Cai, Yifan

Email: fyc1007261@live.com | Homepage: http://www.yifancai.tech

EDUCATION

Shanghai Jiao Tong University — Shanghai
B.S. Software Engineering; System Software Track — Expected in June 2020

- Overall GPA: 3.77 / 4.30 (88.19/100); Major GPA: 3.88 / 4.30 (89.95/100)
- Honors/Awards: SJTU Scholarships (2016-2017 & 2017-2018), Huawei Scholarships (2017-2018)
- Selected Courses: Operating Systems (4.0), Distributed Systems (4.0), Computer Architecture (4.0), etc.

PUBLICATIONS

- **Rethinking Data Management Systems for Disaggregated Data Centers. [Under Review]**
  Qizhen Zhang, **Yifan Cai**, Sebastian Angel, Ang Chen, Vincent Liu and Boon Thau Loo.
  Submitted for *Conference on Innovative Data Systems Research (CIDR), 2020.*

- **Intra-day Forecast of Ground Horizontal Irradiance Using Long Short-Term Memory Network (LSTM). [Under Review]**
  Xianglei Huang, Xiuhong Chen, **Yifan Cai**, Haoming Shen, Jiayue Lu
  Submitted for *IEEE Transactions on Sustainable Energy, 2019*

- **Consensus-based Data Statistics in Distributed Network Systems. [Paper]**
  Yifan Cai, Jianping He, Wenbin Yu, and Xinping Guan.
  In the *57th IEEE Conference on Decision and Control, Miami, USA, Dec 2018.*

EXPERIENCE

University of Pennsylvania — PA, US
Research Intern — Jul 2019 – Dec 2019 (Expected)

- Took a first step towards understanding how disaggregated data centers might affect the design of relational databases, discussed the potential advantages and drawbacks in the context of data processing, and outlined research challenges in addressing them.
- Modified the kernel of LegoOS (A disaggregated operating system published in OSDI'18) and deployed it on cloud machines; wrote code of nested loop join, hash join as well as grace hash join and tested their performance on disaggregated operating systems.
- Designed benchmarks and discovered where the performance bottleneck of disaggregated systems is.

University of Michigan — MI, US
Research Intern — Aug 2018 – Sep 2018

- Built an LSTM-based solar forecasting system on keras deep learning library.
- Completed the predicting model, and the error of which was approximately only 10%, smaller than all of those in published papers in some specific conditions.

Lab of System Control and Information Processing — Shanghai, China
Research Assistant — Sep 2017 – Jun 2019

- Design a protocol in distributed network systems to do statistics cooperatively. It is the first protocol proposed to compute the probability density function in a fully distributed way.
- Responsible for algorithm design and optimization, numerical simulations and paper writing.

PROJECTS

Tiger Language Compiler Sep 2018 – Jan 2019

- Developed a compiler for *Tiger* language which can generate assembly code.
- The compiler follows x86-64 calling conventions. Optimizations such as liveness analysis and register allocations are realized.

Distributed File System Sep 2018 – Jan 2019

- Developed a distributed file system which supports basic Linux interfaces (touch, move, write, unlink, etc.).
- The file system works under single server and multiple clients, where distributed locks with local cache are implemented to ensure the atomicity and the lock performance. Replication protocol is also designed and implemented in case of random failures.

Smart Electric Appliance Monitoring and Scheduling System Jun 2018 – Jul 2018

- Collaborated with team members to build a smart electric appliance monitoring and scheduling system.
- Focused on the back-end Java development of the web application, as well as the integrated hardware controlling system written in Python, which collects information from all the sensors and operates the appliances.
- Managed to integrate a series of cross-process communications, including smart scheduling for appliances such as renewable energy simulating, deep learning based solar energy forecasting, and convenient human computer interactions.
- Recommended to the SJTU Software Exhibition, 2018.

B+ Tree Based Database Jul 2017 – Aug 2017

- Developed a B-plus-tree-based database system with a cache system between disk and memory, which supports CRUD operations at a speed of a million operations per second when running with over a million sets of data.

SKILLS

Languages: Chinese (Native); English (TOFEL 111).
Programming Languages: C/C++, Python, Java, JavaScript, SQL