Cai, Yifan

Email: caiyifan@seas.upenn.edu | Homepage: http://www.yifancai.tech/ | Tel: (484) 957-0135

EDUCATION

University of Pennsylvania Philadelphia, PA
Ph.D. Student in Computer and Information Science Sep 2021 - Present
Advisor: Dr. Linh Thi Xuan Phan
Research Interest: Distributed systems and real-time systems

Shanghai Jiao Tong University Shanghai, China
Bachelor of Engineering, Software Engineering, System Software Track Sep 2016 – Jun 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 & 2018-2019)

PUBLICATIONS

Analysis of Long-term Average Behaviors of Probabilistic Task Systems
Yifan Cai, Linh Thi Xuan Phan, and P.S. Thiagarajan.
To appear at *International Conference on Real-Time Networks and Systems (RTNS), 2024*

Object-oriented Unified Encrypted Memory Management for Heterogeneous Memory Architectures
Mo Sha*, **Yifan Cai***, Sheng Wang, Linh Thi Xuan Phan, Feifei Li, and Kian-Lee Tan (* Equal contribution)
In *International Conference on Management of Data (SIGMOD), 2024*

VeriDB: An SGX-based Verifiable Database.
Wenchao Zhou, **Yifan Cai**, Yanqing Peng, Sheng Wang, Ke Ma, and Feifei Li
In *International Conference on Management of Data (SIGMOD), 2021*

Understanding the Effect of Data Center Resource Disaggregation on Production DBMSs.
Qizhen Zhang, **Yifan Cai**, Xinyi Chen, Sebastian Angel, Ang Chen, Vincent Liu, and Boon Thau Loo.
In *46th International Conference on Very Large Data Bases (VLDB), 2020*.

Intra-day Forecast of Ground Horizontal Irradiance Using Long Short-Term Memory Network (LSTM)
Xiuhong Chen, Xianglei Huang, **Yifan Cai**, Haomin Shen, and Jiayue Lu.
In *Journal of the Meteorological Society of Japan*. *Ser. II, 2020, 98(5): 945-957.*

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

Consensus-based Data Statistics in Distributed Network Systems.
Yifan Cai, Jianping He, Wenbin Yu, and Xinping Guan.
In the *57th IEEE Conference on Decision and Control (CDC), 2018*

EXPERIENCES AND PROJECTS

Schedule-based Timing Side-channel Attacks in Real-time IoT devices University of Pennsylvania
Ph.D. Project, *with Dr. Linh Thi Xuan Phan* Nov 2023 – Present
- Designing and implementing an attack that leverages the deterministic nature of real-time systems to infer the timing information of an IoT task.

Fault Detection and Recovery in Geo-distributed Systems University of Pennsylvania
Ph.D. Project, *with Dr. Linh Thi Xuan Phan* Sep 2021 – Present

- Designing protocols for distributed systems that can detect, account for, and recover from Byzantine faults in real time, despite the unreliable nature of the network in geo-distributed systems.

Data Security in Systems with Heterogeneous Memory Alibaba Cloud
Research Intern, *supervised by Dr. Mo Sha and Dr. Sheng Wang* May 2023 – Oct 2023

- Proposed a solution to ensure confidentiality, integrity, and freshness of data in a system with heterogeneous memory architectures, while providing a unified interface to the applications**.**
- Used centralized data structures to ensure data security, achieving up to 6.3x speedup compared to traditional security techniques (e.g., AEAD).

Schedulability Analysis for Probabilistic Real-time Systems University of Pennsylvania
Ph.D. Project, *with Dr. Linh Thi Xuan Phan and Dr. P.S. Thiagarajan* Dec 2022 – Oct 2023

- Proposed a Markov chain-based model to analyze the long-term average of important properties, such as deadline miss ratio and the weakly-hard constraint violation probability of tasks in a real-time system.
- Provided the mathematical basis for the sampling method, which scales well in large task systems.

Data and Execution Integrity for Cloud Database Systems Alibaba Cloud
Research Intern, supervised by Dr. Sheng Wang and Dr. Wenchao Zhou Sep 2020 – May 2021

- Worked on designing and implementing a cloud database providing data and execution integrity that can be verified by the users.
- Improved the performance of the verification process by introducing multiple verifiers and a mechanism to assign workloads to different verifiers. The improvement therefore allowed multiple transactions to be executed concurrently.
- Implemented database operators such as joins and scans with the additional steps needed for verifications, with which the database supported relational queries rather than only key-value lookup.

Database with Resource Disaggregation University of Pennsylvania
Research Intern, *Advisor: Dr. Vincent Liu* Jul 2019 – Dec 2019

- Conducted research on the impact of disaggregated data centers on the design of relational databases.
- Developed microbenchmarks of nested loop join, hash join, and grace hash join operators and investigated the relationship between local memory size, the number of remote memory access and the performance degradation of each operation.
- Added features such as relative paths to LegoOS (a disaggregated operating system) by adding more system calls in order to deploy complex systems such as PostgreSQL and MonetDB on it.
- Analyzed the performance drawbacks of existing disaggregated operating systems for query executions and outlined potential solutions.

Data Aggregation in Distributed Systems Shanghai Jiao Tong University
Research Assistant, *Advisor: Dr. Jianping He* Sep 2017 – Jun 2019

- Invented a protocol in distributed network systems to compute probability density functions in a fully distributed way and enable multiple compute nodes to share their statistics.
- Designed and optimized two probability distribution functions (PDF) to calculate algorithms which run under both public and anonymous network conditions.

SKILLS

Programming Languages:

Most Familiar: C/C++ and Python
Other: Java, JavaScript, SQL