Traffic Engineering. Mid-Term Exam 2016

February 14, 2022

Problem 1 - 5 points

A Web server can be modelled by a M/M/2/3 queueing system, where service requests arrive at rate λ request/second following a Poisson process. The two CPUs of the Web server can operate in two modes: a) low-energy consumption, and b) high-energy consumption. In mode (a), each CPU is able to process requests at rate μ_a requests/second, and in mode (b), each CPU is able to process requests at rate μ_b, with $\mu_b = k\mu_a$, and k a positive integer constant. Both $1/\mu_a$ and $1/\mu_b$ are exponentially distributed. The system operates in low-energy consumption if the two servers are not busy. Otherwise, it operates in high-energy consumption.

Questions:

1. (2 points) Draw the Markov chain that models the system, write all the balance equations and write the expressions to compute the stationary probabilities π_i for all states.
2. (1.5 point) If $\mu_a = \lambda$, and $\mu_b = k\mu_a$, compute the probability that the system is empty for large values of k.
3. (1.5 point) If $\mu_a = \lambda$, and $\mu_b = 2\mu_a$ (i.e., $k = 2$), what the reduction of the blocking probability will be compared with the case when $k = 1$.

Problem 2 - 5 points

1. (2.5 points) Which of the following two cases has a lower end-to-end delay:
   (a) A M/M/1 queue where packets arrive at rate λ and depart at rate μ.
   (b) Two concatenated M/M/1 queues, where packets arrive at rate λ to the first queue. The two queues have a service rate equal to $\mu' = 2\mu$. The propagation delay between the first and second queue is negligible.
2. (2.5 points) What is the value of μ' that makes the end-to-end delay equal in the two cases. Explain the obtained result.

Solution

The Markov chain has 4 states: $\{0,1,2,3\}$, where all forward transitions are λ and backward transitions are respectively $\mu_a, 2\mu_b, 2\mu_b$. Then, the equilibrium distribution is (obtained from the balance equations):

$$\begin{aligned}
\pi_1 &= \frac{\lambda}{\mu_a}\pi_0 = a\pi_0 \\
\pi_2 &= \frac{\lambda}{2\mu_b}\pi_1 = \frac{\lambda}{2\mu_b}\frac{\lambda}{\mu_a}\pi_0 = \frac{\lambda^2}{2k\mu_a^2}\pi_0 = \frac{a^2}{2k}\pi_0 \\
\pi_3 &= \frac{\lambda}{2\mu_b}\pi_2 = \frac{\lambda}{2\mu_b}\frac{\lambda}{2\mu_b}\frac{\lambda}{\mu_a}\pi_0 = \frac{\lambda^3}{4k^2\mu_a^3}\pi_0 = \frac{a^3}{4k^2}\pi_0 \\
\pi_0 &= \frac{1}{1+a+\frac{a^2}{2k}+\frac{a^3}{4k^2}}
\end{aligned}$$

If $\mu_a = \lambda$, and $\mu_b = k\mu_a$, we have that $a = 1$.

$$\pi_0 = \frac{1}{1+a+\frac{a^2}{2k}+\frac{a^3}{4k^2}} = \frac{1}{1+1+\frac{1}{2k}+\frac{1}{4k^2}}$$

For large values of k:

$$\lim_{k\to\infty} \frac{1}{1+1+\frac{1}{2k}+\frac{1}{4k^2}} = \frac{1}{2}$$

When $k = 2$, the blocking probability will be:

$$\pi_b = \frac{\frac{1}{4k^2}}{1+1+\frac{1}{2k}+\frac{1}{4k^2}} = \frac{\frac{1}{16}}{1+1+\frac{1}{4}+\frac{1}{16}} = \frac{\frac{1}{16}}{\frac{37}{16}} = \frac{1}{37}$$

Compared to the case when $k = 1$, where we have:

$$\pi_b = \frac{\frac{1}{4k^2}}{1+1+\frac{1}{2k}+\frac{1}{4k^2}} = \frac{\frac{1}{4}}{1+1+\frac{1}{2}+\frac{1}{4}} = \frac{\frac{1}{4}}{\frac{11}{4}} = \frac{1}{11},$$

we can observe that the reduction is bigger than a factor of 2.

Solution Exercise 2

We know that the delay of an M/M/1 queue is $E[D]_{1q} = \frac{1}{\mu-\lambda}$. Therefore, the delay of two consecutive queues is $E[D]_{2q} = \frac{2}{2\mu-\lambda}$. Comparing both, assuming that $\lambda < \mu$, as otherwise the first case would not be stable, we observe that $\frac{1}{\mu-\lambda} > \frac{1}{\mu-\frac{\lambda}{2}}$, which tells us the second case is faster.

To compute the value of μ' that makes both delays equal, we do:

$$\begin{aligned}
\frac{1}{\mu-\lambda} &= \frac{2}{\mu'-\lambda} \\
\mu'-\lambda &= 2(\mu-\lambda) \\
\mu' &= 2\mu - 2\lambda + \lambda \\
\mu' &= 2\mu - \lambda
\end{aligned}$$

This is a very interesting result. For low λ values, we need μ' values twice higher than μ. However, for λ values close to μ, we have that $\mu' \approx \mu$, which is a non intuitive result at all. It can be justified because when λ tends to μ, $E[D]_{1q}$ behaves asymptotically, and any small variation in that range causes a large difference in delay (the response is not linear).