Duo Zhang

315-949-8633 • duozhang@iastate.edu
https://duozhangisu.github.io/

Education

Ph.D. in Computer Engineering   GPA: 3.84/4.0

Iowa State University, Ames, IA   *Sept. 2018 - Dec. 2023 (expected)*

Major Professor: Dr. Mai Zheng

M.S. in Electrical Engineering   GPA: 3.78/4.0

Syracuse University, Syracuse, NY   *Sept. 2016 - May. 2018*

B.E. in Automation   GPA: 3.5/4.0

Beijing Union University, Beijing, China   *Sept. 2010 - May. 2014*

Research Focus

Data Storage Systems; Data-Intensive Computing; Reliability & Observability

Research Experiences

Research Assistent, Iowa State University

- Designed and simulated NVMe and SCSI bus analyzer by QEMU-KVM
- Conducted in-depth analysis of system failures and Linux kernel patches by Bash script and kernel tailoring
- Designed bug-triggering workloads and pruned Linux kernel to reproduce real system failures
- Built and implemented multiple testing and debugging frameworks by modifying QEMU, Python, and LLVM
- Published conference proceedings and journal articles in top venues such as FAST, SYSTOR, and TOS
- Maintained lab servers, and deployed new devices (e.g., Intel Optane Persistent Memory)

Teaching Experiences

Guest Lecturer & Teaching Assistant, Iowa State University

- CPRE308 Operating Systems Principles and Practice (FA'21, SP'22, FA'22, SP'23, FA'23)
- CPRE563X Advanced Data Storage Systems (SP'22)

Skills

Programming languages

Python, Bash, C language, C++

Tools

Tracer: FTrace, Strace, SystemTap

Framework: MySQL, Redis, Docker, QEMU, LLVM, GDB, Tensorflow, PyTorch

Platforms

Google Cloud, AWS EC2, VMware workstation, CloudLab, Chameleon Cloud

File Systems

Distributed file system: BeeGFS, Lustre, Ceph

Local file system: EXT4, XFS

Projects

Uncovering Vulnerabilities for Persistent Memory Based System
Gitlab: https://git.ece.iastate.edu/data-storage-lab/prototypes/pm-bugs
- Collected 1553 persistent memory-related patches in the Linux kernel source tree by Bash script, and analyzed patches in detail including classification of bug types
- Designed bug-triggering workloads by C language and Bash to reproduce 8 kernel bug cases, which help understand software stack inner dependencies of persistent memory system
- Built and implemented static analysis (LLVM) based persistent memory testing framework for kernel driver security bug detection and identified 29 potential issues including buffer overflow, and non-pointer dereference
- Leveraged virtual machine(QEMU) and record & reply technique to emulate power outage scenario for full-stack crash consistency testing
- Customized target OS through kernel tailoring, and identified 10 unexpected system behaviors of PM software stack(e.g., Redis, Btree, EXT4-dax) through our crash consistency testing framework

Virtual Machine-based System Failures Diagnosis
Github: https://github.com/data-storage-lab/BugBench
- Collected bug cases on the application repo, the Linux kernel Bugzilla by Bash script, and extracted cross-layer features including bug triggering conditions and bug patterns
- Measured and evaluated the debugging observability of FTrace and PANDA via reproduced system failures, and identified their limitations via Bash script and Python
- Modified QEMU to capture device commands and CPU instructions and align them through timestamp by Python to identify critical paths, which effectively narrow down the search space for the root cause (0.06% - 6.2% of the original kernel function trees generated by FTrace)

Uncovering Vulnerabilities in Local File Systems and Parallel File Systems
GitHub: https://github.com/data-storage-lab/BugBench, https://github.com/data-storage-lab/pfault
- Deployed PFSes (e.g., Lustre and BeeGFS) via VMware workstation and Ceph by AWS EC2, and identified system behavior differences through log analysis
- Designed and implemented a fault-injection taxonomy for PFSes, and investigated 28 unexpected system behaviors comprehensively
- Collected and reproduced 59 FS-aware application bug cases by Bash script, and identified general bug patterns and triggering conditions by Python and domain knowledge
- Analyzed configuration dependencies of the local file system (e.g., EXT4 and XFS) by LLVM and Python, and identified 78 potential issues by a configuration-based testing method built by Bash script and Python

Publications

- **"Understanding Persistent-Memory Related Issues in the Linux Kernel".** Om R. Gatla, Duo Zhang, Wei Xu, and Mai Zheng. ACM Transactions on Storage (TOS), 2023

- **"Analyzing Configuration Dependencies of DAX File Systems".** Tabassum Mahmud, Om R. Gatla, Duo Zhang, Carson Love, Ryan Bumann, and Mai Zheng. The 14th Annual Non-Volatile Memories Workshop (NVMW), 2023

- **"ConfD: Analyzing Configuration Dependencies of File Systems for Fun and Profit".** Tabassum Mahmud, Om R. Gatla, Duo Zhang, Carson Love, Ryan Bumann, and Mai Zheng. Proceedings of the 21st USENIX Conference on File and Storage Technologies (FAST), 2023

- **"On the Scalability of Testing the Crash Consistency of PM Systems".** Duo Zhang, Om Rameshwar Gatla, Abdullah Al Raqibul Islam, Dong Dai, and Mai Zheng. The 21th USENIX Conference on File and Storage Technologies (FAST), Work-in-Progress (WiP) & Poster Sessions, 2023

- **"On the Reproducibility of Bugs in File-System Aware Storage Applications".** Duo Zhang, Tabassum Mahmud, Om Rameshwar Gatla, Runzhou Han, Yong Chen and Mai Zheng. 16th International Conference on Networking, Architecture, and Storage (NAS), 2022

- **"Understanding configuration dependencies of file systems".** Tabassum Mahmud, Duo Zhang, Om Rameshwar Gatla, Mai Zheng. 14th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage), 2022 (Best Paper Nominee!)

- **"Benchmarking for observability: The case of diagnosing storage failures".** Duo Zhang, Mai Zheng. BenchCouncil Transactions on Benchmarks (Bench), 2021

- **"A study of persistent memory bugs in the Linux kernel".** Duo Zhang, Om Rameshwar Gatla, Wei Xu, Mai Zheng. Proceedings of the 14th ACM International Conference on Systems and Storage (SYSTOR), 2021

- **"Position: On Failure Diagnosis of the Storage Stack".** Duo Zhang, Om R. Gatla, Runzhou Han, Mai Zheng. The 12th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage), Position Paper & Poster Sessions, 2020

- **"Fingerprinting the Checker Policies of Parallel File Systems".** Runzhou Han, Duo Zhang, Mai Zheng. IEEE/ACM Fifth International Parallel Data Systems Workshop (PDSW), 2020

- **"A Cross-Layer Approach for Diagnosing Storage System Failures".** Duo Zhang, Chander B. Gupta, Mai Zheng, Adam Manzanares, Filip Blagojevic, and Cyril Guyot, The 18th USENIX Conference on File and Storage Technologies (FAST), Work-in-Progress (WiP) & Poster Sessions, 2020

Certificate

AWS Certified Cloud Practitioner

Service

IEEE IPDPS 2021 reviewer

IEEE SECON 2019 reviewer

IEEE INFOCOM 2019 reviewer

IEEE VNC 2018 reviewer

HONORS & AWARDS

HotStorage Best Paper Nominee, ACM Hot Topics in Storage & File Systems Committee

Graduate tuition award, Iowa State University

USENIX Student Grant, USENIX Conference on File and Storage Technologies Committee

Languages

English - Proficient in reading, writing, and speaking
Chinese - Native