CAPACITY PLANNING

24 may 2012

N°1

Name of the student

Users of a data center can ask for documents memorized or in a USER file system or in a Web server. The data center is composed by two local networks connected by a router (R1). The data center is connected to the ISP through a router (R2) that is also connected to the first LAN of the data center. In the first network, that is an Ethernet network, there is a proxy cache node equipped with a CPU (PROXY-CPU) and by a RAID1 system with 2 disks (1+1), able to memorize file coming from Web-servers.

The other network is an FDDI, where there is file server (USER file server) equipped with a 2 CPUs, working in parallel, and a 7 disks RAID5 system. Moreover in the same network there is a Web server characterized by 2 computing nodes that are working in parallel with the same load (there is a load balancer), characterized by a CPU (MAIN-CPU), and by a local file server composed by a CPU (FS-CPU) and by 4 disks working in parallel (hypothesize that the disks have the same data).

In the follow you have to hypothesize for the performance evaluation only the presence of web server requests; instead for the availability all the physical resources.

The users can **read** files from the USER file system or from the Web server system and their incoming rate is 3 req/sec.. 30% of their requests are directed to the file system and the residual 70% to the web server. In case of request to the Web server there is a probability of 50% that the file is in the proxy cache node. The average file dimensions from the file server is 30Kbytes, instead from the web server is of 200Kbytes, instead the requests have a dimension of 400 bytes.

The service demands for the USER file system are: 30 msec for the CPU and 30 msec for the disk subsystem.

The service demands for the Web server are: 50 msec of MAIN-CPU, 30 msec of FS-CPU and 40 msec for a disk .

The service demands for the proxy server node are: 10 msec of PROXY-CPU and disk in case of hit, 20 msec of PROXY-CPU and disk in case of miss.

The Ethernet network has a bandwith of 100 MB/sec and the FDDI of 500 MB/sec. The routers has a delay of 1 msec and the data center is connected to the ISP through a connection of 1Gbit/sec.

- Evaluate the average response time (from the incoming port to the output port of the connection) and identify the performance bottleneck.
- Evaluate in a parametric way the steady state availability of the system supposing the same fault rate for all CPUs (λ_{CPU}), for all disks (λ_{DISK}), for all networks (λ_{NET}) – indicate with (λ_{ROU}) the fault rate of the router. In case of reparation the repairman can repair all the faulty components with a repair rate equal to μ.
- Show the methodology to evaluate the performability of the system (average response time even in case of presence of faults).

CAPACITY PLANNING

24 may 2012

N°2

Name of the student

Users of a data center can ask for documents memorized or in a USER file system or in a Web server. The data center is composed by two local networks connected by a router (R1). The data center is connected to the ISP through a router (R2) that is also connected to the first LAN of the data center. In the first network, that is an Ethernet network, there is a proxy cache node equipped with a CPU (PROXY-CPU) and by a RAID1 system with 2 disks (1+1), able to memorize file coming from Web-servers.

The other network is an FDDI, where there is file server (USER file server) equipped with a 4 CPUs, working in parallel, and a 5 disks RAID5 system. Moreover in the same network there is a Web server characterized by 4 computing nodes that are working in parallel with the same load (there is a load balancer), characterized by a CPU (MAIN-CPU), and by a local file server composed by a CPU (FS-CPU) and by 6 disks working in parallel (hypothesize that the disks have the same data).

In the follow you have to hypothesize for the performance evaluation only the presence of web server requests; instead for the availability all the physical resources.

The users can **read** files from the USER file system or from the Web server system and their incoming rate is 4 req/sec.. 40% of their requests are directed to the file system and the residual 60% to the web server. In case of request to the Web server there is a probability of 50% that the file is in the proxy cache node. The average file dimensions from the file server is 40Kbytes, instead from the web server is of 300Kbytes, instead the requests have a dimension of 300 bytes.

The service demands for the USER file system are: 20 msec for the CPU and 20 msec for the disk subsystem.

The service demands for the Web server are: 20 msec of MAIN-CPU, 20 msec of FS-CPU and 40 msec for a disk .

The service demands for the proxy server node are: 10 msec of PROXY-CPU and disk in case of hit, 20 msec of PROXY-CPU and disk in case of miss.

The Ethernet network has a bandwith of 500 MB/sec and the FDDI of 1000 MB/sec. The routers has a delay of 1 msec and the data center is connected to the ISP through a connection of 2 Gbit/sec.

- Evaluate the average response time (from the incoming port to the output port of the connection) and identify the performance bottleneck.
- Evaluate in a parametric way the steady state availability of the system supposing the same fault rate for all CPUs (λ_{CPU}), for all disks (λ_{DISK}), for all networks (λ_{NET}) – indicate with (λ_{ROU}) the fault rate of the router. In case of reparation the repairman can repair all the faulty components with a repair rate equal to μ.
- Show the methodology to evaluate the performability of the system (average response time even in case of presence of faults).