Gelei Deng

+65-8403-0341, gdeng003@e.ntu.edu.sg

Profile

Ph.D. candidate in Cybersecurity, Computer Science; Certified penetration tester and auditor; Researcher in cybersecurity domain complemented with strong background in electrical and computer engineering. Extremely interested and passionate about cybersecurity. A fast learner and a creative thinker.

Areas of Expertise

- System Security
- Large Language Model Security
- Blockchain Security
- Penetration Testing
- Proficient in Python and Solidity; Familiar with Mainstream Web2 and Web3 Languages including JavaScript, PHP, Java, C#, etc.

Education

Nanyang Technological University (Aug 2020 – Present)

- Ph.D. Computer Science, Cybersecurity
- NTU Cyber Security Lab
- Main Research Interests on System Security, Web Security and Penetration Testing

Singapore University of Technology and Design (May 2015 – Sep 2018)

- Department of Engineering Product Design, B.E. Electrical Engineering
- Singapore Ministry of Education Full Scholarship (SM2) Holder
- Engineering Product Design Track, Honor List

Employment

OpenAI (Jan 2024 – Present)

- Independent Contractor at OpenAI Red Teaming Network.
- Participated in OpenAI led red teaming efforts to assess the risks and safety profile of OpenAI models and systems.
- Evaluating the safety of non-public models generation results.

Quantstamp, Inc. (Oct 2022 – Present)

- Worked as Senior Audit Engineer. Perform security auditors for various blockchain projects, including DeFi, Wallet, DEX, NFT, etc. Identified 200+ vulnerabilities in 20+ DeFi projects with a total asset value of 500M+.
- Currently working as AI Engineer and leading the research work on large language model (LLM) and LLM-based automatic smart contract audits.

Institute for Infocomm Research (I^2R, A*STAR) (Jan 2019 – Jul 2020)

- Research Engineer and Penetration Tester. Perform Penetration Testing and Conduct Research Works for Singapore government-based Agencies.

Academic Research

Cyber Security Lab, NTU (July 2020 - Present)

- Ph.D. student in Cybersecurity
- Main research topics include system security, robotics security, Web3 security, and penetration testing automation.
- Collaborate with industrial partners (Huawei, etc.) to verify system security and stability.

SUTD-MIT International Design Center (2016-2018)

- Work on hardware-oriented digital signal processing
- Leading the research work in IIR filter design
- Involved in IDCT image processing research work

Selected Publications

Highlighted Publications

Gelei Deng, Yi Liu, Víctor Mayoral-Vilches, Peng Liu, Yuekang Li, Yuan Xu, Tianwei Zhang, Yang Liu, Martin Pinzger, Stefan Rass, "PentestGPT: An LLM-empowered Automatic Penetration Testing Tool", Revision in Process in *33rd USENIX Security Symposium (USENIX '24)*. 2024.

Gelei Deng, Yi Liu, Kailong Wang, Yuekang Li, Tianwei Zhang, Yang Liu, "PANDORA: Jailbreak GPTs by Retrieval Augmented Generation Poisoning", In *Workshop on Artificial Intelligence System with Confidential Computing (AISCC)*, Distinguished Paper Award, February, 2024

Gelei Deng, Yi Liu, Yuekang Li, Kailong Wang, Ying Zhang, Zefeng Li, Haoyu Wang, Tianwei Zhang, Yang Liu, "MASTERKEY: Automated Jailbreaking of Large Language Model Chatbots", In *Network and Distributed System Security Symposium (NDSS '24)*. 2024.

Gelei Deng, Zhiyi Zhang, Yuekang Li, Yi Liu, Tianwei Zhang, Yang Liu, Guo Yu, Dongjin Wang, "NAUTILUS: Automated RESTful API Vulnerability Detection", In 3*2nd USENIX Security Symposium (USENIX '23)*. 2023.

Gelei Deng, Guowen Xu, Yuan Zhou, Tianwei Zhang, and Yang Liu, "On the (In)Security of Secure ROS2," In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22)*. Association for Computing Machinery, New York, NY, USA, 739–753.

