

Yiying Zhang

Department of Computer Sciences
1210 West Dayton Street
Madison, WI 53706

yyzhang@cs.wisc.edu
www.cs.wisc.edu/~yyzhang

RESEARCH INTERESTS

Areas: Operating Systems, File and Storage Systems, Distributed Systems

Current interests: New storage interfaces, flash and emerging non-volatile memory technologies, software-defined storage

EDUCATION

University of Wisconsin - Madison
Ph.D. in Computer Science
Advisors: Professor Remzi H. Arpaci-Dusseau and Professor Andrea C. Arpaci-Dusseau
Dissertation: De-virtualization in Storage Systems

Madison, WI

Summer 2013 (expected)

University of Florida
M.S. in Computer Engineering

Gainesville, FL

December 2006

Fudan University
B.S. in Computer Science

Shanghai, China

June 2005

RESEARCH EXPERIENCE

File and Storage Systems

File System De-Virtualizer

2012 - present

Designed File System De-Virtualizer (FSDV), a stand-alone tool to remove storage device virtualization by changing file system pointers from virtual to physical addresses. Implemented FSDV with the Linux Ext3 and Ext4 file system. Working on optimizing and evaluating FSDV.

Hardware Prototyping of Nameless Writes

2012

Prototyped the Nameless Writes interfaces and FTL on SATA and the OpenSSD hardware platform. As a first attempt, built the nameless-write FTL in the device firmware and re-engineered SATA for the new interfaces. As a second prototype, built the nameless-write FTL in the kernel block layer and a raw SSD FTL in the device firmware.

Storage-level Cache Warmup

2012

Designed Bonfire, a system for accelerating storage-level cache warmup, after analyzing a set of block-level data-center traces that provides insights into heuristics of cache warmup. Implemented and evaluated Bonfire with both simulation and real systems.

De-indirection with Nameless Writes

2010-2011

Designed a new interface, Nameless Writes, to remove the need for indirection in flash-based SSDs. Nameless writes allow the device to allocate physical address of a write and then inform the file system of the physical address. Ported the Linux Ext3 file system to use nameless writes and built an emulated nameless-writing device. Demonstrated through experiments that nameless writes reduce both space and time overheads of indirection.

Solid-State Cache

2011

Built an emulator for Solid-State Cache, a flash device designed for caching that leverages cache behavior during garbage collection.

Duplicate-aware Disk Arrays

2010

Proposed Duplicate-Aware Disk Arrays (DADA), which instead of removing duplicates, keeps track of block duplication and uses duplicate contents to improve the reliability and availability

of storage arrays. Analyzed different file server contents for duplicate-aware usage. Built and evaluated DADA as a user-level software RAID mirror.

Correlated Failure of Flash-based Arrays

2009-2010

Designed, simulated, and evaluated Warped Mirror, an approach that separates correlated device failures in flash-based arrays in a predictive way by carefully introducing slightly heavier dummy writes on one device to cause it to fail sooner and by slowing down I/O rates on the healthy devices during recovery.

Green and High Performance Computing

Green Computing on Condor

2009

Designed algorithms to reduce energy consumption of server clusters by dynamically powering individual machines using online policies generated by system status and workload characteristics. Built and evaluated a simulator and a real scheduling system on Condor.

Scheduling with Configurable Processors

2008

Designed, implemented, and evaluated algorithms that dynamically change configurable multi-core processors and schedule workloads for better performance and system utilization.

Machine Learning and Bioinformatics

Rule Discovery in Biological Data

2005-2006

Investigated machine learning and statistical techniques to learn the cleavage properties of HIV-1 virus protein sequences and other properties of DNA and RNA sequences.

