Queuing Theory 2014 - Exercises

Ioannis Glaropoulos

February 13, 2014

1 Probability Theory and Transforms

1.1 Exercise 1.2

X is a random variable chosen from X_1 with probability a and from X_2 with probability b. Calculate $E[X]$ and σ_X for $\alpha = 0.2$ and $b = 0.8$. X_1 is an exponentially distributed r.v. with parameter $\lambda_1 = 0.1$ and X_2 is an exponentially distributed r.v. with parameter $\lambda_2 = 0.02$. Let the r.v. Y be chosen from D_1 with probability α and from D_2 with probability b, where D_1 and D_2 are deterministic r.v.s. Calculate the values D_1 and D_2 so that $E[X] = E[Y]$ and $\sigma_X = \sigma_Y$.
Solution: a) We directly apply the **conditional expectation** formula:

$$E[X] = \alpha E[X_1] + bE[X_2].$$

We can do this since the expectation is a raw moment – not central. The proof is straightforward: we have

$$\begin{aligned}
&f_X(x) = \alpha f_{X_1}(x) + b f_{X_2}(x) \rightarrow \\
&\rightarrow E[X] = \textstyle\int_0^\infty x f_X(x)dx = \alpha \int_0^\infty x f_{X_1}(x)dx + b \int_0^\infty x f_{X_2}(x)dx = \\
&\quad = \alpha E[X_1] + bE[X_2].
\end{aligned}$$

We then replace the given data

$$E[X] = \alpha\frac{1}{\lambda_1} + b\frac{1}{\lambda_2} = 0.2\frac{1}{0.1} + 0.8\frac{1}{0.02} = 42. \tag{1}$$

We can not calculate the variance (or the standard deviation) in the same way, since this is a central moment. Instead, we proceed with calculating the expected square of the r.v. X, which is a **raw** moment:

$$\begin{aligned}
&E[X^2] = \textstyle\int_0^\infty x^2 f_X(x)dx = \alpha \int_0^\infty x^2 f_{X_1}(x)dx + b \int_0^\infty x^2 f_{X_2}(x)dx = \\
&\quad = \alpha E[X_1^2] + bE[X_2^2].
\end{aligned}$$

Replacing the data we get

$$E[X] = \alpha\frac{2}{\lambda_1^2} + b\frac{2}{\lambda_2^2} = 0.2\frac{2}{0.1^2} + 0.8\frac{2}{0.02^2} = 4040. \tag{2}$$

Finally, we use the relation between the expectation, square mean and variance

$$\sigma_X^2 = E[X^2] - [E[X]]^2 = 4040 - 42^2 \rightarrow \sigma_X = 47.70. \tag{3}$$

b) We have $E[Y] = \alpha d_1 + bd_2$ and $E[Y^2] = \alpha d_1^2 + bd_2^2$. So the system of equations becomes

$$\begin{aligned}
&0.2d_1 + 0.8d_2 = 42, \\
&0.2d_1^2 + 0.8d_2^2 = 4040
\end{aligned} \tag{4}$$

Solving this 2 by 2 non-linear system we obtain the solution. Notice that because of the second order of the equation we may **in general** have more than one solutions.

1.2 Exercise 1.3

X is a discrete stochastic variable, $p_k = P(X = k) = \frac{a^k}{k!}e^{-a}, k = 0, 1, 2, ...$ and a is a positive constant.

a) Prove that $\sum_{k=0}^{\infty} p_k = 1$.

b) Determine the z-transform (generating function) $P(z) = \sum_{k=0}^{\infty} z^k p_k$.

c) Calculate $E[X], \mathrm{Var}[X]$ and $E[X(X-1)...(X-r+1)], r = 1, 2, ...$ with and without using z-transforms.

