15.072 Queues: Theory and Application

HW 4 Solutions

Apr 24, 2006

Problem 1. Consider a G/M/1 queueing system where the distribution of the interarrival times is a mixture of two exponential distributions with parameters $\lambda_1 = 1$ and $\lambda_2 = 2$ and probabilities $p_1 = .3, p_2 = 1-p_1$. The service rate is assumed to be $\mu = 2$. Determine the probability distribution of the waiting time in steady state.

Solution:
For a $G/M/1$ queue the waiting time distribution is given by $\mathbb{P}(W = 0) = 1 - \sigma$ and $\mathbb{P}(W > t) = \sigma e^{-\mu(1-\sigma)t}$. where $0 < \sigma < 1$ solves $\sigma = A^*(\mu(1 - \sigma))$. Here,

$$\begin{aligned}
A^*(s) &= \mathbb{E}[e^{-sA}] \\
&= p_1 \frac{\lambda_1}{\lambda_1 + s} + p_2 \frac{\lambda_2}{\lambda_2 + s} \\
&= \frac{1}{10}\left(\frac{3}{s+1} + \frac{14}{s+2}\right) \\
&= \frac{2 + 1.7s}{(s+1)(s+2)}
\end{aligned}$$

Let $y \triangleq \mu(1 - \sigma)$ Hence, $\sigma = 1 - \frac{y}{\mu}$. Then,

$$\begin{aligned}
\sigma &= A^*(\mu(1-\sigma)) \\
\Rightarrow 1 - \frac{y}{2} &= \frac{2 + 1.7y}{(y+1)(y+2)} \\
\therefore (y+1)(y+2)(y-2) + 4 + 3.4y &= 0 \\
\therefore y^3 + y^2 - 0.6y &= 0
\end{aligned}$$

The above equation has roots $y = 0, \frac{-1+\sqrt{3.4}}{2}, \frac{-1-\sqrt{3.4}}{2}$. The root that leads to an admissible value of σ is $\frac{\sqrt{3.4}-1}{2}$. The corresponding value of σ is $\frac{5-\sqrt{3.4}}{4} \approxeq 0.7890$. Thus the waiting time distribution is $\mathbb{P}(W = 0) = 1 - \sigma$ and $\mathbb{P}(W > t) = \sigma e^{-\mu(1-\sigma)t}$ with $\sigma = \frac{5-\sqrt{3.4}}{4} \approxeq 0.7890$.

Problem 2. Consider a G/M/1 queueing system with interarrival times A and service rate μ. Prove that the expected steady state number of customers in the system observed by an arriving customer diverges to infinity in the limit as μ approaches $1/\mathbb{E}[A]$.

HINT: express the expected number in the system in terms of σ and prove that σ approaches unity as $\mu \to 1/\mathbb{E}[A]$. Use the second order Taylor expansion to establish this fact.

Solution:
We know that the steady state number of customers observed at the arrival instance is $\mathbb{E}[L^-] = \frac{\sigma^*}{1-\sigma^*}$, where σ^* is the unique solution of $\sigma^* = A^*(\mu(1-\sigma^*))$ in the open interval $(0,1)$. Thus it suffices to show that $\sigma^* \to 1$ as $\mu \to 1/\mathbb{E}[A]$, or equivalently $\rho \to 1$. Consider

$$G(\sigma) = \mathbb{E}[e^{-\mu(1-\sigma)A}] = \int_0^\infty e^{-\mu(1-\sigma)s} dA(s).$$

We computed in the class the first two derivatives of G at $\sigma = 1$:

$$\left.\frac{dG}{d\sigma}\right|_{\sigma=1} = \int_0^\infty (-\mu)(-s) dA(s) = \mu\mathbb{E}[A] = \frac{1}{\rho}$$
$$\left.\frac{d^2G}{d\sigma^2}\right|_{\sigma=1} = \int_0^\infty \mu^2 s^2 dA(s) = \mu^2\mathbb{E}[A^2].$$

Now we take constant c such that

$$\frac{c\mu^2\mathbb{E}[A^2]}{2} - \frac{1}{\rho} > 0 \tag{1}$$

when $\rho > 1/2$. For example, simply take $c > 4/(\mu^2\mathbb{E}[A^2])$.

