3.38

(a) Let N_{Qk} = number in queue for priority k

W_k = queueing (waiting) time for priority k

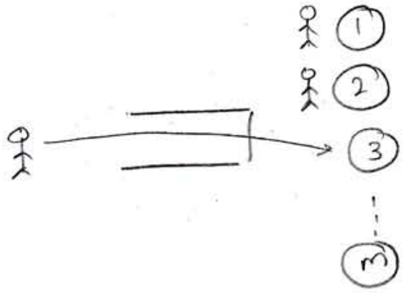
R = residual service time

ρ_k = $\frac{\lambda_k}{\mu}$ = utilization for priority k

N = number of customers in the system

* First, we evaluate $E[R]$. Consider the two following cases:

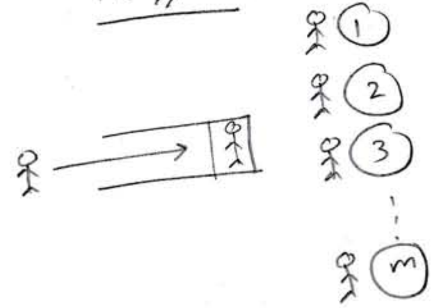
$N < m$



an arriving customer goes immediately to service i.e. $R=0$

$E[R \mid N<m] = 0$

$N \geq m$



- in this case, the residual service time is the time until one server empties
- Let $\tilde{T}_i$, $i=1,\dots,m$, be the service time of customer at server i

$\tilde{T}_i$: exponential, mean $= \frac{1}{\mu}$

$$R = \min\{\tilde{T}_1, \dots, \tilde{T}_m\}$$

$$\begin{aligned} \mathbb{P}(R \leq t) &= \mathbb{P}(\min\{\tilde{T}_i\} \leq t) \\ &= 1 - \mathbb{P}(\min\{\tilde{T}_i\} > t) \\ &= 1 - \prod_{i=1}^{m} \mathbb{P}(\tilde{T}_i > t) \\ &= 1 - (e^{-\mu t})^m \\ &= 1 - e^{-m\mu t} \end{aligned}$$

more accurately, this should be written as $\mathbb{P}(R \leq t \mid N \geq m)$

$$E[R \mid N \geq m] = \frac{1}{m\mu}$$

$$E[R] = \underbrace{\mathbb{P}(N<m)}_{1-P_Q}\ \underbrace{E[R\mid N<m]}_{0} + \underbrace{\mathbb{P}(N\geq m)}_{P_Q}\ \underbrace{E[R\mid N\geq m]}_{\frac{1}{m\mu}}$$

$$\therefore\ E[R] = \frac{P_Q}{m\mu}$$

* Next, we evaluate $E[W_k]$

priority class 1 $\quad E[W_1] = E[R] + \frac{1}{m\mu} E[N_{Q1}]$

Intuitively speaking,

$$\text{avg waiting time} = \text{avg residual time} + \underbrace{\text{avg time it takes to serve the customers of priority 1 in the queue}}_{=\ \frac{\text{avg \# of customers} \times \text{avg service time for each}}{m}}$$

(we divide by m because the servers work in parallel)

From Little's law, we have

$$E[N_{Q1}] = \lambda_1 E[W_1]$$

Therefore, $E[W_1] = \dfrac{E[R]}{1-\rho_1}$, $\rho_1 = \dfrac{\lambda_1}{m\mu}$

priority class 2 $\quad E[W_2] = E[R] + \frac{1}{m\mu} E[N_{Q1}] + \frac{1}{m\mu} E[N_{Q2}] + \frac{1}{m\mu}\lambda_1 E[W_2]$

Using Little's theorem, we have $E[N_{Q1}] = \lambda_1 E[W_1]$ and $E[N_{Q2}] = \lambda_2 E[W_2]$

$$\therefore\ E[W_2] = E[R] + \rho_1 E[W_1] + \rho_2 E[W_2] + \rho_1 E[W_2]$$

$$\therefore\ E[W_2] = \frac{E[R] + \rho_1 E[W_1]}{1-\rho_1-\rho_2}\ ,\quad \rho_1 = \frac{\lambda_1}{m\mu} \text{ and } \rho_2 = \frac{\lambda_2}{m\mu}$$

$$\therefore\ E[W_2] = \frac{E[R]}{(1-\rho_1)(1-\rho_1-\rho_2)}$$

priority class k $\quad E[W_k] = \dfrac{E[R]}{(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)}$ where $\rho_k = \dfrac{\lambda_k}{m\mu}$

and $E[R] = \dfrac{P_Q}{m\mu}$ (P_Q as defined in sec. 3.4.1 with $\rho = \rho_1+\rho_2+\cdots+\rho_n$)

