6.033
COMPUTER SYSTEMS ENGINEERING
RECITATION 7: RAID

JOHN WANG

1. Introduction to RAID

Application has reads and writes to the RAID controller. Then the RAID controller does the actual reads and writes to the actual disk. It makes a bunch of disks look like a single disk. You can think of each disk as a private variable to the RAID controller, and RAID abstracts everything away.

The motivation for RAID:

- Reliability. You want to be able to retain access to your data even when you have failures. Most people who care about reliability are implementing servers.
- Cost. You can buy a lot of cheap disks for a cost much lower than a single fault tolerant disk.
- Performance. Increase disk throughput. For example, making games better.

1.1. **Reliability.** This is improved via redundancy.

1.2. **Performance.** One basic mechanism to improve performance: interleaving. Take data that you would put on just one disk and put it on multiple disks so you can read and write data concurrently on many disks.

2. Levels of RAID

2.1. **RAID 0.** Simplest approach is to alternate the physical addresses on the physical disk. For example, if you have m disks, then you place the virtual memory location of the kth location on the $k \pmod m$th disk.

If you don't make an effort to make concurrency opportunities, you tend not to get concurrency. This is purely a performance optimization, and no redundancy improvements. You get a little bit of concurrency because you expect to be able to access multiple items at the same time from different disks in a concurrent fashion.

2.2. **RAID 1.** You have two copies of the same disk. This improves redundancy by creating a new disk with the same information. You can improve performance on reads because if you are reading a chunk of data, you can break that chunk up into different start places and read these smaller chunks concurrently.

Alternatively, you can create more than just two copies of the data. This improves the chances that at least one disk has not failed, but improves the chances that any given disk fails.

Performance: Let's say c is the time for one rotation. Then the expected delay for a single disk is just $c/2$ because the head could be at any point away from the sector. The expected delay of a full read operation in RAID 1 is just $E[\max\{delay_1, delay_2\}]$. Thus, the expected delay time increases as you have more redundancy. Slowdown factor is s is somewhere in $[1, 2]$. This is because the worst expected delay you can get is c.

2.3. **RAID 2.** Using error correcting codes. Let's say you have a 7 bit string 7654321. Three of these bits are check bits, and the other four bits actually store the data values. You compute the check bits and make sure the data bits are correct.

$$p_1 = XOR(d_1, d_3, d_4) \tag{1}$$
$$p_2 = XOR(d_1, d_2, d_4) \tag{2}$$
$$p_3 = XOR(d_1, d_2, d_3) \tag{3}$$

If all of the parity bits p_1, p_2, p_3 pass, then we probably have the correct answer. Each column has at least 2 check bits. For example d_2 occurs in both p_2 and p_3. This code only works if there is at most 1 bit that has been flipped to the wrong value.

How do we map this encoding onto a RAID scheme? You can think of each data bit as a disk.

2.4. **RAID 3.** Instead of using the more expensive error correcting code, we have any number of data disks and a single check disk which is just the XOR of all the data disks. If the check disk is incorrect, then you know that one of the data disks is incorrect as well.

2.5. **RAID 4.** Parity bits are stored on a single data.

2.6. **RAID 5.** If you have some disks, you alternative responsibility of data and check disks for different addresses. You rotate the information of the parity information across the disk. It is unlikely that all the blocks you choose will have the same check disks so hopefully you can spread the check bits evenly across the disks. This prevents your single check disk from failing and losing your redundancy.

2.7. **Recursive RAID.** You have a bunch of RAID 0 arrays. Then you combine the entire set of RAID 0 arrays using RAID 1. This is called RAID 0+1.