Cai, Yifan

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

EDUCATION

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)
- Huawei Scholarships (2017-2018)

University of Pennsylvania — Philadelphia, PA
Ph.D. Student in Computer and Information Science — Sep 2021 - Present

Advisor: Dr. Linh Thi Xuan Phan

Research Interest: System security and real-time systems

PUBLICATIONS

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

Understanding the Effect of Data Center Resource Disaggregation on Production DBMSs. [Paper]
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, Jiayue Lu.
In *Journal of the Meteorological Society of Japan*, 2020.

Rethinking Data Management Systems for Disaggregated Data Centers. [Paper]
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. [Paper]
Yifan Cai, Jianping He, Wenbin Yu, and Xinping Guan.
In the *57th IEEE Conference on Decision and Control, Miami, USA,* December 2018.

EXPERIENCES AND PROJECTS

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
- Categorized database queries by the operators used and analyzed to which degree resource disaggregation would lead to the performance degradation of each operator

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

LSTM-based Solar Energy Forecasting University of Michigan
Research Intern, *Advisor: Dr. Xianglei Huang* Aug 2018 – Sep 2018

- Built an LSTM-based solar forecasting system using the Keras deep learning library
- Designed the prediction model with an approximate error of only 10%, which was less than all other prediction models in published papers based on the same dataset

SKILLS

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

蔡一凡

邮箱: caiyifan@seas.upenn.edu | 主页: http://www.yifancai.tech | 电话：13857154817

教育背景

宾夕法尼亚大学 美国宾夕法尼亚州费城
计算机科学博士在读 2021 年 9 月至今
指导老师：Linh Thi Xuan Phan 和 Andreas Haeberlen
研究方向：分布式系统安全性和可靠性

上海交通大学 中国上海
工学学士，软件工程专业，系统软件方向 2020 年 6 月毕业
综合 GPA: 3.77 / 4.30 (88.19/100); **专业课 GPA:** 3.88 / 4.30 (89.95/100)

荣誉:

- 上海交通大学校级奖学金 (2016-2017, 2017-2018 & 2018-2019)
- 华为奖学金 (2017-2018)

已发表论文

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

Understanding the Effect of Data Center Resource Disaggregation on Production DBMSs. [Paper]
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, Jiayue Lu.
In *Journal of the Meteorological Society of Japan*, 2020.

Rethinking Data Management Systems for Disaggregated Data Centers. [Paper]
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. [Paper]
Yifan Cai, Jianping He, Wenbin Yu, and Xinping Guan.
In the *57th IEEE Conference on Decision and Control, Miami, USA,* December 2018.

研究和项目经历

宾夕法尼亚大学
研究助理，指导老师：Linh Thi Xuan Phan 和 Andreas Haeberlen 2017.09 – 2019.06

- 设计协议，使得其能够在对于延迟非常敏感的系统中检测到拜占庭错误，并在短时间内恢复。

阿里云达摩院数据库与存储实验室
研究实习生，指导老师：汪晟 周文超 2020.09 – 2021.05

- 设计和实现云数据库，使其提供可由用户验证的数据和执行完整性；
- 通过引入多个验证器和一种将工作负载分配给不同验证器的机制，改进了验证过程的性能；
- 实现了常见的数据库运算，例如 join 和 scan，使得数据库支持关系查询而不仅仅是键值查找。

宾夕法尼亚大学

研究实习生，指导老师：Vincent Liu 2019.07 – 2019.12

- 进行了关于 resource disaggregation 架构对关系数据库设计的影响的研究；
- 开发了嵌套循环连接，哈希连接和宽限度哈希连接运算符的微基准测试程序，并研究了 CPU 端本地内存大小，远程内存访问次数和各个运算的性能下降之间的关系；
- 通过添加更多系统调用以在其上部署诸如 PostgreSQL 和 MonetDB 之类的复杂系统，从而增加了诸如至 LegoOS（分类操作系统）的相对路径之类的功能；
- 分析了现有的支持 resource disaggregation 操作系统的在执行数据库查询语句时的性能缺陷，并提出了潜在的解决方案；
- 对各个数据库中的运算符进行了分类，并分析了 resource disaggregation 的程度对各种运算的影响。

上海交通大学系统控制与信息处理实验室

研究助理，指导老师：何建平 2017.09 – 2019.06

- 在分布式网络系统中发明了一种协议，以完全分布式的方式计算概率密度函数，并使多个计算节点可以共享其统计信息；
- 设计和优化了两求个概率分布函数（PDF）的算法，以使之在公共和匿名网络条件下均能运行。

密西根大学

研究实习生，指导老师：Xianglei Huang 2018.08 – 2018.09

- 使用 Keras 深度学习库构建了基于 LSTM 的太阳预报系统；
- 设计的预测模型的近似误差仅为 10%，这比基于相同数据集的已发表论文中的所有其他预测模型要小。

技能

程序语言: C/C++, Python, Java, JavaScript, SQL