

Juncheng Yang

☎ (+1) 404-285-5231 | ✉ juncheny@cs.cmu.edu | 🏠 <http://jasony.me> | 📄 github.com/1a1a11a

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

Education

Ph.D. in Computer Science (advisor: Rashmi Vinayak)

CARNEGIE MELLON UNIVERSITY, COMPUTER SCIENCE DEPARTMENT

Pittsburgh, U.S.A

Aug. 2018 - Present

M.S. in Computer Science (advisor: Ymir Vigfusson)

EMORY UNIVERSITY, DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

Atlanta, U.S.A

Jan. 2015 - Dec. 2016

M.S. in Chemistry (advisor: Craig L. Hill)

EMORY UNIVERSITY, DEPARTMENT OF CHEMISTRY

Atlanta, U.S.A

Aug. 2013 - Jun. 2015

B.S. in Chemistry (advisor: Ying Wang)

NANJING UNIVERSITY, DEPARTMENT OF CHEMISTRY AND CHEMICAL ENGINEERING (TOP 2 IN CHINA)

Nanjing, China

Sept. 2009 - Jun. 2013

Publication and Presentation

UNDER SUBMISSION

1. Juncheng Yang, Anirudh Sabnis, Daniel S. Berger, K. V. Rashmi, Ramesh Sitaraman. “C2DN: How to Code on the Edge for Content Delivery.” *submitted to USENIX Symposium on Networked System Design and Implementation (NSDI’20)*.
2. Saurabh Kadekodi, Francisco Maturana, K. V. Rashmi, Gregory Ganger, Juncheng Yang, Suhas Jayaram Subramanya. “Pacer: eliminating transition overload for device-adaptive redundancy.” *submitted to USENIX Conference on File and Storage Technologies (FAST’20)*.

PEER REVIEWED PUBLICATIONS

1. 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 (SOCC), 2018*.
2. Jinfei Liu, Juncheng Yang, Li Xiong, Jian Pei. “Secure and Efficient Skyline Queries on Encrypted Data.” *IEEE Transactions on Knowledge and Data Engineering (TKDE), 2018*.
3. 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), 2018*.
4. Juncheng Yang, Reza Karimi, Trausti Saemundsson, Avani Wildani, Ymir Vigfusson. “MITHRIL Mining Sporadic Associations for Cache Prefetching.” *ACM Symposium on Cloud Computing (SOCC), 2017*.
5. Jinfei Liu, Juncheng Yang, Li Xiong, Jian Pei. “Secure Skyline Queries on Cloud Platform.” *IEEE International Conference on Data Engineering (ICDE), 2017*.
6. Helgi Sigurbjarnarson, Petur Orri Ragnarsson, Juncheng Yang, Ymir Vigfusson, Mahesh Balakrishnan. “Enabling Space Elasticity in Storage Systems.” *ACM International Systems and Storage Conference (SYSTOR), 2016. (Best student paper)*.

PRESENTATION AND TALK (PAPER TALK EXCLUDED)

1. Juncheng Yang, Reza Karimi, Avani Wildani, Ymir Vigfusson, “A Simple Cache Prefetching Layer Based on Block Correlation”. *Usenix Conference on File and Storage Technologies (FAST), 2017*. (10 min WiP talk and Poster)
2. Juncheng Yang, Reza Karimi, Ymir Vigfusson, “Mithril: Mining Block Correlation for Cache Prefetching”. *Usenix Symposium on Operating System Design and Implementation (OSDI), 2016*. (Poster)

Research Experience

RESEARCH IN CACHING

C2DN: How to Code on the Edge for Content Delivery *submitted to NSDI'20*

- Illustrated unavailability in CDN edge clusters are common and redundancy is necessary for high availability and stringent SLO requirement.
- Identified the opportunity and the potential challenges of using erasure coding in CDN edge clusters, demonstrated in both theory and simulation that the benefits outweigh the challenges.
- Designed Coded CDN (C2DN) and built a prototype in Golang on top of Apache traffic server.
- Setup experiment on AWS and evaluated using two traces collected from two Akamai edge clusters.
- Compared to state-of-the-art CDN, C2DN reduces midgress bandwidth by up to 30%, while providing better tail latency because of better load balancing.

MITHRIL: Mining Block IO Associations for Cache Prefetching *SOCC'17*

- Proposed a general lightweight history-based cache prefetching algorithm that effectively discovers associations between blocks/objects in modern caching workloads.
- Implemented MITHRIL in C and demonstrated that it provides up to seven times hit ratio improvement over LRU and state-of-the-art prefetching algorithms.
- Analyzed and proved the source of good performance of MITHRIL - hit ratio improvement on mid-frequency blocks/objects.

