

THANH D. DO

thanhdo@cs.wisc.edu

<http://www.cs.wisc.edu/~thanhdo>

922 Eagle Heights, Apt C, Madison, WI 53705. Phone: (608) 320-4949

Areas of Expertise

File/Storage Systems, Operating Systems, and Distributed Systems
Current: Cloud Computing, focusing on reliability and performance

Education

University of Wisconsin, Madison

2009–present

Ph.D. in Computer Sciences, *July 2014 (expected)*. GPA: 3.77

Thesis: *Towards Reliable Cloud Systems: Managing Silent and Performance Failures*

M.S. in Computer Sciences, *December 2013*

Advisors: Remzi H. Arpaci-Dusseau, Andrea C. Arpaci-Dusseau and Haryadi S. Gunawi

Hanoi University of Technology, Hanoi, Vietnam

2001–2006

B.S. in Computer Sciences, *Summa cum laude*, ranked 1/300

Won Intel and Yamaha scholarship for highest GPA among 300 CS majors

Work Experience

Engineering Intern (Google Inc., Madison, WI, *Summer 2012*)

- Built a framework to evaluate availability SLA for 10000+ jobs running in 5 Google clusters containing thousands of machines over a 6-month period. The framework analyzes 100GB+ of logs of job events and calculates availability of the jobs in order to find ones that violate predefined SLA. This information helps pinpoint clusters that do not meet availability requirement. (Mentor: Florentina Popovici).

Research Assistant (ADSL, UW-Madison, *2010 – present*; UCARE, UChicago, *Summer 2013 - present*)

- Built a model checking framework that checks MapReduce’s performance properties automatically. Found more than 20 bugs that lead to severe job execution slowdown. The framework provides insight into making MapReduce more performance-robust.
- Built an analysis and testing framework that injects slow I/Os on 5 cloud systems (Hadoop, HDFS, HBase, Cassandra, and Zookeeper). Measured the impact of slow hardware on these systems. Unearthed design flaws in 15 protocols in which slow hardware can severely impact an operation, a node, or an entire cluster.
- Designed and implemented selective and lightweight versioning (SLEEVE), a new approach that efficiently hardens cloud systems against fail-silent behaviors that result from memory corruption and software bugs. Applied SLEEVE to build HARDDFS, a hardened version of HDFS, and showed that HARDDFS is more robust than HDFS: it handles 90% of the fail-silent faults that result from random memory corruption and correctly detects and recovers from 100% of 78 targeted corruptions and 5 real-world bugs.
- Built a cloud recovery testing framework that systematically tests recovery in the face of multiple failures. Integrated the framework to HDFS, explored over 40,000 failure scenarios, wrote 74 specifications, found 16 new bugs, and reproduced 51 old bugs.

Volunteer Experience (Shorewood School CS club, Madison, WI, *Fall 2012 – Spring 2013*)

- Instructed a CS club of 20 kids (4th graders and up) how to program in Scratch.

Teaching Assistant (UW-Madison, *2009 – 2010*)

- Instructed a class of 30 students in “Introduction to Programming” using Java.
- Built a framework in Python that automatically grades students’ projects in “Undergrad OS”.

Publications

REFEREED CONFERENCE/WORKSHOP PUBLICATIONS

1. **Thanh Do**, Mingzhe Hao, Tanakorn Leesatapornwongsa, Tiratat Patana-anake, Haryadi S. Gunawi. Limplock: Understanding the Impact of Limpware on Scale-out Cloud Systems. *Proceedings of the 2013 ACM Symposium on Cloud Computing (SoCC '13)*.
2. **Thanh Do**, Haryadi S. Gunawi. The Case for Limping-Hardware Tolerant Clouds. *Proceedings of the 5th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud '13)*.
3. **Thanh Do**, Tyler Harter, Yingchao Liu, Haryadi S. Gunawi, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau. HARDFS: Hardening HDFS with Selective and Lightweight Versioning. *Proceedings of the 9th Conference on File and Storage Technologies (FAST '13)*.
4. Haryadi S. Gunawi, **Thanh Do**, Pallavi Joshi, Peter Alvaro, Joseph M. Hellerstein, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, Koushik Sen, Dhruba Borthakur. FATE and DESTINI: A Framework for Cloud Recovery Testing. *Proceedings of the 8th USENIX Symposium on Networked Systems Design and Implementation (NSDI '11)*.
5. Haryadi S. Gunawi, **Thanh Do**, Pallavi Joshi, Joseph M. Hellerstein, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, Koushik Sen. Towards Automatically Checking Thousand of Failures with Micro-specifications. *Proceedings of the 6th Workshop on Hot Topics in System Dependability (HotDep '10)*.
6. **Thanh Do**, Suhil Rawshdeh, Weisong Shi. ptop: A process-level power profiling tool. *Poster Session - the 2nd Workshop on Power Aware Computing and Systems (HotPower '09)*.

OTHER PUBLICATIONS

7. **Thanh Do**, Haryadi S. Gunawi. Impact of Limpware on HDFS: A Probabilistic Estimation. *CoRR, Nov 2013, abs/1311.3322*.
8. Haryadi S. Gunawi, **Thanh Do**, Joseph M. Hellerstein, Ion Stoica, Dhruba Borthakur, Jesse Robbins. Failure as a Service (FaaS): A Cloud Service for Large-Scale, Online Failure Drills. *UC-Berkeley Technical Report, 2011*.

Talks

1. Towards Reliable Cloud Systems. *Microsoft Gray Systems Lab*, Apr. 2014.
2. Understanding the Impact of Limpware on Scale-Out Cloud Systems. *Google Inc.*, Feb. 2014.
3. Understanding the Impact of Limpware on Scale-Out Cloud Systems. *Wisconsin SyNS '13*, Nov. 2013.
4. Understanding the Impact of Limpware on Scale-Out Cloud Systems. *SoCC '13*, Oct. 2013.
5. HARDFS: Hardening HDFS with Selective and Lightweight Versioning. *FAST '13*, Feb. 2013.
6. SLEEVE: Hardening Cloud Systems with Selective and Lightweight Versioning. *WISDoM*, Nov. 2012.
7. FATE and DESTINI: A Framework for Cloud Recovery Testing. *Wisconsin Systems Retreat*, Apr. 2011.
8. Towards Automatically Checking Thousand of Failures with Micro-specifications. *HotDep '10*, Oct. 2010.

Courses Taken at UW-Madison

Advanced Operating Systems (CS-736)
Distributed Systems (CS-739)
Topics in Database Management Systems (CS-764)
Computer System Performance Evaluation and Modeling (CS-737)
Analysis of Software Artifacts (CS-706)
Construction of Compilers (CS-701)

References

Remzi H. Arpaci-Dusseau, Ph.D.
Professor of Computer Science
University of Wisconsin, Madison
1210 W. Dayton St.
Madison, WI 53706
Phone: (608) 263-7764
Email: remzi@cs.wisc.edu

Andrea C. Arpaci-Dusseau, Ph.D.
Professor of Computer Science
University of Wisconsin, Madison
1210 W. Dayton St.
Madison, WI 53706
Phone: (608) 265-6013
Email: dusseau@cs.wisc.edu

Haryadi S. Gunawi, Ph.D.
Professor of Computer Science
University of Chicago
Ryerson Hall 259
1100 E. 58th St.
Chicago, IL 60637
Phone: (773) 702-5772
Email: haryadi@cs.uchicago.edu

Florentina Popovici, Ph.D.
Google Inc.
10 N. Livingston St.
Suite 201
Madison, WI 53703
Phone: (608) 695-0707
Email: florentina@google.com