Yifei Liu

Stony Brook, NY 11794 • yifeliu@cs.stonybrook.edu • Cell: +1 631-710-8377
github.com/Yifei-Liu • https://www.fsl.cs.stonybrook.edu/~yifei/ • linkedin.com/in/yifei-liu

Education

Stony Brook University — Stony Brook, NY
- Ph.D. in Computer Science (Advisor: Prof. Erez Zadok) — 08/2019 – Present
- CGPA: 3.93 / 4.0
- *Relevant Courses*: CSE506 Operating Systems; CSE512 Machine Learning; CSE532 Theory of Database Systems; CSE548 Analysis of Algorithms

Huazhong University of Science and Technology — Wuhan, China
- M.Eng. in Computer System Architecture (Advisor: Prof. Ke Zhou) — 09/2016 – 06/2019

Huazhong Agricultural University — Wuhan, China
- B.Eng. in Computer Science and Technology — 09/2012 – 06/2016

Experience

File systems and Storage Lab (FSL), Stony Brook University — Stony Brook, NY
Graduate Research Assistant — 05/2020 – Present
- Analyzed the effectiveness of code coverage and proposed input and output coverage in the context of file system testing
- Applied differential testing and model checking to test Linux kernel file systems thoroughly and automatically
- Designed and evaluated multi-tier caching systems with intelligent MRC point selection to identify optimal cache configurations efficiently

Samsung Semiconductor, Inc. — San Jose, CA
Storage Systems Architect Intern — 05/2022 – 08/2022
- Implemented the custom PostgreSQL plans and paths to offload aggregate operations to SmartSSD, for enhancing the performance of queries
- Investigated the feasibility of offloading PostgreSQL aggregate and join operations to computational storage drive SmartSSD 2.0

Stony Brook University — Stony Brook, NY
Graduate Teaching Assistant — 08/2019 – 05/2020
- CSE376 Advanced Systems Programming in Unix/C (S'21, S'20)
- CSE306 Operating Systems (F'19)

Wuhan National Laboratory for Optoelectronics — Wuhan, China
Research Assistant — 09/2016 – 06/2019
- Used deep learning hash to design and implement a metadata system to integrate high-precision and low-latency content-based semantic queries in storage systems
- Proposed a framework for assessing image "dark data" (unstructured, untapped data) based on a novel semantic hash ranking (SHR) algorithm
- Performed theoretical analysis on hash-based graphs to facilitate rank algorithms and graph database operations

Tencent Cloud   Shenzhen, China
Backend Developer Intern   12/2015 – 08/2016
- Predicted disk failures with disk data collected via machine learning algorithms to achieve high precision and recall
- Built infrastructure for collecting long-term disk S.M.R.A.T. data from over 10,000 servers in Tencent data centers
- Tested various cache replacement policies on Tencent application traces

Publications

Google Scholar Profile: scholar.google.com/citations?user=WNu87vQAAAAJ

Journal Articles

[1] Ke Zhou, Yangtao Wang, Yu Liu, Yujuan Yang, Yifei Liu, Guoliang Li, Lianli Gao, and Zhili Xiao. "A Framework for Image Dark Data Assessment." *World Wide Web*, 2020.

[2] Yu Liu, Yangtao Wang, Ke Zhou, Yujuan Yang, and Yifei Liu. "Semantic-aware Data Quality Assessment for Image Big Data." *Future Generation Computer Systems*, 2020.

Conference and Workshop Papers

[1] Yifei Liu, Gautam Ahuja, Geoff Kuenning, Scott Smolka, and Erez Zadok. "Input and Output Coverage Needed in File System Testing." In *the 15th ACM Workshop on Hot Topics in Storage and File Systems (HotStorage)*, Boston, MA, 2023. (To Appear)

[2] Wei Su, Yifei Liu, Gomathi Ganesan, Gerard Holzmann, Scott Smolka, Erez Zadok, and Geoff Kuenning. "Model-Checking Support for File System Development." In *the 13th ACM Workshop on Hot Topics in Storage and File Systems (HotStorage)*, Virtual, 2021.

[3] Yu Liu, Hong Jiang, Yangtao Wang, Ke Zhou, Yifei Liu, and Li Liu. "Content Sifting Storage: Achieving Fast Read for Large-scale Image Dataset Analysis." In *the 57th Design Automation Conference (DAC)*, San Francisco, CA, 2020.

