Yiwen Zhang

(319) 855-0995 | yiwenzhg@umich.edu | 1841 Shirley Ln, Apt A3, Ann Arbor, MI 48105

Research Interests

I am interested in kernel-bypass networks and systems with a focus on making data transfer fast and in a fair fashion in datacenters. My current research focuses on how to solve performance anomalies in RDMA NICs with software-only solutions in an efficient, fair, and easy-to-deploy manner.

Education

University of Michigan — Ann Arbor, MI
Ph.D. Candidate, Computer Science and Engineering — 2018 - Present
Masters, Electrical Engineering — April 2018
Bachelors, Electrical Engineering — December 2016

Industrial Experience

Research Intern, Google — Sunnyvale, CA, Summer 2020

Congestion Control Team, NetInfra. Worked on multi-tenancy support for applications in datacenters.

Software Engineering Intern, Google — Sunnyvale, CA, Summer 2019

Congestion Control Team, NetInfra. Worked on datacenter bandwidth allocation for ultra-low RPC latencies.

Performance Modeling Intern, ARM — Chandler, AZ, Summer 2017

Performance Modeling Group. Worked on modeling ARM's next-generation CPU.

Research Experience

Research Assistant, Computer Science and Engineering, University of Michigan — 2018 - Present
Advisor: Prof. Mosharaf Chowdhury

RDMA Performance Isolation

- Discovered and analyzed performance anomalies in various RDMA implementations across a wide range of RNICs.
- Designed a solution to provide predictable latencies and fair RNIC resource sharing via message-level shaping, initiator-based resource mediation, and passive system-wide latency monitoring.
- Achieved great performance improvement for both InfiniBand and RoCE with 40/100 Gbps RNICs.

Multi-Tenancy Support via Distributed QoS Management

- Analyzed how QoS usage impacts application performance isolation
- Designed a distributed QoS enforcement algorithm to achieve desired application performance
- Ongoing work in preparation for a conference paper in 2022

Efficient Model Inference Pipelines with Model Multiplexing

- Studied existing solutions in model inference and identified new research opportunities
- Build an event-based simulator to analyze performance of different inference system design
- Ongoing work in preparation for a conference paper in 2020

Efficient Message Queuing System in RDMA

- Identify performance bottleneck in state-of-the-art MQ systems to motivate an RDMA-based design
- Designed and implemented a message queuing system in RDMA
- Code open-sourced at https://github.com/SymbioticLab/rdmaMQ

Memory Disaggregation

- Deployed various applications (VoltDB, Memcached, Powergraph and Spark) to evaluate Infiniswap
- Contributed to most of the Evaluation section in our NSDI'17 paper

TEACHING EXPERIENCE (AT UNVERSITY OF MICHIGAN)

Graduate Student Instructor, EECS489: Computer Networks — Fall 2018
Wrote assignment auto-grader and led discussion.

Graduate Student Instructor, EECS582: Big Data Systems and Applications — Fall 2017
Designed programming assignments on big-data systems and wrote solutions

PUBLICATIONS

1. Justitia: Software Multi-Tenancy in Hardware Kernel-Bypass Networks. **NSDI'**22
   Yiwen Zhang, Yue Tan, Brent Stephens, Mosharaf Chowdhury
2. NetLock: Fast, Centralized Lock Management Using Programmable Switches. ***SIGCOMM****'20*
   Zhuolong Yu, **Yiwen Zhang**, Vladimir Braverman, Mosharaf Chowdhury, Xin Jin
3. Performance Isolation Anomalies in RDMA. ***KBNets****'17*
   Yiwen Zhang, Jucheng Gu, Youngmoon Lee, Mosharaf Chowdhury, Shin Kang
4. Efficient Memory Disaggregation with INFINISWAP. **NSDI**'17
   Jucheng Gu, Youngmoon Lee, **Yiwen Zhang**, Mosharaf Chowdhury, Shin Kang.

AWARDS

- **Ph.D. Student Fellowship**, Computer Science and Engineering, University of Michigan — 2018 - 2019