Juncheng Yang

☎ (+1) 404-285-5231 | ✉ juncheny@cs.cmu.edu | ⌂ http://junchengyang.com

“Learn something about everything, learn everything about something.”

Education

Ph.D. in Computer Science, Carnegie Mellon University *Pittsburgh, U.S.A*
Computer Science Department, advisor: Rashmi Vinayak *Aug. 2018 - Present*

M.S. in Computer Science, Emory University *Atlanta, U.S.A*
Department of Mathematics and Computer Science, advisor: Ymir Vigfusson *Jan. 2015 - Dec. 2016*

M.S. in Chemistry, Emory University *Atlanta, U.S.A*
Department of Chemistry, advisor: Craig L. Hill *Aug. 2013 - Jun. 2015*

B.S. in Chemistry, Nanjing University *Nanjing, China*
Department of Chemistry and Chemical Engineering, advisor: Ying Wang *Sept. 2009 - Jun. 2013*

Selected Publications

NSDI’24	Yazhuo Zhang* (mentored student), Juncheng Yang*, Yao Yue, Ymir Vigfusson, K. V. Rashmi. **“SIEVE is Simpler than LRU: an Efficient Turn-Key Eviction Algorithm for Web Caches.”** *The 21st USENIX Symposium on Networked System Design and Implementation*.
SOSP’23	Juncheng Yang, Yazhuo Zhang, Ziyue Qiu, Yao Yue, K. V. Rashmi. **“FIFO Queues are All You Need for Cache Eviction.”** *ACM Symposium on Operating System Principles*.
HotOS’23	Juncheng Yang, Ziyue Qiu, Yazhuo Zhang, Yao Yue, K. V. Rashmi. **“FIFO Can be Better than LRU: the Power of Lazy Promotion and Quick Demotion.”** *The 19th Workshop on Hot Topics in Operating Systems*.
FAST’23	Juncheng Yang, Ziming Mao, Yao Yue, K. V. Rashmi. **“GL-Cache: Group-level learning for efficient and high-performance caching.”** *The 21st USENIX Conference on File and Storage Technologies*.
NSDI’22	Juncheng Yang, Anirudh Sabnis, Daniel S. Berger, K. V. Rashmi, Ramesh K. Sitaraman. **“C2DN: How to Harness Erasure Codes at the Edge for Efficient Content Delivery.”** *19th USENIX Symposium on Networked Systems Design and Implementation*.
NSDI’21	Juncheng Yang, Yao Yue, K. V. Rashmi. **“Segcache: memory-efficient and high-throughput DRAM cache for small objects.”** *18th USENIX Symposium on Networked Systems Design and Implementation*. **Community (Best Paper) Award**.
OSDI’20	Juncheng Yang, Yao Yue, K. V. Rashmi. **“A Large Scale Analysis of Hundreds of In-memory Cache Clusters at Twitter.”** *14th USENIX Symposium on Operating Systems Design and Implementation*. **Invited fast track submission to TOS’21**.
SOCC’17	Juncheng Yang, Reza Karimi, Trausti Saemundsson, Avani Wildani, Ymir Vigfusson. **“Mithril Mining Sporadic Associations for Cache Prefetching.”** *ACM Symposium on Cloud Computing*.
Eurosys’23	Ziyue Qiu, Juncheng Yang, Juncheng Zhang, Cheng Li, Xiaosong Ma, Qi Chen, Mao Yang, Yinlong Xu. **“FrozenHot Cache: Rethinking Cache Management for Modern Hardware.”** *The European Conference on Computer Systems*.
SOCC’23	Yazhuo Zhang, Rebecca Isaacs, Yao Yue, Juncheng Yang, Lei Zhang, Ymir Vigfusson. **“Latenseer: Causal Modeling of End-to-End Latency Distributions by Harnessing Distributed Tracing.”** *ACM Symposium on Cloud Computing*.
VLDB’23	Tianyu Zhang, Kaige Liu, Jack Kosaian, Juncheng Yang, K. V. Rashmi. **“Efficient Fault Tolerance for Recommendation Model Training via Erasure Coding.”** *49th International Conference on Very Large Database*.
SOSP’21	Sara McAllister, Benjamin Berg, Julian Tutuncu-Macias, Juncheng Yang, Sathya Gunasekar, Jimmy Lu, Nathan Beckmann, Gregory R. Ganger. **“Kangaroo: Caching Billions of Tiny Objects on Flash.”** *28th ACM Symposium on Operating Systems Principles*. **Best Paper Award, invited fast-track to TOS’22**.
OSDI’20	Saurabh Kadekodi, Francisco Maturana, Suhas Jayaram Subramanya, Juncheng Yang, K. V. Rashmi, Gregory R. Ganger. **“PACEMAKER: Avoiding HeART Attacks in Storage Clusters with Disk-adaptive Redundancy.”** *14th USENIX Symposium on Operating Systems Design and Implementation*.