RESEARCH IN STORAGE SYSTEMS AND REDUNDANCY

Pacer: eliminating transition overload for device-adaptive redundancy *submitted to FAST'20*

- This project proposed techniques in reducing overheads of tuning redundancy in large scale storage systems and make redundancy adaption based on disk AFR practical.
- Worked on cross disk group stripe MTDDL analysis.

Enabling Space Elasticity in Storage Systems (Best student paper) *SYSTOR'16*

- This project proposed motifs abstraction for file system to enable storage elasticity by allowing applications to describe how soft state can be regenerated.
- Worked on running experiments written in C++ with thrift for RPC and FUSE for user-space filesystem.
- Diagnosed and identified problem about latency spikes during file regeneration due to unnecessary disk reads, wrote a simple in-memory server to fix the problem.

RESEARCH IN DATABASES AND DATA MANAGEMENT

Mutant: Balancing Storage Cost and Latency in the Cloud *SOCC'18*

- The project proposed and designed mutant, a layer for LSM-tree based database to achieve balance between cost and latency by exploiting temporal locality in query.
- Conducted experiments on RocksDB with YCSB workload and real-world *QuizUp* workload to show the effectiveness of mutant.

Skyline Diagram: Finding the Voronoi Counterpart for Skyline Queries *ICDE'18*

- This project defined a novel structure, skyline diagram enabling fast skyline query after pre-computation.
- Designed and implemented all experiments and proposed parallel algorithms for fast computation.

Secure Skyline Queries on Cloud Platform *ICDE'17*

- This project proposed a novel Paillier-based fully secure dominance protocol with no information leakage that can be used as a building block for constructing encrypted database query.
- Designed and implemented a simulation system based on the proposed protocol in C, parallelized the computation using POSIX threads and obtained a sub-linear parallel performance.
- Designed and implemented a distributed computation protocol for even faster computation.
- Proposed two optimizations, optimal partitioning and lazy merge, and proved the correctness using both theory and experiments.

Work Experience

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 geo-database.
- Deployed a LSTM based OCR engine and developed software for potential recognition error crowd-sourcing and LSTM model training sample production.

Selected Honors & Awards

| | | |
|------|------------------------------------------------------------------------------------------|---------------------------|
| 2018 | AWS Research Grant | |
| 2017 | SOCC'17 travel grant | <i>Santa Clara</i> |
| 2016 | SYSTOR'16 Best Student Paper | <i>Haifa, Israel</i> |
| 2013 | Emerson Fellowship The only one in the department. | <i>Emory University</i> |
| 2013 | Best Thesis Award 5/3000 in the university, 1/200 in the department. | <i>Nanjing University</i> |
| 2012 | "Person of the Year" Nomination 100 nominations among all Chinese undergraduates. | <i>China</i> |
| 2012 | Third Place Green Tech International Competition. | <i>Taiwan, China</i> |
| 2012 | Academic Excellence Award the 5th National Undergraduate Innovation Forum. | <i>Beijing, China</i> |
| 2008 | First Award in National Chemistry Olympiad | |

Open Source Contributions

mimircache a Python Platform for Cache Performance Analysis, released under GPLv3

CORE DEVELOPER

Mar. 2016 - Present

- Allow developers to analyze cache performance using traces efficiently in Python with intensive computation in C back-end.
- Support visualization of different cache replacement algorithms and cache time-varying behavior.
- Used by *CloudPhysics Inc.*, *Akamai* and students from *Stony Brook* and *CMU*.

Service & Activities

| | |
|-----------|--------------------------------------------------------------------------|
| 2019 | Reviewer Transactions on Parallel and Distributed Systems (TPDS) |
| 2018 | Shadow PC Eurosys'18 |
| 2017 | External Reviewer ACM Symposium on Cloud Computing (SoCC'17) |
| 2016 | External Reviewer ACM Symposium on Cloud Computing (SoCC'16) |
| 2013-2015 | Project Manager Chinese Students & Scholars Union at Emory (CSUE) |
| 2011-2013 | Co-Founder Technical Support Volunteer Team in Nanjing University |

Selected Courses

- Advanced OS and Distributed Systems, System Programming, Operating System, Advanced Computer System, Advanced Database System, Computer Security
- Practical information and coding theory for computer systems
- Data Mining, Machine Learning, Artificial Intelligence, Algorithms, Theory of Computing
- Natural Language Processing, Data Privacy and Security

Teaching Experience

| | | |
|------|------------------------------------------------------|---------------------------|
| 2017 | Guest lecturer CS584 Advanced Computer System | <i>Emory University</i> |
| 2017 | Teaching assistant CS453 Computer Security | <i>Emory University</i> |
| 2014 | Lab instructor Chem142 General Chemistry II | <i>Emory University</i> |
| 2013 | Lab instructor Chem141 General Chemistry I | <i>Emory University</i> |
| 2012 | Teaching assistant Modern Website Programming | <i>Nanjing University</i> |