TONG WU

tongwu@princeton.edu
https://tongwu2020.github.io/tongwu/

RESEARCH INTEREST

(Trustworthy) Machine Learning, Security, Meta Learning, and Computer Vision

EDUCATION

Princeton University *August 2021 - May 2026*
Ph.D. in Electrical and Computer Engineering
Advisor: Prateek Mittal

Washington University in St. Louis *August 2018 - May 2021*
B.S./M.S. in Computer Science; Second Major in Mathematics GPA: 4.0/4.0
Advisor: Yevgeniy Vorobeychik

DePauw University, College of Liberal Arts *August 2016 - May 2018*
B.A. in Pre-Engineering; Minor in Mathematics GPA: 3.94/4.0

RESEARCH & PROFESSIONAL EXPERIENCE

Princeton University *August 2021 - May 2022*
Research Assistant *Princeton, NJ*

- Developed a new threat model utilizing rotation transformations as a trigger to deploy backdoor attack.
- Present a detailed analysis of the rotation poisoned model and demonstrated that standard data augmentations, although mitigating the effect at the backdoor angle, introduce new vulnerabilities.
- Illustrated that deploying rotation backdoor attacks in the physical world, for both image classification and object detection tasks, is easily accessible and raised a new real-world security issue.
- *Key words*: Backdoor poisoning attacks; Spatial transformation; Backdoor trigger design.

NEC Laboratories America, Inc. *May 2021 - August 2021*
Research Intern *Princeton, NJ*

- Proposed a model personalization (meta-learning) framework for event detection of dialysis patients.
- Adapted covariance transfer and adversarial attacks to do OOD detection in few-shot learning, which achieves more human interpretability and mitigates the miss data issues.
- *Key words*: Meta-learning; Model personalization; OOD detection; Event detection.

Washington University in St. Louis *Dec 2018 - May 2021*
Research Assistant *St. Louis, MO*

- Studied the problem of defending deep neural network approaches from physically realizable attacks and demonstrated that the state-of-the-art robust models exhibit limited effectiveness.
- Proposed a new abstract model, ROA, where an adversary places a small crafted rectangle that fools the image classifier, and adversarial training using ROA achieved much better robustness than all SOTA.
- Designed optical lens which assists the adversarial attacks via coded defocus while maintaining stealthy.
- Demonstrated that such lens could be easily deployed in real world by evaluating the performance under various lens' positions, quantization constraints and noise inside lens.
- Illustrated the robustness of sensor fusion models against image-only and LiDAR-only attacks.
- Developed gradient-based camera-and-LiDAR combined adversarial attack on fusion methods.

· *Key words*: Physically realizable adversarial attacks; ML security; Camera-and-LiDAR fusion.

University of California, Berkeley *Dec 2020 - May 2021*
Research Assistant *Remote*

· Developed a parameter-efficient defending method against data poisoning and backdoor attacks, where the residual neural network adapts the test samples during inference.
· *Key words*: Test-time Adaptation; Poisoning and Backdoor Attacks;

University of Toronto *May 2020 - August 2020*
Research Assistant *Remote*

· Illustrated the performance degradation of adversarial attacks reconstructed from spectrogram to audio via Griffin-Lim and true-phase inverse short-time Fourier transform algorithms.
· *Key words*: Adversarial attacks on audio classification; Griffin-Lim algorithm; Fourier transform.

PUBLICATIONS

1. Shaojie Wang, **Tong Wu**, Ayan Chakrabarti, Yevgeniy Vorobeychik. Adversarial Robustness of Deep Sensor Fusion Models. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* , 2022.
2. Adith Boloor, **Tong Wu**, Patrick Naughton, Ayan Chakrabarti, Xuan Zhang, Yevgeniy Vorobeychik. Can Optical Trojans Assist Adversarial Perturbations? In *IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2021.
3. **Tong Wu**, Liang Tong, Yevgeniy Vorobeychik. Defending Against Physically Realizable Attacks on Image Classification. In *International Conference on Learning Representations (ICLR)*, 2020. **Spotlight Presentation**.

PREPRINTS

1. **Tong Wu**, Tianhao Wang, Vikash Sehwag, Saeed Mahloujifar, Prateek Mittal. Just Rotate it: Deploying Backdoor Attacks via Rotation Transformation . In *arXiv Preprint*, 2022.

PATENTS

1. Yevgeniy Vorobeychik, **Tong Wu**, Liang Tong. Systems and Methods for Defending against Physical Attacks on Image Classification. US Patent App. 17/214,071, 2021.

REVIEWING

- *Journals*
  - International Journal of Computer Vision (IJCV)
- *Conferences*
  - AAAI Conference on Artificial Intelligence (AAAI'21)
  - IEEE Symposium on Security and Privacy (IEEE S&P'21)
  - CVPR Workshop on Adversarial Machine Learning in Real-World Computer Vision Systems and Online Challenges (AML-CV'21)
  - Winter Conference on Applications of Computer Vision (WCAV'22)
  - ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22)
  - International Conference on Learning Representations (ICLR'22)

– Neural Information Processing Systems (Neurips'22)

TEACHING EXPERIENCE

- Teaching Assistant of CSE 417 of Introduction to Machine Learning (Spring 2019, Fall 2019, Spring 2020, Spring 2021), Washington University in St. Louis.

HONORS & AWARDS

- Princeton First Year Fellowship, 2021
- Research Excellence Award at Washington University, 2021
- AAMAS 2021 Student Scholarship, 2021
- Washington University Graduate Affiliation Scholarship, 2019, 2020, 2021
- Member of Tau Beta Pi Association, 2019, 2020, 2021
- Washington University Undergraduate Research Conference Travel Award, 2020
- DePauw University Merit Scholarship, 2016, 2017, 2018
- DePauw Dean's List, 2016, 2017, 2018
- Michigan Competition MATH Challenge 3/74, 2018

SKILLS

Python: Numpy, PyTorch, Tensorflow. R. C/C++. Java. Matlab