**SOLUTIONS**

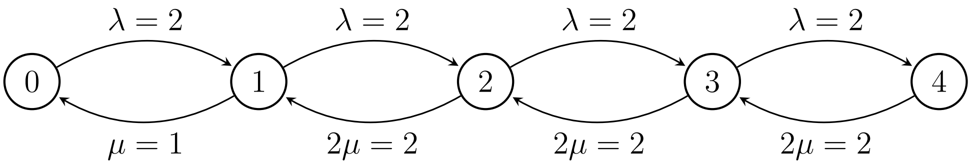
**Final exam Stochastic Modelling**

December 21, 2022

**Question 1.** Consider an $M/M/2/4$ system where customers (attempt to) arrive at rate $\lambda = 2$ and each of the two servers has service rate $\mu = 1$. Recall that, next to the two spots at the servers, this system only has two spots in the waiting room. Those customers which, upon arrival, do not find a free server or a free spot in the waiting room are lost.

**(a) [2pt]** Formulate a CTMC based on which you can answer the subsequent parts of the question.

**Solution** $L(t)$ = number of customers in the system at time $t$ is a CTMC with the transition diagram




**(b) [5pt]** What is the fraction of time that each of the following situations occur: (i) the system is empty, (ii) the waiting room is full?

**Solution** The question is what is (i) $p_0^{occ}$, (ii) $p_4^{occ}$. Since $L(\cdot)$ is an irreducible CTMC on a finite state space, the occupancy distribution exists and solves the balance and normalization equations:

$$\begin{cases} p_0 * 2 = p_1 * 1, & \text{balance for state } 0 \\ p_1 * 2 = p_2 * 2, & \text{balance for set } \{0,1\} \\ p_2 * 2 = p_3 * 2, & \text{balance for set } \{0,1,2\} \\ p_3 * 2 = p_4 * 2, & \text{balance for set } \{0,1,2,3\} \\ \sum_{i=0}^4 p_i = 1. \end{cases}$$

From the balance equations it follows that $p_1 = p_2 = p_3 = p_4 = 2p_0$ and then the normalization equation gives $p^{occ} = (\frac{1}{9}, \frac{2}{9}, \frac{2}{9}, \frac{2}{9}, \frac{2}{9})$.

I.e. the answers are (i) 1/9, (ii) 2/9.

**(c) [2pt]** What fraction of customers are lost and why?

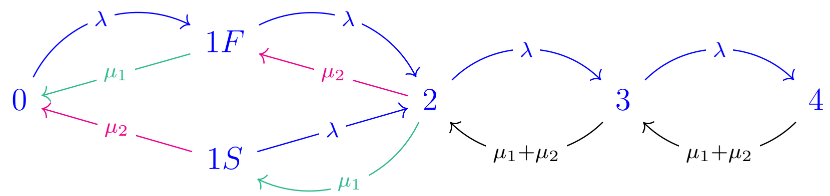
**Solution** By PASTA, it is $p_4^{occ} = 2/9$.

**(d) [2pt]** What is the time-average number of customers in the waiting room?

**Solution** It is $EL^q = p_3^{occ} * 1 + p_4^{occ} * 2 = 6/9 = 2/3$.

**(e) [4pt]** Now assume that the 2 servers are not identical: one is faster and the other is slower. The service times at the fast server are distributed exponentially with rate $\mu_1 = 1.5$, and at the slow server - exponentially with rate $\mu_2 = 0.5$. A customer only has a choice which server to go to if that customer arrives into an empty system; in that case the customer goes to the fast server. In all other cases, the customer goes to whichever server is available at the moment. Model this new situation as a CTMC.
*Hint:* this new situation requires one state more than the original situation.

