Ren Pang

Resume

Education

2019–present **PhD Informatics**, *College of Information Sciences and Technology*, Pennsylvania State University, (Advisor: Ting Wang).

2018–2019 (transferred) **PhD Computer Science**, *Department of Computer Science and Engineering*, Lehigh University, (Advisor: Ting Wang).

2014–2018 **BSc. Mathematics**, *Department of Mathematics*, Nankai University.

Research Interests

Deep Learning Security; Adversarial Robustness; Trojan Backdoor

Publications

ACM CCS'20 **A Tale of Evil Twins: Adversarial Inputs versus Poisoned Models**.
Ren Pang, Hua Shen, Xinyang Zhang, Shouling Ji, Yevgeniy Vorobeychik, Xiapu Luo, Alex Liu, Ting Wang
ACM SAC Conference on Computer and Communications (CCS)

ACM KDD'20 **AdvMind: Inferring Adversary Intent of Black-Box Attacks**.
Ren Pang, Xinyang Zhang, Shouling Ji, Xiapu Luo, Ting Wang
ACM International Conference on Knowledge Discovery and Data Mining (KDD)

AAAI'21 **i-Algebra: Towards Interactive Interpretability of Neural Nets**.
Xinyang Zhang, **Ren Pang**, Shouling Ji, Fenglong Ma, Ting Wang
AAAI Conference on Artificial Intelligence

USENIX Security'21 **Graph Backdoor**.
Zhaohan Xi, **Ren Pang**, Shouling Ji, Ting Wang
USENIX Security Symposium

USENIX Security'22 **On the Security Risks of AutoML (conditionally accepted)**.
Ren Pang, Zhaohan Xi, Shouling Ji, Xiapu Luo, Ting Wang
USENIX Security Symposium

✉ *rbp5354@psu.edu* • *ain-soph.github.io*

Preprints

2020 **TROJANZOO: Everything you ever wanted to know about neural backdoors (but were afraid to ask)**.
Ren Pang, Zheng Zhang, Xiangshan Gao, Zhaohan Xi, Shouling Ji, Cheng Peng, Ting Wang

Research Experiences

2018 **Adversarial Vulnerabilities in Neural Networks**.
My work explores the mutual reinforcement effects between two attack vectors in deep learning: adversarial inputs and poisoned models, and designs a unified framework to control the trade-offs. The framework can be easily extended to the backdoor scenario and lead to a new powerful attack (IMC).

2019 **Detection of Black-box Adversarial Attacks**.
The project develops a novel estimation model to infer the adversary intent of black-box adversarial attacks.

2020 **Backdoors in Neural Networks**.
We construct a universal platform (TrojanZoo) that contains state-of-the-art works on backdoor attacks and defenses, evaluate the performance of different methods under the same metrics, and explore the underlying mechanism of backdoors in neural networks.

2021 **Vulnerabilities and Robustness in AutoML**.
We propose that Neural-Architecture-Search(NAS) algorithms introduce vulnerabilities of different kinds of attacks. Compared with human-designed models, the DARTS-like models tend to be more sensitive against PGD, TrojanNN, Membership Inference, etc.

GitHub

Owner **TrojanZoo**.
`https://github.com/ain-soph/trojanzoo`
TrojanZoo provides a universal pytorch platform to conduct security researches (especially backdoor attacks/defenses) of image classification in deep learning. All my research studies are using this powerful library. Docs and unit-tests are still in development.

Owner **trojanzoo-sphinx-theme**.
`https://github.com/ain-soph/trojanzoo-sphinx-theme`
I modify pytorch-sphinx-theme, remove auxiliary items and make it a generalized theme. It supports easy customization in your project without touching the package contents.

Contributor **Open-source Projects**.
torchvision; pytorch-sphinx-theme; matplotlib; sphinxcontrib-katex; DI-star

Teaching Experience

2018 Fall **Teaching Assistant**, *CSE 017: Structured Programming and Data Structures*, Lehigh University (Instructor: Jeff Heflin).

Technical Skills

Language	Python; Java; C++; MatLab; Bash; LaTeX; HTML
Package	pytorch; matplotlib; sphinx; jinja; pytest
Tools	Auto CI; Docker; GitHub Actions