Other Publication

Kunsheng Tang, Wenbo Zhou, Jie Zhang, Aishan Liu, **Gelei Deng**, Shuai Li, Peigui Qi, Weiming Zhang, Tianwei Zhang, Nenghai Yu, "GenderCARE: A Comprehensive Framework for Assessing and Reducing Gender Bias in Large Language Models", In *ACM Conference on Computer and Communications Security (CCS)*, October, 2024

Yi Liu, Yuekang Li, **Gelei Deng**, Felix Juefei-Xu, Yao Du, Cen Zhang, Chengwei Liu, Yeting Li, Lei Ma, Yang Liu, "ASTER: Automatic Speech Recognition System Accessibility Testing for Stutterers", In *38th IEEE/ACM International Conference on Automated Software Engineering (ASE 2023)*. 2023

Yi Liu*, **Gelei Deng***, Zhengzi Xu, Yuekang Li, Yaowen Zheng, Ying Zhang, Lidao Zhao, Tianwei Zhang, Yang Liu, "Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study", https://arxiv.org/abs/2305.13860

Yuan Xu, Xingshuo Han, **Gelei Deng**, Jiwei Li, Tianwei Zhang, Yang Liu, "SoK: Rethinking Sensor Spoofing Attacks against Robotic Vehicles from a Systematic View," In *8th IEEE European Symposium on Security and Privacy (Euro S&P 23)*. 2023

Yi Liu, Yuekang Li, **Gelei Deng**, Yang Liu, et al., "Morest: Model-based RESTful API Testing with Execution Feedback," *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*, 2022, pp. 1406-1417.

Gelei Deng, Yuan Zhou, Yuan Xu, Tianwei Zhang, and Yang Liu, "An Investigation of Byzantine Threats in Multi-Robot Systems," *24th International Symposium on Research in Attacks, Intrusions and Defenses (RAID '21)*, New York, NY, USA, 17–32.

Gelei Deng, Stefanie Yu Xingjie and Huaqun Guo, "Efficient Password Guessing based on a Password Segmentation Approach," *GLOBECOM*, Waikoloa, Hawaii, 2019.

Yuan Xu, **Gelei Deng**, Tianwei Zhang, Han Qiu, Yungang Bao, "Novel denial-of-service attacks against cloud-based multi-robot systems," *Information Sciences*, Volume 576, 2021, Pages 329-344,

Luying Zhou, Huaqun Guo and **Gelei Deng**, "A fog computing based approach to DDoS mitigation in IIoT systems," *Computer & Security*, vol. 85, pp. 51-62, 2019

Certificates, Awards and Activities

Projects

PentestGPT: An LLM-empowered Automatic Penetration Testing Tool

- Open-source tool with more than 6k stars on GitHub: http://pentest-gpt.com/
- Various collaboration with industrial partners and potential VC investments.
- Paper in revision process of USENIX Security 2024; Arxiv version available online: https://arxiv.org/pdf/2308.06782.pdf

CVEs and Vulnerability Identification

More than verified 10 CVEs including: CVE-2021-39114, CVE-2021-30224, CVE-2021-37392, CVE-2021-37393, CVE-2021-37476, CVE-2021-37477

Multiple vulnerabilities and bugs confirmed by vendors including: Atlassian Confluence, Apache Magento, Bitbucket, SEO Panel, Spree Commerce, etc.

Certificates

Offensive Security Certified Professional (OSCP) (2020)

- The most recognized penetration testing certificate
- Achieved with full mark (105 over 100, 5 points bonus)

BlackHat Advanced Infrastructure Hacking Completion

- BlackHat Asia, Singapore, 2019

Awards

SUTD Y2015 Student Honor List (2018)

- Awarded to be one of the students with best overall performance and grade
- 12 out of more than 400

The Most Amazing SUTD Student Works of the Year (2018) – Drone ranger

- 5 selected projects out of more 100 student projects in 2018

Additional Information

- Fluent in English and Mandarin (spoken & written)
- Actively participating varieties of sports activities