Talha Waheed

twaheed2@illinois.edu | https://talha.cs.illinois.edu

Education

University of Illinois Urbana-Champaign 2022 – Present
PhD, Computer Science | Expected Graduation: May 2027 *Advisors: Brighten Godfrey and Radhika Mittal*

Lahore University of Management Sciences (LUMS) 2018 – 2022
BS, Computer Science

Publications

Multi-party Load Balancing in the Cloud
Talha Waheed, Sachin Ashok, Brighten Godfrey, Radhika Mittal, and Rayadurgam Srikant.
Under submission.

Coal Not Diamonds: How Memory Pressure Falters Mobile Video QoE
Talha Waheed, Zahaib Akthar, Ihsan Ayyub Qazi, and Zafar Ayyub Qazi.
ACM CoNEXT 2022.

Experience

ByteDance June 2025 – Aug 2025
Research Scientist Intern, Traffic Infrastructure Global Engineering *Bellevue, WA*

- Laid the foundational R&D work, and built an MVP, for an SRE AI Assistant for BGE (ByteDance Global Edge) and TLB (Toutiao Load Balancer).
- Built and curated MCP Servers to integrate real-time observability data, and integrated a knowledge base with internal documentation, enabling RAG-powered AI Assistant workflows.

Research Projects

Multi-party Load Balancing in the Cloud
Advisors: Brighten Godfrey and Radhika Mittal

- Designed *MPLB*, a global load-balancing solution that optimizes performance, reduces load imbalance across the cluster, and ensures cluster-wide fair resource distribution.
- Addressed critical challenges in global load balancing, including fast and dynamic adaptations to changing workloads, ensuring compatibility with diverse heterogeneous applications, and enforcing stable fair shares of resource allocation across the cluster.
- Implemented *MPLB* on *Envoy*, deployed it on *Kubernetes*, and evaluated it on a comprehensive sweep of possible topologies and load distributions on both microservice-based and monolithic applications.

Impact of Memory Pressure on Mobile Video Streaming QoE
Advisors: Zafar Qazi, Ihsan Qazi, and Zahaib Akhtar

- Performed a user study of memory usage patterns through `SignalCapturer`, an IRB-approved, privacy-compliant Android application I developed.
- Measured mobile video QoE under memory pressure by creating video playback server setups through DASH and Exoplayer, and developing a remote Node.js server to apply and persist memory pressure on the device.
- Analyzed system-level traces using Perfetto and found that frequent CPU preemption by the disk I/O daemon `mmcqd` and increased CPU usage by the memory reclaim daemon `kswapd` reduced CPU availability for video playback, degrading video performance under memory pressure.

Teaching Assistantships

Distributed Systems | Profs. Indranil Gupta, Aishwarya Ganesan & Ram Kesavan Fall 2025
Cloud Networking | Prof. Brighten Godfrey Spring 2025
Network-Centric Computing | Prof. Zafar Qazi Spring 2022
Data Structures | Prof. Ihsan Qazi Spring 2021

Notable Technical Skills

Languages: Go, Python, C++, C, JavaScript, Java
Frameworks/Libraries: MCP, LangChain, Kubernetes, Istio, Envoy, Gurobipy, Android Studio, FFmpeg

Awards

Sohaib and Sara Abbasi Computer Science Fellowship | University of Illinois Urbana-Champaign Aug 2022 – Present
Award of High Distinction | LUMS May 2022