Check: substitute $m=1$,

$$\rho_k = \frac{\lambda_k}{\mu}$$

$$E[W_k] = \frac{E[R]}{(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)}$$

$$P_Q = 1-P_0 = 1-(1-\rho) = \rho = \rho_1+\rho_2+\cdots+\rho_n = \sum_{i=1}^{n}\frac{\lambda_i}{\mu}$$

$$E[R] = \frac{P_Q}{1\times\mu} = \sum_{i=1}^{n}\lambda_i\frac{1}{\mu^2}$$

Compare that with what we get by considering exponential service time in the M/G/1 expressions

$$E[W_k] = \frac{E[R]}{(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)}$$

$$E[R] = \frac{1}{2}\sum_{i=1}^{n}\lambda_i\overline{X^2}$$

$X \sim$ exponential with mean $= \frac{1}{\mu}$

$$\Rightarrow E[X^2] = \frac{2}{\mu^2}$$

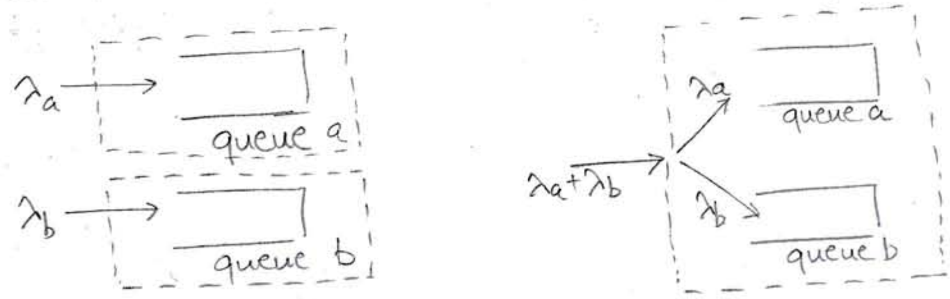
$$\therefore E[R] = \sum_{i=1}^{n}\lambda_i\frac{1}{\mu^2}$$ ✓ (expressions match)

3.38

(b) Define $W_{(k)}$ = average time in queue averaged over the first k priorities

Obviously, $W_{(1)} = W_1$

<u>Note</u> Consider two types of traffic: type a and type b

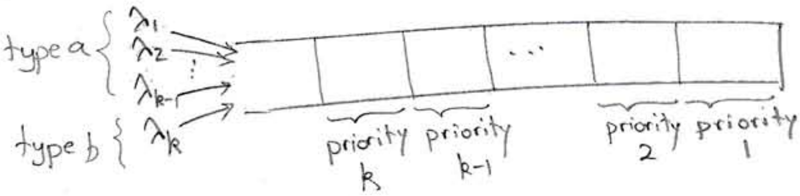


$N_{Qa} = \lambda_a W_a$, $N_{Qb} = \lambda_b W_b$ $\qquad N_Q = (\lambda_a + \lambda_b) W$

$\because N_Q = N_{Qa} + N_{Qb}$

$\therefore (\lambda_a + \lambda_b) W = \lambda_a W_a + \lambda_b W_b$ $\qquad$ (Recall prob 3.9) $\qquad$ #

We will use Little's theorem in a similar manner.



We have $\left(\sum_{i=1}^{k-1} \lambda_i + \lambda_k\right) W_{(k)} = \left(\sum_{i=1}^{k-1} \lambda_i\right) W_{(k-1)} + \lambda_k W_k$ $\qquad k = 1, \dots, n$

$\therefore W_k = \frac{1}{\lambda_k}\left[\left(\sum_{i=1}^{k} \lambda_i\right) W_{(k)} - \left(\sum_{i=1}^{k-1} \lambda_i\right) W_{(k-1)}\right]$

$W_{(k)}$ is the average waiting time of an M/M/m with arrival rate $\lambda = \sum_{i=1}^{k} \lambda_i$ and mean service time $\frac{1}{\mu}$