Yifei Liu

Email: 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
- Developed a model checking framework to test Linux 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

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 and graph algorithms

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

Publications

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

Journal Articles

[1] Tyler Estro, Mário Antunes, Pranav Bhandari, Anshul Gandhi, Geoff Kuenning, Yifei Liu, Carl Waldspurger, Avani Wildani and Erez Zadok. "Accelerating Multi-Tier Storage Cache Simulations Using Knee Detection." ***Performance Evaluation***, 2024.

[2] 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.

[3] 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, Manish Adkar, Gerard Holzmann, Geoff Kuenning, Pei Liu, Scott Smolka, Wei Su and Erez Zadok. "Metis: File System Model Checking via Versatile Input and State Exploration." In *the 22nd USENIX Conference on File and Storage Technologies* ***(FAST '24)***, Santa Clara, CA, 2024.

[2] Tyler Estro, Mário Antunes, Pranav Bhandari, Anshul Gandhi, Geoff Kuenning, Yifei Liu, Carl Waldspurger, Avani Wildani and Erez Zadok. "Guiding Simulations of Multi-Tier Storage Caches Using Knee Detection." In *the 31st International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems* ***(MASCOTS '23)***, Stony Brook, NY, 2023.

[3] 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 '23)***, Boston, MA, 2023.

[4] 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 '21)***, Virtual, 2021.

[5] 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 '20)***, San Francisco, CA, 2020.

[6] 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 '19)***, Chengdu, China, 2019. **(Best Paper Runner-Up)**

[7] 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 '19)***, Chengdu, China, 2019.

[8] 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 '18)***, Seattle, WA, 2018.

[9] 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 '15)***, Istanbul, Turkey, 2015.

Posters

[1] Yifei Liu, Gerard Holzmann, Geoff Kuenning, Scott Smolka, and Erez Zadok. "The Case for Model Checking Emerging File Systems." In *the Poster Presentation of the 17th USENIX Symposium on Operating Systems Design and Implementation* ***(OSDI '23)***, Boston, MA, 2023.

[2] Yifei Liu, Gerard Holzmann, Geoff Kuenning, Scott Smolka, and Erez Zadok. "Exploring File Systems for Input Coverage." In *the Poster Presentation of the 21st USENIX Conference on File and Storage Technologies* ***(FAST '23)***, 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. Granted: 10/05/2021.

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

Service

Journal Reviewer

- ACM Transactions on Architecture and Code Optimization (TACO)

Artifact Evaluation Committee

- USENIX OSDI '23, USENIX ATC '23

Talks

- ***Metis: File System Model Checking via Versatile Input and State Exploration***, FAST 2024, Santa Clara, CA.
- ***Input and Output Coverage Needed in File System Testing***, HotStorage 2023, Boston, MA.
- ***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***, HotStorage 2021, Virtual. (Joint talk with Wei Su)
- ***OS Support for File System Model Checking***, Graduate Research Day 2021, Stony Brook, NY.

Skills

Programming Languages

- Familiar (5+ years of experience): C, C++, Python, Bash
- Intermediate (1–4 years): MATLAB, SQL, Java, Cypher, JavaScript, Prolog

Technologies

- **Databases:** MySQL (3 years), Neo4j (2 years), PostgreSQL (< 1 year), HBase (< 1 year)
- **File and Storage:** Linux VFS (4 years), OpenStack Swift (2 years), HDFS (< 1 year)
- **Operating Systems:** Linux (9 years), Linux kernel development (4 years)
- **Big Data:** Hadoop (1 year), Spark (< 1 year)
- **Tools:** Git (6 years), LaTeX (5 years), Makefile (5 years), GDB (4 years), Vim (3 years), Elasticsearch (1 year), CMake (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 few 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

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.