YI ZENG

(858)-952-2135 ◇ yizeng@vt.edu
Google Scholar ◇ Github ◇ LinkedIn◇ Webpage

EDUCATION

Virginia Polytechnic Institute and State University (Virginia Tech) *May. 2021 - May. 2026*
Ph.D. Student in Computer Engineering
Advisor: *Prof.* **Ruoxi Jia**

University of California - San Diego (UCSD) *Aug. 2019 - Mar. 2021*
Master of Science in Machine Learning and Data Science, Electrical and Computer Engineering

Xidian University (XDU) *Sep. 2015 - Jun. 2019*
Bachelor of Engineering in Electrical and Information Engineering
Senior Thesis: Deep Learning Based Network Encrypted Traffic Classification and Intrusion Detection
Advisor: *Prof.* **Huaxi Gu**

SPECIAL AREAS

AI Security, **Deep Learning**, **Adversarial Machine Learning**, **Data Security**, **Backdoor Attacks**

WORK EXPERIENCE

Sony Corporation of America, NY, USA *May. 2022 - Present*
AI Research Intern @ Privacy-Preserving Machine Learning Team
Towards meta-robust training against general dataset corruptions from a security perspective.

Jacobs School of Engineering, UCSD, CA, USA *Aug. 2019 - Mar. 2021*
Research Assistant Volunteer @ Adaptive Computing and Embedded Systems Lab
Rethinking the adversarial robustness of hyperdimensional computing.

College of Electrical Engineering, Columbia University, NY, USA *Mar. 2018 - Nov. 2018*
Visting Scholar @ Signal Processing & Communications Lab
Towards practical defenses against adversarial attacks via automatic evaluation and input augmentations.

College of Electrical Engineering, XDU, Shaanxi, China *Sep. 2015 - Jun. 2019*
Research Assistant @ State Key Lab of Integrated Service Networks
Deep-learning-based network traffic decryption and intrusion detection.

HONORS & AWARDS

- **Best Paper Award**, 20th ICA3PP. *2020*
- **Outstanding Senior Thesis Award**, Xidian University. *2019*
- **Outstanding Academic Scholarship**, Xidian University. *2015, 2016, 2017, 2018*

INVITED TALKS

- **Online Talk on *Trojaning Advanced AI and Countermeasures***. AI TIME. *Jun. 2022*
- **Online Talk on *Advanced Backdoor Attacks in Deep Learning***. CSIG, BAAI, Meituan. *Aug. 2021*

TECHNICAL STRENGTHS

Programming: Python, Matlab, C/C++, HTML
Frameworks: Tensorflow, Pytorch, Numpy, Cleverhans, Foolbox, SciPy, Scikit-learn

PROFESSIONAL SERVICE

Conference Reviewer: CVPR-22, NeurIPS-22, ICML-22, ECCV-22
Conference PC Member: AAAI-22, KSEM-22, KSEM-21, EUC-21, IEEE ISPA-21, ICA3PP-20
Journal Reviewer: IEEE TDSC, IEEE TII

CONFERENCE PUBLICATIONS

(i) **Adversarial Unlearning of Backdoors via Implicit Hypergradient**
Yi Zeng, Si Chen, Won Park, Z. Morley Mao, Jin Ming and Ruoxi Jia
International Conf. on Learning Representations (ICLR), 2022.

(ii) **Rethinking the Backdoor Attacks' Triggers: A Frequency Perspective**
Yi Zeng*, Won Park*, Z. Morley Mao and Ruoxi Jia
International Conf. on Computer Vision (ICCV), 2021.

(iii) **DeepSweep: An Framework for Mitigating DNN Backdoor Attacks using Data Augmentation**
Han Qiu, **Yi Zeng**, Shangwei Guo, Tianwei Zhang, Meikang Qiu and Bhavani Thuraisingham
In Proceeding of the ACM Asia Conf. on Computer and Communications Security (AsiaCCS), 2021.

(iv) **Fine-tuning Is Not Enough: A Simple yet Effective Watermark Removal Attack for DNN Models**
Shangwei Guo, Tianwei Zhang, Han Qiu, **Yi Zeng**, Tao Xiang and Yang Liu
In Proceeding of the International Joint Conf. on Artificial Intelligence (IJCAI), 2021.

(v) **Defending Adversarial Examples in Computer Vision based on Data Augmentation Techniques**
Yi Zeng, Han Qiu, Gerard Memmi and Meikang Qiu
Best Paper of International Conf. on Algo & Archit for Parallel Processing (ICA3PP), 2020.

