

Xu He

✉ xu.he0105@hotmail.com
🌐 xuhe.info
🔗 github.com/Viewer-HX
☎ +1 571-685-3344
☎ +86 18362972726

Education

- **PhD, Information Technology - George Mason University** Sep 2019 - Dec 2024
Research Focus - AI Security; Binary Analysis; Machine Learning for Cybersecurity. GPA: 4.0/4.0
- **MS, Computer Science - Nanjing University of Posts & Telecom.** Sep 2016 - May 2019
Concentration - Machine Learning; AI-driven Data Mining; Security Analytics. GPA: 87.9/100
- **BS, Communication Engineering - Nanjing University of Posts & Telecom.** Sep 2012 - Jun 2016
Outstanding Thesis Award; Honors College Graduate (Top 5/77) GPA: 87.5/100

Research & Work Focus

- **AI Security & Robustness:** LLM redteam & guardrail, secure AI Agent design
- **AI for Security:** AI-driven vulnerability discovery, Anomaly/Intrusion detection, Program analysis, Automated program repair

Experience

- **VISA Inc.** Ashburn, VA
Senior Data Scientist, Cybersecurity Analytics & AI Innovation Jan 2025 - Present
 - **Generative AI Guard Service:** Architect and deploy a Generative AI Guard Service to protect GenAI-based applications, including LLM input/output filtering and tool protection (MCP/skill scanning and runtime checks).
 - **Secure Agent/Harness Service:** (1) Design and deploy an RL-based System Prompt Hardener Service to improve agent robustness against adversarial inputs. (2) Design and implement an agent runtime protection system, including endpoint monitoring, graph-based anomaly detection, and hook-based intervention.*PhD Intern, Cybersecurity Analytics & AI Innovation* May 2024 - Aug 2024
 - **AI-powered Deepfake Defense System:** Design and develop a DL-based defense solution to mitigate deepfake attacks on voice-based authentication systems.
 - **AI Agent System for Cybersecurity Operations:** Architect and develop an intelligent Subject Matter Expert (SME) agent system to automate threat analysis and metric summarization.
- **Center for Secure Information System (CSIS), GMU** Fairfax, VA
Graduate Research Assistant Sep 2019 - Dec 2024
 - **AI-Enhanced Binary Security Analysis:** Develop ML-driven static analysis for binary code understanding, including novel graph neural networks for security patch detection and compiler provenance identification.
 - **LLM-based Automated Program Repair:** Design and implement PathFix by combining LLMs with constraint solving for automated program repair.
 - **AI-driven Network Traffic Optimization:** Apply RL to adaptively tune Linux network parameters, including intelligent buffer management to improve network performance.
- **NIO USA Inc.** San Jose, CA
PhD Intern, Automotive Cybersecurity Software Development May 2022 - Aug 2022 / May 2023 - Aug 2023
 - **AI-Enhanced Intrusion Detection Systems:** Design and implement an IDS for remote vehicle security services by integrating ML algorithms for real-time threat detection.
 - **Automotive Security Infrastructure:** Develop an IAM system for in-vehicle software services, including anomaly detection for unauthorized access patterns in connected vehicle environments.

Professional Services & Recognition

- **Academic Paper Reviewer:** **Conference:** CCS; USENIX Security; NDSS; ACSAC; AsiaCCS; CNS; ICICS; INFOCOM; DSN; MILCOM. **Journal:** IEEE TIFS; ACM DTRAP; Springer WINE; Elsevier Computer Communications.
- **Artifact Evaluation:** NDSS; USENIX Security; ACSAC; CCS.
- **Security Research Impact:** CVE Discovery (CVE-2022-47085); RUSTSEC Advisories (2023-0046, 2023-0047).

Honors & Awards

- **CISO Award*2**, Visa USA Inc, Oct. 2025 / Apr. 2026.
- **Doctoral Research Scholarship**, George Mason University, Jun. 2023.
- **Student Grant Award**, IEEE INFOCOM 2022, Apr. 2022.
- **Outstanding Graduate Thesis Award**, Nanjing University of Posts & Telecom - Top 5/77 graduates, Jun. 2016.
- **Honorable Mention Award**, American Mathematical Contest In Modeling (COMAP), Feb. 2015.

