Xida Ren

Expert in x86, RISC, Systems, Security, and Reliability. Fullstack / Cloud / Machine Learning Hobbiest

Education

2019 - 2023 PhD in Computer Architecture; University of Virginia

GRE: Verbal: 167, Quant: 168 / GPA: 4.0 / All but Thesis

2016-2019 BSc, Computer Science and Mathematics; College of William and Mary 3.8

■ Skills

Tools: Expertise in: C++, Python, C, Zig, SQL, MatLab, Haskell Familiar with: Pandas, Numpy, TensorFlow,

PyTorch, sklearn, Linux

Specialities: CPU/GPU Performance, Machine Learning, Queuing Theory, Risk, Statistics **Interests:** Security, Hardware Accelerators, Formal Verification, Hearthstone, FengShui

Experience

2023: MevLink (Informal Consulting for Low-Latency Etherium Data Company)

- · Cut version interval by half by consulting with team to implement test-driven development and agile practices
- Cut RLPx cryptographic overhead by 10x using LLVM and hand-coded AVX512 Keccak-256 implementation.
- Lowered data latency (company's key competitive advantage) for distributed Eth data system by 2x in my first week by
 mastering Ziglang from scratch to optimizing a 70,000 LoC codebase.

Aug 2022-Jan 2023: Intel Labs, Research Intern

- Achieved <3% CPI and <10% MPKI error while speeding up benchmarking by 10^6x using SimPoints, measured over SPEC 2017 x86 workloads
- Sped up new-workload SimPoint generation 200x by sampling CPU performance counters to generate SimPoints, avoiding simulation and instrumentation overheads.
- Use differential privacy to enable trace-sharing across organizational boundaries without concern for leaking sensitive IP.

May 2022 - Aug 2022: NXP Semiconductors, Hardware Security ML Engineer

- Achieved 85% accuracy rate in generalizing detectors to zero-day attacks on edge hardware by training unsupervised ML algorithms.
- Compared supervised and unsupervised algorithms for Spectre Side-Channe Detection, including VAEs, logistic regression, perceptrons, time-convolutional neural networks, decision trees, k-nearest neighbors, random forests, and support vector machines.
- Designed system in close collaboration with the VP of Edge Software for usable tradeoffs between detection accuracy and performance.

Aug 2019-May 2023: UVA Computer Science Department, PhD Candidate (Exited All-But-Thesis)

- Unearthed two critical vulnerabilities in modern x86 processors, enhancing security and integrity, published in ISCA 2021
- Wrote LLVM / MLIR compiler passes for discovering vulnerabilities and optimizing / studying Machine Learning Models.
- Applied formal verification techniques to harden machine learning models against adversarial attacks.

Aug 2020 - Nov 2020: Lawrence Berkeley National Lab, Research Intern

- Contributed to the PARADISE++ Project, spearheading the development of a memory subsystem for a parallel discreteevent simulator with C++ and Valgrind.
- Engineered an optimistically synchronized parallel discrete-event simulator memory subsystem, elevating project efficiency and output.

■Selected Projects

2023 LLM Quick Serverless Inference Project – Achieving Rapid Cold-Starts with Large Language Models

- Implement REST-based Upload-Your-Own-Model serverless inference infrastructure using FastAPI and Docker
- Achived cold-start latency 4x faster than industry standard (15s, down from 67s-3m59s for GPT Neo
 1.3b) using AWS SageMaker (compute), AWS S3 (model caching), Docker-Compose (orchestration), and Ansible (deployment)
- 2017 Int'l Genetically Engineered Machines Contest (iGEM 2017) Won Best Model award by modeling genetic circuit behaviors and designing novel gene expression rate control methods.
- 2017 React-HexLife Conway's Game of Life on Hexagonal Grid, Available on Github Developed a distinctive version of Conway's Game of Life using a hexagonal grid, innovating on a classic concept using React.js Implement robust full-stack solution with NodeJS for backend development Revisiting and enhancing the project with "bun" framework

■ Contact