Wen Li

PhD Candidate in Computer Science, Washington State University, Pullman, WA 99163
li.wen@wsu.edu | https://awen-li.github.io | https://github.com/awen-li

Research Interest

- **General Area:** *Software Engineering and Security in Compiler & Language Runtimes* (e.g., JVM, CPython, Machine learning framework), *Multi-language Systems* (e.g., Java/C, Python/C software).
- **Topics:** *Program analysis* (e.g., static control/data flow analysis, dynamic information flow analysis), *Holistic fuzzing* (e.g., language-extensible, whole system coverage measurement), *Collaborative fuzzing* (e.g., multiple-layer fuzzing).
- **Overview:** Focus on providing practical support to ensure a series of reliable, scalable, and efficient program analyses and security techniques for multi-language software and the host language runtimes.

Education

Ph.D., Washington State University, USA Aug 2019 - May 2024

- Major: Computer Science
- Dissertation: Run-time Analysis and Security of Multi-Language Systems
- Advisor: Dr. Haipeng Cai (`https://chapering.github.io`)

BSc, Huazhong University of Science and Technology, China Sep 2003 - Jun 2007

- Major: Computer Science

Awards

2023 Outstanding Research Assistant, EECS of Washington State University, Pullman, US

2015 Excellent New Employee, HUAWEI Research Institute, Wuhan, China

2013 Excellent Employee, ZTE Research Institute, Nanjing, China

2008 Excellent Employee, NEUSOFT, Shenyang, China

Research Highlight

Highlights

- A decade-long industrial experience in software design and development.
- Progressive research experience in program analysis, compiler/language runtime testing, and fuzzing for software security.
- Major publications (first-author) are accepted top-tier software engineering and security conferences, including to ESEC/FSE, USENIX Security, and CCS.
- Discover 14 previously unknown security vulnerabilities.

First-Author Publications

1. PyRTFuzz: Detecting Bugs in Python Runtimes via Two-Level Collaborative Fuzzing.
   Wen Li, Haoran Yang, Long Cheng, Xiapu Luo, Haipeng Cai.
   In ACM SIGSAC Conference on Computer and Communications Security(CCS), pp. 1645-1659. 2023.
   Paper:`https://dl.acm.org/doi/pdf/10.1145/3576915.3623166`

2. PolyFuzz: Holistic Greybox Fuzzing of Multi-Language Systems.
   Wen Li, Jinyang Ruan, Guangbei Yi, Long Cheng, Xiapu Luo, Haipeng Cai.
   In 32nd USENIX Security Symposium, pages 1379–1396, Anaheim, CA, August 2023.
   Paper: `https://www.usenix.org/system/files/sec23summer_411-li_wen-prepub.pdf`

3. PolyCruise: A Cross-Language Dynamic Information Flow Analysis.
   Wen Li, Ming Jiang, Xiapu Luo, Haipeng Cai.
   In 31st USENIX Security Symposium, pages 2513–2530, Boston, MA, August 2022.
   Paper: `https://www.usenix.org/system/files/sec22fall_li-wen.pdf`

4. On the Vulnerability Proneness of Multilingual Code.
   Wen Li, Li Li, Haipeng Cai.
   In ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software

Engineering (ESEC/FSE), pages 847–859, 2022.
Paper: https://dl.acm.org/doi/10.1145/3540250.3549173

5. (Journal) How are Multilingual Systems Constructed: Characterizing Language Use and Selection in Open-Source Multilingual Software.
Wen Li, Austin Marino, Haoran Yang, Na Meng, Li Li, Haipeng Cai.
ACM Transactions on Software Engineering and Methodology (TOSEM), 45 pages, 2023.
Paper: https://dl.acm.org/doi/pdf/10.1145/3631967

Tool Demos & Dataset

1. PolyFax: A Toolkit for Characterizing Multi-Language Software.
Wen Li, Li Li, Haipeng Cai.
In ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), Tool Demos, pages 1662–1666, 2022.
Paper: https://dl.acm.org/doi/10.1145/3540250.3558925

2. AndroCT: Ten Years of App Call Traces in Android.
Wen Li, Xiaoqin Fu, Haipeng Cai.
In IEEE/ACM Working Conference on Mining Software Repository (MSR), Data showcase, pages 570–574, 2021.
Paper: https://ieeexplore.ieee.org/document/9463081

3. PCA: Memory Leak Detection using Partial Call-Path Analysis.
Wen Li, Haipeng Cai, Yulei Sui, David Manz.
In ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), Tool Demos, pages 1621–1625, 2020.
Paper: https://dl.acm.org/doi/10.1145/3368089.3417923

