# 1 Avodat Bayit for storage systems course

# 2 Solve 10 of the following

1)a) Read the paper "RAID: High performance, reliable secondary storage" By Katz, Gibson, Chen, Lee and Patterson. (google the first 3 listed authors and pick the first item). Describe the main characteristics of RAID 1,3,4,5. What is the difference between RAID 4 and RAID 5.

b) Assume that we employ physical mirroring to a system of $n$ disks. What is the size $G$ of the RAID group. Assume that we use random mirroring in which any two disks share a file, what is $G$?

c) Explain the small write penalty for parity protected drives. Do bit interleave systems (RAID 3) suffer from it ? explain.

2) Suppose that MTTR is very large (there is no repairman) how much time will pass until data in a system with two identical disks is lost?

3) We have 4 disks in a left symmetric parity RAID5 group similar to that shown in figure 4 of the paper and below. The disks are numbered DO,D1,D2,D3. Suppose the stripe size is of 8 blocks. Thus the configuration looks like

0 1 2 p

4 5 p 3

8 p 6 7

p 9 10 11

Where each number represents a stripe of 8 consecutive blocks, i.e. blocks 0-7 are in stripe 0, 8-15 are in stripe 1 etc. Find the location of data block 199 ( blocks are numbered starting with zero ). Find the location of the parity block which protects block 199. Find the location of a general block

x. Location means which disk and which block on the disk.

4) Suppose you have a physically mirrored system and that all requests are read requests. Suppose that you can choose probabilities $q_i$ and randomly choose to service the requests to file $i$ from the first disk with probability $q_i$. Show that the worst possible choice in terms of total seek activity is when all $q_i$ are 0 or all $q_i$ are 1.

5)a) Assume that requests arrive at a disk with exponentially distributed interarrival times , and rate 1. assume the disk serves the requests with an exponentially distributed service time , with rate 3. Assume the disk allows at most $K$ requests to queue. compute the queue length distribution $N$

b) Assume now that the service rate is 1 instead of 3. Compute the average queue length.

c) Assume now that the interarrival rate is still exponential with rate 1, there is no limit on the queue length. Compute the average wait time if the service time is exponentially distributed with rate 3. Also compute the average wait time in case the service time is 1/3 deterministically.

6) We Showed in class that if $G$ is the graph of active videos in a mirrored system we can show all videos if $G$ does not have components with two cycles. Prove that we will still be able to show all movies after the loss of any given disk only if all components of $G$ are trees.
7)a) Find the optimal cache policy for the PQRS model assuming you have infinite bandwidth to bring things into cache at any given moment.

b) Estimate the hit (or miss) ratio for the optimal policy in the same way we estimated it for the universal static policy (use Stirling's formula)

8) Consider a system with two hosts, exponential inter arrival times, a heavy-tailed service time with parameter $\alpha = 1.2$, shortest job of size 1 and largest job of size $10^6$. Assume that the utilization for a single server is 1

a) Compute the waiting time assuming random request allocation to the servers.

b) Let $1 < c < 10^6$ and consider the policy which sends jobs of size smaller than $c$ to the first server and jobs larger than $c$ to the second server. Compute the relevant utilizations and then the average waiting time.

c) Optimize the calculations of (b) over all values of $c$.

9) There are 2 disk types. Disk C fails unformly in the range $[0, R_C]$ and disk D fails uniformly in the range $[0, R_D]$

A) Compute in terms of $R_C$ and $R_D$ the probability of a disk of type C to fail before a disk of type D.

B) If we have a system with 20 disks of type C and $R_C = 1$ what is the average time until one of the disks fails.

C) in the system of part B) if a first disk failure occured immediately at time 0 and it takes the technician 0.2 time to repair the disk what is the probability of data loss.

10) Show that for any $\varepsilon > 0$, with probability approaching 1 there is no increasing subsequence of size $(e+\varepsilon)\sqrt{n}$ in a random permutation. Hint: Show that the expected number of increasing subsequences of size $k$ is $B(n,k)/k!$ where $B(n,k)$ is the binomial coefficient. Then use Stirling.

11) Show that if the number of edges in a random graph is $cn$, with $c < 1/2$ and $n$ the number of vertices then with probability approaching 1, the graph contains no "Handcuffs".

12)a) Compute the stationary probabilities $p_n$ for queue length in an M/M/1 queue wher the arrival and service rates are queue length dependent, namely we have $\lambda_n$ and $\mu_n$.

b) show that for the exponential distribution $E(S^2)/E(S)^2 = 2$.