(vi) **Model Uncertainty for Annotation Error Correction in DL Based Intrusion Detection System**
Wencheng Chen, Hongyu Li, **Yi Zeng**, Zichang Ren and Xingxin Zheng
IEEE International Conf. on Smart Cloud (IEEE SmartCloud), IEEE, 2019.

(vii) **Using Adversarial Examples to Bypass Deep Learning Based URL Detection System**
Wencheng Chen, **Yi Zeng** and Meikang Qiu
IEEE International Conf. on Smart Cloud (IEEE SmartCloud), IEEE, 2019.

(viii) **End-to-End Network Traffic Classification System With Spatio-Temporal Features Extraction**
Yi Zeng, Zihao Qi, Wencheng Chen and Yanzhe Huang.
IEEE International Conf. on Smart Cloud (IEEE SmartCloud), IEEE, 2019.

(ix) **Time-Division based Scheduling Scheme for Hybrid Optical/Electrical Data Center Network**
Shangqi Ma, Xiaoshan Yu, Kun Wang, **Yi Zeng** and Huaxi Gu
International Conf. on Optical Communications and Networks (ICOCN), IEEE, 2019.

(x) **V-PSC: A Perturbation-Based Causative Attack Against DL Classifiers' Supply Chain in VANET**
Yi Zeng, Meikang Qiu, Jingqi Niu, Yanxin Long, Jian Xiong and Meiqin Liu
IEEE International Conf. on Embedded and Ubiquitous Computing (IEEE EUC), IEEE, 2019.

(xi) **DeepVCM: A Deep Learning Based Intrusion Detection Method in VANET**
Yi Zeng, Meikang Qiu, Dan Zhu, Zhihao Xue, Jian Xiong and Meiqin Liu
IEEE International Conf. on High Performance and Smart Computing (IEEE HPSC), IEEE, 2019.

(xii) **Joint Energy and Spectrum Efficient Virtual Optical Network embedding in EONs**
Wenting Wei, Huaxi Gu, Achille Pattavina, Jiru Wang and **Yi Zeng**
IEEE International Conf. on High Performance Switching and Routing (IEEE HPSR), IEEE, 2019.

(xiii) **Senior2local: A Machine Learning Based Intrusion Detection Method for VANETs**
Yi Zeng, Meikang Qiu, Zhong Ming and Meiqin Liu
International Conf. on Smart Computing and Communication (SmartCom), Springer, 2018.

JOURNAL PUBLICATIONS

(i) **Adaptive Backdoor Trigger Detection in Edge-Deployed DNNs in 5G-Enabled IIoT Systems**
Yi Zeng, Ruoxi Jia and Meikang Qiu
IEEE Transactions on Industrial Informatics, 2021.

(ii) **An Effective and Efficient Preprocessing-based Approach to Mitigate Advanced Adversarial Attacks**
Han Qiu*, **Yi Zeng***, Qinkai Zheng, Tianwei Zhang, Meikang Qiu and Bhavani Thuraisingham
IEEE Transactions on Computers, 2020.

(iii) **Optimizing Energy and Spectrum Efficiency of Virtual Optical Network Embedding in Elastic Optical Networks**
Wenting Wei, Huaxi Gu, Achille Pattavina, Jiru Wang and **Yi Zeng**
Optical Switching and Networking, 2019.

(iv) **Deep Learning Based Network Encrypted Traffic Classification and Intrusion Detection Framework**
Yi Zeng, Huaxi Gu, Wenting Wei and Yantao Guo
IEEE Access, 2019.

BOOK PUBLICATIONS

(i) **Research and Technical Writing for Science and Engineering**
Meikang Qiu, Han Qiu and **Yi Zeng**
CRC Press, 2022.

MANUSCRIPTS

(i) **NARCISSUS: A Practical Clean-Label Backdoor Attack with Limited Information**
Yi Zeng*, Minzhou Pan*, Hoang Anh Just, Lingjuan Lyu, Meikang Qiu and Ruoxi Jia
Preprint on the arXiv, 2022.

(ii) **A Unified Framework for Task-Driven Data Quality Management**
Tianhao Wang, **Yi Zeng**, Ming Jin and Ruoxi Jia
Preprint on the arXiv, 2021.

(iii) **FenceBox: A Platform for Defeating Adversarial Examples with Data Augmentation Techniques**
Han Qiu, **Yi Zeng**, Tianwei Zhang and Meikang Qiu
Preprint on the arXiv, 2020.