[4] Yu Liu, Yangtao Wang, Ke Zhou, Yujuan Yang, Yifei Liu, Jingkuan Song, and Zhili Xiao. "A Framework for Image Dark Data Assessment." In *the 3rd APWeb-WAIM joint conference on Web and Big Data (APWeb-WAIM)*, Chengdu, China, 2019. **(Best Paper Runner-Up)**

[5] Yangtao Wang, Yu Liu, Yifei Liu, Ke Zhou, Yujuan Yang, Jiangfeng Zeng, Xiaodong Xu, and Zhili Xiao. "Analysis and Management to Hash-Based Graph and Rank." In *the 3rd APWeb-WAIM joint conference on Web and Big Data (APWeb-WAIM)*, Chengdu, China, 2019.

[6] Jianxiao Liu, Zonglin Tian, Yifei Liu, and Liang Zhao. "Research of Web Service Recommendation Using Bayesian Network Reasoning." In *the 15th International Conference on Services Computing (SCC)*, Seattle, WA, 2018.

[7] Pujuan Shi, Yihang Fang, Chengda Lin, Yifei Liu, and Ruifang Zhai. "A New Line Detection Algorithm - Automatic Measurement of Character Parameter of Rapeseed Plant by LSD." In *the 4th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, Istanbul, Turkey, 2015.

Posters

[1] Yifei Liu, Gerard Holzmann, Geoff Kuenning, Scott Smolka, and Erez Zadok. "Exploring File Systems for Input Coverage." In *the Poster Session of the 21st USENIX Conference on File and Storage Technologies (FAST)*, Santa Clara, CA, 2023.

Patents

[1] Ke Zhou, Yu Liu, Yujuan Yang, Hua Wang, Chunhua Li, Yangtao Wang, Yifei Liu. Method for valuation of image dark data based on similarity hashing. U.S. Patent US11,138,479B2, Filed: 07/30/2019. Issued: 10/05/2021.

[2] Ke Zhou, <u>Yifei Liu</u>, Yu Liu, Yangtao Wang, Yujuan Yang. Image query method and system based on content semantic metadata. Chinese Patent CN110413807B, Filed: 06/24/2019. Issued: 04/20/2021.

Service

Artifact Evaluation Committee
- USENIX OSDI '23, USENIX ATC '23

Talks

- *Model-Checking Support for File System Development*, Dutch Model Checking Day 2022, Eindhoven University of Technology. (Joint talk with Prof. Scott Smolka)
- *Model-Checking Support for File System Development*, ACM HotStorage 2021, Virtual. (Joint talk with Wei Su)
- *OS Support for File System Model Checking*, Computer Science Graduate Research Day 2021, Stony Brook, NY.

Skills

Programming Languages
- Familiar ($\geq$ 4 years of experience): C, C++, Python, Bash
- Intermediate (1 ~ 3 years): MATLAB, SQL, Java, Cypher
- Basic ($\leq$ 1 year): JavaScript, Prolog

Technologies
- **Databases:** MySQL (3 years), Neo4j (2 years), PostgreSQL (< 1 year), HBase (< 1 year), DB2 (< 1 year)
- **File and Storage:** Linux VFS (3 years), OpenStack Swift (2 years), HDFS (< 1 year)
- **Operating Systems:** Linux (8 years), Linux kernel development (3 years)
- **Big Data:** Hadoop (1 year), Spark (< 1 year)
- **Tools:** Git (5 years), LaTeX (4 years), Makefile (4 years), Vim (3 years), GDB (2 years), Elasticsearch (1 year), CMake (1 year), Scikit-learn (1 year)

Human Languages
- Chinese (Native), English (Fluent)

Projects

IOCov: Input and Output Coverage for File System Testing   2022 – Present
- A framework to compute syscall input and output coverage of file system test suites

VeriFS: a Novel File System for Seamless Model Checking Integration   2020 – Present
- An in-memory FUSE file system to save and restore its entire file system state

MCFS: Model Checking File Systems   2020 – Present
- A model-checking framework to test file systems thoroughly with minimal constraints

MTCache: Multi-Tier Caching Simulation and Optimization   2020 – 2022
- A point selection method to identify optimal multi-tier caching configurations

SwiftGraph: Bringing Deep Learning Hash to Storage Systems   2017 – 2019
- A system middleware for content-based semantic queries in storage systems

Disk Failure Prediction in Tencent Cloud   2015 – 2016
- A pipeline to collect disk information and predict disk failures by machine learning

Contest Awards

- Finalist, Interdisciplinary Contest in Modeling (MCM/ICM), USA, 2015.
- First Prize, National Postgraduate Mathematic Contest in Modeling, China, 2014.
- First Prize, Contemporary Undergraduate Mathematical Contest in Modeling, China, 2014.