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*
Doctor of Philosophy in Computer Engineering
University of California - San Diego (UCSD) *Aug. 2019 - Mar. 2021*
Master of Science in Machine Learning and Data Science
Xidian University (XDU) *Sep. 2015 - Jun. 2019*
Bachelor of Engineering in Electrical and Information Engineering

HONORS & AWARDS

- **Best Paper Award**, 20th International Conf. on Alg.o & Archit. for Parallel Processing (ICA3PP), 2020;
- **Best Degree Paper Award**, Xidian University, Top 2%, 2019;
- **Outstanding Academic Scholarship**, Xidian University, Top 10%, year 2015, 2016, 2017, 2018;

SELECTED PUBLICATIONS & MANUSCRIPTS

(i) **Adversarial Unlearning of Backdoors via Implicit Hypergradient**
Yi Zeng, Si Chen, Won Park, Z. Morley Mao, Jin Ming and Ruoxi Jia
International Conference 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 Conference on Computer Vision (ICCV), 2021.

(iii) **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.

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

(v) **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 Conference on Computer and Communications Security (AsiaCCS), 2021.

(vi) **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 Conference on Artificial Intelligence (IJCAI), 2021.

(vii) **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.

(viii) **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.

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

(x) **End-to-End Network Traffic Classification System With Spatio-Temporal Features Extraction**
Yi Zeng, Zihao Qi, Wencheng Chen and Yanzhe Huang.
IEEE International Conference on Smart Cloud (IEEE SmartCloud), 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 Intl Conference on High Performance and Smart Computing (IEEE HPSC), IEEE, 2019.

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

EXPERIENCE HIGHLIGHTS

Bradley Department of ECE, Virginia Tech, VA, USA *May. 2021 - Present*
Graduate Research Assistant @ Responsible Data Science Lab **4** Publications & Manuscripts
College of Electrical Engineering, Columbia University, NY, USA *Mar. 2018 - Mar. 2021*
Visting Student Scholar @ Signal Processing & Communications Lab **11** Publications & Manuscripts
College of Electrical Engineering, XDU, Shaanxi, China *Sep. 2015 - Jun. 2019*
Research Assistant @ State Key Lab of Integrated Service Networks **4** Publications & Manuscripts

MINORS & SPECIAL AREAS

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

TECHNICAL STRENGTHS

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

PROFESSIONAL SERVICE

Reviewer: International Conference on Machine Learning (ICML-22)
Reviewer: IEEE/CVF, Conf. on Computer Vision and Pattern Recognition (CVPR-22)
PC Member: Elsevier, 36th AAAI Conf. on Artificial Intelligence (AAAI-22)
PC Member: Springer, 20th International Conf. on Algorithms and Architectures for Parallel Processing
PC Member: IEEE, 17th International Conf. on Embedded and Ubiquitous Computing
PC Member: Springer, 15th International Conf. on Knowledge Science, Engineering and Management

SELECTED PROJECTS

Project ① (2021): Investigating General Robustness Issues in the Context of Backdoor Attacks on Deep Learning Models *Advisor:* ***Prof. Z. Morley Mao*** *&* ***Prof. Ruoxi Jia***

- The frequency-domain was identified as an underutilized domain for backdoor attacks and defenses: a supervised frequency-based backdoor detector with an average detection efficacy of 98.5% on unseen triggers was proposed; a novel backdoor trigger invisible in both the image and frequency domains was proposed, capable of breaking previous defenses.
- Proposed a minimax formulation of backdoor removal and an implicit hypergradient based method is proposed to solve the minimax. Theoretically, we show convergence and generalizability. Empirically, we show that our method is the only generalizable defense on 7 different triggers, 3 attack patterns, on 2 datasets; the proposed method is also more efficient (average 7.35s for one-target attacks) and more robust to poison rate and the number of available clean samples.

Project ② (2020): Mitigating White-box Adversarial Attacks toward Deep Learning Models with Preprocessing-only Techniques as the Defense. *Advisor:* ***Dr. Han Qiu*** *&* ***Prof. Meikang Qiu***

- Developed the first preprocessing-only adversarial defense method that demonstrates robustness against advanced interactive adversarial attacks (BPDA and EOT) on ImageNet, allowing the attack success rate of the most advanced adaptive attack to remain below 7% even after unlimited rounds of $l_2 = 0.05$ bounded attack.
- Proposed a lightweight preprocessing framework able to provide real-time adversarial attack mitigation. The framework achieved averaging 50% better performance regarding lower attack success rates than others.

Project ③ (2020): Research on Developing Preprocessing-based Techniques to Mitigate Backdoor Attacks in DNNs. *Advisor:* ***Dr. Han Qiu*** *&* ***Prof. Tianwei Zhang***

- 64 existing preprocessing methods were thoroughly investigated on mitigating six different backdoor attacks.
- Proposed DeepSweep, a comprehensive backdoor defense method that is the first to take into account invisible backdoor attacks and successfully reduces the attack success rate of various advanced attacks to less than 18% in attack agnostic settings.

Project ④ (2019): Designing of Light-weight Network Traffic Classification/Identification Methods only Requires Raw Packets Based on Deep Learning Techniques. *Advisor:* ***Prof. Huaxi Gu***

- Developed an Encrypted Traffic Classification (ETC) and Intrusion Detection (ID) method based on CNN, LSTM, and SAE, outperforming published methods by 13.49 % on ETC's F1 and 12.15% on ID's F1.
- Proposed a Spatio-Temporal network traffic examination method based on 1D-CNN and LSTM, which attained an averaging accuracy of 99.98% on 2 public datasets.