

# Chong Xiang

🌐 <https://xiangchong.xyz>    ✉ [cxiang@princeton.edu](mailto:cxiang@princeton.edu)

## Education

---

### Princeton University

Princeton, NJ

*Ph.D. Student, Department of Electrical and Computer Engineering*

*Sept. 2019 - Present*

- Research Focus: Machine Learning Security
- Advisor: Prof. Prateek Mittal

### Shanghai Jiao Tong University

Shanghai, China

*B.S., School of Electronic Information and Electrical Engineering*

*Sept. 2015 - June 2019*

- Major: Information Security
- Advisor: Prof. Haojin Zhu

## Publications

---

- **Chong Xiang**, Alexander Valtchanov, Saeed Mahloujifar, Prateek Mittal, “ObjectSeeker: Certifiably Robust Object Detection against Patch Hiding Attacks via Patch-agnostic Masking”, arXiv 2202.01811.
  - Proposed a certifiably robust object detection defense against patch hiding attacks via a patch-agnostic pixel masking technique
- **Chong Xiang**, Saeed Mahloujifar, Prateek Mittal, “PatchCleanser: Certifiably Robust Defense against Adversarial Patches for any Image Classifier”, in 31<sup>th</sup> *USENIX Security Symposium (USENIX Security 2022)*.
  - Proposed a certifiably robust image classification technique against adversarial patch attacks that is compatible with any state-of-the-art classification model
- **Chong Xiang**, Prateek Mittal, “DetectorGuard: Provably Securing Object Detectors against Localized Patch Hiding Attacks”, in 2021 *ACM Conference on Computer and Communications Security (CCS 2021)*. (Acceptance rate: 196/879=22.2%)
  - Proposed the first provably robust defense for object detectors against patch hiding attacks
- **Chong Xiang**, Arjun Nitin Bhagoji, Vikash Sehwal, Prateek Mittal, “PatchGuard: A Provably Robust Defense against Adversarial Patches via Small Receptive Fields and Masks”, in 30<sup>th</sup> *USENIX Security Symposium (USENIX Security 2021)*. (Acceptance rate: 246/1295=19.0%)
  - Proposed a defense framework for provably robust image classification against adversarial patch attacks via small receptive fields and secure feature aggregation
- **Chong Xiang**, Prateek Mittal, “PatchGuard++: Efficient Provable Attack Detection against Adversarial Patches”, in *ICLR 2021 Workshop on Security and Safety in Machine Learning Systems*. (Travel Award)
  - Proposed an efficient feature-space attack detection defense against adversarial patch attacks
- **Chong Xiang**, Charles R. Qi, Bo Li, “Generating Adversarial 3D Point Clouds”, in 2019 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019)*. (Acceptance rate: 1294/5160=25.1%)
  - Proposed the first adversarial example attacks for 3D point cloud data
- **Chong Xiang**, Xinyu Wang, Qingrong Chen, Minhui Xue, Zhaoyu Gao, Haojin Zhu, Cailian Chen, Qiuhua Fan, “No-Jump-into-Latency in China’s Internet! A Hop Count Based IP Geo-localization Approach”, in 27<sup>th</sup> *IEEE/ACM International Symposium on Quality of Service (IWQoS 2019)*. (Acceptance rate: 42/153=27.4%)

- Proposed to use hop counts instead of RTT for IP geo-localization in China's Internet
  - **Chong Xiang**, Qingrong Chen, Minhui Xue, Haojin Zhu, "AppClassifier: Automated App Inference on Encrypted Traffic via Meta Data Analysis", in *2018 IEEE Global Communications Conference (GLOBECOM 2018)*. (Acceptance rate: 999/2562=39.0%)
    - Proposed an encrypted traffic analysis method for real-world Android application inference
- 
- Vikash Sehwal, Saeed Mahloujifar, Sihui Dai, Tinashe Handina, **Chong Xiang**, Mung Chiang, Prateek Mittal, "Robust Learning Meets Generative Models: Can Proxy Distributions Improve Adversarial Robustness?" in *International Conference on Learning Representations (ICLR 2022)*.
    - Proposed to use data from proxy distributions to improve model robustness against adversarial examples
  - Saeed Mahloujifar, **Chong Xiang**, Vikash Sehwal, Sihui Dai, Prateek Mittal, "Robustness from Perception", in *ICLR 2021 Workshop on Security and Safety in Machine Learning Systems*.
    - Proposed a framework to use perceptual metrics for robust ML model predictions
  - Lei Zhang, Yan Meng, Jiahao Yu, **Chong Xiang**, Brandon Falk, Haojin Zhu, "Voiceprint Mimicry Attack Towards Speaker Verification System in Smart Home", in *IEEE International Conference on Computer Communications (INFOCOM 2020)*. (Acceptance rate: 268/1354=19.8%)
    - Proposed an adversarial example attack against audio-based speaker verification systems
  - Qingrong Chen, **Chong Xiang**, Minhui Xue, Bo Li, Nikita Borisov, Dali Kaafar, Haojin Zhu, "Differentially Private Data Sharing: Sharing Models versus Sharing Data", in *CCS 2019 Workshop on Privacy Preserving Machine Learning (PPML 2019)*.
    - Proposed differentially private methods for privacy-preserving data/model sharing