Solution a) We have

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{a^k}{k!}e^{-a} = e^{-a}\sum_{k=0}^{\infty} \frac{a^k}{k!} = e^{-a}e^{a} = 1.$$

Notice this useful and well-known infinite series summation.

b) We replace the definition of the mass function and gradually have:

$$P(z) = \sum_{k=0}^{\infty} z^k \frac{a^k}{k!}e^{-a} = e^{-a}\sum_{k=0}^{\infty} z^k \frac{a^k}{k!} = e^{-a}\sum_{k=0}^{\infty} \frac{(za)^k}{k!} = e^{-a}e^{az} = e^{-a(1-z)}.$$

c) First, we try without the z-transform, i.e. using the definitions in the probability domain. We start from the third sentence, using the **definition** of expectation:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \tag{5}$$

$$\begin{aligned} E[X(X-1)...(X-r+1)] &= \textstyle\sum_{k=0}^{\infty} k(k-1)...(k-r+1)p_k = \\ &= \textstyle\sum_{k=0}^{\infty} k(k-1)...(k-r+1)\frac{a^k}{k!}e^{-a} = \sum_{k=0}^{\infty} \frac{a^k}{(k-r)!}e^{-a} = \\ &= e^{-a}a^r \textstyle\sum_{k=0}^{\infty} \frac{a^{(k-r)}}{(k-r)!} = a^r e^{-a}e^{a} = a^r. \end{aligned}$$

Then clearly, we have (by setting $r = 1$) $E[X] = a^1 = a$. And, finally,

$$\mathrm{Var}[X] = E[X^2] - [E[X]]^2 = E[X^2] - a^2 = E[X(X-1)] + E[X] - a^2 = a^2 + a - a^2.$$

We try, now, with the z-transform. We differentiate r times the definition of the z-transform:

$$\frac{d^r}{dz^r}P(z) = \frac{d^r}{dz^r}\sum_{k=0}^{\infty} z^k p_k = \sum_{k=0}^{\infty} k(k-1)...(k-r+1)z^{k-r}p_k$$

If we replace $z = 1$ we get

$$\left.\frac{d^r}{dz^r}P(z)\right\}_{z=1} = E[X(X-1)...(X-r+1)].$$

We, then, calculate,

$$\left.\frac{d^r}{dz^r}P(z)\right\}_{z=1} = a^r e^{-a(1-1)} = a^r.$$

1.3 Exercise 1.4

X_i's are independent Poisson distributed random variables, thus, $p_k = \frac{a_i^k}{k!}e^{-a_i}$, $k = 0, 1, 2, ...$, and each $a_i, i = 1, 2, ..., n$ is a positive constant. Give the probability distribution function of $X = \sum_{i=1}^{n}$.

Solution: This problem indicates the usefulness of the z-transform in the calculation of the distribution of the sum of variables. We have proven that **the ZT of the sum of independent random variables is the product of their individual z-transforms**. Thus,

$$P(z) = \prod_{i=1}^{n} P_i(z) = \prod_{i=1}^{n} e^{-a_i(1-z)} = e^{\sum_{i=1}^{n} -a_i(1-z)} = e^{-\alpha(1-z)},$$

where $\alpha = \sum_{i=1}^{n} -a_i$. This proves that the distribution is also Poisson with parameter α, i.e. the sum of parameters. The proof is based on the uniqueness of z-transform[1]. As a result, the distribution function will be

$$p_X(k) = \frac{\alpha^k}{k!}e^{-\alpha}$$

1.4 Exercise 1.5

X is a positive stochastic continuous variable with probability distribution function (PDF)

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-ax}, & x \geq 0. \end{cases}$$

a) Give the probability density function $f(x) = dF(x)/dx$.

b) Give $\overline{F}(x) = P(X > x)$.

c) Calculate the Laplace Transform $f^*(s) = E[e^{-sX}] = \int_0^\infty e^{-sx} f(x)dx$.

d) Calculate the expected values $m = E[X], E[X^k]$, $k = 0, 1, 2, ...$, the variance σ_X^2, the standard deviation σ_X and the coefficient of variation $c = \sigma/m$, with and without the transform $F^*(s)$.