PUBLICATIONS

“Warped Mirrors for Flash,” **Yiying Zhang**, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, *Proceedings of the 29th IEEE Conference on Massive Data Storage (MSST '13)*

“Warming up Storage-Level Caches with Bonfire,” **Yiying Zhang**, Gokul Soundararajan, Mark W. Storer, Lakshmi N. Bairavasundaram, Sethuraman Subbiah, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, *Proceedings of the 11th Conference on File and Storage Technologies (FAST '13)*

“Getting Real: Lessons in Transitioning Research Simulations into Hardware Systems,” Mohit Saxena, **Yiying Zhang**, Michael M. Swift, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, *Proceedings of the 11th Conference on File and Storage Technologies (FAST '13)*

“De-indirection for Flash-based SSDs with Nameless Writes,” **Yiying Zhang**, Leo Prasath Arulraj, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, *Proceedings of the 10th Conference on File and Storage Technologies (FAST '12)*

“FlashTier: a Lightweight, Consistent and Reliable Storage Cache,” Mohit Saxena, Michael M. Swift, **Yiying Zhang**, *Proceedings of the 7th European Conference on Computer Systems (EuroSys '12)*

“Duplication Aware Disk Array,” **Yiying Zhang**, Vijayan Prabhakaran, *Microsoft Technical Report (MSR-TR-2012-127)*

“Specificity Rule Discovery in HIV-1 Protease Cleavage Site Analysis,” Hyeoncheol Kim, **Yiying Zhang**, Yong-Seok Heo, Heung-Bum Oh, Su-Shing Chen, *Computational Biology and Chemistry* 32(1): 72-79 (2008)

“Predictability of Rules in HIV-1 Protease Cleavage Site Analysis,” Hyeoncheol Kim, Tae-Sun Yoon, **Yiying Zhang**, Anupam Dikshit, Su-Shing Chen, *Proceedings of the 2006 International Conference on Computational Science (ICCS 06)*

WORKS SUBMITTED OR IN PREPARATION

“A File System De-virtualizer for Non-Volatile Memory,” **Yiying Zhang**, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, *Submitted to the 1st Workshop on Interactions of NVM/Flash*

with Operating-Systems and Workloads with SOSP'13 (INFLOW'13)

PATENTS

“Duplicate-Aware Disk Arrays,” Vijayan Prabhakaran, Yiying Zhang, US Patent #20120226936 A1 filed by Microsoft

“System and Method for an Efficient Cache Warm-up,” Lakshmi N. Bairavasundaram, Gokul Soundararajan, Mark W. Storer, Yiying Zhang, Patent filed Dec 2012 by NetApp

TALKS

“De-virtualization in Storage Systems”, UCSD, San Diego, CA, May 2013

“Warming up Storage-Level Caches with Bonfire”, FAST'13, San Jose, CA, Feb. 2013

“De-virtualization for Flash-based SSDs”, WISDoM Workshop, Madison, WI, Nov. 2012

“Warming up Storage-level Caches with Bonfire”, NetApp, Sunnyvale, CA, Aug. 2012

“De-indirection for Flash-based SSDs with Nameless Writes”, Google, Madison, WI, Mar. 2012

“De-indirection for Flash-based SSDs with Nameless Writes”, Microsoft Research, Mountain View, CA, Feb. 2012

“De-indirection for Flash-based SSDs with Nameless Writes”, FAST'11, San Jose, CA, Feb. 2012

“Duplicate-aware Disk Arrays”, Microsoft Research, Mountain View, CA, Aug. 2010

WORKING EXPERIENCE

Summer Intern, **NetApp, Advanced Technology Group**, Sunnyvale, CA, Summer 2012
Mentors: Gokul Soundararajan, Mark W. Storer
Storage-level cache warmup project

Summer Intern, **Microsoft Research**, Mountain View, CA, Summer 2010
Mentor: Vijayan Prabhakaran
Duplicate-aware disk array project

Summer Intern, **Microsoft**, Business Intelligence Group, Shanghai, China, Summer 2008
Designed and implemented a prototype to integrate the MS Integration Service into the MS SQL Server environment and a tool to publish and execute SQL queries on collaboration web services.

Software Design Engineer, **Innovative Scheduling**, Gainesville, FL, Jan. to Aug. 2007
Developed a new product, Locomotive Planning Optimizer, which optimizes locomotive routing and scheduling using multiple integer programming, network flow, and graph theories.

SERVICES AND OTHER ACTIVITIES

Technical Reviewer: Transactions on Computers (TOC), Transactions on Storage (TOS)
External Reviewer: HotOS 2011 External Reviewer, SOSP 2009 Shadow PC
Conferences Attended: FAST (2012, 2011, 2010), OSDI 2010, SOSP 2009
Member: ACM, USENIX