SOCC'18 Hobin Yoon, Juncheng Yang, Sveinn Fannar Kristjansson, Steinn E. Sigurdarson, Ymir Vigfusson, Ada Gavrilovska. **"Mutant: Balancing Storage Cost and Latency in LSM-Tree Data Stores."** *ACM Symposium on Cloud Computing*.

ICDE'18 Jinfei Liu, Juncheng Yang, Li Xiong, Jian Pei, Jun Luo. **"Skyline Diagram: Finding the Voronoi Counterpart for Skyline Queries."** *IEEE International Conference on Data Engineering*.

ICDE'17 Jinfei Liu, Juncheng Yang, Li Xiong, Jian Pei. **"Secure Skyline Queries on Cloud Platform."** *IEEE International Conference on Data Engineering*.

SYSTOR'16 Helgi Sigurbjarnarson, Petur Orri Ragnarsson, Juncheng Yang, Ymir Vigfusson, Mahesh Balakrishnan. **"Enabling Space Elasticity in Storage Systems."** *ACM International Systems and Storage Conference*. Best Student Paper Award.

Invited Talk

1. FIFO queues are all you need for cache eviction.
    - VMware, 2023
    - Alluxio, 2023
    - Microsoft Research Asia, 2023
    - Kuaishou, 2023
    - University of Science and Technology of China, 2023
    - Tsinghua University, 2023
2. LESSCache: LEarned Segment-Structured cache.
    - Meta, 2023
    - VMware, 2022
3. Ubiquitous caching: building efficient distributed and in-process caching. *QCon SF*, 2022.
4. Segcache: a memory-efficient and high-throughput DRAM cache for small objects.
    - Oracle, 2023
    - Alluxio, 2022
    - UMich seminar, 2021
5. Caching on PMEM: an iterative approach. *SNIA SDC keynote talk*, 2020.

Selected Honors & Awards

2023	**Machine Learning and System Rising Star**	
2023	**Google Cloud Research Innovator**	
2021	**SOSP'21 Best Paper Award**	
2021	**NSDI'21 Community (Best Paper) Award**	
2016	**SYSTOR'16 Best Student Paper**	
2013	**Emerson Fellowship** The only one in the department.	*Emory University*
2013	**Best Thesis Award** 5/3000 in the university, 1/200 in the department.	*Nanjing University*
2012	**"Person of the Year" Nomination** 100 nominations among all Chinese undergraduates.	
2008	**First Award in National Chemistry Olympiad**	

Funding and grants

2023	**Google Cloud Innovator grant** $10,000
2020-2022	**Meta Fellowship**
2018	**AWS research grant** $10,000

Service & Activities

2023	**Reviewer** IEEE Access
2023	**Reviewer** ACM Transactions on Storage (TOS)
2023	**Artifact Reviewer** Sixth Conference on Machine Learning and Systems (mlsys'23)
2022	**Reviewer** Transactions on Cloud Computing (TCC)
2023	**Artifact Evaluation Reviewer** Journal of Systems Research (JSys)
2019	**Reviewer** Transactions on Parallel and Distributed Systems (TPDS)
2018	**Shadow PC** Eurosys'18
2016, 2017	**External Reviewer** ACM Symposium on Cloud Computing (SOCC'16, SOCC'17)

