Hongwei Li

* Department of Computer Science, University of California Santa Barbara, CA, United States 93106
(+1)765-543-8337 hongwei@ucsb.edu 3rdn4li.github.io

Education

University of California, Santa Barbara Sep 2024 – Present
Ph.D. Student in Computer Science, advised by Wenbo Guo Santa Barbara, CA, USA

Purdue University Sep 2023 – May 2024
First-Year Ph.D. Student in Computer Science, advised by Wenbo Guo West Lafayette, IN, USA

Shanghai Jiao Tong University Sep 2020 – Jun 2023
M.Eng. in Electronic Information Engineering Shanghai, China

Shanghai Jiao Tong University Sep 2016 – Jun 2020
B.A. in French; B.Eng. in Information Engineering (Dual Degree) Shanghai, China

Awards

DARPA AIxCC Finalist (Top 7) Dec 2023 – Aug 2025
Core member of the Shellphish team

- Core member of the patching group, focusing on automated vulnerability patching.
- Contributed key modules to the root-cause analysis engine.
- Fine-tuned custom LLM for vulnerability detection.

SBFT 2024 Fuzzing Competition (Top 1) Jun 2023 – Jan 2024
Co-led the project

- Built a collaborative fuzzer augmented with a multi-armed bandit algorithm.
- Achieved best performance across all evaluation metrics, including highest mutant kills, mutation score, and mutant coverage.

Publications

[**ICML'25**] **Hongwei Li**, Yuheng Tang, S Wang, Wenbo Guo, "PatchPilot: A Cost-Efficient Software Engineering Agent with Early Attempts on Formal Verification", In *Proceedings of the Forty-second International Conference on Machine Learning*, Vancouver, Canada, 2025.
A novel approach to automated software patching using AI agents with refinement and formal verification capabilities, achieving the best performance among open-source agents on SWE-bench at the time of publication.

[**ICSE/SBFT'24**] Wenxuan Shi, **Hongwei Li**, Jiahao Yu, Wenbo Guo, Xinyu Xing, "BandFuzz: A Practical Framework for Collaborative Fuzzing with Reinforcement Learning", In *Proceedings of the 17th ACM/IEEE International Workshop on Search-Based and Fuzz Testing*, Lisbon, Portugal, 2024.
A practical framework that leverages reinforcement learning to improve collaborative fuzzing effectiveness.

[**Computers & Security'22**] Jingcheng Yang, **Hongwei Li**, Shuo Shao, Futai Zou, Yue Wu, "FS-IDS: A framework for intrusion detection based on few-shot learning", *Computers & Security*, 122: 102899, 2022.
A framework for intrusion detection based on few-shot learning techniques.

Preprints

Yuheng Tang, **Hongwei Li**, Kaijie Zhu, Yuan Yang, Yangruibo Ding, Wenbo Guo, "Co-PatcheR: Collaborative Software Patching with Component(s)-specific Small Reasoning Models", arXiv preprint, 2025.
A collaborative patching system with small and specialized reasoning models for individual components, achieving 46% resolved rate on SWE-bench-Verified with only 3 × 14B models.

Tianneng Shi, Jingxuan He, Zhun Wang, Linyu Wu, **Hongwei Li**, Wenbo Guo, Dawn Song, "Progent: Programmable Privilege Control for LLM Agents", arXiv preprint, 2025.
The first privilege control mechanism for LLM agents providing fine-grained constraints over tool calls to ensure security while preserving utility.

Skills

Proficient: Python, C/C++, Generic/Directed/Kernel Fuzzing, User Space Binary Exploitation
Familiar: Smart Contract Fuzzing, White-box Web Application Testing
Exposure to: Reverse Engineering, Kernel Exploitation, Static Analysis (Codeql)