Xuhao Luo

201 N Goodwin Ave, Urbana, IL, 61801
(217) 377-2021 ⋄ xuhaol2@illinois.edu ⋄ LinkedIn

Education

University of Illinois Urbana-Champaign Aug, 2021 - Now
Ph.D. Student in Computer Science

University of California San Diego Sep. 2019 - Mar. 2021
M.S. in Computer Science, *GPA: 3.82/4.00*

University of Science and Technology of China Sep. 2015 - Jun. 2019
B.S. in Applied Physics

Research Publication

· Zhiyuan Guo*, Yizhou Shan*(co-first author), **Xuhao Luo**, Yutong Huang, Yiying Zhang, **Clio: A Hardware-Software Co-Designed Disaggregated Memory System** (***Preprint***) In-submission

Experience

University of Illinois Urbana-Champaign May. 2021 - Now
Research Assistant, advised by Prof. Tianyin Xu *Urbana, IL*

· Building a framework to implement and reason about fail-slow tolerant distributed systems in an easy and effective way.
· Implementing a light-weight user-space thread library with cooperative task scheduling using **C++ Coroutine**.
· Introducing `event` abstraction and `wait()` API for better management of waiting points globally.

Microsoft Research Jun. 2020 - Sep. 2020
Research Intern *Beijing, China*

· Designed and implemented task scheduling and dispatching system for distributed machine learning using **C++**.
· Designed and implemented **CUDA**-based high-performance inter-GPU communication channel for distributed ML within a large-scale GPU cluster.
· Multi-GPU collective operation(AllReduce, AllGather, Broadcast) throughput outperforms Nvidia NCCL by at most 18.4% under the same system setting.

University of California San Diego Sep. 2019 - Dec. 2020
Research Assistant, advised by Prof. Yiying Zhang *La Jolla, CA*

· Worked on building FPGA-based disaggregated memory system.
· Designed and implemented a *go-back-N* based reliable network stack on both FPGA and host Linux server to support high-performance reliable network communication. Using kernel-bypass to achieve high-throughput and low-latency.
· Designed and implemented an RPC-semantic connectionless network stack to improve scalability, with a delay-based congestion control.
· Achieved μs-level latency and 10Gbps(limited by hardware interface) throughput at rack scale.

Agora.io Jul. 2019 - Sep. 2019
Software Engineer Intern *Shanghai, China*

· Participated in the development of CapSync, a distributed capability negotiation system for synchronizing media capability info between users, implemented with **C++** and **libevent**.

Projects

Distributed Messaging System Apr. 2020 - Jun. 2020
Project for CSE223, Distributed System

· Built a distributed messaging system patterned on Kafka using **Go**. Provided messaging service via `Append()` and `Get()` APIs. Implemented *Topic* and *Partition* abstraction for replication management with **Zookeeper**.

Fault-tolerant Distributed Storage System Sep. 2019 - Dec. 2019
Project for CSE224, Networked System

· Implemented a cloud-based file storage system patterned on Dropbox. Used multiple servers for duplicated file storage. Achieved consistence and fault-tolerance mechanism using **Raft** consensus algorithm.

Skills

Language C/C++, Python, Go, Rust, Haskell, OpenCL, Verilog
Tools/Framework TensorFlow, Docker, Zookeeper, LLVM, Google Test