Juncheng **Yang**

(+1) 404-285-5231 | jasonyang@cmu.edu | http://jasony.me | github.com/1a1a11a

"Learn something about everything, learn everything about something."

Education

Ph.D. in Computer Science — *Pittsburgh, U.S.A*
Carnegie Mellon University, Department of Computer Science — *Aug. 2018 - Present*

Emory University, Research Assistant of Prof. Ymir Vigfusson — *Atlanta, U.S.A*
Department of Mathematics and Computer Science, Simbiosys Lab — *Jan. 2017 - PRESENT*

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

M.S. in Chemistry — *Atlanta, U.S.A*
Emory University, Department of Chemistry — *Aug. 2013 - Dec. 2014*

B.S. in Chemistry — *Nanjing, China*
Nanjing University, Department of Chemistry and Chemical Engineering (Top 2 in China) — *Sept. 2009 - Jun. 2013*

Research Experience

Research in Smart Cache

The Missing Dimension in the Cache — *Emory University*
Advisor: Prof. Ymir Vigfusson — *Aug. 2017 - Present*

Work in Progress about cache partitioning. Collaboration with CachePhysics Inc.

Dynamic Cache Boundary Resizing — *Emory University*
Advisor: Prof. Ymir Vigfusson — *Jun. 2017 - Present*

Work in Progress about boundary re-sizing in cache hierarchy. Collaboration with Akamai.

Mithril: Mining Block IO Associations for Cache Prefetching ***SoCC'17*** — *Emory University*
Advisor: Prof. Ymir Vigfusson — *May. 2016 - Feb. 2017*

- Proposed a general lightweight history-based cache prefetching algorithm that effectively discovers associations between blocks/objects in modern 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 good performance of Mithril comes from capture of mid-frequency blocks.

Research in Storage Systems

Enabling Space Elasticity in Storage Systems ***SYSTOR'16*** — *Emory University*
Advisor: Prof. Ymir Vigfusson, Collaborators: Helgi Sigurbjarnarson (University of Washington), Petur O. Ragnarsson (Reykjavik University) — *Feb. 2016 - May. 2016*

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

Research in Databases

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

Advisor: Prof. Ymir Vigfusson, Collaborators: Hobin Yoon (Georgia Tech) *May. 2017 - Present*

- 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.
- I conducted experiments on RocksDB with YCSB workload and real-world *QuizUp* workload to show the effectiveness of mutant.

Research in Database and Data Management

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

Advisor: Prof. Li Xiong, Collaborators: Jinfei Liu (Emory University & Georgia Tech) *May. 2016 - Dec. 2016*

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

Secure Skyline Queries on Cloud Platform ***ICDE'17*** *Emory University*

Advisor: Prof. Li Xiong, Collaborators: Jinfei Liu (Emory University & Georgia Tech) *Dec. 2015 - May. 2016*

- 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.
- I 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.
- I designed and implemented a distributed computation protocol for even faster computation.
- I proposed two optimizations, optimal partitioning and lazy merge, and proved the correctness using both theory and experiments.

Publication and Presentation

Peer Reviewed Publications

1. Juncheng Yang, Reza Karimi, Trausti Saemundsson, Avani Wildani, Ymir Vigfusson."Mithril Mining Sporadic Associations for Cache Prefetching." *ACM Symposium on Cloud Computing (SoCC)*, **2017**.
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. Jinfei Liu, Juncheng Yang, Li Xiong, Jian Pei. "Secure Skyline Queries on Cloud Platform." *IEEE International Conference on Data Engineering (ICDE)*, **2017**.
5. 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**).

Paper under Review Or In Progress

6. Hobin Yoon, Juncheng Yang, Ymir Vigfusson, Ada Gavrilovska. "Mutant: Balance Cost and Latency in the Cloud." *going to submit to Very Large Data Bases (VLDB)*.

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)

Work Experience

Software Engineer @ Emory Center for Digital Scholarship (ECDS) *Emory University*

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 under client-server model for potential recognition error crowd-sourcing and LSTM model training sample production.
- Developed an early error-detection mechanism enabling highly accurate OCR result.

Open Source Contributions

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

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.
- Packaged into docker container and used by *CloudPhysics Inc.*, *Akamai* and students from *Stony Brook* and *CMU*.

