Homework 2

CS 547

Due Friday, Feb. 10, in class

Check My Math

I measured a system with one disk for a long period of time and collected the following data:

- Average service time for a disk access = 5 ms
- Average number of disk accesses per job = 2
- Average number of jobs in the system at any moment = 12
- Average residence time of a job in the system = 1 second

Did I make a mistake?

Web Server Benchmarks

In a typical web server benchmark, a group of simulated clients will submit requests to the server. After receiving a response to its request, a client will "think" for some time before submitting another request.

In a Web server benchmark run with 40 clients, the average system residence time for a web page request was 1.5 seconds. The throughput at the disk on the web server was 16 accesses per second, and each request required, on average, two disk accesses.

What was the average client think time, in seconds?

On average, how many clients were thinking at any particular moment?

Asymptotic Bounds for a Storage System

Consider a storage system with 2 disks and a single CPU serving a closed population of customers. The measured parameters of the system are as follows:

- $\overline{s}_{cpu} = 1$ ms
- $\overline{V}_{cpu} = 2$ visits per request
- $\overline{s}_{disk1} = 20$ ms
- $\overline{V}_{disk1} = .65$ visit per request
- $\overline{s}_{disk2} = 25$ ms
- $\overline{V}_{disk2} = .35$ visit per request
- $N = 25$ customers

- $\overline{Z} = 5$ ms

Is the load on the disks balanced? What is the bottleneck resource?

Derive bounds on the throughput and residence time for the storage system as it's currently configured.

Suppose we double the number of customers in the system. What effect does this have on throughput and residence time?

Modification Analysis

Suppose we modify the storage system described in the previous problem by adding an in-memory buffer to serve reads more efficiently. Now, when the system performs a disk access, it will *read ahead* by a certain amount and store the extra data in the buffer. Due to sequential locality, we should be able to serve future requests out of the buffer, rather than reading from the disk.

After analyzing workloads, we believe that 75% of read requests will be satisfied from the buffer with an average service time of only 100 μs, but the need to read ahead will increase the average time for a disk access by 20%. Requests that are satisfied from the buffer do not need to perform a disk access.

What are the new average service times and visit counts at the two disks?

If all other parameters in the system remain the same, what effect will this modification have on throughput and response time? Does it change the system's bottleneck?

Is N too large or too small? Find a good operating value for the number of customers.