

**CSE 451: Operating Systems
Autumn 2008**

**Redundant Arrays of Inexpensive Disks
(RAID)**

Hank Levy

The challenge

- Disk transfer rates are improving, but much less fast than CPU performance
- We can use multiple disks to improve performance
 - by *striping* files across multiple disks (placing parts of each file on a different disk), we can use parallel I/O to improve access time
- Striping reduces reliability
 - 100 disks have 1/100th the MTBF (mean time between failures) of one disk
- So, we need striping for performance, but we need something to help with reliability / availability
- To improve reliability, we can add redundant data to the disks, in addition to striping

11/24/2008

2

Refresher: What's parity?

1 0 1 1 0 1 1 0 1

- To each byte, add a bit set so that the total number of 1's is even
- Any single missing bit can be reconstructed
- More complex schemes (e.g., based on Hamming codes) can detect multiple bit errors and correct single bit errors. Called ECC (error correcting code) memory.

11/24/2008

3

RAID

- A RAID is a **Redundant Array of Inexpensive Disks**
- Disks are small and cheap, so it's easy to put lots of disks (10s to 100s) in one box for increased storage, performance, and availability
- Data plus some redundant information is striped across the disks in some way
- How striping is done is key to performance and reliability

11/24/2008

4

Some RAID tradeoffs

- Granularity
 - fine-grained: stripe each file over all disks
 - high throughput for the file
 - limits transfer to 1 file at a time
 - course-grained: stripe each file over only a few disks
 - limits throughput for 1 file
 - allows concurrent access to multiple files
- Redundancy
 - uniformly distribute redundancy information on disks
 - avoids load-balancing problems
 - concentrate redundancy information on a small number of disks
 - partition the disks into data disks and redundancy disks

11/24/2008

5

RAID Level 0

- RAID Level 0 is a non-redundant disk array
- Files are striped across disks, no redundant info
- High read throughput
- Best write throughput (no redundant info to write)
- Any disk failure results in data loss

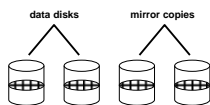


11/24/2008

6

RAID Level 1

- RAID Level 1 is mirrored disks
- Files are striped across half the disks
- Data is written to two places – data disks and mirror disks
- On failure, just use the surviving disk
- 2x space expansion

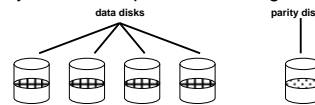


11/24/2008

7

RAID Levels 2, 3, and 4

- RAID levels 2, 3, and 4 use ECC (error correcting code) or parity disks
 - E.g., each byte on the parity disk is a parity function of the corresponding bytes on all the other disks
- A read accesses all the data disks
- A write accesses all the data disks plus the parity disk
- On disk failure, read the remaining disks plus the parity disk to compute the missing data

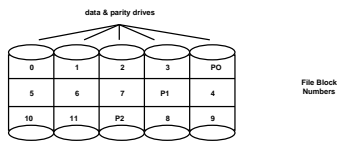


11/24/2008

8

RAID Level 5

- RAID Level 5 uses block interleaved distributed parity
- Like parity scheme, but distribute the parity info (as well as data) over all disks
 - for each block, one disk holds the parity, and the other disks hold the data
- Significantly better performance
 - parity disk is not a hot spot



11/24/2008

9

RAID Level 6

- Basically like RAID 5 but with replicated parity blocks so that it can survive two disk failures.
- Useful for larger disk arrays where multiple failures are more likely.
- RAID 10 – striping plus mirroring
- RAID 50 – RAID 5 plus mirroring
- RAID xx – something for you to invent ☺

11/24/2008

10

Example RAID Storage



- Promise 3U rack-mountable 16-disk RAID Storage System
- Hot swappable drives
- Dual controllers with 4 host interface ports for reliability
- Can be ganged together into larger units

11/24/2008

11