# RAID6-based Distributed Storage System Implementation

Kai Liu
School of Computer Engineering
Nanyang Technological University
Singapore 639798
kliu006@e.ntu.edu.sg

Bingshui Da
School of Computer Engineering
Nanyang Technological University
Singapore 639798
da0002ui@e.ntu.edu.sg

Jianglei Han
School of Computer Engineering
Nanyang Technological University
Singapore 639798
jhan011@e.ntu.edu.sg

## I. Introduction

In distributed systems, data is partitioned and stored in storage devices that are separated from each other in location. For example, in a small scale distributed system, a file could be stored in multiple hard disks. In large distributed systems, data segments could be stored in different data centers located in separate geographical locations. When a particular data is being accessed by an user, the system resembles data from partitions and response to the request. Logically, serving read and write request to a distributed file should be no different from accessing a local file, from user's perspective. Thus, storage resiliency is one of the most important considerations when designing a reliable and fault-tolerant distributed storage system.

The reliability is related to I/O speed and latency user may experience. It is subjected to system architecture, quality of hardware and communication channels.

The fault-tolerance of a system is system's capability of handling situations when the data segment needed is unavailable. Individual disk availability in a norm instead of an exception in distributed systems. Except hardware difficulties, a storage node goes unavailable during upgrading or when the workload is exceptionally high.

## II. Background

RAID

### A. RAID

### B. GS

### C. Reed-Solomen Coding

## III. Implementations

### A. System Architecture

### B. Pseudo Code

## IV. Experiments

Compared 2+2, 3+2, 4+2, 5+2, 6+2

### A. Storage

### B. Recoverability

Raid-6 recoverability theory

### C. Speed

## V. Conclusion

to conclude