Yuanli Wang

(+1) 612-963-5189 | yuanliw@bu.edu | https://pentium3.github.io/

EDUCATION

Boston University Boston, MA
PhD in Computer Science *09/2021 – Present*
- Advisor: Vasiliki Kalavri

University of Minnesota, Twin Cities Minneapolis, MN
Master of Science in Computer Science *09/2018 – 06/2021*
- Advisor: Abhishek Chandra

Hefei University of Technology Anhui, China
Bachelor of Science in Computer Science *09/2013 – 06/2017*

RESEARCH EXPERIENCE

A Disaggregated Stream Processing System for zero-downtime reconfiguration
Advisor: Prof. Vasiliki Kalavri, Boston University *05/2024 – Present*
- Designing a novel fully disaggregated architecture for data stream processing systems that completely decouples state management from computation, allowing workers to access both local and remote states transparently.
- Designing non-blocking lazy state migration protocols, enabling zero-downtime reconfiguration.

Automatic Task Placement for Data Stream Processing System
Advisor: Prof. Vasiliki Kalavri, Boston University *09/2021 – 05/2024*
- Conducted an empirical evaluation study on auto-scaling of cloud-based data stream analytics services and quantified their cost of over-provisioning.
- Proposed *CAPSys*, an adaptive resource controller for dataflow stream processors, that considers auto-scaling and task placement in concert.
- Designed an adaptive task placement policy that ensures compute, disk, and network resource-intensive tasks are balanced across available resources. Achieved orders of magnitude lower computing time and up to 6× higher throughput with fixed resources, compared to the state-of-the-art work.

Data Heterogeneity for Performance and Reliability in Federated Learning
Advisor: Prof. Abhishek Chandra, University of Minnesota *03/2020 – 05/2021*
- Investigated the impact of data heterogeneity in client selection and fault tolerance of federated learning training on edge computing environments.
- Designed client selection policy for federated learning system in order to improve training performance without reducing accuracy.
- Implemented a distributed federated learning system on top of PyTorch with the proposed policy. Reduced training time by 18%-38% without the loss of accuracy.

PUBLICATIONS

- **Yuanli Wang***, Lei Huang*, Zikun Wang, Vasiliki Kalavri, and Abraham Matta. CAPSys: Contention-aware task placement for data stream processing. Proceedings of the Twentieth European Conference on Computer Systems (EuroSys 2025). 2025 (To appear)
- **Yuanli Wang**, Baiqing Lyu, Vasiliki Kalavri. The *Non-Expert Tax*: Quantifying the cost of auto-scaling in Cloud-based data stream analytics. International Workshop on Big Data in Emergent Distributed Environments (BiDEDE 2022). 2022
- Esmail Asyabi, **Yuanli Wang**, John Liagouris, Vasiliki Kalavri, Azer Bestavros. A New Benchmark Harness for Systematic and Robust Evaluation of Streaming State Stores. Proceedings of the Seventeenth European Conference on Computer Systems (EuroSys 2022). 2022

- Joel Wolfrath, Nikhil Sreekumar, Dhruv Kumar, **Yuanli Wang**, Abhishek Chandra. HACCS: Heterogeneity-Aware Clustered Client Selection for Accelerated Federated Learning. 36th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2022). 2022
- **Yuanli Wang**, Joel Wolfrath, Nikhil Sreekumar, Dhruv Kumar, Abhishek Chandra. Accelerated Training via Device Similarity in Federated Learning. The 4th International Workshop on Edge Systems, Analytics and Networking (EdgeSys 2021). 2021
- Andrew Yoo, **Yuanli Wang**, Ritesh Sinha, Shuai Mu, Tianyin Xu. Fail-slow fault tolerance needs programming support. The 18th Workshop on Hot Topics in Operating Systems (HotOS XVIII). 2021
- **Yuanli Wang**, Dhruv Kumar, Abhishek Chandra. Exploiting Data Heterogeneity for Performance and Reliability in Federated Learning. Poster in the Fifth ACM/IEEE Symposium on Edge Computing. 2020

PROFESSIONAL EXPERIENCE

Apple | *Data Processing Platform Intern* 05/2023 – 08/2023

- Investigate on technological solutions to support data lineage tracking for Apple's AIML data processing platform.
- Integrate OpenLineage framework with Flink to track data lineage of Flink jobs. Implemented new visitors for Kafka and Iceberg data connectors in Flink.
- Fixed bugs in OpenLineage framework and contribute back to open source repository.

PingCAP | *Database Engineer Intern* 05/2019 – 08/2019

- Worked on AutoTiKV project from scratch: used machine learning to tune a database system under user-specific workloads.
- Implemented a Gaussian Process Regression Model to predict the performance of RocksDB(the core storage engine in TiKV) under different knob configurations. Achieved 1.3x lower latency under several types of workloads without human guidance.

INVITED TALKS

Towards a cost-efficient and QoS-aware self-managed stream processing system
- Meta 07/2022

CAPSys: Contention-aware task placement for data stream processing
- Tufts University 10/2024

SKILLS

Programming: C++, Python, Java, Go, Rust, Shell, SQL, Docker, Git, Flink, Beam, Kafka, RocksDB
Cloud infrastructures: AWS, Google Cloud, Microsoft Azure

PROFESSIONAL SERVICE

- **Program Committee:** EuroSys 2022 (Shadow PC), IMC 2022 (Shadow PC)
- **Artifact Evaluation Committee:** SOSP 2021, SIGCOMM 2021, MLSys 2023
- **Reviewer:**ICDCS 2024 (sub-reviewer), Future Generation Computer Systems, Internet of Things Journal

TEACHING

Teaching Fellow, CAS CS 551 Streaming and Event-driven Systems, Boston University Spring 2024
Teaching Fellow, CAS CS 210 Computer Systems, Boston University Fall 2022
Teaching Assistant, CSCI 5105 Distributed Systems, University of Minnesota Spring 2021
Teaching Assistant, CSCI 5103 Operating Systems, University of Minnesota Fall 2020

SELECTED AWARDS

Conference Student Grant: NSDI'21, OSDI'21
Bronze Medal, China Collegiate Programming Contest 10/2015