Solution: a) For the calculation of $f(x)$ we just need to differentiate:

$$f(x) = dF(x)/dx = d(1 - e^{-ax})/dx = ae^{-ax}.$$

b) The complementary PDF is simply given as

$$\overline{F}_X(x) = P(X > x) = 1 - P(X \leq x) = 1 - F_X(x) = e^{-ax}.$$

c) Calculation of the Laplace Transform with simple integration

$$f^*(s) = \int_0^\infty e^{-sx} f(x)dx = \int_0^\infty e^{-sx} ae^{-ax} dx = a\int_0^\infty e^{-x(s+a)}dx = \frac{a}{s+a}.$$

d) We proceed first, without the help of Laplace transforms, using the definition of the expectation

$$E[X^0] = \int_0^\infty x^0 f(x)dx = \int_0^\infty f(x)dx = 1.$$

[1]or the 1-1 correspondence between the mass function and the ZT

$E[X^k] = \int_0^\infty x^k f(x)dx = \int_0^\infty x^k a e^{-ax}dx = a\frac{-1}{a}\int_0^\infty x^k (e^{-ax})'dx =$
$= k\int_0^\infty x^{k-1}e^{-ax}dx = \frac{k}{a}\int_0^\infty x^{k-1}ae^{-ax}dx = \int_0^\infty x^{k-1}f(x)dx =$
$= \frac{k}{a}E[X^{k-1}].$

This is a recursive formula that enables the calculation of any moment. We have:

$$E[X^k] = \frac{k}{a}E[X^{k-1}] = \frac{k}{a}\frac{k-1}{a}E[X^{k-2}] = \frac{k}{a}\frac{k-1}{a}..\frac{1}{a}E[X^0] = \frac{k}{a}\frac{k-1}{a}..\frac{1}{a} = \frac{k!}{a^k}$$

which gives, simply, $E[X] = 1/a$, for $k = 1$. The variance is calculated through the usual formula, and the raw moments are taken from above:

$$\sigma^2 = E[X^2] - [E[X]]^2 = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = 1/a^2.$$

so the standard deviation is simply the square root of the variance, $1/a$, and the coefficient of variation is 1. Notice that this is special for the exponential distribution.

We try, now, with the help of the Laplace transforms.

$$E[X^k] = (-1)^k \frac{d^k}{ds^k} f^*(s) = (-1)^k \frac{d^k}{ds^k}\frac{a}{s+a} = \frac{(-1)^k ak!}{(s+a)^{k+1}}.$$

We find this formula by differentiating k times the Laplace transform and replacing $s = 0$. The rest follows with simple replacement $k = 1, 2, ...$

1.5 Exercise 1.6

X_i's are independent, exponentially distributed random variables with a mean value of $1/a, a > 0, i = 1, 2, ..., n$. Calculate $P(X \le x)$ and $P(X \ge x)$ where

a) $X = \min(X_1, X_2, ..., X_n)$,

b) $X = \max(X_1, X_2, ..., X_n)$.

Solution: a) The key point in this exercise is the fact that the random variables are independent (mutually independent). We gradually have:

$P(X \le x) = P(\min(X_1, X_2, ..., X_n) \le x) = 1 - P(\min(X_1, X_2, ..., X_n) > x)$
$= 1 - P(X_1 > x, X_2 > x, ..., X_n > x) = 1 - \prod_{i=1}^n P(X_i > x)$
$= 1 - \prod_{i=1}^n e^{-ax} = 1 - e^{-\sum_{i=1}^n ax} = 1 - e^{-nax}$

This shows that the minimum of exponentially distributed random variables is also an exponential variable and its rate is the sum of the individual rates.

b) Similar calculations:

$P(X \le x) = P(\max(X_1, X_2, ..., X_n) \le x) = P(X_1 \le x, X_2 \le x, ..., X_n \le x)$
$= \prod_{i=1}^n P(X_i \le x) = \prod_{i=1}^n (1 - e^{-ax}) = (1 - e^{-ax})^n.$

Cleary, the variable X is, now, not exponential.