Claim 1. *Take $\sigma_0 = 1 - c(1-\rho)$. Then $G(\sigma_0) > \sigma_0$ when ρ is sufficiently close to* 1.

Before proving the claim let us see how it implies the result. We have established that $G(\sigma)$ is a strictly convex function. Therefore $G(\sigma) > \sigma$ for all $\sigma < \sigma_0$ and the unique solution σ^* to $G(\sigma) = \sigma$ must be in the region $(\sigma_0, 1)$. But $\sigma_0 \to 1$ as $\rho \to 1$, implying $\sigma^* \to 1$ and we are done.

Proof of the claim: Consider a second order Taylor expansion of $G(\sigma)$ around $\sigma = 1$. We have

$$\begin{aligned} G(\sigma_0) &= 1 + (\sigma_0 - 1)\frac{1}{\rho} + (\sigma_0 - 1)^2 \frac{\mu^2\mathbb{E}[A^2]}{2} + o((\sigma_0 - 1)^2) \\ &= 1 - c(1-\rho)\frac{1}{\rho} + c^2(1-\rho)^2 \frac{\mu^2\mathbb{E}[A^2]}{2} + o((1-\rho)^2) \end{aligned}$$

Now consider

$$
\begin{aligned}
\frac{G(\sigma_0)-\sigma_0}{(1-\rho)^2} &= \frac{c(1-\rho)(-\frac{1}{\rho}+1)+c^2(1-\rho)^2\frac{\mu^2\mathbb{E}[A^2]}{2}+o((1-\rho)^2)}{(1-\rho)^2} \\
&= \frac{c(1-\rho)^2\left(c\frac{\mu^2\mathbb{E}[A^2]}{2}-\frac{1}{\rho}+o(1)\right)}{(1-\rho)^2} \\
&= c\frac{\mu^2\mathbb{E}[A^2]}{2}-\frac{1}{\rho}+o(1).
\end{aligned}
$$

When ρ approaches unity, certainly $\rho > 1/2$. Then from (**??**) and since $o(1)$ is a value approaching zero as $\rho \to 1$, we obtain that the expression above is positive for ρ sufficiently close to unity. This completes the proof. □

Problem 3. Suppose two call centers A and B with m_A and m_B agents respectively, serve a demand from the same pool. The demand is Poisson with rate λ. The service rate in the two centers is identical μ. Upon arrival of a call the scheduler needs to route the call to one of the two centers. It is too costly to make the routing decision based on the number of calls in progress. As a result the decision needs to be done in an oblivious way. Only the the values of λ, μ, m_A, m_B are known, and with some probability p $(1-p)$ each call is routed to call center A (call center B), independently from everything else.

1. Propose a routing scheme $p = p(\lambda, m_A, m_B, \mu)$ which minimizes the steady state cost given as $\mathbb{P}(W_A > 0) + \mathbb{P}(W_B > 0)$, where W_A, W_B are waiting times in the call centers A, B, respectively.
2. Suppose $\lambda = 6400, \mu = 10, m_A = 300, m_B = 400$. Find approximately the optimal p up to 2 decimal points.

   HINT: Standard normal table can be found on the internet ...

Solution:

We will use Halfin-Whitt approximations. For a probability p of routing the call to center A, the center A sees Poisson arrivals with rate $p\lambda$ and the call center B sees Poisson arrivals with rate $(1-p)\lambda$. Then $\beta_A(p) \stackrel{\triangle}{=} \sqrt{m_A}(1-\frac{p\lambda}{m_A\mu})$ and $\beta_B(p) \stackrel{\triangle}{=} \sqrt{(m_B)}(1-\frac{(1-p)\lambda}{m_B\mu})$. Then using the Halfin-Whitt approximations,

(a)

$$
\begin{aligned}
C(p) &= \mathbb{P}(W_A > 0) + \mathbb{P}(W_B > 0) \\
&= \frac{1}{1+\sqrt{2\pi}\beta_A(p)\Phi(\beta_A(p))e^{\frac{\beta_A^2(p)}{2}}} + \frac{1}{1+\sqrt{2\pi}\beta_B(p)\Phi(\beta_B(p))e^{\frac{\beta_B^2(p)}{2}}}
\end{aligned}
$$