Publications

1. **BenchChecker: Assessing the Credibility of Bug-Fixing Benchmarks for LLMs**
Di Wu, **Xu He**, Shu Wang, Kun Sun
The 35th USENIX Security Symposium (USENIX Security 2026)
(Acceptance Rate: Unknown)
2. **CULPA: Universal Detection of Memory-Safety Bugs in Unsafe Rust Through the Lens of Safety Requirements**
Hung-Mao Chen, Bo Lu, **Xu He**, Xiaokuan Zhang, Kun Sun
The 35th USENIX Security Symposium (USENIX Security 2026)
(Acceptance Rate: Unknown)
3. **COGNITION: From Evaluation to Defense against Multimodal LLM CAPTCHA Solvers**
Junyu Wang, Changjia Zhu, Yuanbo Zhou, Lingyao Li, **Xu He**, Junjie Xiong
The 35th USENIX Security Symposium (USENIX Security 2026) (Acceptance Rate: Unknown)
arXiv preprint arXiv:2512.02318 (2026)
4. **Tracing MCP Security Vulnerabilities: Constructing and Analyzing an MCP Security Vulnerability Dataset**
Xu He, Sagar Patel, Yan Zhai
The 1st Workshop on Datasets for Software Security (co-located with IEEE S&P 2026)
5. **SentinelAgent: Graph-based Anomaly Detection in Multi-Agent Systems**
Xu He, Di Wu, Yan Zhai, Kun Sun
Secure Agents for Generative Artificial Intelligence Workshop (Co-located with IEEE S&P 2026)
arXiv preprint arXiv:2505.24201 (2025)
6. **PathFix: Automated Program Repair with Expected Path**
Xu He, Shu Wang, Kun Sun
The IEEE Secure Development Conference (SecDev 2025)
(Acceptance Rate: 21/66=31.8%)
7. **TYPEPULSE: Detecting Type Confusion Bugs in Rust Programs**
Hung-Mao Chen, **Xu He**, Shu Wang, Xiaokuan Zhang, Kun Sun
The 34th USENIX Security Symposium (USENIX Security 2025)
(Acceptance Rate: 407/2385=17.06%)
8. **BinGo: Identifying Security Patches in Binary Code with Graph Representation Learning**
Xu He, Shu Wang, Pengbing Feng, Xinda Wang, Shiyu Sun, Qi Li, Kun Sun
ACM ASIA Conference on Computer and Communications Security (ASIACCS 2024)
(Acceptance Rate: 65/301 = 21%)
9. **What IF Is Not Enough? Fixing Null Pointer Dereference With Contextual Check**
Yunlong Xing, Shu Wang, Shiyu Sun, **Xu He**, Kun Sun, Qi Li
The 33rd USENIX Security Symposium (USENIX Security 2024)
(Acceptance Rate: 98/515 = 19.0%)

10. **BinProv: Binary Code Provenance Identification without Disassembly**
Xu He, Shu Wang, Pengbin Feng, Kun Sun, Haining Wang, Qi Li, Songqing Chen
The 25th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2022)
(Acceptance Rate: $35/139 = 25.1\%$)
11. **Consistency is All I Ask: Attacks and Countermeasures on the Network Context of Distributed Honeypots**
Songsong Liu, Pengbin Feng, Jiahao Cao, **Xu He**, Tommy Chin, Kun Sun, and Qi Li
The 19th Conference on Detection of Intrusions and Malware & Vulnerability Assessment (DIMVA 2022)
(Acceptance Rate: $11/39 = 28.2\%$)
12. **Auter: Automatically Tuning Multi-layer Network Buffers in Long-Distance Shadowsocks Networks**
Xu He, Jiahao Cao, Shu Wang, Kun Sun, Lisong Xu, Qi Li
The 41st IEEE International Conference on Computer Communications (INFOCOM 2022)
(Acceptance Rate: $225/1129 = 19.9\%$) (**Student Grant Award**)
13. **When the Differences in Frequency Domain are Compensated: Understanding and Defeating Modulated Replay Attacks on Automatic Speech Recognition**
Shu Wang, Jiahao Cao, **Xu He**, Kun Sun, Qi Li
The 27th ACM Conference on Computer and Communications Security (CCS 2020)
(Acceptance Rate: $121/715 = 16.9\%$)
14. **RusTEE: Developing Memory-Safe ARM TrustZone Applications**
Shengye Wan, Mingshen Sun, Kun Sun, Ning Zhang, **Xu He**
2020 Annual Computer Security Applications Conference (ACSAC 2020)
(Acceptance Rate: $70/302 = 23\%$)
15. **An Automatic Annotation Method for Discovering Semantic Information of Geographical Locations from Location-Based Social Networks**
Zhiqiang Zou, **Xu He**, A-xing Zhu
2019 ISPRS International Journal of Geo-Information
16. **Enhancing the Impression on Cities: Mining Relations of Attractions with Geo-Tagged Photos**
Zhiqiang Zou, **Xu He**, Xingyu Xie, Qunying Huang
2018 IEEE Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing (UIC-ATC 2018)
17. **Sparse Representation of Sensor Network Signals Based on the K-SVD Algorithm**
Zhiqiang Zou, **Xu He**, Yinxia Wang, Jiagao Wu
2018 ACM International Conference on Modelling, Analysis, and Simulation of Wireless and Mobile Systems (MSWIM 2018)