

Zhiyuan Yu

yu.zhiyuan@wustl.edu
https://zh1yu4nyu.github.io

RESEARCH INTERESTS

Cyber-physical Security: Cyber-physical system security; AI/ML security & privacy; Medical security;

EDUCATION

Ph.D. in Computer Science Expected 2025

Advisor: Dr. Ning Zhang

Washington University in Saint Louis, MO

B.S. in Electrical Engineering

2015-2019

Huazhong University of Science and Technology, Wuhan, China

PUBLICATIONS

Conference

1. Zhiyuan Yu, Xiaogeng Liu, Shunning Liang, Zach Cameron, Chaowei Xiao, Ning Zhang
Don't Listen To Me: Understanding and Exploring Jailbreak Prompts of Large Language Models
33rd USENIX Security Symposium (USENIX Sec), 2024
***Distinguished Paper Award**
Artifact Badges: **Available, Functional, Results Reproduced**
2. Han Liu, Yuhao Wu, Zhiyuan Yu, Ning Zhang
Please Tell Me More: Privacy Impact of Explainability through the Lens of Membership Inference Attack
45th IEEE Symposium on Security and Privacy (IEEE S&P), 2024
3. Zhiyuan Yu, Shixuan Zhai, Ning Zhang
AntiFake: Using Adversarial Audio to Prevent Unauthorized Speech Synthesis
ACM Conference on Computer and Communications Security (CCS), 2023
***Winner of the Federal Trade Commission (FTC) Voice Cloning Challenge, 2024.**
***Reported by National Public Radio (NPR). "Worried about AI hijacking your voice for a deepfake? This tool could help"**
4. Zhiyuan Yu, Yuanhaur Chang, Shixuan Zhai, Nicholas Deily, Tao Ju, XiaoFeng Wang, Uday Jammalamadaka, Ning Zhang
XCheck: Integrity Verification for 3D Printed Patient-Specific Devices via Computing Tomography
32nd USENIX Security Symposium (USENIX Sec), 2023
***Distinguished Artifact Award**
Artifact Badges: **Available, Functional, Results Reproduced**
5. Zhiyuan Yu, Yuanhaur Chang, Ning Zhang, Chaowei Xiao
SMACK: Semantically Meaningful Adversarial Audio Attack
32nd USENIX Security Symposium (USENIX Sec), 2023
Artifact Badges: **Available, Functional, Results Reproduced**
6. Zhiyuan Yu, Yuhao Wu, Ning Zhang, Chenguang Wang, Yevgeniy Vorobeychik, Chaowei Xiao
CodeIPrompt: Intellectual Property Infringement Assessment of Code Language Models
40th International Conference on Machine Learning (ICML), 2023

7. Han Liu, Yuhao Wu, Zhiyuan Yu, Yevgeniy Vorobeychik, Ning Zhang
SlowLiDAR: Increasing the Latency of LiDAR-Based Detection Using Adversarial Examples
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023
8. Zhiyuan Yu, Yuhao Wu, Ning Zhang, Chenguang Wang, Yevgeniy Vorobeychik, Chaowei Xiao
Poster: Intellectual Property Infringement Assessment of Code Language Models
44th IEEE Symposium on Security and Privacy (IEEE S&P), 2023
9. Zhiyuan Yu, Zhuohang Li, Yuanhaur Chang, Skylar Fong, Jian Liu, Ning Zhang
HeatDeCam: Detecting Hidden Spy Cameras via Thermal Emissions
ACM Conference on Computer and Communications Security (CCS), 2022
10. Han Liu, Zhiyuan Yu, Mingming Zha, XiaoFeng Wang, William Yeoh, Yevgeniy Vorobeychik, Ning Zhang
When Evil Calls: Targeted Adversarial Voice over IP Network
ACM Conference on Computer and Communications Security (CCS), 2022
11. Ao Li, Marion Sudvarg, Han Liu, Zhiyuan Yu, Chris Gill, Ning Zhang
PolyRhythm: Adaptive Tuning of a Multi-Channel Attack Template for Timing Interference
IEEE Real-Time Systems Symposium (RTSS), 2022
12. Huifeng Zhu, Zhiyuan Yu, Weidong Cao, Ning Zhang, Xuan Zhang
PowerTouch: A Security Objective-Guided Automation Framework for Generating Wired Ghost Touch Attacks on Touchscreens
IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2022
13. Brian Tung, Zhiyuan Yu, Ning Zhang
Towards Automated Computational Auditing of mHealth Security and Privacy Regulations
ACM Conference on Computer and Communications Security (CCS), 2021
14. Wei Yan, Huifeng Zhu, Zhiyuan Yu, Fatemeh Tehranipoor, John Chandy, Ning Zhang, Xuan Zhang
Bit2RNG: Leveraging Bad-page Initialized Table with Bit-error Insertion for True Random Number Generation in Commodity Flash Memory
IEEE International Symposium on Hardware Oriented Security and Trust (HOST), 2020

Journal

1. Zhiyuan Yu, Zack Kaplan, Qiben Yan, and Ning Zhang
Security and Privacy in the Emerging Cyber-Physical World: A Survey
IEEE Communications Surveys and Tutorials (COMST), 2021
2. Yuanzheng Li, Tianyang Zhao, Chang Liu, Yong Zhao, Zhiyuan Yu, Kaicheng Li, Lei Wu
Day-ahead Coordinated Scheduling of Hydro and Wind Power Generation System Considering Uncertainties
IEEE Transactions on Industry Applications, 2019

AWARDS AND HONORS

- Machine Learning and Systems Rising Star, 2024
- USENIX Security Distinguished Paper Award, 2024
- USENIX Security Distinguished Artifact Award, 2023
- USENIX Security Student Travel Grant, 2023
- ACM CCS Student Travel Grant, 2022

Invited Talks

- “Safeguarding Voices via Adversarial Examples: Defense and Way Forward in the Era of GenAI”
International Speech Communication Association Webinar Jan 15th, 2024

TEACHING

Washington University in St. Louis

- CSE 569S: Recent Advances in Computer Security and Privacy Spring 2022

Role: Teaching Assistant, Student Body: Undergraduate/Graduate = 3/29

Evaluation: Overall : 6.54 (Department Average: 5.44), Inclusive : 6.74 (Department Average: 6.02)

- CSE 569S: Recent Advances in Computer Security and Privacy Spring 2024

Role: Guest Lecturer, Lecture Topic: “Security and Safety in Generative AI”

ACADEMIC ACTIVITIES AND SERVICES

Conference Organization

- Web Chair, 10th ACM Workshop on Moving Target Defense (MTD’23)

Journal/Conference Review

- IEEE/ACM Transactions on Networking
- ACM Conference on Computer and Communications Security (CCS)
- ACM ASIA Conference on Computer and Communications Security (ASIACCS)
- IEEE European Symposium of Security and Privacy (EuroS&P)
- ISOC The Network and Distributed System Security Symposium (NDSS)
- IEEE International Conference on Computer Communications (INFOCOM)
- International Conference on Pattern Recognition (ICPR)
- IEEE Winter Conference on Applications of Computer Vision (WACV)
- ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)

Volunteering

- Student Volunteer, ACM Conference on Computer and Communications Security (CCS) 2022