COURSE 436
PERFORMANCE ANALYSIS
P.G. HARRISON
Solutions to coursework 1
Spring term 2010
Unassessed

Solutions

1. The bottleneck in a network of service centres is the one with the maximum demand, i.e. average amount of service required by all tasks in unit time. It can be shown that demand is proportional to server utilisation.

   (a) In a closed network, i.e. one with a constant population of tasks, K, what happens to the location of the bottleneck as K increases?

   (b) What happens to the bottlenecks utilisation as $K \to \infty$?

   (c) What happens to the population at non-bottleneck servers as $K \to \infty$?

   (d) Which servers should you speed up to best improve performance?

   (e) What should the server utilisations be for optimal performance?

   Solution 1 *(a) Demand per task depends only on workload and so the ordering of nodes by workload by task satays the same as K increases. Hence bottleneck stays in the same place.*

   (b) As $K \to \infty$ at least one node saturates and so has utilisation 1. This is the largest possible utilisation, so the bottleneck must get utilisation 1.

   (c) If there is a single bottleneck, the other nodes must get utilisation < 1, so have finite queues with probability 1

   (d) The bottleneck, since its maximum throughput is its service rate (utilisation approaches 1 as $K \to \infty$) and this limits the performance of the rest of the network; this is clarified in the last lecture's notes.

   (e) You can always increase performance by upgrading the bottleneck. Hence optimum performance when all servers are equal bottlenecks, i.e. all have the same demand (λ_i/μ_i).

2. Consider a simple queue with arrival rate λ, mean service time m and standard deviation of service time σ.

   (a) If σ is always proportional to m, what would you expect to happen to response time if the arrival rate and service rate both double?

   (b) If σ doubles whilst λ and m remain fixed, what would you expect to happen to average response time?

Solution 2

(a) *Response time should halve, since, if it depends on no other parameters, all we have done is to change the basic time unit – counting half seconds instead of seconds.*

(b) *Now we're not just changing the time unit. When service times are highly variable response time might increase or decrease for a particular visit to the queue. What do you think would happen to mean response time? Would you like to use a highly erratic system or a consistent one? Most people prefer the latter and this is backed up in the M/G/1 queue analysis.*

3. Consider a queue in steady state with arrival rate λ, service rate μ and utilisation U. How is the *throughput* related to the arrival rate? Prove that $U = \lambda/\mu$.

Solution 3 *Arrival rate = throughput in steady state, so* λ = *utilisation* $\times$ *service rate.*

4. Suppose that the amount of time that a light-bulb works before burning itself out is exponentially distributed with mean ten hours. Suppose that a person enters a room in which such a light-bulb is burning. If this person desires to work five hours, then what is the probability that he will complete his work without the bulb burning out? What can be said if the distribution is not exponential?

Solution 4 *We write* X *for the random variable of the remaining lifetime of the bulb. Thus, since we know that the lifetime is exponentially distributed parameter* $\lambda = \frac{1}{10}$, *i.e. mean* $\frac{1}{\lambda} = 10$, *write:*

$$
\begin{aligned}
P(X > t + 5 | X > t) &= \frac{P(X > t \wedge X > t + 5)}{1 - F(t)} \\
&= \frac{P(X > t + 5)}{1 - F(t)} \\
&= \frac{1 - F(t + 5)}{1 - F(t)}. \\
&= \frac{1 - (1 - e^{-\lambda(t+5)})}{1 - (1 - e^{-\lambda(t)})} \\
&= \frac{e^{-\lambda(t+5)}}{e^{-\lambda t}} \\
&= e^{-\lambda 5} \\
&= e^{-\frac{1}{2}}
\end{aligned}
$$

Perfectly acceptable and shorter solution invoking the memoryless property of the exponential distribution:

$$P(X > 5) = 1 - F(5) = 1 - (1 - e^{-5\lambda}) = e^{-\frac{1}{2}}$$

If the lifetime distribution F(t) were not exponential, then we would calculate the probability in the following way:

$$\begin{aligned} P(X > t + 5 | X > t) &= \frac{P(X > t \wedge X > t + 5)}{1 - F(t)} \\ &= \frac{1 - F(t + 5)}{1 - F(t)}. \end{aligned}$$

Notice that in order to calculate the latter, we need to keep the information about t, which was unnecessary in the previous case.

5. Given two exponentially distributed random variables X_1 and X_2, determine the probability that one is smaller than the other?

