

CAPACITY PLANNING

19 july 2012

Name of the student

Exercise n. 1

Evaluate the reliability, the steady state availability and the safety of a Web server system, whose requests are served by a 2 servers, each server is composed by three computing nodes (each computing node is composed of a CPU and of a RAM) working in parallel and whose output is given by a voter and by a file server composed by a CPU and by RAID 1 system with four disks (2 +2). All the component are connected to two Ethernets there are working in parallel and whose output are given by a voter. The Ethernets are connected to internet by a route.

Hypothesize that the faults happen according to an exponential distribution with rate equal to λ_{CPU} , λ_{RAM} , λ_{DISK} , λ_{BUS} , λ_{ROUTER} , λ_{VOTER} , λ_{VOTER} , with covering factors equal to C_{CPU} , C_{RAM} , C_{DISK} , C_{BUS} , C_{ROUTER} , (the covering factor for the voters is equal to one) and with the same repair rate (μ) for all kind of components. A single "repair technician for component type" is available and the repair rate is independently by the number of fault occurrences of the same type.

N.B. Advice: evaluate each global index analyzing by single subsystem.

Exercise n. 2

Evaluate the average response time and throughput for a system composed of three servers working in parallel with a finite queue (at most 5 users in the system), with a finite number of users (number of users equal 8), given that the think time for each user is 20 sec and the service rate is 10 sec, hypothesing that the server can be faulty and repaired, the faults happen with an exponential distribution with rate equal to $\lambda_{server} = 1 \text{ fault}/12 \text{ months}$ and with a repair rate $\mu_{server} = 1 \text{ repair}/12 \text{ hours}$.

Exercise n.3

A web site receives 50 requests per second. The workload is equally distributed across 20 identical servers of a cluster. The maximum number of concurrent requests for a single server is 4, hence when there are 4 processing requests the server rejects a new arriving request. On average a request requires 8 CPU msec and 5 disk reads. A disk read requires 10 msec. Calculate the average request service time, the throughput and the percentage of rejected requests.

Moreover, what will be the performance improvement if the disk system has a cache with an access time of 100 μsec with a probability of hit equal to 70%?