

he returns to me
prove oral solution to

27

CAPACITY PLANNING

27 folio 11 APRIL 2014

Student's name and surname

ROBERTO PADULA 1186308

A web site is connected to a 1 Gbps Ethernet, which is connected to an ISP through a router. The router has a latency of $200\mu\text{sec}/\text{packet}$ and connects the Ethernet to the ISP through a line of 1 Gbps full duplex. The web site is realized with 7 computing node (a CPU with a main memory and a local disk, used only for the OS's purpose) and with a file system, realized with a CPU and with a RAID 4 with 8 (4+4) disks. The incoming rate is 100 requests/sec and each HTTP request for a file is of 400 bytes, the dimension for the requested file is 80 Kbytes, the load is equally distributed among the server nodes and the RAM is used as cache for the requested file. The P_{hit} is the probability of hit.

Each computing node can contemporaneously manage at most 4 users, instead the file system can manage an indefinite number of users.

The CPU service demand for each request is 10 msec in case of hit (file stored in the local main memory) and 30 msec in case of miss, the CPU-file-system service demand for each request is 5 msec, instead the service time of the each single disk for 10 Kbytes is 10 msec.

Compute the average response time in case all the components are fault-free, hypothesizing that the probability of hit is 30%. The lost requests are not taken into account in the evaluation of the response time.

Moreover evaluate the availability of the system hypothesizing the following parameters (the system can work even in degraded mode):

MTTF (MTTR) for the CPU with its main memory: 4 years (1 week)

MTTF (MTTR) for the Ethernet: 8 years (3 weeks)

MTTF (MTTR) for each disk: 3 year (4 weeks)

MTTF (MTTR) for the router: 15 year (1 week)

Finally identify all the acceptable configurations and show the methodology to evaluate the average response time given the presence of faults and reparations.