Teaching Experience

2022	**Guest lecturer** 15612 Intro to Computer System	*Carnegie Mellon University*
2022	**Teaching assistant** 15712 Advanced and Distributed Operating Systems	*Carnegie Mellon University*
2020	**Teaching assistant** 15746 Storage Systems	*Carnegie Mellon University*
2017	**Guest lecturer** CS584 Advanced Computer System	*Emory University*
2017	**Teaching assistant** CS453 Computer Security	*Emory University*
2013, 2014	**Lab instructor** General Chemistry I and II	*Emory University*
2012	**Teaching assistant** Modern Website Programming	*Nanjing University*

Mentees

2021-2023	Jonathan Chiu (CMU undergraduate)
2022	Ziming Mao (Yale undergraduate, UC Berkeley Ph.D.)
2022-2023	Yazhuo Zhang (Emory Ph.D.)
2022-2023	Ziyue Qiu (CMU Ph.D.)
2023	Bob Chen (CMU undergraduate)
2023	Frank Chen (CMU undergraduate)
2023	Emily Zhang (CMU undergraduate)
2023	Yiyan Zhai (CMU undergraduate)
2023	Parinay Chauhan (IIT undergraduate)

Work Experience

Intern @ Twitter

JVM off-heap caching for reduced memory footprint and more predictable service, manager: Yao Yue *May 2022 - July 2022*

- Worked with ads ranking team to better understand how local JVM cache can reduce ads serving latency and why the size of JVM caches bottlenecks existing service.
- Explored options for enabling large caches for JVM-based services and chose to build an off-heap JVM cache library.
- Designed and built JSegcache — a Java library that uses JNI on top of Segcache (rust-based). I further explored several optimizations to improve JSegcache throughput and scalability.

Software Engineer Intern @ Cloudflare

Content delivery performance analysis and origin egress reduction, manager: Aki Shugaeva *June 2021 - Aug 2021*

- Analyzed the traffic of different search engine crawlers and showed that 1) they frequently crawl unchanged content, 2) close to 80% of newly published content takes more than one day to be crawled and indexed. This analysis motivated a cross-functional project between Cloudflare and Bing/Apple/Yandex/Baidu for efficient crawling and faster indexing. In addition, designed signals for discovering new and updated content on the edge.
- Designed an algorithm that discovers cacheable content in dynamic traffic (HTTP responses that were not cached). Deployed the detector in Kubernetes and made several discoveries that reduced close to 100 Gbps egress bandwidth for customers.
- Analyzed the effectiveness and performance of content delivery cache in over 200 edge clusters using data from (1) Clickhouse (SQL), (2) Thanos/Prometheus, and (3) error logs printed from Nginx. In addition, added more logging and tests to Nginx and deployed them to production to gain further insights.

Researcher @ Twitter

MEMORY-EFFICIENT AND HIGH-THROUGHPUT IN-MEMORY CACHING, MANAGER: YAO YUE *Feb 2020 - Nov 2020*

- Helped investigate cache-related sev incidents caused by client timeouts and cache out-of-memory (OOM).
- Built a pipeline to collect and process 100s TB logs directly from production in-memory caches (Twemcache and customized Redis) clusters and stored the logs in Hadoop HDFS.
- Wrote scripts in C/C++ and Python to analyze the collected logs on 10s of physical nodes in a distributed fashion. Published several insights about building better in-memory caching systems (OSDI'20).
- Designed a storage component for in-memory cache (Segcache), which reduces the DRAM usage of in-memory caches at Twitter by 60% with slightly higher single-core throughput. In addition, Segcache archives close-to-linear core scalability.
- Benchmarked NIC performance on cache workloads and showed that the Intel E800 series NIC with Application Device Queue (ADQ) reduces tail latency by 90%, and helped with Intel-Twitter product launch co-advertisement.

Software Engineer @ Emory Center for Digital Scholarship (ECDS)

ATLANTA EXPLORER, MANAGER: MICHAEL PAGE *Sept 2015 - Dec 2016*

- Collaborated on building a 3D model and visualization tool for exploring historic Atlanta from 1880-1930.
- Proposed and developed a novel workflow for information extraction from old city directories into a geo-database.
- Deployed an LSTM-based OCR engine and developed software for potential recognition error crowd-sourcing and LSTM model training sample production.