Zhisheng Hu

3761 Watkins Dr.
Riverside, CA 92507

Email : hzsxiaoyi@gmail.com
Mobile : +1-814-321-4649

About me

Ph.D. in AI Security, Senior Security Scientist at Baidu USA.

Education

- **The Pennsylvania State University** PA, USA
  Ph.D Candidate in Electrical Engineering (Advisor: Dr. Minghui Zhu) *Aug. 2014 – Aug. 2019*
- **Sun Yat-sen University** Guangzhou, China
  Bachelor of Engineering in Communication Engineering *Sep. 2009 – July. 2013*

Experience

- **Senior Security Scientist** Baidu USA
  Research and development in autonomous driving security and AI security *Oct. 2019 - Present*
- **Summer Intern** JD.com Silicon Valley R&D Center
  Research and develop bot detection framework based on deep learning *May. 2018 - Aug. 2018*
- **Graduate Research Assistant** The Pennsylvania State University
  Design reinforcement learning algorithms for adaptive cyber defense *Aug. 2014 - Aug. 2019*

Recent Projects

- **Research and development in autonomous driving security and AI security** *Oct. 2019 - Present*
  - Lead the Project PASS (https://theprojectpass.org/), an open platform to efficiently validate and verify safety and security risks in autonomous driving (AD) systems.
  - Lead the AutoDriving CTF@DEFCON 29 (https://autodrivingctf.org/), which reveals unforeseeable threats to AD systems through hands-on challenges.
  - Apply modern software bug finding techniques to generate real-world transferrable critical test cases for AD systems. The cases help us identify logical bugs in multiple AD systems including latest Tesla FSD.
  - Improve robustness of object detectors through adversarial training in practice.
  - Develop an efficient adversarial patch recognition technique using AI model interpretation.
  - Develop the first open-source robustness benchmark (https://github.com/advboxes/perceptron-benchmark) for computer vision DNN models.
- **Deep learning on Adversarial Attacks and Defenses** *May. 2018 - Oct. 2018*
  - Design hybrid neural networks for malicious Android applications data detection
  - Generate advanced CAPTCHA with adversarial examples
  - Mitigate adversarial example effects on image classifiers

- **Deep learning on ROP attacks** *Aug. 2017 - May. 2018*
  - Customize convolutional neural networks for gadget chains classification
  - Design a tool for attackers to predict which gadget chains can bypass control flow integrity (CFI)
- **Reinforcement learning algorithms on zero-day continuous attacks** *Aug. 2014 - Aug. 2019*
  - Design reinforcement learning algorithms based on Partially Observable Markov Decision Processes (POMDP) to defend against external intrusions under uncertainties
  - Design reinforcement learning algorithms to mitigate memory corruption attacks
  - Design game-theoretic reinforcement learning algorithms to identify optimal defense actions over unreliable ICT systems

Selective Publications

C1 Z. Zhong, **Z. Hu** and X. Chen, “Quantifying DNN Model Robustness to the Real-World Threats,” *DSN*, pp. 150-157, June 2020.

C2 **Z. Hu** and Z. Zhong, “Towards Practical Robustness Improvement for Object Detection in Safety-critical Scenarios” *MLHat@SIGKDD* , August 2020.

C3 **Z. Hu**, S. Guo, Z. Zhong, K. Li, “Coverage-based Scene Fuzzing for Virtual Autonomous Driving Testing,” *arXiv preprint*, June 2021.

C4 **Z. Hu**, S. Guo, Z. Zhong, K. Li, “Disclosing the Fragility Problem of Virtual Safety Testing for Autonomous Driving System,” *ISSRE*, pp. 387-392, October 2021.

C5 **Z. Hu**, J. Shen, S. Guo, X. Zhang, Z. Zhong, A. Q. Chen and K. Li, “PASS: A System-Driven Evaluation Platform for Autonomous Driving Safety and Security,” *AutoSec*, April 2022. (**Best Short Paper Award**)

C6 **Z. Hu**, S. Guo, and K. Li, “Disclosing the Pringles Syndrome in Tesla FSD Vehicles,” *AutoSec*, April 2022.

C7 Z. Zhong, **Z. Hu** S. Guo, X. Zhang, Z. Zhong, B. Ray, “Towards Practical Robustness Improvement for Object Detection in Safety-critical Scenarios” *ISSTA accepted*, July 2022.

J1 **Z. Hu**, P. Chen, M. Zhu, P. Liu, “A co-design adaptive defense scheme with bounded security damages against Heartbleed-like attacks,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4691-4704, 2021.

Honors & Awards

- **Final Reward in Competition on Adversarial Attacks and Defenses (GeekPwn)** *2018*

Academic Services

- **Conference review**: MTD 2016, DSN 2016, MTD 2017, ACC 2017, Securecomm 2018, MTD 2019, SPAI 2020, AutoSec 2021
- **Journal review**: IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Services Computing, IEEE Transactions on Emerging Topics in Computing, IET Information Security, Journal of Computer Security, ACM Transactions on Cyber-Physical Systems, Scientific Reports - Nature