Yue Gao

+1 (608) 733-8789 gy@cs.wisc.edu pages.cs.wisc.edu/~gy ygao234 Lodour Madison, WI

Research Interests

Trustworthy Machine Learning (adversarial robustness, black-box evasion attacks and defenses)
System Security (machine learning systems, web-based applications and services)

Education

University of Wisconsin–Madison Madison, WI
Ph.D. Candidate in Computer Science *Sep 2018 – May 2024 (expected)*

- Advisor: Prof. Kassem Fawaz
- Thesis: *Characterizing the Limitations of Defenses in Adversarial Machine Learning*

Shanghai University Shanghai, China
B.S. in Computer Science and Technology *Sep 2014 – Jul 2018*

- Major GPA: 3.99/4.00 (ranked 1/292)
- Advisor: Prof. Xiaodong Yue
- Thesis: *A Deep Neural Network based Image Compression Method*

Work Experience

Research Assistant @ University of Wisconsin–Madison Madison, WI
Advised by Prof. Kassem Fawaz *Nov 2018 – present*

- Explore the weaknesses of evasion attacks and defenses for ML-based systems.
- Improve the security analysis of ML-based and web-based systems in black-box settings.

Research Intern @ Microsoft Research Redmond, WA
Mentored by Dr. Jay Stokes and Dr. Emre Kiciman *Jun 2021 – Sep 2021*

- Explore data-centric solutions for backdoor attacks on language models with domain knowledge.
- Design auditing frameworks for the continual update of backdoor-free language models.

Research and Development Intern @ TuCodec Shanghai, China
Mentored by Dr. Chunlei Cai *Jan 2018 – Jul 2018*

- Optimize learning-based image compression algorithms.
- Develop DNN-based applications on mainstream operating systems and deploy them to cloud services.
- Winner of the CVPR 2018 Challenge on Learned Image Compression.

Selected Projects

The Role of Randomization in Adversarial Robustness *Feb 2022 – May 2022*

- Characterize the limitations of using randomization to defend ML models.
- Theoretically explain the source of robustness for randomized defenses against evasion attacks.

Trustworthy Machine Learning in Real-World Systems *Sep 2020 – Jan 2021*

- Explore the security of ML systems under threats from multiple components.
- Propose plug-and-play techniques to enable system-level black-box attacks.
- Demonstrate amplified threats from the interplay between multiple vulnerabilities.

Defending against Evasion Attacks in Multimodal Scenarios (Collaborative) *Since 2019 (semiannual)*

- Improve adversarial robustness with physical constraints.
- Design defenses for multimodal tasks (e.g., remote sensing satellites and autonomous driving).
- Develop a usable code base for team members with varying tracks and technical backgrounds.

Publications

Conference

[1] I Know Your Triggers: Defending Against Textual Backdoor Attacks With Benign Backdoor Augmentation
Yue Gao, Jack W. Stokes, Manoj Prasad, Andrew Marshall, Kassem Fawaz, and Emre Kiciman
IEEE Military Communications Conference (MILCOM), 2022

[2] On the Limitations of Stochastic Pre-processing Defenses
Yue Gao, Ilia Shumailov, Kassem Fawaz, and Nicolas Papernot
Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS), 2022

[3] Rethinking Image-Scaling Attacks: The Interplay Between Vulnerabilities in Machine Learning Systems
Yue Gao, Ilia Shumailov, and Kassem Fawaz
Proceedings of the 39th International Conference on Machine Learning (ICML), 2022
Oral Presentation (Top 2%)

[4] Experimental Security Analysis of the App Model in Business Collaboration Platforms
Yunang Chen*, **Yue Gao***, Nick Ceccio, Rahul Chatterjee, Kassem Fawaz, and Earlence Fernandes
31st USENIX Security Symposium (USENIX Security), 2022
Bug Bounty ($1500)

Workshop

[1] Variational Autoencoder for Low Bit-rate Image Compression
Lei Zhou*, Chunlei Cai*, **Yue Gao**, Sanbao Su, and Junmin Wu
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018
Winner of the first Challenge on Learned Image Compression

Preprints

[1] Analyzing Accuracy Loss in Randomized Smoothing Defenses
Yue Gao*, Harrison Rosenberg*, Kassem Fawaz, Somesh Jha, and Justin Hsu
arXiv, 2020

Selected Honors & Awards

Slack Bug Bounty: Medium Severity, $1500 — 2022
Top 10% Reviewers Award: NeurIPS — 2022
CVPR Competition Winner: Challenge on Learned Image Compression — 2018
National Scholarship: China — 2017
Top 100 Elite Collegiate Award: China Computer Federation — 2017
Scholarship for Exceptional Leadership: Shanghai University — 2017
City Scholarship: Shanghai — 2016
Outstanding Student Award: Shanghai University — 2016
Outstanding Volunteer Award: ACM ICPC Asia Regional Contest — 2016
Scholarship for Exceptional Innovation: Shanghai University — 2016
Scholarship for Exceptional Academic Achievements: Shanghai University — 2015 – 2018
Bronze Prize for Programming Contest: ACM ICPC Asia East-Continent Final Contest — 2015
Bronze Prize for Programming Contest: ACM ICPC Asia Shanghai Regional Contest — 2015

Professional Activities

Reviewer: NeurIPS and ICML — 2022 – 2023
External Reviewer: USENIX Security Symposium — 2021 – 2022
External Reviewer: IEEE Symposium on Security and Privacy — 2021 – 2022
External Reviewer: ACM Conference on Computer and Communications Security — 2019
Team Leader: Collegiate ICPC Team at Shanghai University — 2016 – 2017

TALKS

1. **On the Limitations of Stochastic Pre-processing Defenses** *Oct 2022*
   University of Southern California (virtual)
2. **The Interplay Between Vulnerabilities in Machine Learning Systems** *Sep 2022*
   University of Michigan
3. **Experimental Security Analysis of the App Model in Business Collaboration Platforms** *Aug 2022*
   USENIX Security 2022
4. **The Interplay Between Vulnerabilities in Machine Learning Systems** *Jun 2022*
   ICML 2022

TEACHING AND MENTORING

Teaching Assistant: CS 368 (C++ for Java Programmers), University of Wisconsin–Madison	Fall 2018
Guest Lecturer: Advanced Algorithms & Data Structures, Shanghai University	2015 – 2017
Problem Designer: Undergraduate Programming Contests, Shanghai University	2015 – 2017
Student Mentor: Undergraduate Computer Science Coursework, Shanghai University	2015 – 2017

TECHNICAL SKILLS

Python	Research (2018 – present), System Optimization (2018), Backend Development (2016 – 2017).
PyTorch	Research (2019 – present), Distributed Training (2020 – 2022).
Docker	Research (2018 – present), Computing Cluster (2017 – 2018).
C / C++	Kernel Development (2019), System Optimization (2018), Programming Contest (2014 – 2018).
TensorFlow	Service Deployment (2018).
Java EE	Backend Development (2016).

ARTICLES AND MEDIA COVERAGE

CleverHans. Can stochastic pre-processing defenses protect your models?	2022
USENIX login. Experimental Security Analysis of the App Model in Business Collaboration Platforms	2022
Wired. Slack's and Teams' Lax App Security Raises Alarms	2022