CS 547 Lecture 5: Storage System Example

Daniel Myers

Single Node Examples

Consider a node in a storage system with one CPU, two disks, and the following measurements:

- $\overline{s}_{CPU} = 2$ ms
- $\overline{s}_{disk1} = 50$ ms
- $\overline{s}_{disk2} = 30$ ms
- $\overline{V}_{disk1} = .375$ visits per request
- $\overline{V}_{disk2} = .625$ visits per request

Is the load on the disks balanced?
The load is balanced if both disks have the same demand.

$$\begin{aligned}\overline{D}_{disk1} &= \overline{D}_{disk2} \\ \overline{V}_{disk1}\overline{s}_{disk1} &= \overline{V}_{disk2}\overline{s}_{disk2} \\ (.375)(50) &= (.625)(30) \\ 18.75 &= 18.75\end{aligned}$$

Yes, the load is balanced.

Find an upper bound on throughput.
Use the bottleneck bound. The demand at the disk is the bottleneck.

$$\begin{aligned}\Lambda &\le \frac{1}{\overline{D}_{max}} \\ &\le \frac{1}{.001875 \text{ s}} \\ &\le 53 \text{ requests / sec}\end{aligned}$$

Suppose $\Lambda = 50$ requests per second. What is the CPU utilization?
This is a direct application of the Utilization Law.

$$\begin{aligned}U_{CPU} &= \Lambda_{CPU}\overline{s}_{CPU} = (50)(.002) \\ &= .10\end{aligned}$$

Suppose the average time for a request is 50 ms. What is the average time spent waiting?

We've been given $\overline{R} = 50$ ms. The average waiting time is the total residence time minus the total expected time spent in service, which is simply the sum of the demands.

$$\begin{aligned}\overline{W} &= \overline{R} - (\overline{D}_{CPU} + \overline{D}_{CPU} + \overline{D}_{CPU}) \\ &= 50 - (2 + 18.75 + 18.75) \\ &= 10.5 \text{ ms}\end{aligned}$$

What is the average number of requests at the storage node at any particular time?
This is a direct application of Little's result.

$$\begin{aligned}\overline{N} &= \Lambda\overline{R} \\ &= 50 * .05 \\ &= 2.5\end{aligned}$$

Closed System Examples

Suppose we group K nodes into a closed storage system with N total customers. Assume the demands at each node and each disk are balanced.

What are the average number of visits made per request to each CPU and each disk?
If there are K nodes, then there are K CPUs and $2K$ disks. Each request visits one of the K CPUs, followed by one of the $2K$ disks, so the expected visit counts are simply

$$\begin{aligned}\overline{V}_{CPU} &= \frac{1}{K} \\ \overline{V}_{disk} &= \frac{1}{2K}\end{aligned}$$

Let $\overline{s}_{CPU} = 2$ ms, $\overline{s}_{disk} = 50$ ms, and $\overline{Z} = 100$ ms. Calculate bounds on thoughput as a function of the number of customers if there are 100 nodes.
We need to calculate both throughput bounds for a closed system. First, calculate the total demand across the entire system, which is the sum of the demands at all the CPUs and all the disks.

$$\begin{aligned}\overline{D}_{total} &= K\overline{V}_{CPU}\overline{s}_{CPU} + 2K\overline{V}_{disk}\overline{s}_{disk} \\ &= \overline{s}_{CPU} + \overline{s}_{disk} \\ &= 52 \text{ ms}\end{aligned}$$

This makes sense: each request has to visit one CPU and one disk, so the average service accumulated in the entire system is just the sum of the service times at each of the two resources. The first throughput bound is

$$\begin{aligned}\Lambda &\leq \frac{N}{\overline{D}_{total} + \overline{Z}} \\ &\leq \frac{N}{.152} \text{ requests per sec.}\end{aligned}$$

The other bound is due to the maximum demand at any indvidual resource. The maximum demand is at the disk, where

$$\begin{aligned}\overline{D}_{max} &= \frac{1}{2K}\overline{s}_{disk} \\ &= \frac{1}{200}(50) \\ &= .25 \text{ ms.}\end{aligned}$$

The bottlneck bound is given by

$$\begin{aligned}\Lambda &\leq \frac{1}{\overline{D}_{max}} \\ &\leq 4000 \text{ requests per sec.}\end{aligned}$$

Which bound applies depends on the number of customers in the closed system.

Find a good value for the number of customers in the system.
The feasible operating point N^* is obtained by setting the two bounds equal and solving for N.

$$\begin{aligned}N^* &= \frac{\overline{D}_{total} + \overline{Z}}{\overline{D}_{max}} \\ &= 608\end{aligned}$$

Find a bound on the average residence time in the storage system if $N = 1000$.
Since $N > N^*$, the bottleneck bound applies and $\overline{D}_{max}$ is the constraint on throughput. The bound comes from using the maximum throughput in Little's result for a closed system.

$$\begin{aligned}\overline{R} &\geq \frac{N}{\Lambda_{max}} - \overline{Z} \\ &\geq \frac{1000}{4000} - .100 \\ &\geq .15\end{aligned}$$