

22 September 2011

Student's name

ice ??

A web server platform is composed of two web servers (that see the same load), which are just a simple CPU and a RAM memory (that acts as cache), and by a file system, which is realized by a CPU and by four twin disks (they have the same data). The platform is connected to a 100 Mbps Ethernet, that is connected to an Internet Service Provider by a router, the router has a latency of  $40 \mu\text{sec}/\text{packet}$  and connects the Ethernet to the ISP through a line of 1 Mbps (that is equally divided for the incoming and outgoing traffic). There are two kind of HTTP requests to the web server: the first ones, that is the 60% of the total, requires files of 200 KB and to manage it the CPU time is 10 msec in case of HIT and 20 msec in case of MISS, the latter, instead, requires files of 100 KB and to manage it the CPU time is 5 msec in case of HIT and 10 msec in case of MISS. The CPU time of the file system to manage a request is 1 msec for each 20KB of data, instead the service time of the disk is 10 msec for each 20 KB of data. The probability of HIT is 30% and the HTTP incoming rate requests is 5 req/sec.

Evaluate the average response time for the incoming HTTP requests and identify the performance bottleneck.

Moreover evaluate the availability of the system in a parametric way (assuming, then, to know the MTTF and the MTTR of each kind of component of the system).

→ m  $10^{-3}$   
m  $10^{-6}$