   Solution 5 *Assume that you have two independent exponential random variables X_1 and X_2 with rates λ_1 and λ_2 respectively. Thus:*

$$\begin{aligned} P(X_1 < X_2) &= \int_0^\infty P(X_1 < X_2 | X_2 = x) \lambda_2 e^{-\lambda_2 x} dx \\ &= \int_0^\infty P(X_1 < x) \lambda_2 e^{-\lambda_2 x} dx \\ &= \int_0^\infty (1 - e^{-\lambda_1 x}) \lambda_2 e^{-\lambda_2 x} dx \\ &= \int_0^\infty \lambda_2 e^{-\lambda_2 x} dx - \lambda_2 \int_0^\infty e^{-(\lambda_1 + \lambda_2)} dx \\ &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \end{aligned}$$

6. The times taken for transactions from sites A and B are exponentially distributed with means $\frac{1}{\lambda}$ and $\frac{1}{\mu}$ minutes respectively on the host computer system. If two transactions T_A and T_B arrive from A and B respectively and their service is started immediately in both cases, what is the probability that T_B finishes first? What if they are not both started at the same time?

Solution 6 *Let X_A and X_B be the random variables for the lifetimes of the transactions T_A and T_B respectively. Thus, by using the solution of question (2), we have*

$$P(X_B < X_A) = \frac{\mu}{\lambda + \mu}.$$

No change if the transactions do not start at the same time, provided one has not finished before the other starts, by the memoryless property.

7. Customers arrive to a supermarket according to a Poisson process with rate λ per hour. Suppose that two customers arrive during the first hour. Find the probability that

   - both arrived in the first 20 minutes;
   - at least one of them arrived in the last 30 minutes.

Solution 7 *Let the random variable $N(1) = 2$ mean that two customers have arrived in an hour, and $N(\frac{1}{3} = 2)$ mean that both customers have arrived in the first twenty minutes.*

$$\begin{aligned}
P(N(\tfrac{1}{3}) = 2|P(N(1) = 2)) &= \frac{P(N(\frac{1}{3} = 2 \wedge N(1) = 2))}{P(N(1) = 2)} \\
&= \frac{P(N(\frac{1}{3} = 2 \wedge (N(1) - N(\frac{1}{3})) = 0))}{P(N(1) = 2)} \\
&= \frac{P(N(\frac{1}{3} = 2 \wedge N(\frac{2}{3}) = 0))}{P(N(1) = 2)} \\
&= \frac{P(N(\frac{1}{3} = 2)P(N(\frac{2}{3}) = 0)}{P(N(1) = 2)} \\
&= \frac{\dfrac{e^{-\lambda\frac{1}{3}}((\frac{\lambda}{3})^2)e^{-\frac{\lambda 2}{3}}}{2!}}{\dfrac{e^{-\lambda}\lambda^2}{2!}} \\
&= \tfrac{1}{9}
\end{aligned}$$

For part (2) the answer is quite similar with $N(\frac{1}{2}) = 1$ meaning that one *customer arrived in half an hour.*

$$
\begin{aligned}
P(N(\tfrac{1}{2}) \geq 1 | N(1) = 2) &= 1 - P(N(\tfrac{1}{2}) = 0 | N(1) = 2) \\
&= 1 - \frac{P(N(\frac{1}{2}) = 0 \wedge N(1) = 2)}{P(N(1) = 2)} \\
&= 1 - \frac{P(N(\frac{1}{2}) = 0 \wedge (N(1) - N(\frac{1}{2})) = 2)}{P(N(1) = 2)} \\
&= 1 - \frac{P(N(\frac{1}{2}) = 0 \wedge N(\frac{1}{2}) = 2)}{P(N(1) = 2)} \\
&= 1 - \frac{\frac{e^{-\frac{\lambda}{2}} e^{-\frac{\lambda}{2}} (\frac{\lambda}{2})^2}{2!}}{\frac{e^{-\lambda}\lambda^2}{2!}} \\
&= \tfrac{3}{4}.
\end{aligned}
$$

8. Let T_n be the instant of the nth arrival in a Poisson process with rate λ. Show that the distribution function of T_n, $F_n(x)$ is given by:

$$F_n(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}$$

Solution 8 *If T_n is the instant of the nth arrival, then:*

$$
\begin{aligned}
P(T_n \leq t) &= P(N(t) \geq n) \\
&= 1 - P(N(t) < n).
\end{aligned}
$$

Since $P(N(t) < n) = P(N(t) = 0 \vee N(t) = 1 \vee \ldots \vee N(t) = n - 1)$ and the events $N(i)$ are disjoint, $(i = 0, 1, \ldots, n - 1)$, then we have:

$$F_n(x) = 1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}.$$