Xie, Minghao

phone: (+1) 831-515-8143
e-mail: mhxie@ucsc.edu

Summary

A PhD candidate conducting research on predictable, high-performance storage systems within data centers at Center for Research in Storage Systems (CRSS).

Education

University of California, Santa Cruz, California — Sep 2018 – Now
- Ph.D. Student, Computer Engineering (GPA: 3.92)
- Conducting data-center research advised by Chen Qian and Heiner Litz

Sichuan University, Chengdu, China — Sep 2014 – Jul 2018
- B.Eng., Computer Science (GPA: 3.73, top 2% with honors)
- Top 4 in the Top-Notch Talent Program (National Honor Program)

Selected Publication

1. (Accepted) Liu, Y., Shi, S., **Xie, M.**, Litz H., & Qian, C. SMASH: Flexible, Fast, and Resource-efficient Placement and Lookup of Distributed Storage. In conference of the ACM special interest group for the computer systems performance evaluation community (SIGMETRICS '23).
2. Shi, S., Yu, Y., **Xie, M.**, Li X., Li, X., Zhang, Y. & Qian, C. Concury: A Fast and Light-weight Software Cloud Load Balancer. In conference of the ACM Symposium on Cloud Computing 2020 (SOCC '20).
3. **Xie, M.**, Qian, C., Litz, H. ReFlex4ARM: Supporting 100GbE Flash Storage Disaggregation on ARM SoC. OCP Future Tech-nologies Symposium 2020. Poster session in San Jose, CA (OCP '20).

Technical Skills

Programming Language: **Python**, **C/C++**, **Rust**, Java, matlab, Perl
System: Operating System, TCP/IP Networking, Flash Storage
DevOps: Git, Mercury, Terraform, Docker, Kubernetes, AWS, Meson
Library: DPDK, SPDK, ZeroMQ, Thrift, Django, twisted, OpenCV, qt5
Miscellaneous: OOD, SQL, KV Store, MQ, MVC, Async

Industrial Experiences

Disaggregated Storage for Embedding Table — July – Sep 2021
Company: Facebook, Position: Research Engineering Intern, CEA
- Analyzed the bottleneck of internal **PyTorch**-based SparseNN workloads
- Implemented a userspace NBD driver sped up by 1.7x˜6.7x in **C++**

Transport Optimization for Storage Disaggregation — June – Sep 2022
Company: Meta Platform, Position: Research Engineering Intern, CEA
- Optimized and achieved 2.6x gain on throughput, 63% reduction in latency
- Removed the performance bottleneck and scaled throughput linearly

Selected Recent Projects

LESS: A Latency-aware Ephemeral Storage System for Serverless Computing — Sep 2019 – June 2023
Advisor: Heiner Litz and Chen Qian, Funding: CRSS & NSF
- Built a disaggregated server w/ **SPKD** & **DPDK** at 1MIOPS per SSD
- Designed an SLO-aware flow-based scheduler for serverless jobs using **Ray**, achieving 14% ˜37% TCO savings compared other SLO-enforced schedulers
- Developed μ-second level **Cython**-based **asynchronous** storage library
- Implemented & deployed cloud-native workloads with **Terraform** on **AWS**