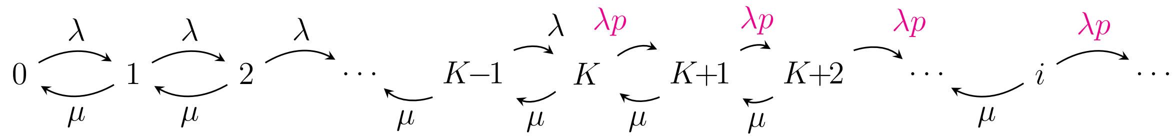
**Solution** Let $X(t)$ be the number of customers in the system at time $t$ if that number is different than 1. In case of exactly 1 customer in the system let $X(t)$ be the occupied server at time $t$. Then $X(t)$ is a CTMC with the transition diagram




**Question 2.** Consider a channel that transmits packages one at a time and has an infinite buffer (i.e., an infinite queue for packages). The packages are generated according to a Poisson process of rate $\lambda$; all transmission times are independent and distributed exponentially with rate $\mu$. There is also a *patience threshold* $K$, meaning: if a new package is generated when there are already $K$ or more packages at the channel (including the package currently in transmission), then the new package remains at this channel with probability $p$ or gets rerouted elsewhere with probability $1 - p$.

**(a) [4pt]** Argue that the number of packages at the channel (in total in transmission and in the buffer) is a CTMC.

**Solution** $L(t)$ = number of packages at the channel at time $t$ is a CTMC with the transition diagram




**(b) [2pt]** Argue intuitively what is the stability condition for this CTMC. In particular, does the specific value of $K$ matter?

Intuitively, this CTMC is stable when, in large states, the growth rate is slower than the decay rate. Hence, the stability condition is

$$\lambda p < \mu$$

and the specific value of $K$ does not matter.

**(c) [5pt]** Find the limit and occupancy distribution. In particular, derive that

$$p_0^{lim} = p_0^{occ} = \left[ \sum_{i=0}^{K-1} (\lambda/\mu)^i + \frac{(\lambda/\mu)^K}{1 - \lambda p/\mu} \right]^{-1}.$$

*Hint:* For normalisation, group states $i \geq K$.

**Solution** Below we find a solution to balance and normalization equations. This solution is both $p^{lim}$ and $p^{occ}$ since $L(t)$ is an irreducible CTMC.

The system for $p^{lim}$ and $p^{occ}$ is,

$$\begin{cases} \text{balance for sets } \{0, \ldots, i-1\}: \\ p_{i-1} * \lambda = p_i * \mu, \quad i = 1, \ldots, K, \\ p_{i-1} * \lambda p = p_i * \mu, \quad i = K+1, K+2, \ldots \\ \text{normalization: } \sum_{i=0}^{\infty} p_i = 1. \end{cases}$$

Hence, for $i = 1, \ldots, K$,

$$p_i = \tfrac{\lambda}{\mu} p_{i-1} = (\tfrac{\lambda}{\mu})^2 p_{i-2} = \ldots = (\tfrac{\lambda}{\mu})^i p_0 \quad \text{(also true for } i = 0\text{)},$$

and for $i \geq K+1$,

$p_i = \frac{\lambda p}{\mu} p_{i-1} = (\frac{\lambda p}{\mu})^2 p_{i-2} = \ldots = (\frac{\lambda p}{\mu})^{i-K} p_K = (\frac{\lambda p}{\mu})^{i-K} (\frac{\lambda}{\mu})^K p_0$ (also true for $i = K$).

Now we follow the hint for the normalization equation, plug the black and red relations in there, and get

$$1 = \sum_{i=0}^{\infty} p_i = \sum_{i=0}^{K-1} p_i + \sum_{i=K}^{\infty} p_i = p_0 \Bigg( \sum_{i=0}^{K-1} (\tfrac{\lambda}{\mu})^i + (\tfrac{\lambda}{\mu})^K \underbrace{\sum_{i=K}^{\infty} (\tfrac{\lambda p}{\mu})^{i-K}}_{=\sum_{j=0}^{\infty}(\frac{\lambda p}{\mu})^j = \frac{1}{1-\lambda p/\mu}} \Bigg).$$

To summarize, the last derivation implies that $p_0^{occ} = p_0^{lim}$ are as given in the question, and

$$\text{for } i = 0, 1, \ldots, K, \quad p_i^{occ} = p_i^{lim} = (\tfrac{\lambda}{\mu})^i p_0^{occ},$$
$$\text{for } i \geq K+1, \quad p_i^{occ} = p_i^{lim} = (\tfrac{\lambda}{\mu})^i p^{i-K} p_0^{occ}.$$

