Hojin Park

hojinp.github.io

Email: hojinp@andrew.cmu.edu

Mobile: +1-617-642-1884

RESEARCH INTEREST

Distributed systems, Public cloud storage, Storage systems, Cache, Deep learning systems

EDUCATION

Carnegie Mellon University (CMU)   Sep. 2019 – Present.
   Ph.D. Student in Computer Science Department   *Pittsburgh, PA*

Seoul National University (SNU)   Mar. 2013 – Feb. 2019
   B.S. in Electrical and Computer Engineering   *Seoul, Republic of Korea*
   Graduated with Summa Cum Laude (GPA: 4.21 / 4.30)

Korea Science Academy of KAIST   Mar. 2010 – Mar. 2013
   Math & Science specialized high school   *Busan, Republic of Korea*

RESEARCH EXPERIENCE

Parallel Data Lab, CMU   Sept. 2019 - Present.
Graduate Research Assistant (Advisor: Prof. George Amvrosiadis, Prof. Greg Ganger)   *Pittsburgh, PA*

- **Macaron:** multi-cloud/region aware cache auto-configuration system

   ○ I am developing a system that optimizes the cache configuration in runtime to minimize the total cost of accessing data in a remote cloud or region.

- **Mimir:** a tool that finds the cost-efficient cloud storage configuration for storage systems

   ○ I developed Mimir, a system that helps users to make optimal decisions when composing distributed storage systems in the public cloud.

   ○ Mimir lets users enter a set of SLOs and outputs the most cost-efficient cloud resources configuration that minimizes the overall cost paid by the user.

- **Burstable storage in public clouds:** proposes a new way of exploiting burstable storage service cost-efficiently

   ○ I examined how burstable storage can be leveraged to reduce cost and/or improve performance for three use cases with different data-longevity requirements: traditional persistent storage, caching, and ephemeral storage

   ○ I found that by aggressively exploiting burstable storage service in public clouds, it is possible to increase storage throughput by up to 100x at a cost increase of only 10-40%.

VMware Research   Summer, 2023
Intern - Member of Technical Staff (Mentor: Adriana Szekeres)   *Remote*

- **Trace analysis:** Generated object storage access traces for cross-cloud/region caching research and analyzed data access patterns

CORE group, Microsoft Research   Summer, 2021
Research Intern (Mentor: Ishai Menache)   *Virtual*

- **VM packing:** Researched a deep learning model that optimizes the VM packing efficiency in a large cluster

Software Platform Laboratory, SNU   Jan. 2017 – Aug. 2019
Research Intern (Advisor: Prof. Byung-Gon Chun)   *Seoul, Republic of Korea*

- **Parallax:** a tool for automatic parallelization of deep learning training

   ○ Transforms a single-GPU deep learning model for distributed execution, handling correctness and scalability.

   ○ I used Parallax to explore two distributed training designs: Parameter Server and AllReduce.

   ○ I implemented four deep learning models with each distributed architecture to better understand these designs.

   ○ I ran experiments to evaluate these models on Parallax, in terms of correctness, scalability, and optimization.

- **Cruise:** a distributed machine learning framework with automatic system configuration

- Optimizes a system by adjusting worker/server assignment to homogeneous clusters at runtime.
- I enabled Cruise to work with heterogeneous cluster sets by implementing a custom linear-programming-based solver to optimize a generalized cost model.
- I implemented a Gradient Boosting Tree (GBT) application on top of Cruise.

Virtual Machine and Optimization Laboratory, SNU Jan. 2018 – July. 2018
Research Intern (Advisor: Prof. Soo-Mook Moon) *Seoul, Republic of Korea*

- **GitChain:** a distributed version control system using blockchain

  - Uses a public ledger to save version controlled repositories in InterPlanetary File System (IPFS).
  - I designed and implemented blockchain-related components of the system.
  - I implemented basic Git functions, such as push, pull, and clone, on the IPFS.

