

Research Interests

Performance and observability at scale, modern high-performance I/O interfaces (io_uring, libfabric, SPDK), streaming data workflows, versatile and expressive abstractions.

Education

- **Carnegie Mellon University (Parallel Data Lab)** Pittsburgh, United States
Ph.D. in Computer Engineering (Advisor: George Amvrosiadis) August 2018 - Present
- **IIIT Hyderabad** Hyderabad, India
B.Tech. in Computer Science and Engineering July 2011 - April 2015

Research

- **Realtime Control and Observability for Parallel Apps (Ongoing)**. Our experience with AMR codes motivated the need for real-time, programmable, and interactive observability for *bulk synchronous parallel* (BSP) applications. Currently, richer visibility over large-scale BSP application performance relies on analyzing large traces, such as OTF2 in HPC, and JSON in Pytorch/Kineto. These traces are large, expensive to parse and analyze, and delay feedback. We are building a real-time observability and control plane to dynamically orchestrate telemetry using declarative interfaces over modern high-performance fabrics.
- **Profiling, Tuning, and Optimizing Placement in AMR Codes (CLUSTER '25)**. AMR (*Adaptive Mesh Refinement*) codes are a class of scientific codes whose computational characteristics track the simulated phenomena. As a result, they exhibit volatility that needs to be aggressively and efficiently load-balanced to maintain a high cluster utilization. We studied such codes using fine-grained telemetry and designed placement policies that use their characteristics observed at runtime to balance different placement constraints. We also investigated and tuned out other sources of jitter, such as network drivers, and developed an analytical model to reason about task ordering in such codes.
- **CARP (SC '24)**. CARP (*Continuous Adaptive Range Partitioner*) is an adaptive single-pass range partitioner for efficient ingestion from large parallel applications. Unlike prior approaches that require expensive postprocessing to organize data for range queries, CARP intercepts application writes and shuffles data to create a partially ordered data layout with *nearly-optimal* query performance. CARP works with arbitrarily skewed and dynamic attribute distributions, and ingests at the maximum rate permitted by the parallel filesystem, making it up to 5× faster than prior work.

Experience

- **Carnegie Mellon University** Pittsburgh, United States
Graduate Student (Expected graduation: December 2025) August 2018 - Present
 - Working on a variety of problems in high-performance computing applications — online data indexing, placement and load-balancing, tuning and observability in large scientific codes ($O(100K)$ cores)
 - Teaching Assistant for 15-719/18-709 *Advanced Cloud Computing (Spring 2020)*
 - Teaching Assistant for 18-746 *Storage Systems (Fall 2023)*
- **Los Alamos National Laboratory/New Mexico Consortium** Los Alamos, United States
Summer Intern. See CARP May 2019 - Aug 2019
- **Microsoft Research** Bangalore, India
Research Fellow August 2017 - July 2018
 - Prototyped static analysis tools to enforce enterprise security and privacy policies

- Found a security vulnerability in an internal data lake as part of a red team exercise
- Prototyped a distributed ledger protocol for mobile apps (published as *Blockene*, OSDI 2020)
- **Directi** Mumbai, India
Software Developer (Flock) *July 2015 - July 2016*
 - Developed multiple user-facing features for Flock (a Javascript-based workplace collaboration app), revamped the build pipeline, and initiated migration to a more modern cross-platform framework.
- **Samsung R&D Institute** Bangalore, India
Summer Intern *May 2014 - July 2014*
 - Developed a real-time acoustic localization pipeline using a Kinect microphone array in C

Awards & Achievements

- *Benjamin Garver Lamme/Westinghouse Graduate Fellowship* for CMU academic year 2021-22
- *Commendation Certificate for Contribution to Web Services* at IIIT Hyderabad
- *Dean's List* for 5/8 semesters at IIIT Hyderabad
- Qualified for *ACM ICPC — Amritapuri Regionals* in 2013

Publications

1. **A. Jain**, C. Cranor, Q. Zheng, D. Manno, G. Amvrosiadis, and G. Grider. Lessons from Profiling and Optimizing Placement in AMR Codes. In *2025 IEEE International Conference on Cluster Computing (CLUSTER)*, Edinburgh, UK. IEEE, Sept. 1–30, 2025
2. **A. Jain**, C. D. Cranor, Q. Zheng, B. W. Settlemyer, G. Amvrosiadis, and G. A. Grider. CARP: Range Query-Optimized Indexing for Streaming Data. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC '24*, pages 1–19, Atlanta, GA, USA. IEEE Press, Nov. 2024
3. D. Kim, **A. Jain**, Z. Liu, G. Amvrosiadis, D. Hazen, B. Settlemyer, and V. Sekar. Unleashing In-network computing on scientific workloads, 2020. arXiv: 2009.02457
4. Q. Zheng, C. D. Cranor, **A. Jain**, G. R. Ganger, G. A. Gibson, G. Amvrosiadis, B. W. Settlemyer, and G. Grider. Streaming Data Reorganization at Scale with DeltaFS Indexed Massive Directories. *ACM Transactions on Storage*, 2020
5. Q. Zheng, C. D. Cranor, **A. Jain**, G. R. Ganger, G. A. Gibson, G. Amvrosiadis, B. W. Settlemyer, and G. Grider. Compact Filters for Fast Online Data Partitioning. In *Proceedings - IEEE International Conference on Cluster Computing, ICC, 2019*
6. A. Mehra, S. Lokam, **A. Jain**, M. Sivathanu, S. Singanamalla, and J. O'Neill. Vishrambh: Trusted philanthropy with end-to-end transparency. In *HCI for Blockchain, Proceedings of the A CHI 2018 workshop on Studying, Critiquing, Designing and Envisioning Distributed Ledger Technologies*, 2018

Skills

- *Programming languages:* C++, Python, Rust
- *Frameworks/libraries:* MPI, Apache Arrow, Datafusion, Pandas, Pytorch Kineto
- *Systems & I/O:* Linux, gdb, git, perf, eBPF, userspace I/O (*verbs, libfabric, io_uring, SPDK etc.*)

Last updated: September 18, 2025