**(d) [4pt]** Derive the fraction of rerouted packages *based on your answer to (c)*. In particular, if you obtained the correct answer in (c), you should obtain the fraction of rerouted packages

$$\Pi_{loss} = (1-p)\frac{(\lambda/\mu)^K}{1-\lambda p/\mu} p_0^{occ}.$$

**Solution** Proportion $p_i^{occ}$ of packages are generated when there are $i$ other packages in the system, by PASTA. Out of packages generated into the system with $i \geq K$ other packages, proportion $1-p$ get rerouted. Hence, the proportion of rerouted packages is given by

$$\Pi_{loss} = (1-p)\sum_{i=K}^{\infty} p_i^{occ} \quad \text{now use the red relation from (c)}$$
$$= (1-p) \underbrace{\left( \sum_{i=K}^{\infty} (\tfrac{\lambda p}{\mu})^{i-K} \right)}_{=\frac{1}{1-\lambda p/\mu} \text{ from (c)}} (\tfrac{\lambda}{\mu})^K p_0.$$

I.e. the formula for $\Pi_{loss}$ is indeed as given.

**Question 3.** Jobs arrive at a server according to a Poisson process of rate $\lambda = 1/3$. For 1/4 of the jobs, their service times have a normal $N(4, 1^2)$

distribution. The remaining 3/4 of the jobs require a fixed service time 2.

**(a) [5pt]** Assume the FIFO discipline at the server and find the average waiting time $\mathbb{E}W$.

**Solution** This is an $M/G/1$ system with arrival rate and service time

$$\lambda = 1/3, \quad B = \begin{cases} N(4, 1^2), & \text{wp } 1/4, \\ 2, & \text{wp } 3/4. \end{cases}$$

Since the order of service is FIFO, the Pollaczek-Khinchine formula applies,

$$\mathbb{E}W = \frac{\rho}{1-\rho} \cdot \frac{\mathbb{E}(B^2)}{2\mathbb{E}(B)}.$$

We have

$$\begin{aligned} \mathbb{E}(B) &= \tfrac{1}{4} * \overbrace{\mathbb{E}(N(4, 1^2))}^{4} + \tfrac{3}{4} * 2 = 5/2, \\ \mathbb{E}(B^2) &= \tfrac{1}{4} * \underbrace{\mathbb{E}(N(4, 1^2))^2}_{=\mathbb{V}+(\mathbb{E})^2=1^2+4^2=17} + \tfrac{3}{4} * 2^2 = 29/4, \\ \rho &= \lambda\mathbb{E}(B) = 1/3 * 5/2 = 5/6, \end{aligned}$$

and hence

$$\mathbb{E}W = \frac{5/6}{1/6} \cdot \frac{29/4}{2 * 5/2} = 29/4 = 7.25.$$

**(b) [3pt]** The system under consideration is the result of pooling. Previously, the varying-size jobs and fixed-size jobs were served at two separate servers, each twice as slow as the server in the present system. Has the pooling been beneficial for the fixed-sized jobs? Why yes or why not?

**Solution** Prior to pooling, the fixed-sized jobs at their own server formed an $M/G/1$ model with arrival rate and service times

$$\lambda_f = 1/3 * 3/4 = 1/4, \quad B_f = 2 * 2 = 4.$$

This system had load $\rho_f = \lambda_f \mathbb{E}B_f = 1/4 * 4 = 1$ and hence it was unstable. The pooled system is stable and hence the pooling has been beneficial to the fixed-sized jobs.

For questions (c) and (d) assume that, in the pooled system, the service discipline is not FIFO but instead the $N(4, 1^2)$-sized jobs have a non-preemptive priority over the jobs of fixed size 2.

**(c) [2pt]** Is the Pollaczek-Khinchine formula applicable in this new situation in order to find the average waiting time across all jobs together? Why yes or why not?

**Solution** The new discipline is size-based and hence the PK formula is not applicable anymore.

**(d) [5pt]** Determine the average waiting time $\mathbb{E}W_1$ of the high-priority jobs by doing Mean Value Analysis for this group.

