

Xiao Zhang

☎ (+1) 737-382-2777 | ✉ zx123@utexas.edu | 🏠 timez-zx.github.io | 📷 Timez-zx

ABOUT ME

I am a Ph.D. student at UT Austin advised by Prof. Daehyeok Kim. I build networked and distributed systems for AI and latency-critical applications. My research spans datacenter networks and edge computing, with a current focus on making AI workloads predictable and efficient—from the network fabric up to GPU resources. My goal is to build practical systems that bridge real-world deployment with the infrastructure needs of modern AI.

EXPERIENCE

Google Cloud

Software Engineering Intern

Sunnyvale, CA

June 2026 - Present

- Developing network telemetry for LLM workload analysis

University of Texas at Austin (with Prof. Daehyeok Kim)

Ph.D. Student

Austin, TX

Sept. 2024 - Present

- Designed and built a private 5G testbed with GPU-accelerated edge servers for studying AI inference latency over cellular networks
- Conducted large-scale measurements using AWS Wavelength Zones and Verizon 5G to identify latency variability in edge AI applications
- Identified key performance bottlenecks in wireless scheduling and GPU resource contention; currently developing cross-layer resource management techniques for predictable edge inference latency

University of Pennsylvania (with Prof. Vincent Liu)

Visiting Student

Philadelphia, USA

July. 2023 - Present

- Beaver: Practical Partial Snapshots for Distributed Cloud Services

Shanghai Jiao Tong University (with Prof. Shizhen Zhao)

Master Student

Shanghai, China

Sep. 2021 - July. 2023

- Flattened Clos Plus (FC+): Near-optimal topology-routing co-design free of deadlocks for RoCE-based expander networks
- Flattened Clos (FC): Deadlock-free topology-routing co-design for RoCE-based expander networks

HONORS AND AWARDS

2025 Amazon AI Fellowship, Awardee

Austin, TX

2021 Outstanding Graduate of Shanghai, Awardee

Shanghai

2020 Liu Yongling Scholarship, Awardee

Shanghai

EDUCATION

University of Texas at Austin

Ph.D. in Computer Science

Austin, Texas

Sept. 2024 - Present

- Advisor: Prof. Daehyeok Kim
- Research focus: Networked & distributed systems, AI infrastructure, Edge computing

Shanghai Jiao Tong University
M.E. in Communication Engineering

Shanghai, China
Sept. 2021 - May. 2024

- Thesis title: FC+: Near-optimal Deadlock-free Expander Data Center Networks

Shanghai Jiao Tong University
B.E. in Information Engineering

Shanghai, China
Sept. 2017 - June. 2021

- Thesis title: Design of Robust and Efficient Edge Server Placement and Server Scheduling Policies

PUBLICATION

- **Xiao Zhang**, Daehyeok Kim. [Enabling SLO-Aware 5G Multi-Access Edge Computing with SMEC](#). In Proceedings of 23rd USENIX Symposium on Networked Systems Design and Implementation (**NSDI**), May 2026.
- Liangcheng Yu, **Xiao Zhang**, Haoran Zhang, John Sonchack, Dan Ports, Vincent Liu. [Beaver: Practical Partial Snapshots for Distributed Cloud Services](#). In Proceedings of 18th USENIX Symposium on Operating Systems Design and Implementation (**OSDI**), July 2024.
- **Xiao Zhang**, Peirui Cao, Yongxi Lyu, Qizhou Zhang, Shizhen Zhao, Xinbing Wang, Chenghu Zhou. [FC+: Near-optimal Deadlock-free Expander Data Center Networks](#). In Proceedings of 21st IEEE International Symposium on Parallel and Distributed Processing with Applications (**ISPA**), December 2023.
- Shizhen Zhao*, Qizhou Zhang*, Peirui Cao, **Xiao Zhang**, Xinbing Wang, Chenghu Zhou. [Flattened Clos: Designing High-performance Deadlock-free Expander Data Center Networks Using Graph Contraction](#). In Proceedings of 20th USENIX Symposium on Networked Systems Design and Implementation (**NSDI**), 2023.
- Shizhen Zhao*, **Xiao Zhang***, Peirui Cao, Xinbing Wang. [Design of Robust and Efficient Edge Server Placement and Server Scheduling Policies](#). In Proceedings of IEEE/ACM 29th International Workshop on Quality of Service (**IWQoS**), 2021.

SKILLS

Programming	C/C++, Python, Verilog, VHDL, Matlab
Tools	eBPF, DPDK, FPGA, Network Simulator 3 (NS-3), AWS
Languages	English, Chinese