Juncheng Yang

http://jasony.me

Email : peter.waynechina@gmail.com

Education

- **Ph.D., Carnegie Mellon University** — Pittsburgh, PA, Aug. 2018 – Present
- **M.S., Emory University** — Atlanta, GA, Aug. 2015 – Dec. 2016
- **B.S., Nanjing University** — Nanjing, China, Aug. 2009 – Dec. 2013

Skills

- **Languages**: Python, C/C++, Go, Bash, Javascript, Java
- **Primary area**: Caching (CDN and in-memory), storage system, distributed system, networking
- **Secondary area** :data mining and data management, machine learning, deep learning (tensorflow)
- **Other** : version control (git), CI (travis), SQL, AWS/Azure, large system (≥200 nodes) deployment experience, distributed web crawling (≥1M RPS)

Selected Projects

- **Erasure Coding on the Edge for Content Delivery, submitted to NSDI'20** — Pittsburgh, PA
  Research in CDN caching — *Sept 2018 - Nov 2019*
  - Identified the opportunity and the potential challenges of using erasure coding in current CDN edge clusters, demonstrated in both theory and simulation that the benefits outweigh the drawbacks.
  - Designed Coded CDN (C2DN), built a simulator in C++ and built a prototype in Golang on top of Apache traffic server.
  - Setup multiple 10-node edge clusters on AWS and evaluated using two Akamai edge-cluster workloads.
  - 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 I/O Associations for Cache Prefetching, SOCC'17** — Atlanta, GA
  Research in caching — *Apr 2016 - Sept 2017*
  - 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. As a result, microbenchmark showed an 26% reduction in latency with moderate CPU usage increase.
  - Analyzed and proved the source of good performance - hit ratio improvement on mid-frequency blocks/objects.

Other Projects

- **Pacer**: Designed an adaptive storage redundancy scheme and eliminated transition overload.
- **Enabling Space Elasticity in Storage Systems**: Designed a new storage system with elasticity by storing unused file remotely, best student paper in SYSTOR'16.
- **Mutant: Balancing Storage Cost and Latency in the Cloud**: Introduced dynamic SSTable movement between different storage devices (HDD/SSD), built on top of RocksDB, published in SOCC'18.
- **Skyline Diagram: Finding the Voronoi Counterpart for Skyline Queries**: Defined a novel structure, skyline diagram enabling fast skyline query after pre-computation, published in ICDE'18.
- **Secure Skyline Queries on Cloud Platform**: Proposed a novel Paillier-based fully secure dominance protocol for constructing encrypted database query, built an emulation system in C, published in ICDE'18.
- **NLP**: Extracted keyphrases constructed fingerprints for scientific documents.
- **Social Network Analysis**: Collected, cleaned and integrated data from different sources, used graph theory to explore the relationships between groups and predicted potential relationship using similarity scores.

Experience

- **Emory Center for Digital Scholarship (ECDS)** — Atlanta, GA
  Software Engineer (Part time) — *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, developed software for crowd-sourcing recognition error and training sample production.
- **mimircache - a Python Platform for Cache Performance Analysis, released under GPLv3**
  Core developer
  - Allow users to analyze cache performance using traces efficiently in Python with intensive computation in C backend.
  - Support comparing different cache replacement algorithms and visualizing time-varying caching behavior.
  - Used by *CloudPhysics Inc.*, *Akamai* and students from *Emory University*, *Stony Brook University*, *UMass* and *CMU*.