**Solution** The MVA equations are:

$$\begin{cases} \text{Little's law} & EL_1^q &= \lambda_1 * \mathbb{E}W_1, \\ \text{arrival relation} & EW_1 &= \rho * \mathbb{E}R + EL_1^q * \mathbb{E}(N(4, 1^2)). \end{cases}$$

In the Little's law, we use the arrival rate of high-priority customers $\lambda_1 = 1/3 * 1/4 = 1/12$.

In the arrival relation, we use
- the total load $\rho = 5/6$ (from (a)) which is also the fraction of time the server is busy,
- the remaining service time in progress $\mathbb{E}R = \dfrac{\mathbb{E}B^2}{2\mathbb{E}B} = \dfrac{29/4}{2 * 5/2}$ (from (a)).

The logic of the arrival relation is as follows:
- with probability $\rho$, a newly arriving high-priority customer finds the server occupied (by PASTA) and has to wait for the residual service time in progress $\mathbb{E}R$,
- the service in progress can be a low- or high- priority customer, in either case it is not going to be interrupted, that is why $\mathbb{E}R$ is based on all of the customers, both high- and low- priority together,
- after waiting for the remaining service in progress, the new high-priority customer has to wait for full service times of all high-priority customers in the queue.

We plug in the Little's law into the arrival relation and get

$\mathbb{E}W_1 = \rho * \mathbb{E}R + \lambda_1 * \mathbb{E}W_1 * \mathbb{E}(N(4, 1^2))$,

$$\mathbb{E}W_1 = \frac{\rho}{1 - \lambda_1 * \mathbb{E}(N(4, 1^2))} * \mathbb{E}R = \frac{5/6}{1 - 1/12 * 4} * \frac{29/4}{2 * 5/2} = \frac{5/6}{2/3} * \frac{29/4}{5} = \frac{29}{16}.$$

**FORMULA SHEET**

**Erlang distribution.** If $S_n$ has an Erlang$(n, \mu)$ distribution, then

$$P(S_n > t) = \sum_{k=0}^{n-1} e^{-\mu t} \frac{(\mu t)^k}{k!} \qquad \text{and} \qquad f_{S_n}(t) = \mu\, e^{-\mu t} \frac{(\mu t)^{n-1}}{(n-1)!}.$$

**Residual time till next event.** Let $X$ be a generic inter-event time and $R$ the residual time till next event. Then

$$P(R \leq x) = \frac{1}{E(X)} \int_0^x P(X > u)\, \mathrm{d}u \qquad \text{and} \qquad E(R) = \frac{E(X^2)}{2E(X)}.$$

**M/G/1 queue.** The waiting time $W$ under FIFO and the busy period $BP$ under work-conserving disciplines satisfy

$$E(W) = \frac{\rho}{1-\rho} \frac{E(B^2)}{2E(B)} = \frac{1}{2} \frac{\rho}{1-\rho}(1 + c_B^2)E(B), \quad \text{where } c_B^2 = \frac{V(B)}{(E(B))^2}$$

$$E(BP) = \frac{E(B)}{1-\rho}.$$

**M/M/$c$ queue.** The probability of waiting $\Pi_W$, waiting time $W$ and sojourn time $S$ satisfy

$$\Pi_W = \frac{(c\rho)^c/c!}{(1-\rho)\sum_{i=0}^{c-1}(c\rho)^i/i! + (c\rho)^c/c!},$$

$$E(W) = \Pi_W \frac{1}{c\mu(1-\rho)} \qquad \text{and} \qquad P(W > t) = \Pi_W\, e^{-c\mu(1-\rho)t},$$

$$P(S > t) = \frac{\Pi_W}{1 - c(1-\rho)} e^{-c\mu(1-\rho)t} + \left(1 - \frac{\Pi_W}{1 - c(1-\rho)}\right) e^{-\mu t}.$$

**M/G/$c$/$c$ queue.** The blocking probability is

$$B(c, a) = \frac{a^c/c!}{\sum_{i=0}^{c} a^i/i!} \qquad \text{with } a = \lambda\, E(B) = c\rho.$$