**Cai, Yifan**

Email: [caiyifan@seas.upenn.edu](mailto:caiyifan@seas.upenn.edu) | Homepage: <https://www.yifancai.tech/> | Tel: (+86) 13857154817

**EDUCATION**

**University of Pennsylvania** PA, US

*Ph.D. student in Computer and Information Science* Sep2020 - present

**Shanghai Jiao Tong University** Shanghai, China

*Bachelor of Engineering, Software Engineering*  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 & 2018-2019)

**PUBLICATIONS**

[**VeriDB: An SGX-based Verifiable Database**](https://www.yifancai.tech/publication/SIGMOD2021/)

Wenchao Zhou, **Yifan Cai**, Yanqing Peng, Sheng Wang, Ke Ma, and Feifei Li.

To appear at *ACM SIGMOD International Conference on Management of Data, SIGMOD 2021.*

**Understanding the Effect of Data Center Resource Disaggregation on Production DBMSs.** [**[PDF]**](https://www.cis.upenn.edu/~sga001/papers/ddc-vldb20.pdf)

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).** [**[PDF]**](https://www.jstage.jst.go.jp/article/jmsj/98/5/98_2020-048/_pdf/-char/en)

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.** [**[PDF]**](http://www.yifancai.tech/files/DDCCIDR20.pdf)

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.** [**[**](http://www.yifancai.tech/files/cdc18_main.pdf)**PDF]**

**Yifan Cai**, Jianping He, Wenbin Yu, and Xinping Guan.

In the *57th IEEE Conference on Decision and Control, Miami, USA,* December 2018.

**EXPERIENCE**

**DAMO Academy, Alibaba Cloud Intelligence, Alibaba Group** Hangzhou, China

*Research Intern*, *Advisor: Dr. Wenchao Zhou and Dr. Sheng Wang* Jul 2020 – Apr 2021

* Conducted research on a verifiable database system that runs on an untrusted cloud service provider.
* Implemented trusted operators and trusted data structures to the sgx-based database prototype.
* Designed and implemented methods such as seprating the read/write sets to improve the performance significantly.
* Designed and implemented new methods to guarantee both durability and verifiability of data.

**University of Pennsylvania** PA, US

*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

**Lab of System Control and Information Processing, SJTU** Shanghai, China

*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

**University of Michigan** MI, US

*Research Intern*, *Advisor: Dr. Xianglei Huang*  Aug 2018 – Sep 2018

* Built an LSTM-based solar forecasting system
* Designed and implemented the prediction model with an approximate error of only 10%.

**SKILLS**

**Programming Languages:** C/C++, Python, Java, JavaScript, SQL

**Languages**: Mandarin (Native), English (TOFEL 111)

**蔡一凡**

邮箱: [fyc1007261@live.com](mailto:fyc1007261@live.com) | 主页: <http://www.yifancai.tech> | 电话：13857154817

**教育背景**

**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**

**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.

To appear at *46th International Conference on Very Large Data Bases (VLDB), 2020*

**Rethinking Data Management Systems for Disaggregated Data Centers.** [**[Paper]**](http://www.yifancai.tech/files/DDCCIDR20.pdf)

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**](http://www.yifancai.tech/files/cdc18_main.pdf)**]**

**Yifan Cai**, Jianping He, Wenbin Yu, and Xinping Guan.

In the *57th IEEE Conference on Decision and Control, Miami, USA,* December 2018

**研究经历**

**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。

**项目经历**

**分布式文件系统**

* 用Fuse开发了一个分布式文件系统，该系统可以支持基本的Linux接口（touch，move，write，unlink等）并可以支持多个客户端一起使用；
* 使用本地系统缓存实现分布式锁，以实现原子性并达到高性能
* 在发生故障的情况下，使用heartbeat和notification实施备份恢复，以提高整个系统的可用性

**Tiger语言编译器**

* 按照x86-64调用约定，为Tiger语言开发了C编译器以生成汇编代码
* 通过进行活跃性分析和寄存器分配来优化编译器的性能

**智能电器监控调度系统**

* 与团队成员合作，构建智能的电器监控和调度系统，该系统可以从所有传感器收集信息并操作电器；
* 用Spring boot框架开发了Web应用程序的后端，并将其与硬件用C ++和Python集成在一起，
* 集成了一系列跨过程通信，用于智能设备调度，基于深度学习的太阳能预测以及便捷的人机交互，并确保了它们之间的一致性
* 参加2018年上海交大软件展览并获奖

**基于B+树的数据库**

* 开发了一个基于B +树的数据库系统，当运行超过100万行数据时，可以以每秒多于100万次操作的速度支持CRUD（创建，读取，更新和删除）操作；
* 设计并实现了磁盘和内存之间的缓存系统，以提高读取和写入数据的性能。

**技能**

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

**语言:** 英语（TOFEL 111分）

**技能证书:** Deep Leaning Specializations (by Coursera), Machine Learning (by Coursera), The Bits and Bytes of Computer Networking (by Coursera).