IEOR 151 – Lecture 18
Review of Queuing Theory II

1 Little's Law

Suppose that we define the following variables

- L – average number of customers in system;
- λ – average arrival rate;
- W – average time in the system.

Then a useful relationship for queues is *Little's Law*, which states that

$$L = \lambda W.$$

To see why this relationship is useful, consider the M/M/1 queue from last lecture. There we showed that the average number of customers in the system is given by

$$L = \rho/(1-\rho), \text{ for } \rho = \lambda/\mu,$$

where λ is the average arrival rate and μ is the average service rate. Using Little's Law, we have that the average time in the system for a customer in an M/M/1 queue is given by

$$W = L/\lambda = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda}.$$

1.1 Variants of Little's Law

There are other versions of Little's Law. Suppose that we define the following additional variables

- L_q – average number of customers waiting to be served;
- L_s – average number of customers being served;
- W_q – average time in the queue waiting to be served;
- W_s – average service time.

Then, we also have

$$\begin{aligned} L_q &= \lambda W_q \\ L_s &= \lambda W_s. \end{aligned}$$

Also, note that because the average time in the system is the average time spent waiting to be served and the average serving time, we have that

$$W = W_q + W_s$$

Again returning to the example of the M/M/1 queue from last lecture, we have that

$$\begin{aligned} L_s &= \lambda/\mu = \rho \\ W_q &= W - 1/\mu = \frac{\mu - (\mu - \lambda)}{(\mu - \lambda)\mu} = \frac{\rho}{\mu - \lambda} \\ L_q &= \lambda W_q = \frac{\rho}{1/\rho - 1} = \frac{\rho^2}{1 - \rho}. \end{aligned}$$

2 M/M/s Queue with s Lines

Now, we turn our attention to a Markovian queue with s servers, average arrival rate λ, and average service rate for each queue of μ. First, we examine the situation in which there is a single line for each server. This is the situation at, for instance, Safeway. In our model, we will assume that each customer randomly chooses a line. Then, this is simply s distinct M/M/1 queues with average arrival rate λ/s and average service rate of μ, for each queue. Using the results for M/M/1 queues we have that

$$\begin{aligned} L &= s\left(\frac{\lambda/(s\mu)}{1 - \lambda/(s\mu)}\right) = \frac{\lambda/\mu}{1 - \lambda/(s\mu)} \\ W &= \frac{1}{\mu - \lambda/s}. \end{aligned}$$

3 M/M/s Queue with One Line

Now, we turn our attention to a Markovian queue with s servers, average arrival rate λ, and average service rate for each queue of μ. Here, we examine the situation in which there is a single line. This is the situation at, for instance, Fry's Electronics or the baggage check-in line for an airline at the airport. Some math gives that

$$\begin{aligned} L &= \frac{\rho}{1 - \rho} C(s, \rho) + s\rho \\ W &= \frac{C(s, \rho)}{s\mu - \lambda} + \frac{1}{\mu}, \end{aligned}$$

where

$$C(s,\rho) = \frac{\left(\frac{(s\rho)^s}{s!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{s-1}\frac{(s\rho)^k}{k!} + \left(\frac{(s\rho)^s}{s!}\right)\left(\frac{1}{1-\rho}\right)}$$

is the probability that all servers are occupied. An approximation is that

$$W_q \approx \frac{\left(\frac{\lambda}{s\mu}\right)^{\sqrt{2(s+1)}-1}}{s(1-\frac{\lambda}{s\mu})} \cdot \frac{1}{\mu}$$

It is interesting to compare an M/M/s queue with one line to an M/M/s queue with s lines. The time spent in an M/M/s queue with s lines is longer than the time spent in an M/M/s queue with one line. The intuition is that having just one line allows for greater utilization of all s servers.

4 M/M/∞ Queue

Some systems are modeled using an infinite number of servers. In this model, the service rate is state-dependent and is given by $n\mu$ where n is the number of customers in line, and μ is the service rate for a single customer. Some calculations give that $L = \lambda/\mu$ and $W = 1/\mu$.

5 More Information and References

The material in these notes follows that of the course textbook "Service Systems" by Mark Daskin and of the Wikipedia article on "Poison process".