SCHOLARSHIPS & AWARDS

International Graduate Student Scholarship Sep. 2019 - Aug. 2024
Full tuition, insurance, and living expenses (5 years) *Korea Foundation for Advanced Studies*

Blockchain Technology Competition Jul. 2018
Two-person team won first prize ($3,000), with GitChain project *LINE, KIISE*

Undergraduate Study Scholarship Feb. 2017 - Dec. 2018
Full tuition and stipend ($2,500/semester) *Kwanjeong Educational Foundation*

Academic Excellence Scholarship Jun. 2013 - Dec. 2014
Full tuition *SNU*

PROGRAMMING SKILLS

- Languages: C/C++, Java, Python, SQL

- Multicore/GPU Libraries: OpenCL, CUDA, MPI, OpenMP

- Other: Tensorflow, Horovod, Gurobi (ILP)

PUBLICATIONS AND PREPRINTS

[1] Hojin Park, Ziyue Qiu, Gregory R. Ganger, George Amvrosiadis. Reducing cross-cloud/region costs with the auto-configuring MACARON cache. *SOSP 2024*, November 2024.

[2] Hojin Park, Gregory R. Ganger, George Amvrosiadis. Mimir: Finding Cost-efficient Storage Configurations in the Public Cloud. *SYSTOR 2023*, June 2023.

[3] Hojin Park, Gregory R. Ganger, George Amvrosiadis. More IOPS for Less: Exploiting Burstable Storage in Public Clouds. *HotCloud 2020*, July 2020.

[4] Woo-Yeon Lee, Yunseong Lee, Joo Seong Jeong, Gyeong-In Yu, Joo Yeon Kim, Hojin Park, Beomyeol Jeon, Wonwook Song, Gunhee Kim, Markus Weimer, Brian Cho, Byung-Gon Chun. Automating System Configuration of Distributed Machine Learning. *ICDCS 2019*, March 2019.

[5] Soojeong Kim, Gyeong-In Yu, Hojin Park, Sungwoo Cho, Eunji Jeong, Hyeonmin Ha, Sanha Lee, Joo Seong Jeong, Byung-Gon Chun. Parallax: Sparsity-aware Data Parallel Training of Deep Neural Networks. *EuroSys' 19*, March 2019.

[6] Soojeong Kim, Eunji Jeong, Joo Seong Jeong, Gyeong-In Yu, Hojin Park, Byung-Gon Chun. Auto-Parallelizing Deep Learning for Multi-machine, Multi-GPU Environments. *Workshop on AI Systems at Symposium on Operating Systems Principles (SOSP)*, October 2017.

TEACHING

Carnegie Mellon University
Storage Systems (15-746) Fall 2022
Advanced Cloud Computing (15-719) Spring 2022

TALKS

MACARON: Multi-cloud/region Aware Cache Auto-ReconfiguratiON

- The 30th ACM Symposium on Operating Systems Principles (SOSP)	November 2024
- CMU Parallel Data Lab Retreat	October 2024
- CMU Advanced Cloud Computing Course Guest Lecture	April 2024
- Alluxio & Uber Data Infra Meetup	January 2024
- CMU Parallel Data Lab Retreat	November 2023

Toward cost-efficient storage systems and data transfer in public clouds

- Salesforce Database Team Reading Group	January 2024

Mimir: Finding Cost-efficient Storage Configurations in the Public Cloud

- The 16th ACM International Systems and Storage Conference (SYSTOR)	June 2023
- CMU Advanced Cloud Computing Course Guest Lecture	April 2023
- CMU Parallel Data Lab Retreat	November 2022

MENTORING

Saileshwar Karthik (CMU Information Networking Institute masters student)	2025
Mohit Gaggar (CMU Information Networking Institute masters student)	2025
Fulun Ma (CMU Computational Data Science masters student)	2024
Somansh Satish (CMU Computational Data Science masters student)	2023
Anurag Choudhary (CMU Computational Data Science masters student)	2023
Midhush Manohar Thevendria Karthic (CMU Computational Data Science masters student)	2023
Shalini Shukla (CMU ECE masters student)	2022
Hao Yang Lu (CMU SCS masters student)	2022