Haoran Wan

35 Olden Street, Princeton, NJ, 08540 ◇ Email: haoran.w@princeton.edu ◇ Mobile: 609 366 6317 ◇ Homepage ◇ Google Scholar

Education

- **Princeton University**, Ph.D. Candidate in Computer Science, Advisor: Kyle Jamieson — Jul. 2023 - Present
  Teaching Assistant: COS 418 Distributed Systems ’25 Fall
- **Nanjing University**, M.S. in Computer Science, Advisor: Wei Wang — Sep. 2019 - Jun. 2023
  Outstanding Graduate Students Award ’21; Principal Special Scholarship ’19
- **University of Electronic Science and Technology of China**, B.Eng in Networking Engineering — Sep. 2015 - Jul. 2019
  China National Scholarship ’17; 2nd People’s Scholarship ’16, ’18

Skills Summary

- **Languages:** C++ (C++14/17, STL, Multithreading), Python (PyTorch, NumPy, Pandas), Go, Java, Bash.
- **Systems & Networking:** Linux Kernel Networking, WebRTC Internals, TCP/IP Optimization, Socket Programming, eBPF, Docker.
- **ML Infrastructure:** Apache TVM, CUDA, Model Quantization, Hugging Face Transformers.
- **Wireless & Hardware:** srsRAN, SDR (USRP), 5G NR/LTE Protocol Stack, Android System Tracing.

Technical Projects

Princeton University | *Advisor: Kyle Jamieson* — July 2023 - Present

- **Project L4Span**: Ultra-Low Latency Network Stack Optimization | C++, Linux, Congestion Control
  - **Engineered a low-latency zero-copy packet processing module** within an open-source 5G base station (srsRAN), utilizing C++ to integrate L4S (Low-Latency, Low-Loss, Scalable) standards and explicit congestion notification (ECN).
  - **Optimized critical data paths** to support advanced congestion control algorithms (TCP Prague, BBRv2&3, SCReAM), reducing packet sojourn time by **95%** while maintaining near line-rate throughput.
  - **Designed a real-time channel capacity predictor** that dynamically adjusts buffer management strategies based on sub-millisecond network telemetry, handling distinct flow requirements for classic (CUBIC) and low-latency traffic.
- **Project Athena & Domino**: Cross-Layer Network Telemetry System | C++, WebRTC, Data Analysis
  - **Engineered a distributed telemetry pipeline** that achieves millisecond-precision synchronization between low-level cellular logs (Layer 1) and application-level video performance data (Layer 7).
  - **Instrumented the WebRTC C++ native codebase**, modifying the Google Congestion Control (GCC) module to expose real-time internal state (bandwidth estimation, jitter buffer delay, frame freeze counts) for granular performance debugging
  - **Developed a root-cause analysis engine** that correlates network impairments with quality drops, identifying protocol-level bottlenecks and defining pacing optimizations that reduce video jitter by **50%**.
- **Project NR-Scope**: High-Throughput Network Telemetry Engine | C++, Multithreading, Real-Time Systems
  - **Architected a real-time telemetry extraction engine** capable of decoding and processing physical layer control signals within a strict **0.5 ms** transmission window.
  - **Engineered a custom C++ thread pool and worker queue system** to parallelize signal decoding, enabling real-time introspection of downlink/uplink resource allocation without blocking the critical path.
  - **Open-sourced the codebase** (GitHub), establishing it as a foundational telemetry tool now adopted by multiple external research teams for 5G protocol benchmarking and root-cause analysis.

Nanjing University | *Advisor: Wei Wang* — Sept. 2019 - Jun. 2023

- **Project ALT:** End-to-End Deep Learning Complier | C++, Python, Apache TVM, CUDA
  - **Architected a high-performance ML compiler** that unifies graph-level data layout and operator-level loop optimizations, resolving the ”phase ordering” problem in traditional compilation stacks.
  - **Engineered a custom Auto-Tuning Engine** on top of Apache TVM, implementing a cost model that searches for optimal tensor layouts (NCHW/NHWC) and loop tiling strategies simultaneously.
  - **Achieved 1.5x inference speedup** on single operators (Conv2D, MatMul) and 1.4x end-to-end latency reduction on ResNet, BERT, MobileNets models by generating hardware-aware CUDA/LLVM code.
- **mmSilent & Acoustic Sensing** | PyTorch, Signal Processing, Android
  - **Built a multimodal sensor fusion pipeline** synchronizing 60GHz mmWave radar signals with video streams, utilizing a custom Transformer backend for real-time inference.
  - **Developed a distributed acoustic ranging protocol** on Android/Linux, achieving sub-millimeter localization accuracy.

Selected Publications

- **[ACM CoNEXT ’25]** L4Span: Spanning Congestion Signaling over NextG Networks for Interactive Applications. **Haoran Wan**, Kyle Jamieson
- **[ACM IMC ’25]** Automated, Cross-Layer Root Cause Analysis of 5G Video-Conferencing Quality Degradation. Fan Yi, **Haoran Wan**, Kyle Jamieson, Oliver Michel
- **[ACM CoNEXT ’24]** NR-Scope: A Practical 5G Standalone Telemetry Tool. **Haoran Wan**, Xuyang Cao, Alexander Marder, Kyle Jamieson
- **[ACM HotNets ’24]** Athena: Seeing and Mitigating Wireless Impact on Video Conferencing and Beyond. Fan Yi, **Haoran Wan**, Kyle Jamieson, Jennifer Rexford, Yaxiong Xie, Oliver Michel
- **[ACM EuroSys ’23]** ALT: Boosting Deep Learning Performance by Breaking the Wall between Graph and Operator Level Optimizations. Zhiying Xu, Jiafan Xu, Hongding Peng, Wei Wang, Xiaoliang Wang, **Haoran Wan**, Haipeng Dai, Yixu Xu, Hao Cheng, Kun Wang, Guihai Chen
- **[ACM IMWUT ’23]** mSilent: Towards General Corpus Silent Speech Recognition using COTS mmWave Radar. Shang Zeng, **Haoran Wan**, Shuyu Shi, Wei Wang
- **[ACM IMWUT ’22]** VECTOR: Velocity Based Temperature-field Monitoring with Distributed Acoustic Devices. **Haoran Wan**, Lei Wang, Ting Zhao, Ke Sun, Shuyu Shi, Haipeng Dai, Guihai Chen, Haodong Liu, Wei Wang. **Distinguished Paper Award.**