Co-Author Publications

1. Language-Agnostic Dynamic Analysis of Multilingual Code: Promises, Pitfalls, and Prospects.
Haoran Yang, **Wen Li**, Haipeng Cai.
In ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), Ideas, Visions and Reflections, pages 1621–1626, 2022.
Paper: https://dl.acm.org/doi/10.1145/3540250.3560880

2. (Journal) Seads: Scalable and Cost-Effective Dynamic Dependence Analysis of Distributed Systems via Reinforcement Learning.
Xiaoqin Fu, Haipeng Cai, **Wen Li**, Li Li.
ACM Transactions on Software Engineering and Methodology (TOSEM), 30(1): 1–45. 2020. (impact factor 2.5; journal-first paper).
Paper: https://dl.acm.org/doi/10.1145/3379345

3. Towards Learning Visual Semantics.
Haipeng Cai, Shiv Raj Pant, **Wen Li**.
In ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), Visions and Reflections, pages 1537–1540, 2020.
Paper: https://dl.acm.org/doi/10.1145/3368089.3417040

Security Findings

- 14 previously unknown vulnerabilities are reported with CVEs:
  2023: CVE-2023-36632.
  2022: CVE-2022-34075, CVE-2022-34074, CVE-2022-34073, CVE-2022-34072, CVE-2022-34070.
  2021: CVE-2021-41499, CVE-2021-41498, CVE-2021-41500, CVE-2021-41497, CVE-2021-34141, CVE-2021-41496, CVE-2021-41495, CVE-2021-33430.
- Security threats in multi-language systems (e.g., Numpy) are critical.

Professional Experience

Aug 2019 – Present **Research Assistant on Program Analysis and Software Security, Washington State University**

– Research on Security Analysis of Compiler & Language Runtime: *PyRTFuzz (CCS 2023)*.

My Role: Project Leader

Design and develop a two-layer collaborative fuzzing framework for Python interpreter and runtimes. Especially propose a novel SLang-based application generation technique to generate APPs boasting diverse control flow complexities and domain-specific features.

– Research on Security Analysis of Multi-language Systems: *PolyCruise (USENIX Security 2022)* & *PolyFuzz (USENIX Security 2023)*.

My Role: Project Leader
Design and develop a scalable and efficient cross-language dynamic information flow analysis framework (i.e., *PolyCruise*) on top of a customized intermediate presentation, especially supporting languages of [C, Python].
Design and develop a holistic grey-box fuzzer (i.e., *PolyFuzz*) for multi-language software, supporting effective seed generation with sensitivity analysis. Three languages [C, Python, Java] are supported to demonstrate language extensibility.

– Empirical Study of Security Proneness of Multi-language Systems: *ESEC/FSE 2022*.

My Role: Project Leader
Design and develop a toolkit for security issue analysis in multi-language software in open-source community and mine the associations between multi-language combinations and security vulnerabilities.

Jan 2015 – Nov 2017 **Module Design Engineer, HUAWEI Research Institute**

My Role: Technical leader of the system support team (15 people).

– Responsible for the technical design and difficult problem tackling.

– Development environment: C, Linux 32/64, X86/ARM.

– Design, development, and commercial deployment of multiple requirements for ONT products worldwide (market share: 65%).

– Design and development of the ONT device virtualization/IoT device management plug-ins.

– Served as an internal Open-Source Committer at HUAWEI.

Jun 2010 – Dec 2014 **System Engineer, ZTE Research Institute**

My Role: Technical leader of the DPI team (30 people).

– Responsible for technical design and difficult problem tackling.

– Development environment: C, Linux 32/64, X86/MIPS/ARM/DSP.

– Design and lead the development of a high-performance regular compilation and matching engine, network flow classification system and security components (e.g., online Spam/DDoS detecting & blocking). Applications include core networks, high-performance routers, and home gateways that are commercially available worldwide.

Jun 2007 – Apr 2010 **Software Engineer, NEUSOFT**

My Role: Team leader of the GIS map team (10 people).

– Responsible for team resource management and technical design.

– Design and lead the development of compiler of GIS map for vehicle navigation in BMW, GM, HONDA.

– Development environment: C, Linux 32/64, X86.

– Participate and pass CMMI level 3 and BMW SPICE level 2 certifications as a team leader.

Proposal and Grant Experience

Play helpful roles in two research proposals, significantly contributing to writing the preliminary results.