Selected Courses

- System Programming, Operating System, Advanced Computer System, Advanced Database System, Computer Security
- Computer Networking (self-study), Computer Architecture (self-study)
- 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	*Emory University*
2017	**Teaching assistant** CS453 Computer Security	*Emory University*
2014	**Lab instructor** Chem142 General Chemistry II	*Emory University*
2013	**Lab instructor** Chem141 General Chemistry I	*Emory University*
2012	**Teaching assistant** Modern Website Programming	*Nanjing University*

Selected Honors & Awards

2018-2019	**AWS research grant ($15000)**	*Emory University*
2017	**SOCC'17 travel grant ($800)**	*Santa Clara*
2013-2014	**Emerson Fellowship ($4000)** The only one in the department.	*Emory University*
2013	**Best Thesis Award** 5/3000 in the university, 1/200 in the department.	*Nanjing University*
2013	**Outstanding Award** "Challenge Cup" Innovated Science and Technology Competition.	*Nanjing, China*
2012	**"Person of the Year" Nomination** 100 nominations among all Chinese undergraduates.	*China*
2012	**Third Place** Green Tech International Competition.	*Taiwan*
2012	**Academic Excellence Award** the 5th National Undergraduate Innovation Form.	*Beijing, China*
2010 & 2011	**Technological Innovation Award**	*Nanjing University*
2010 & 2011	**Outstanding Student Award**	*Nanjing University*
2010	**Outstanding Leadership Award** For acting as league branch secretary.	*Nanjing University*
2008	**First Award in National Chemistry Olympiad**	

Service & Activities

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)	*Emory University*
2011-2013	**Co-Founder** Technical Support Volunteer Team	*Nanjing University*

Projects

Operating System

Emory University

HOCA, a Kernel for Motorola 6800 Processor — *Jan. 2016 - May. 2016*

- Wrote a micro kernel for Motorola 6800 processor, including process scheduling, memory management.
- Wrote round-robin for process scheduling, SCAN(elevator) algorithm for disk scheduling.
- Used segmentation and paging in memory management to allow logical separation of memory segments and resolve fragmentation.
- Wrote virtual memory to enable memory protection and sharing, used floppy as swap device and clock algorithm as replacement algorithm.

Natural Language Processing (NLP)

Emory University

Keyphrases Extraction and Fingerprints Construction from Scientific Documents — *Sep. 2015 - Dec. 2015*

- Modified Stanford and Berkeley dependency parser to extract the noun phrases (NP) from scientific papers.
- Generated sorted and pruned candidate list of keyphrases by ranking extracted NPs using TF-IDF.
- Used ranked keyphrases as document fingerprint for measuring document similarity based on Word2Vec, which shows great improvement compared with using TF-IDF on keyphrases or using Word2Vec on full text.

Reference Finder

Find reference info for research papers, http://findref.xyz — *May. 2015 - Sept. 2015*

- Built a web service using Flask giving researchers important information about references in a paper.
- Designed the workflow with a manager server receiving and analyzing PDF files, then using remote procedure calls(RPC) to have worker servers conduct real-time query and crawling.
- Analyzed pdf files and extracted references from given any research paper in Java.
- Crawled paper information from publisher and stored in database (SQLite and MySQL).
- Will build our own research paper database for searching instead of querying search engine, and use machine learning to analyze reference string.

Find My Advisor

A Search Engine about Professors — *May. 2015 - Apr. 2016*

- Wrote a web crawler using scrapy framework to crawl professors' information.
- Designed strategies for efficient locating homepages of different departments beginning from university main webpage, uncovered faculty page with the help of APIs from *import.io*, then extracted professor information.
- Will use crawled data to provide a search engine for searching professors.

Social Network Analysis

Emory University

A Study of Relationship between Students/Faculty Involved in Sustainability — *Feb. 2015 - Oct. 2015*

- Cooperated with graduate sustainability group for efficient data collection in the survey.
- Cleaned and integrated data from different sources, built a bipartite graph matching unrecognized names and known names, then use edit distance to evaluate the similarity between names and a priority queue to link the most possible matching.
- Used graph theory to explore the relationships between people and how they were connected, based on this, discovered the most important factor in connecting different people and provided suggestions for better cooperation between people.
- Analyzed dissimilarity and similarity (like Euclidean distance and cosine similarity) between different people and predicted new friendship and new trends in the social network.