

XIAOSU, LYU

PhD Senior Systems Engineer | High-Performance Networking & Cloud-Native Infrastructure

+1 202-250-4112 • Chandler, AZ 85248 (Open to Relocation) • lxiaosu22@gmail.com • LinkedIn:

linkedin.com/in/xiaosu-lyu-21b72b162/ • Github: github.com/lyuxiaosu

Summary

Systems Researcher & Senior Engineer with 13 years of C++ expertise in **high-performance** infrastructure. PhD candidate specializing in **virtualization (Wasm/Enclaves)**, **kernel-bypass** networking (**RDMA/DPDK**), and concurrent tasks **scheduling**. Deep understanding of **Linux internals, kernel/user-space boundaries, interrupts, DMA and system-level debugging**. Proven track record in identifying performance bottlenecks across the CPU-Memory-IO stack and optimizing Kernel data paths for **low-latency** execution. Seeking to leverage deep expertise in system bring-up and cluster-scale performance tuning to architect the software foundation for next-generation, high-performance AI computing infrastructure.

SKILLS

Languages: C, C++, Go, Java, Python, Perl, Shell

Systems & High-Performance: Operating System, WebAssembly (Wasm) Runtimes, Linux Kernel, Real-time Scheduling, Interrupts, RDMA, DPDK, Multithreading/Concurrency, Kernel-bypass, memory management.

Cloud & Infrastructure: Kubernetes (K8s), KNative, FaaS (OpenFaaS/OpenWhisk), Istio, Microservices, AWS (Nitro Enclaves), CDN, Prometheus, Grafana.

Networking: TCP/IP, UDP, IPv4/v6, Congestion Control, SDN, UDT, ethernet, DMA

Tools: Git, CMake, CI/CD, Ansible, Docker, Perf, GDB, Valgrind, TensorFlow.

EDUCATION

The George Washington University

Washington, D.C., U.S.

Ph.D. Computer Science (Computer Architecture Networks and Distributed Computing, GPA:3.87/4.0) May 2026

The George Washington University

Washington, D.C., U.S.

M.S. Computer Science (Distributed System and Machine Learning, GPA:3.86/4.0)

May 2019

Shenyang Jianzhu University

Shenyang, China

B.E. Computer Science and Technology (Overall GPA: 86/100 Rank: 6th/175)

July 2006

RESEARCH PROJECT

The George Washington University

Washington, D.C., U.S.

Research Assistant - Cloud Systems Lab

May 2019 to present

- Architected a scalable Wasm FaaS platform, achieving strong **isolation** and high **density** for **sub-millisecond** functions.
- Developed **kernel-bypass RDMA/DPDK communication layers** for ultra-low latency inter-function calls.
- Developed user-space interrupt-driven schedulers to schedule concurrent tasks and minimize deadline violations.

Independent study - Lab for Intelligent Networking and Computing

October 2017 to September 2018

Predicted task execution times with deep learning to optimize Hadoop task scheduling and minimize total execution cost.

WORK EXPERIENCE

ARM

Austin, U.S.

Systems Research Engineer - Recurring PhD Intern

May 2021 to August 2023

- Architected a secure Wasm FaaS within **AWS Nitro Enclaves**, cutting cold-start latency by **17x** while maintaining hardware-level isolation.
- Integrated Wasm with **RDMA-based** Remote Function Calls, enabling robust isolation with marginal overhead.
- Built pluggable **real-time schedulers (EDF,SRSF)** for **DAG-based** functions, ensuring **deadline compliance** for concurrent tasks.

Bukahudong Technologies Co., Ltd (Early-stage startup)

Beijing, China

Core Software Engineer - Streaming Server R&D

July 2015 to February 2016

- Architected and managed full-lifecycle **streaming clusters** for video publishing, playback, and recording.

ChinaCache

Beijing, China

Senior Network Engineer - North American R&D

April 2013 to July 2015

- Optimized **TCP/IP congestion control**, reducing packet loss and boosting throughput by **20%** under 3G networks.
- Designed a **real-time TCP bandwidth estimation** module in Squid, improving content delivery on handheld devices.
- Ported UDT (reliable UDP) into Squid, accelerating **large-scale LAN/WAN data transfer**.
- Developed and maintained **large-scale live streaming clusters** for publishing, playback, time-shift, and recording.
- Built a **Nginx module** for real-time FLV live streaming, enabling low-latency video publishing and playback.

Beijing QuanShi Co., Ltd	Beijing, China
<i>Senior Network Engineer - Base Platform R&D</i>	October 2011 to April 2013
• Implemented UDP congestion control for a video conferencing system, improving throughput by 30%.	
Beijing FastWeb Technology Co., Ltd	Beijing, China
<i>R&D Engineer - Network Planning R&D</i>	March 2011 to September 2011
• Designed a distributed system to monitor CDN node availability and aggregate client results.	
Tianjin National Cybernet Security, Ltd	Beijing, China
<i>R&D Engineer - Information Security R&D</i>	February 2009 to March 2011
• Developed proxy servers (HTTP, HTTPS, SOCKS) and engaged in network security R&D.	
Watchdata Technologies, Ltd	Beijing, China
<i>Intern - Security technology R&D</i>	May 2008 to January 2009
• Researched and implemented data encryption, authentication, and digital signatures with OpenSSL.	

PUBLICATIONS

- **Lyu, Xiaosu**, Emil Abbasov, Sean McBride, Gabriel Parmer, and Timothy Wood. "SledgeScale: Load-Aware Dispatch and Deadline-Driven Scheduling for Scalable, Dense Serverless Computing in Edge Data Centers." In *Proceedings of the Tenth ACM/IEEE Symposium on Edge Computing*, pp. 1-17. 2025.
- **Lyu, Xiaosu**, Emil Abbasov, Gabriel Parmer, and Timothy Wood. "A High-Density, Deadline-Aware, and Scalable Serverless Platform for Sub-Millisecond Functions at the Edge." In *2025 IEEE Cloud Summit*, pp. 35-38. IEEE, 2025.
- **Xiaosu Lyu**, "Balancing Three Important Goals for Runtimes - Isolation, High Performance, and Resource Efficiency," 2022 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C), 2022, pp. 60-62, doi: 10.1109/ACSOSC56246.2022.00031.
- **Xiaosu Lyu**, Ludmila Cherkasova, Robert Aitken, Gabriel Parmer, and Timothy Wood. 2022. Towards efficient processing of latency-sensitive serverless DAGs at the edge. In *Proceedings of the 5th International Workshop on Edge Systems, Analytics and Networking (EdgeSys '22)*. Association for Computing Machinery, New York, NY, USA, 49–54. <https://doi.org/10.1145/3517206.3526274>
- Viyom Mittal, Shixiong Qi, Ratnadeep Bhattacharya, **Xiaosu Lyu**, Junfeng Li, Sameer G. Kulkarni, Dan Li, Jinho Hwang, K. K. Ramakrishnan, and Timothy Wood. 2021. Mu: An Efficient, Fair and Responsive Serverless Framework for Resource-Constrained Edge Clouds. In *Proceedings of the ACM Symposium on Cloud Computing (SoCC '21)*. Association for Computing Machinery, New York, NY, USA, 168–181. <https://doi.org/10.1145/3472883.3487014>
- **Xiaosu Lyu**, Gaojun Liu. Research the application of Visual Database technology in telecom transmission network management, The Thirteenth National Conference on Youth Communication Proceedings, volume one, 60:259-262, Oct. 2008