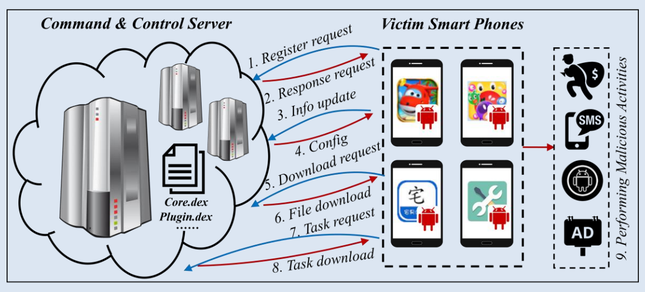
Our goal is to design and develop intelligent and resilient techniques at both feature and model levels against adversarial malware attacks.

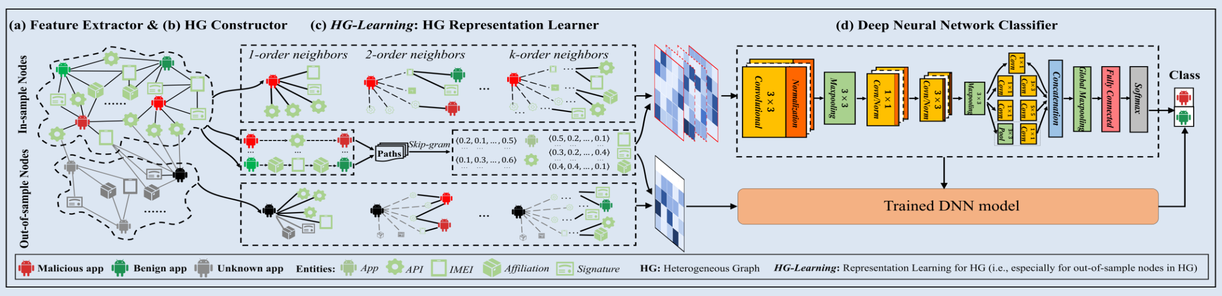
Built upon the PI's long-term and strong collaborations with cybersecurity industry partners (e.g., Comodo), we propose to **address the key challenges for modern malware detection** by answering following three questions:

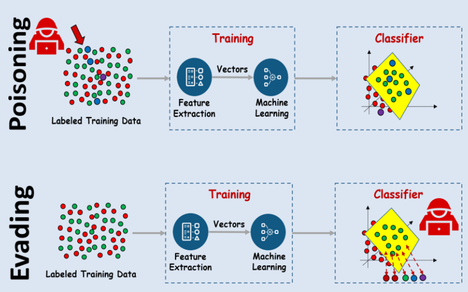
1. Besides file contents, what kinds of newly novel features can be used for malware detection?
2. How to construct an effective model to detect the unknown malware utilizing both content- and relation-based features?
3. How to develop resilient techniques that are robust and secure against adversarial attacks?

Solution (technical approach, key innovations, new contributions)

To combat the evolving malware attacks, with the support of this project, we were the first to propose novel heterogeneous graph based models to comprehensively characterize higher-level semantic relationships among different files within the ecosystem, based on which we have developed a series of intelligent detection systems (e.g., HinDroid, AiDroid, αCyber, SecureDroid) against adversarial malware attacks. Our work in intelligent malware detection (HinDroid) received the prestigious ***ACM SIGKDD 2017 Best Paper Award*** and ***ACM SIGKDD 2017 Best Student Paper Award*** (Applied Data Science Track), and our work in adversarial machine learning won the ***AICS 2019 Challenge Problem Winner*** and the ***IEEE EISIC 2017 Best Paper Award***. The PI also recently received the ***IJCAI 2019 Early Career Spotlights***.







Selected Publications: (selected from 30+ publications)

1. **Yanfang Ye**, Tao Li, Donald Adjeroh, S. Sitharama Iyengar. "A Survey on Malware Detection Using Data Mining Techniques", ACM Computing Surveys, Vol. 50, No. 41, 2017. (94 citations)
2. **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)*
3. 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**.
4. Shifu Hou*, **Yanfang Ye**, Yangqiu Song, Melih Abdulhayoglu. "HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network", ACM SIGKDD, 2017. **SIGKDD 2017 Best Paper Award** and **SIGKDD 2017 Best Student Paper Award**, (Applied Data Science track*) (9.2% acceptance rate for oral)*
5. Lingwei Chen*, Shifu Hou*, **Yanfang Ye**. "SecureDroid: Enhancing Security of Machine Learning-based Detection against Adversarial Android Malware Attacks", ACSAC, 2017. *(19.7% acceptance rate)*
6. Lingwei Chen*, **Yanfang Ye**, Thirimachos Bourlai. "Adversarial Machine Learning in Malware Detection: Arms Race between Evasion Attack and Defense", European Intelligence and Security Informatics Conference (EISIC), Pages: 99-106, 2017. **IEEE EISIC 2017 Best Paper Award**. *(25% acceptance rate)*

Scientific Impacts and Broad Impacts to Society and other Domains

- **Scientific Impacts.** This project provides an effective way to identify different threats to trustworthiness caused by malware, and create a resilient platform at both feature and model levels against adversarial malware attacks.
- **Societal Impacts.** The proposed techniques have been incorporated into popular commercial cybersecurity products such as Comodo AntiVirus that protect millions of users worldwide. It benefits the society at large by making cyberspace more secure and resilient to cyber-attacks.
- **Impacts to other Domains.** The developed techniques are designed to be arms race capable so that they can also be used in other security domains, such as anti-spam, fraud detection, and counter-terrorism.

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.

PI Ye: STEM-Engineering Challenge Camp (All Female), 2016; Summer Coding Camps, 2017-18.

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

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