One should choose that value of p that minimizes the above cost function. The allowable range for p are those values for which $\beta_A(p), \beta_B(p) > 0$ i.e., $1 - \frac{m_B \mu}{\lambda} < p < \frac{m_A \mu}{\lambda}$ and also $0 \le p \le 1$. As the function $C(p)$ is convex, the minimizing p^* will be interior and can be found by first order conditions i.e., solving $C'(p) = 0$.

(b) The admissible range for p is $(\min(0, 1 - \frac{m_B \mu}{\lambda}), \max(1, \frac{m_A \mu}{\lambda}))$ i.e., $(0.375, 0.46875)$. The value of the cost function for p in this range is tabulated in Table **??**. We see the optimal value p^*

p	$\beta_A(p)$	$\beta_B(p)$	$\mathbb{P}(W_A > 0)$	$\mathbb{P}(W_B > 0)$	$C(p)$
0.375	3.464	0.000	0.000	1.000	1.000
0.380	3.279	0.160	0.001	0.814	0.814
0.385	3.095	0.320	0.001	0.654	0.655
0.390	2.910	0.480	0.002	0.520	0.522
0.395	2.725	0.640	0.004	0.407	0.411
0.400	2.540	0.800	0.006	0.315	0.321
0.405	2.356	0.960	0.011	0.240	0.250
0.410	2.171	1.120	0.017	0.180	0.197
0.415	1.986	1.280	0.028	0.132	0.160
0.420	1.801	1.440	0.043	0.096	0.139
0.425	1.617	1.600	0.066	0.068	0.134
0.430	1.432	1.760	0.098	0.048	0.145
0.435	1.247	1.920	0.141	0.033	0.174
0.440	1.062	2.080	0.200	0.022	0.222
0.445	0.878	2.240	0.276	0.014	0.291
0.450	0.693	2.400	0.375	0.009	0.384
0.455	0.508	2.560	0.498	0.006	0.504
0.460	0.323	2.720	0.651	0.004	0.655
0.465	0.139	2.880	0.837	0.002	0.839
0.470	-0.046	3.040	1.059	0.001	1.060

Table 1: Variation of $C(p)$ with p

correct to two decimal places is 0.42.

Problem 4. Consider k identical call centers A which are in a Halfin-Whitt regime. Each of them is characterized by parameters $m, \mu, \lambda = m\mu - \beta\mu\sqrt{m}$, where, as usual, m, μ, λ stand for the number of agents, service rate and the arrival rate. Suppose we merge these k call centers into one call center kA. Prove that the probability of waiting $\mathbb{P}(W > 0)$ in the merged call center converges to zero geometrically fast as a function of k. Namely, $\mathbb{P}(W > 0) \le \delta^k$ for some $\delta < 1$. Obtain a bound on δ in terms of the parameter β of the original call center.

Solution:
We have

$$\begin{aligned}
\beta_k &= \sqrt{km}(1 - \frac{k\lambda}{km\mu}) \\
&= \sqrt{k}\sqrt{m}(1-\rho) \\
&= \sqrt{k}\beta
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{P}(W_k > 0) &= \frac{1}{1 + \sqrt{2\pi}\beta_k \Phi(\beta_k) e^{\frac{\beta_k^2}{2}}} \\
&= \frac{1}{1 + \sqrt{2\pi}\beta\sqrt{k}\Phi(\beta\sqrt{k}) e^{\frac{\beta^2 k}{2}}} \\
&\leq \frac{1}{\sqrt{2\pi}\sqrt{k}\beta\Phi(\beta)}(e^{-\beta^2})^k \\
&\leq (e^{-\beta^2})^k \quad ; \ldots k > \frac{1}{(\beta\Phi(\beta))^2}
\end{aligned}$$

Thus $\mathbb{P}(W < 0)$ converges to 0 geometrically fast. A bound for δ, the rate of convergence, is $\delta \leq e^{-\beta^2}$, $\beta > 0$.