Mi Zhang

Email: mzhang@cse.cuhk.edu.hk
Page: https://millyz.github.io

Education

The Chinese University of Hong Kong — Aug 2015 - Now

- Ph.D candidate, Department of Computer Science & Engineering.
- Research Area: Distributed Storage Systems (Erasure Coding, Reliability Analysis)
- Advised by Prof. Patrick P. C. Lee.

Shandong University — Sep 2010 - Jun 2014

- Bachelor of Engineering, Department of Software Engineering.
- GPA: 91.35/100, Ranking: 3/259

Publications

Conference:

- **A Simulation Analysis of Reliability in Erasure-Coded Data Centers**
  Mi Zhang, Shujie Han and Patrick P. C. Lee.
  *The 36th IEEE International Symposium on Reliable Distributed Systems (**SRDS**), 2017*
  (AR: 24/72 = 33.3%)

Journal:

- **Optimal Repair Layering for Erasure-Coded Data Centers: From Theory to Practice**. Yuchong Hu, Xiaolu Li, **Mi Zhang**, Patrick P. C. Lee, Xiaoyang Zhang, Pan Zhou, and Dan Feng.
  *ACM Transactions on Storage (**TOS**), 13(4), pp. 33:1-33:24, November 2017.*

Selected Projects

SimEDC — SRDS '17

- SimEDC is a comprehensive discrete-event simulator that characterizes the reliability of an erasure-coded data center. We can study how various erasure code constructions, chunk placement schemes, data center topologies, and statistical behaviors of failure/repair patterns affect the overall storage reliability in hierarchical data centers using SimEDC.
- We conduct extensive reliability analysis and find that hierarchical placement generally achieves higher reliability than flat placement due to the reduction of cross-rack repair traffic, even though its reliability degrades in the presence of correlated failures.
- The source code of SimEDC implementation is available at http://adslab.cse.cuhk.edu.hk/software/simedc.

DoubleR — TOS '17

- DoubleR is a practical repair layering framework, which improves the recovery performance by partitioning a repair operation into inner-rack and cross-rack layers.
- We implement and deploy DoubleR atop the Hadoop Distributed File System (HDFS) and show that DoubleR maintains the theoretical guarantees and improves the repair performance of regenerating codes in both node recovery and degraded read operations.
- The source code of DoubleR prototype is available at http://adslab.cse.cuhk.edu.hk/software/doubler.

Selected Awards

CUHK Postgraduate Studentship	2015 - Now
Outstanding Graduate of Shandong Province and Shandong University	2014
Google Excellence Scholarship	2013
National Scholarship of China	2012/2011
The First-grade Scholarship for Outstanding Students of Shandong University	2013/2012/2011

Teaching

Teaching Assistant of *Introduction to Cloud Computing and Storage*	Fall 2017/2015
Teaching Assistant of *Data Communication and Computer Networks*	Spring 2017/2016
Teaching Assistant of *Operating Systems*	Fall 2016

Skills

Programming Languages: Python, Java, C/C++
Data Processing Systems: Hadoop, Spark