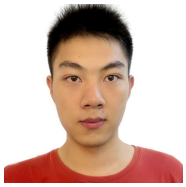


# Jialai Wang (汪嘉来)

Tsinghua University
Beijing, China

+86-15927318278
wang-jl22@mails.tsinghua.edu.cn
wangjialai97@gmail.com


## Education

- **Tsinghua University** *Aug. 2019 - July, 2024 (Expected)*
  *Ph.D. Candidate in Cyberspace Security in Inst. of Network Science and Cyberspace* Beijing, China
  ***Advisor:*** *Prof. Zongpeng Li*

- **Huazhong University of Science and Technology (graduated with the highest honor)** *Sep. 2015 - June 2019*
  *B.Eng. @ School of Computer Science & Technology* Wuhan, China
  ***Advisor:*** *Prof. Deqing Zou*

## Research Domain

I have dedicated my research to artificial intelligence (AI) security and interpretability. My expertise encompasses common AI attack methodologies, including adversarial, backdoor, and bit-flip attacks, along with their respective defense strategies. I also focus on leveraging interpretability to enhance model performance. I possess a multidisciplinary background, with a research focus that spans AI security in many domains, such as computer vision, malware detection, and binary similarity detection.

## Publication

- **[USENIX Security '23] Jialai Wang**, Ziyuan Zhang, Meiqi Wang, Han Qiu, Tianwei Zhang, Qi Li, Zongpeng Li, Tao Wei, Chao Zhang. Aegis: Mitigating Targeted Bit-flip Attacks against Deep Neural Networks. This paper proposes the first defense mechanism against targeted bit-flip attacks on DNNs. The core idea involves designing a dynamic multi-exit mechanism and a robust training strategy against bit-flip attacks. Our work effectively mitigates all existing targeted bit-flip attacks, successfully reducing the attack success rate to below 5%. Our work is highly scalable, does not affect model performance, and can be easily deployed in various models.
- **[Design Automation Conference '23] Jialai Wang**, Wenjie Qu, Yi Rong, Han Qiu, Qi Li, Zongpeng Li, Chao Zhang. MPass: Bypassing Learning-based Static Malware Detectors. This paper proposes a hard-label black-box attack for ML-based malware detectors on the Windows platform. We design a problem-space interpretability method to locate critical positions of malware, apply adversarial modifications to such positions, and utilize a runtime recovery technique to preserve the functionality. Our work significantly outperforms SOTA attacks, and can effectively attack all existing open-source malware detectors and five well-known commercial antivirus products.
- **[ISSTA '22] Jialai Wang**, Han Qiu, Yi Rong, Hengkai Ye, Qi Li, Zongpeng Li, Chao Zhang. BET: Black-box Efficient Testing for Convolutional Neural Networks. This paper proposes a novel black-box testing method for CNNs, utilizing inherent weaknesses of CNNs to test target models, efficiently uncovering wrong behaviors in models. Our work significantly outperforms existing white-box and black-box testing methods considering the effective error-inducing inputs found in a fixed query/inference budget. We further show that the error-inducing inputs found by us can be used to fine-tune the target model, improving its accuracy by up to 3%. This work can be applied to model testing in typical scenarios such as facial recognition, object detection, and malware identification, thereby enhancing model robustness.
- **[Journal of Computer Research and Development '21] Jialai Wang**, Chao Zhang, Xuyan Qi, Yi Rong. A Survey of Intelligent Malware Detection on Windows Platform. This paper mainly introduces current work related to intelligent (learning-based) malware detection, which includes the main parts required for intelligent detection processes. Specifically, we have systematically explained and classified related work for intelligent malware detection in this paper, which includes the features commonly used in intelligent detection, how to perform feature processing, the commonly used classifiers in intelligent detection, and the main problems faced by current malware intelligent detection. Finally, we summarize the full paper and clarify the potential future research directions, aiming to contribute to the development of intelligent malware detection.

## Under review (First Author)

- **Improving ML-based Binary Function Similarity Detection by Assessing and Deprioritizing CFGs**
  *Accept Conditional on Major Revision, USENIX Security, 2024*

- **Black-Box Efficient Testing for Convolutional Neural Networks across Diverse Domains**
  *TDSC*

## Patent

- **Neural Network Test Method and Device, Equipment and Storage Medium** (**CN115374898B**, **2023**)
  *Chao Zhang*, ***Jialai Wang***, *Qi Li*

## Experience And Competition

- **Tencent Company** *June 2021 - August 2021*
  *Research Assistant @ Technology and Engineering Group* Beijing, CHINA

- **National College Student Information Security Contest** *July 2018*
  *We won the first prize* Wuhan, CHINA

- **GeekPwn Adversarial Patch Competition** *October 2019*
  *I'm the captain, and we won third place* Shanghai, CHINA

- **BCTF (Baidu Company)** *September 2020*
  *I'm the captain, and we won third place* Online

## Honors

- **Longfor Properties Scholarship of Tsinghua University** : Top 1%, one of only 2 recipients across our school.
- **First-Class Excellent Comprehensive Scholarship of Tsinghua University** : Top 1%.
- **Second-Class Excellent Comprehensive Scholarship of Tsinghua University** : Top 5%.
- **Excellent Social Work Scholarship of Tsinghua University** : Top 1%.
- **Outstanding Undergraduate Student of Huazhong University of Science and Technology** : Top 1%, the highest honor awarded to undergraduate students by HUST.
- **Samsung Scholarship of HUST** : Top 1%, one of only 15 recipients across the university.
- **Excellent Graduates of HUST** : Top 1%
- **Academic Excellence Scholarship of HUST** : Top 5%, awarded multiple times.

## Community Service

**Vice President of the Student Union, Institute for Network Sciences and Cyberspace, Tsinghua University**