1 Test 2

1) We have 4 disks in a RAID group , numbered DO,D1,D2,D3 . Suppose we use a RAID5 configuration which is left symmetric (see below) with stripe size of 8 blocks . Left symmetric means that data and parity stripes are arranged as follows .

0 1 2 p

4 5 p 3

8 p 6 7

p 9 10 11

In a repeating fashion . 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 .

2) A disk array is divided into two chunks of equal size with activities $A_1 = 1$ and $A_2 = 3$. The cache has size $S = 1.5$. assuming the cache runs MRU,FIFO and the activities follow the IRM model compute the hit ratio H and q_1, q_2 the relative portions of cache occupied by the chunks . Compute the hit ratio using the hot chunks method . In which formula did we use the fact that the cache is a FIFO cache , and in which one did we use MRU .

3) Describe the operations behind the implementation of LSF and snapshot copy . What are they used for .

2 test , Moed C , spring 2000

Solve all 3 questions !!

You may use a calculator

Time : 2 hours and 15 minutes

1) 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 .

2) a) Show that if (X, d) is a metric space , so is $(X, \sqrt{D})$.

b) Show that if $d_{ij} = |i - j|$ then $\sqrt{d}$ is euclidean .

c) Assume we only want to count the number of seeks between distinct files using the IRM model with observed activity vector $A = (A_1, ..., A_n)$. Give a formula for the number of such seeks S_A , does it satisfy the inequality $S_A \geq S_B + S_C$ where $A = B + C$, all non negative vectors .

3) a) Assume we have an interleaved declustering configuration with N disks . Show that the problem of separating the activity of m randomly chosen active files is equivalent to the problem of wether all the components of a random graph with N nodes and m edges are trees and unicycles .

b) Show that being able to separate activity after the loss of any given disk is equivalent to wether all components of the corresponding graph are trees .

3 test spring 02

1) define or explain (10 points each)

A) Increasing subsequence

B) G/M/5/9

C) Euclidean Metric space

D) LSF

E) Disk scheduling

2) Requests arrive at a disk at a pace of 60 I/O per second. the seek time is linear and it takes 10 mili seconds to seek from one end of the disk to the other. A complete rotation takes 6 miliseconds and the transfer time is zero.

A) Compute the average response time of the disk

B) Compute the average queue length

3) 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 C) 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.

4) A) Write down the IRM estimate for the number of inter file seeks during a time interval in which activity A_i was observed for file i

B) Show that the IRM estimate for the number of seeks is non increasing as measurment points are added.

C) 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.

2) a) Assume uniformity of requests to a disk . Show that the average seek distance is r/3 where r is the disk radius (23)

b) assume that you have two files , each occupying half the disk . Assume that requests to the files are given in an alternating fashion and are uniformly distributed on each file . Compute the average seek in this case . (10)

3) a)Describe the main characteristics of RAID 1,3,4,5 . (15)

b) Explain the small write penalty for parity protected drives . Do bit interleave systems suffer from it ? explain . (10)

4) a) Describe efficient implementations of LRU and FIFO , Including the basic operations when reading and writing . (15)

b) Suppose that two caches A and B are initialized identically and are given the same request stream . Suppose A runs an MRU , LRU combination while B runs an MRU , FIFO combination . Suppose the corresponding heads of cache of A and B are identical at all times . What is A's hit ratio and what is B's . (10)