Kushagra **Shah**

System Design · Machine Learning · Network Security

📱(+49) 15904187116 | ✉ kushagrashah298@gmail.com | 🏠 kushagrashah.github.io | in kushagrashah298

Summary

I aspire to contribute to the field of energy-efficient, high-performance, and data-driven computing. My interests span web security, database systems, computer architecture, and machine learning. I am motivated to develop cutting-edge technologies and embrace new challenges!

Experience

Huawei Technologies
Munich, Germany

Senior Research Engineer | AI4Sec Team — *Aug 2024 - Present*

- Developing an ML-based zero-day phishing detector in C for Next-Gen Firewall SoC devices. Leading the system design and component planning.
- Designing, validating, and benchmarking an end-to-end prototype in a production-like setup. Mentoring an intern on implementation tasks.
- Built a high-performance Python module for brand detection in HTML using the Hyperscan engine. Achieved 96% precision with <1 ms runtime.
- Benchmarked multiple CNN inference pipelines using various libraries, optimized pre-processing and buffering to meet performance targets.
- Conducted in-depth analysis of phishing datasets to guide ML feature development; authored reports and presented findings to stakeholders.
- Provided DevOps support for a demo: phishing simulation, HTML generation, MITM proxy, web scraping bots, CI/CD and testing pipeline setup.

Senior Research Engineer | Storage4AI Team — *Jul 2023 - Jul 2024*

- Researched, prototyped and delivered a vector+scalar composite index for a distributed multi-tenant vector database in Huawei Cloud.
- Led the project, taking full responsibility for development and delivery while incorporating valuable feedback from the team and stakeholders.
- Achieved up to 5x performance improvement compared to the state-of-the-art HNSW index, while maintaining similar accuracy and index size.
- Improved the debuggability of a Spark-based system by collecting relevant statistics on run-time and updating the history server web interface.

Oracle Switzerland
Zurich, Switzerland

Research Assistant | Data Plane Team — *Sep 2022 - Mar 2023*

- Explored various architecture choices to optimize the data load operation in an analytical query engine while collaborating with multiple teams.
- Conducted in-depth hardware performance experiments using various benchmarks to test the viability of the proposed engine architecture.
- Experimented with storage technologies, MySQL features, data storage formats, page organization and code optimization at various levels.
- Developed a prototype which scales with data size and compute, while offering more than 3x performance improvement in the load speed.

Selected Projects

Gradient Compression with New Numerical Encodings
EPFL, Switzerland

Advanced Multiprocessor Architecture Course

- Investigated the effectiveness of gradient compression on DNN models trained on a hardware emulator that uses hybrid block FP encoding.
- Tested the Python design with several image classification experiments (ResNet18 on CIFAR10) with different hyperparameters. Achieved an accuracy of about 94% with HBFP (cf. 94.7% with FP) despite using a lower precision encoding with 4 bits only.
- Implemented an RTL design for the gradient compression block which will serve as the foundation for integrating HBFP on a GPU cluster.

Spectre Attack
EPFL, Switzerland

Advanced Computer Architecture Course

- Demonstrated a micro-architectural side-channel attack capable of stealing data from any memory location using a victim C function.
- Trained the branch predictors to mis-speculate load instructions and delay the branch resolution in order to exploit data residue from the cache.
- Executed a standard prime + probe cache attack to illegally read data from the memory. Various parameters in the algorithm were fine tuned for the machine to provide consistently competitive results.

Skills

Industry Knowledge	System Design, Machine Learning, Network Security, Performance Optimization, Digital Logic Simulation
Framework and Tools	SSH, Git, Docker, Kubernetes, Pytest, Wireshark, FAISS, Hyperscan, Spark, MySQL, TensorFlow, PyTorch
Programming Languages	Python, C, C++, Java, Scala, Golang, Verilog, VHDL, Chisel, Shell, SQL, HTML, JS, CSS
Spoken Languages	English (C1), Hindi (C2), Gujarati (C2), German (A1), French (A1), Mandarin (A1)

Education

École Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland

M.Sc. in Computer Science, Specialization in Data Analytics, **CGPA:** 5.3/6 — *Sep 2020 - Mar 2023*

- Thesis at Oracle, Switzerland: Revisiting Data Ingestion for a Distributed Query Engine
- Research Scholar at the Processor Architecture Laboratory (LAP)

Birla Institute of Technology and Science (BITS) Pilani
Goa, India

B.E. in Electrical and Electronics Engineering, **CGPA:** 9.3/10, **Rank:** 3/83 — *Aug 2016 - May 2020*

- Thesis at NTU, Singapore: Published a Technique for Vendor and Device Agnostic Hardware Area-Time Estimation
- Teaching Assistant for courses: Computer Architecture, Microprocessors and Interfacing, Digital Design