[1] SHF: Small: Practical Dynamic Program Reasoning Across Language Boundaries. 2022 [Funded]
National Science Foundation (`https://www.nsf.gov/awardsearch/showAward?AWD_ID=2146233`)
– PI: Haipeng Cai
– My role: wrote part of the methodology section and polished the proposal under the guidance of Prof. Haipeng Cai.
– Awarded amount: $500K (06/01/2022 - 05/31/2025)

[2] SaTC: CORE: Small: Collaborative Runtime Testing of Interpreted Languages. 2023 [Pending]
National Science Foundation
– PI: Haipeng Cai
– My role: wrote part of the proposed technical work and preliminary results from the PyRTFuzz paper, and polished the proposal under the guidance of Prof. Haipeng Cai.
– Awarded amount: $600K (10/01/2024 - 09/30/2027)

Presentations

08/2023 PolyFuzz: Holistic Greybox Fuzzing of Multi-Language Systems. USENIX Security 2023, Anaheim, CA.

11/2022 On the Vulnerability Proneness of Multilingual Code. ESEC/FSE 2022, Singapore.

11/2022 PolyFax: A Toolkit for Characterizing Multi-Language Software. ESEC/FSE Tool Demo 2022, Singapore.

08/2022 PolyCruise: A Cross-Language Dynamic Information Flow Analysis. USENIX Security 2022, Virtual.

05/2021 AndroCT: Ten Years of App Call Traces in Android. MSR 2021, Virtual.

11/2020 PCA: Memory Leak Detection using Partial Call-Path Analysis. ESEC/FSE Tool Demo 2020, Virtual.

Teaching Experience

Fall 2023 Cpts528: Software Security and Reverse Engineering Teaching assistant

- Holding office hours for question answering.
- Grading assignments and exams for around 40 undergraduate and graduate students.
- Guiding the undergraduates to complete their course projects.

Fall 2020 Cpts322: Software Engineering Principles Teaching assistant

- Holding office hours for question answering.
- Grading assignments and exams for around 50 students.
- Guiding the undergraduates to complete their course projects, including coaching teams on establishing appropriate project milestones, holding weekly meetings, and providing technical instruction to students.

Spring 2020 Cpts317: Automata and Formal Languages Teaching assistant

- Holding office hours for question answering and elaborating on basic principles.
- Grading assignments and exams for around 40 students.

Mentoring

2023 Haoran Yang (haoran.yang2@wsu.edu), Junior PhD in Computer Science, Washington state university

- Guided how to partition a complex research problem into solvable subproblems.
- Guided on designing a sound technique for the target problem.
- Result: Submitted a paper to a top SE venue.

2022 Guangbei Yi (guangbei.yi@wsu.edu), Undergraduate in Computer Science, Washington state university

- Worked as a collaborator in research project PolyFuzz.
- Guided how to design the evaluations for a research problem.
- Guided how to develop efficient and automated test suites.
- Result: The 3rd author of the paper PolyFuzz.

2022 Jinyang Ruan (jinyang.ruan@wsu.edu, now SDE@FNZ), Master in Computer Science, Washington state university

- Worked as a collaborator in research project PolyFuzz.
- Guided how to study and understand a research problem.
- Guided how to design a complex software system with multiple collaborative components.
- Guided how to develop a submodule according to the system design.
- Result: The 2nd author of the paper PolyFuzz.

2020 Alissa Cielecki (acielecki@arcadia.edu, now Tech Specialist@Wharton), Undergraduate in Computer Science, Arcadia University

- Worked as a collaborator on the REU project.
- Guided how to design and implement the tool for data mining of open-source software on GitHub.
- Guided how to write the technical report for the project.
- Result: The REU project was finished successfully.

Professional Services

External reviewer	ESEC/FSE, 2024
External reviewer	Network and Distributed System Security (NDSS), 2024
External reviewer	Transactions on Dependable and Secure Computing (TDSC), 2023
External reviewer	Network and Distributed System Security (NDSS), 2023
External reviewer	ESEC/FSE, 2023

Research Software

[1] PyRTFuzz: A collaborative fuzzing framework for the Python interpreter and runtime libraries.

- `https://github.com/awen-li/PyRTFuzz`

[2] PolyFuzz: A holistic grey-box fuzzing framework for multi-language systems, supporting effective seed generation based on sensitivity analysis. (languages: [C/C++, Python, Java]).

-`https://github.com/awen-li/PolyFuzz`

[3] PolyCruise: Dynamic information flow analysis for multi-language systems, with support languages of [C/C++, Python].

- `https://github.com/awen-li/PolyCruise`

[4] PolyFax: A toolkit for data mining of security-related commits open source community.

- `https://github.com/awen-li/PolyFax`

[5] PCA: A static analysis tool for memory leak detection, targeting large-scale C programs with high efficiency

- `https://github.com/awen-li/PCA`

Wen Li

Washington State University

Last update: Dec 07, 2023