

PENGXIANG HUANG

M.S. student at Northwestern

Evanston, Illinois, United States of America

huangpengxiang70@gmail.com ♦ (+1) 773 960 1268 ♦ pengxiang-huang.github.io/

EDUCATION

Northwestern University

Evanston, United States

Master of Engineering in Electrical and Computer Engineering

Sep 2023 – Jun 2025

- **GPA:** 3.711 / 4.0
- **Courses:** Compilers Constructions, Code Analysis & Transformation, Advanced Computer Architectures, Cryptography
- **Thesis:** Interprocedural data flow analysis generalization for LLVM on dependence graph

University of Minnesota, Twin Cities

Minneapolis, United States

Exchange Program in Computer Science

Sep 2022 – May 2023

- **GPA:** 3.667 / 4.0, Dean's List
- **Courses:** System Security, Computer Architecture, Machine Learning, Computer Networks, Natural Language Processing

The Chinese University of Hong Kong, Shenzhen

Shenzhen, China

Bachelor of Science in Computer Science

Sep 2019 – May 2023

- **GPA:** 3.337 / 4.0, 3.525 / 4.0 (Major), Dean's List (2021-2022)
- **Courses:** Distributed and Parallel Computing, Operating System, Database System, Software Engineering
- **Honor:** First Class Graduated Student
- **Academic Scholarship:** Bowen scholarship (30,000 RMB per year)
- **Activity Scholarship:** Diligentia College Excellent Student Activities Scholarship (3,000 RMB on 2021)
- **Leadership:** Diligentia College Men's Basketball Team Leader (2020-2021)

RESEARCH EXPERIENCE

Data Flow Engine Design on LLVM

Evanston, United States

Student Individual Research, Northwestern ARCANA Laboratory

Dec 2023 – Present

- Implemented a data flow engine within the research compiler NOELLE, Generalizing forward and backward data flow analyses into a unified framework. Enabled data flow analysis over the program dependence graph, which provides the chance of dependence-level data flow analysis.
- Implemented a summary based inter procedural data flow engine, providing a scalable interface in LLVM IR for efficient inter-procedural data flow analysis on parallelization pipelines.

C to Safer Rust

Remote, United States

Remote research collaborator, Purdue Security Laboratory

Jun 2024 – Present

- Developed a novel C to Rust translation tool with inter-procedural semantic analysis and intra-procedural memory collections rewriting based on the AI Language Models. Utilize AI-driven aggressive rewriting capabilities to overcome traditional C-to-Rust conversion limitations in semantic understanding.
- Applied formal verification to ensure the correctness and efficiency of translations, with initial results showing overall 30% of improvement upon state of art methods, planning to submit 25' OOPSLA.

LLM Code Generator Security Analysis

Minneapolis, United States

Research Assistant, UMN Security Laboratory

Mar 2023 – Sep 2023

- Identify potential security vulnerabilities in an LLM-based code generator such as Copilot through analysis of outdated training model datasets. Developed an automated detection and filtering tool (APILOT) to detect and prevent the generation of vulnerable code by the generator.
- Develop automated tools to patch vulnerabilities and ensure safe code generation. APILOT can reduce outdated code recommendations by 89.42% on average with limited performance overhead, showing an average increase of 27.54% in usability. The work is submitted to and under review of 25' IEEE Symposium on Security and Privacy.

Graph Computing

Remote, China

Remote Research Collaborator, Alibaba DAMO academy

Jan 2022 – Aug 2022

- Assisted in the development of a large-scale distributed sub-graph matching system in Alibaba GAIA Engine. Conducted a review of existing semantic graph-based pattern-matching techniques to optimize NP-hard query problems.
- Engaged in the proposal of a solution for building an efficient graph-based catalog to optimize edge intersection and joint operations in query operations. Utilized benchmarks to test and debug the proposed catalog.

PUBLICATIONS

[Link]APILOT: Navigating Large Language Models to Generate Secure Code by Sidestepping Outdated API Pitfalls

Published on *arxiv*, under review of *IEEE Symposium on Security and Privacy* 2025

Sep 2024

WORKING EXPERIENCE

System Security Analysis

Shenzhen, China

System security Engineer, Tencent

May 2023 – Aug 2023

- Incorporated the Cross-Translation-Unit (CTU) function into Tencent's static analysis toolchain to analyze the extensive TDSQL database system, enhancing the Inter-procedural bug detection capabilities and identifying an additional 12 false positives and 8 false negatives.
- Developed a novel analysis toolchain for TDSQL utilizing CodeQL, implementing custom rules to identify bugs associated with path explosion issues, and successfully pinpointed 3 false negative cases.
- Aggregated security vulnerabilities and their corresponding mitigation strategies, and pre-trained the Tencent LLM, HunYuan, for automated security vulnerability detection.

Software Development

Shenzhen, China

Software Engineer, Zhongshen Agricultural Innovation Technology Co., Ltd

Apr 2021 – Aug 2021

- Managed the development and operations of distributed database systems for an artificial intelligence of things (AIoT) agriculture project that won 3rd prize in the 7th China National College Students' "Internet +" Innovation and Entrepreneurship Competition in Guangdong.
- Implemented scalable and reliable database systems, ensuring smooth data flow and accurate data analysis for the AIoT agriculture project. Contributed to the platform's success by developing user flows and wireframes to enhance the user usability.
- Assisted in the design and implementation of different machine learning algorithms for decision-support systems to evaluate products based on the data collected through deployed sensors.

PROJECTS

Code Analysis and Transformation, Compiler Construction

Northwestern University

Sep 2023 – Mar 2024

- Developed a frontend and backend compiler for translating a C-like language into Intel x86 Assembly, incorporating control flow translation, Tensor Flattening, bound inference, instruction selection, peephole optimization, register selection, and assembly generation. Achieved superior performance compared to production compilers, GCC and Clang.
- Designed and implemented a middle-end compiler for optimizing a domain-specific language, CAT, employing techniques such as reaching definition analysis, constant propagation, liveness analysis, dead code elimination, dependence analysis, alias analysis, loop normalization, loop unrolling under LLVM IR. Successfully optimized a large code base execution from 3 hours to just 15 minutes.

System Security Exploit and Simulation in Speculative CPU

University of Minnesota, Twin Cities

Sep 2022 – Dec 2022

- Conducted a comprehensive review and cycle-level simulation using gem5 of existing covert channel attacks due to speculative execution, including Meltdown and Spectre.
- Analyzed the mechanisms of these attacks and evaluated the effectiveness of various mitigation strategies. Assessed the performance trade-offs associated with each strategy and provided recommendations for optimal deployment.

SKILLS, LANGUAGES, INTERESTS

- **Languages:** English (fluent), Madarin (Native speaker)
- **On-Campus Employment:** Northwestern Gym Customer Service Representative, American Red Cross CPR/AED Certificates
- **Soft Skills:** Public Speaking, Communication, Academic Writing, Critical Thinking, Leadership
- **Programming:** C++/C, LLVM, CMake/Make, Python
- **Tools:** Git, Vim, Docker
- **Web Development:** Java, Node.js
- **Interests:** Basketball, Bodybuilding, Hiking, Music