CSG389	**Homework 3**	**Assigned: 3/23/2005**
Spring 2005	CCS, Northeastern University	**Due: 3/30/2005**

Problem 1 – Analyze This.

Consider an interactive system (closed system) with a CPU and two disks with a number of terminals attached. The following measurement data was obtained by observing the system:

Observation interval = 30 minutes

Number of complete transactoins during observation interval = 1600

Number of competions at CPU = 1600

Number of fast disk accesses = 32000

Number of slow disk accesses = 12000

CPU busy time = 1080 seconds

fast disk busy time = 400 seconds

slow disk busy time = 600 seconds

(a) Give asymptotic bounds on throughput and response time as a function of the number of terminals.

(b) Now consider the following modifications to the system:

(i) Move all files to the fast disk.

(ii) Replace the slow disk by a second fast disk.

(iii) Increase the CPU speed by 50% (with the original disks).

(iv) Increase the CPU speed by 50% and balance the disk load across the two fast disks.

For each of the 4 modifications graph the effects on the original system. Show your calculations of D_{max} and D.

Problem 2 – Striping data.

(a) The CEO of a new startup has decided that they need to keep their video data in a redundant format, unfortunately he is too cheap to buy a proper RAID 5 system so he came up with a redundancy plan of his own, which is shown in Table 1. The *B's* represent data blocks and the *P's* represent parity blocks, where P_i is the parity block for data blocks $B_{(4i-3)}$ to $B_{(4i)}$. What problems might arise from this arrangement of data blocks and parity blocks?

Disk 1	Disk 2	Disk 3	Disk 4
B_1	B_2	B_3	B_4
P_1	B_5	B_6	B_7
B_8	P_2	B_9	B_{10}
B_{11}	B_{12}	P_3	B_{13}

Table 1: Redundancy scheme

(b) After much debate the CEO decides to purchase a commercial RAID system using RAID level 1 (mirroring) or RAID level 5 (block interleaved, distributed parity). But now after much research he becomes worried that a power failure during a multi-block disk write may not be atomic.

Suggest a scheme for each (RAID 1 and RAID 5) that would allow the CEO to detect non-atomic writes, and if possible, suggest schemes that would allow them to recover from the failure.

(c) Since RAID 1 is easier for the CEO to comprehend he decides to look into it further. However, he becomes increasingly worried about the mean time to failure (MTTF). His main engineers try to convince him that the mean time to data loss is much greater than the mean time to disk failure.

Suppose RAID 1 is being used with 2 disks and that there is a 5% chance per year that a disk will fail. The only time data can be lost if the second disk fails while the first disk has failed and is being replaced. What is the MTTF of the RAID 1 array (involving data loss) assuming the mean time to repair of a single disk is (MTTR) is 3 hours?

(d) Mr CEO is now thinking that he wants more than one RAID array, since their video streaming business is booming, and two are always better than one. However, he begins to hear rumors that instead of having two individual arrays it may be worthwhile to merge the arrays together in some kind of scheme.

For this example we have 8 disks over which to distribute our parity groups. A parity group is defined as the set of data blocks over which parity is computed, plus the parity block itself, so in the case of RAID 5 a parity group consists of a stripe across the entire array.

(i) Given that the MTTF of a RAID 5 array is: $\frac{MTTF(disk)^2}{N \times (G-1) \times MTTR(disk)}$

- N = # disks in the array
- G = # disks in the parity group
- MTTF(disk) = 200,000 hours
- MTTR(disk) = 3 hours.

What is the MTTF for each of the schemes shown in Table 2?

(ii) Given the following conditions, describe how each scheme in Table 2 performs:

- Average number of disks touched on a single block read (no failures).
- Average number of disks touched on a single block write (no failures).
- Average number of disks touched on a single block read (1 disk failure).
- Maximum number of disk failures without data loss.

Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7	Disk 8
PG_1	PG_1	PG_1	PG_1	PG_1	PG_1	PG_1	PG_1
PG_2	PG_2	PG_2	PG_2	PG_2	PG_2	PG_2	PG_2
PG_3	PG_3	PG_3	PG_3	PG_3	PG_3	PG_3	PG_3

Table 2(a): One Standard RAID 5 Array of 8 Disks

Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7	Disk 8
PG_1	PG_1	PG_1	PG_1	PG_2	PG_2	PG_2	PG_2
PG_3	PG_3	PG_3	PG_3	PG_4	PG_4	PG_4	PG_4
PG_5	PG_5	PG_5	PG_5	PG_6	PG_6	PG_6	PG_6

Table 2(b): Two Standard RAID 5 Arrays of 4 Disks Each

Disk 1	Disk 2	Disk 3	Disk 4	Disk 5	Disk 6	Disk 7	Disk 8
PG_1	PG_1	PG_1	PG_1	PG_2	PG_2	PG_2	PG_2
PG_3	PG_3	PG_4	PG_4	PG_3	PG_3	PG_4	PG_4
PG_5	PG_6	PG_5	PG_6	PG_5	PG_6	PG_5	PG_6
PG_7	PG_8	PG_8	PG_7	PG_7	PG_8	PG_8	PG_7

Table 2(c): One RAID Array with Declustered Parity

Table 2: Different RAID 5 Parity Grouping Schemes. Each PG_i stripe unit contains either data or parity information for parity group i.

(iii) Which scheme balances the load over the disks most evenly in the case of a single disk failure? Explain.

Problem 3 – AIR RAID.

(a) Disk failures are not the only concern of RAID arrays – array controller system crashes can also cause the array to be put in an inconsistent state. (Assume that the system contains NO NVRAM).

(i) Describe a scenario that can leave the array's data in an inconsistent state after a controller crash.

(ii) Describe a brief solution to the scenario you described above.

(b) When performing a small write (to a single unit within the parity stripe) in RAID 5, what is the approximate increase in response time (with no failed disks and no request queueing), as opposed to a single disk?

(c) Choosing the best stripe unit size for a set of disks involves tradeoffs between two conflicting goals. Explain what both of these goals are, and how they conflict.

(d) In terms of queueing times and throughput, how does striping affect performance (increase or decrease) over a single disk? Give a reason why each are affected in this manner.

(e) In a database application where availability and transaction rate are more important than space efficiency, which RAID scheme would you use? Why?

Problem 4 – Cache utilization.

Assume a simplified model of a disk array with a cache that can serve a request in constant time of 75 μs. Given the workload characteristics and the cache size, it was observed that the cache hit rate is 75%. When a cache miss occurs, the data has be to fetched from the logical volume with an average service time of 6 ms with an exponential distribution. For the problem, assume that write are infrequent, always go to the cache and the disk array controller manages to destage dirty blocks from the cache when the disk array is not busy. Hence, the write traffic does not interfere with the cache read misses.

(a) Draw the diagram of this simple model. Assume that each component (i.e., cache and logical volume) is a separate service center.

(b) Given an arrival rate of 4500 I/Os per second, calculate the cache utilization, number of I/Os waiting in the cache's queue, and the average response time for the I/Os that are cache hits.

(c) Given the cache miss ratio of the workload described above, what is the smallest number of disks (without any protection scheme) that would have to comprise the "logical volume" to ensure that they can handle the load? In addition, calculate the average response time for this workload and specify what the utilization of each disk would be.

(d) Now suppose that you want to bound the response time for cache misses to three times the disk service time. How many disks would you have to put into the logical volume in a RAID 10 (1+0) configuration, assuming that the load is spread out evenly among all the disks? What would be the utilization of each disk?