REN PANG

College of Information Sciences and Technology, the Pennsylvania State University
Email: rbp5354@psu.edu *Tel*: (484) 747-2401 *Web*: `https://ain-soph.github.io`

A. Research Interests

My current research focuses on the security of deep learning, including adversarial robustness and neural backdoors. Besides, I'm also interested in exploring security issues in other learning tasks, such as neural architecture search (NAS) and lifelong learning.

B. Education Background

Ph.D., Information Sciences and Technology, Pennsylvania State University	2019–present
Ph.D., Computer Science and Engineering, Lehigh University (transferred)	2018–2019
BSc., Mathematics, Nankai University	2014-2018

C. Publications

1. TrojanZoo: Towards Unified, Holistic, and Practical Evaluation of Neural Backdoors,
   R. Pang, Z. Zhang, X. Gao, Z. Xi, S. Ji, P. Cheng, and T. Wang,
   Proceedings of *the IEEE European Symposium on Security and Privacy* (EuroS&P), 2022.

2. On the Security Risks of AutoML,
   R. Pang, Z. Xi, S. Ji, X. Luo, and T. Wang,
   Proceedings of *the USENIX Security Symposium* (SECURITY), 2022.

3. Graph Backdoor,
   Z. Xi, **R. Pang**, S. Ji, and T. Wang,
   Proceedings of *the USENIX Security Symposium* (SECURITY), 2021.

4. i-Algebra: Towards Interactive Interpretability of Deep Neural Networks,
   X. Zhang, **R. Pang**, S. Ji, F. Ma, and T. Wang,
   Proceedings of *the AAAI Conference on Artificial Intelligence* (AAAI), 2021.

5. AdvMind: Inferring Adversary Intent of Black-Box Attacks,
   R. Pang, X. Zhang, S. Ji, X. Luo, and T. Wang,
   Proceedings of *the ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (KDD), 2020.

6. A Tale of Evil Twins: Adversarial Inputs versus Poisoned Models,
   R. Pang, H. Shen, X. Zhang, S. Ji, Y. Vorobeychik, X. Luo, A. Liu, and T. Wang,
   Proceedings of *the ACM Conference on Computer and Communications Security* (CCS), 2020.

D. Research Projects

1. Auto-Augment and Neural-Architecture-Search(NAS)
   The work aims to build a bridge between Auto-Augment and NAS, which are the 2 main categories in AutoML. We are also interested in the possible security concerns under this new scenario.

2. 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.

3. Vulnerabilities and Robustness in Dataset Condensation
   The project evaluates security concerns during dataset condensation and tries to develop a new attack approach to embed backdoors in the condensed data.

4. 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.

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

6. 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).

E. Open-Sourced Projects

Owner

AlpsPlot
`https://github.com/ain-soph/alpsplot`
My personal python plotting library of alps-lab style using matplotlib.

TrojanZoo: Towards Unified, Holistic, and Practical Evaluation of Neural Backdoors
`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.

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

matplotlib; pytorch_sphinx_theme; sphinxcontrib-katex; torchvision

E. Teaching Assistant Experience

CYBER 497: Machine Learning Security (Penn State), 2020 Spring

CSE 017: Structured Programming and Data Structures (Lehigh), 2018 Fall