## Selected Projects

### Provably Robust Image Classification against Adversarial Patches (2)

Apr. 2021 - Oct. 2021

Advisor: Prof. Prateek Mittal

Princeton University

- Proposed PatchCleanser, a provably/certifiably robust defense against adversarial patch attacks that was compatible with any state-of-the-art image classification models
- Designed an efficient double-masking algorithm to remove all adversarial pixels on the input image and proved its robustness guarantee against any adaptive white-box attacker within the threat model
- Evaluated PatchCleanser across ImageNet, ImageNette, CIFAR-10, CIFAR-100, SVHN, and Flowers-102 datasets, and demonstrated huge improvements in certified robust accuracy and clean accuracy from prior works (e.g., 28.9% to 37.6% top-1 accuracy improvements on ImageNet)

### Provably Securing Object Detectors against Patch Hiding Attacks

Sept. 2020 - Feb. 2021

Advisor: Prof. Prateek Mittal

Princeton University

- Proposed DetectorGuard as the first general framework for provably securing object detectors against patch hiding attacks, where adversarial patches were used to hide victim objects from being detected
- Proposed an objectness explaining strategy to build provably robust object detectors from provably robust image classifiers, which achieved substantial provable robustness at a negligible cost of clean performance
- Applied DetectorGuard to YOLOv4 and Faster R-CNN on PASCAL VOC, MS VOC, and KITTI and demonstrated the first provable robustness against patch hiding attacks with a small (<1%) clean AP drop

### Provably Robust Image Classification against Adversarial Patches (1)

Jan. 2020 - Oct. 2020

Advisor: Prof. Prateek Mittal

Princeton University

- Proposed PatchGuard, a general defense framework against adversarial patch attacks; the cornerstone of PatchGuard was to use CNNs with small receptive fields to bound the number of features corrupted by an adversarial patch; the defense was then translated into a secure feature aggregation problem
- Designed the robust masking defense for secure feature aggregation, which aimed to detect and mask cor-

- rupted features; proved its security guarantee against any attacker within the threat model
- Evaluated PatchGuard across 3 image classification datasets: ImageNet, ImageNette, CIFAR-10, and demonstrated state-of-the-art provable robust accuracy and clean accuracy (2-16% improvements)

### **Adversarial Examples for 3D Point Cloud Data**

*July 2018 - Nov. 2018*

*Advisor: Prof. Bo Li*

*University of Illinois at Urbana-Champaign*

- Proposed the initialize-and-shift algorithm for the 3D adversarial point cloud generation, which addressed the challenges of unfixed data dimensionality and large searching space
- Generated adversarial perturbations, adversarial independent points, adversarial clusters, and adversarial objects for different attack goals; achieved a success rate higher than 99% for all targeted attacks
- Proposed six perturbation metrics tailored to different attack tasks and provided a baseline result for future 3D adversarial example research

### **Miscellaneous**

---

- Review: ICLR, TIFS, TOPS, TPAMI, TIP
- Mentor of Princeton undergraduate students for their independent research *2020-2022*
- Assistant Instructor, COS/ELE 432 Information Security *Spring 2021*
- Graduate Student Mentor, Department of Electrical and Computer Engineering *2020*
- Zhiyuan Honors Scholar with Outstanding Achievement Award (the only awarded student in Class of 2019), Shanghai Jiao Tong University *2020*