2 Balance equations, birth-death processes, continuous Markov Chains

2.1 Exercise 3.2

Consider a birth-death process with 3 states, where the transition rate from state 2 to state 1 is $q_{21} = \mu$ and $q_{23} = \lambda$. Show that the mean time spent in state 2 is exponentially distributed with mean $1/(\lambda + \mu)$.[2]

Solution: Suppose that the system has just arrived at state 2. The time until next "birth" – denoted here as T_B – is exponentially distributed with cumulative distribution function $F_{T_B}(t) = 1 - e^{-\lambda t}$. Similarly, the time until next "death" – denoted here as T_D – is exponentially distributed with cumulative distribution function $F_{T_D}(t) = 1 - e^{-\mu t}$. The random variables T_B and T_D are independent.

Denote by T_2 the time the system spends in state 2. The system will depart from state 2 when the first of the two events (birth or death) occurs. Consequently we have $T_2 = \min\{T_B, T_D\}$. We, then, apply the result from exercise 1.6, that is the minimum of independent exponential random variables is an exponential random variable. We can actually show this:

$$\begin{aligned}
F_{T_2}(t) &= \Pr\{T_2 \le t\} = \\
&= \Pr\{\min\{T_B, T_D\} \le t\} = \\
&= 1 - \Pr\{\min\{T_B, T_D\} > t\} = \\
&= 1 - \Pr\{T_B > t, T_D > t\} = \\
&= 1 - \Pr\{T_B > t\} \cdot \Pr\{T_D > t\} = \\
&= 1 - e^{-\lambda t} \cdot e^{-\mu t} = \\
&= 1 - e^{-(\lambda+\mu)t}
\end{aligned}$$

so T_2 is exponentially distributed with parameter $\lambda + \mu$.

Notice that we can generalize to the case with more than two transition branches. This exercise reveals the property of continuous time Markov chains, that is, the time spent on a state is exponentially distributed.

2.2 Exercise 3.3

Assume that the number of call arrivals between two locations has Poisson distribution with intensity λ. Also, assume that the holding times of the conversations are exponentially distributed with a mean of $1/\mu$. Calculate the average number of call arrivals for a period of a conversation.

Solution: Denote by N_C the number of arriving calls during the period of one conversation. Denote by T the duration of this conversation. Given that $T = t$, $N_C | T = t$ is Poisson distributed with parameter $\lambda \cdot t$ so the probability mass function of the number of calls will be

$$\Pr\{\text{arriving calls within } t = k\} = P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

[2] This exercise is similar to Exercise 6 from Chapter 1: "The minimum of independent exponential variables is exponential."

with an average number of calls: $E[N_C|T=t] = \lambda t$.

Moreover T is exponentially distributed, with parameter μ so the density function will be:

$$f_T(t) = \mu e^{-\mu t}.$$

We apply the conditional expectation formula:

$$E[N_C] = \int_0^\infty E[N_C|T=t] \cdot f_T(t)dt = \int_0^\infty \lambda t \mu e^{-\mu t} dt = \lambda \int_0^\infty t\mu e^{-\mu t} dt = \frac{\lambda}{\mu}.$$

2.3 Exercise 3.4

Consider a communication link with a constant rate of $4.8 kbit/sec$. *Over the link we transmit two types of messages, both of exponentially distributed size. Messages arrive in a Poisson fashion with* $\lambda = 10$ *messages/second. With probability 0.5 (independent from previous arrivals) the arriving message is of type 1 and has a mean length of 300 bits. Otherwise a message of type 2 arrives with a mean length of 150 bits. The buffer at the link can at most hold one message of type 1 or two messages of type 2. A message being transmitted still takes a place in the buffer.*

a) Determine the mean and the coefficient of variation of the service time of a randomly chosen arriving message.

b) Determine the average times in the system for accepted messages of type 1 and 2.

c) Determine the message loss probabilities for messages of type 1 and 2.

Solution:

a) We have a link with a constant transmission rate. So the service time distributions follow the packet length distributions. Consequently, the service times of both packet types are exponential with mean values of

- Type 1: $E[T_1] = \frac{300}{4800} = \frac{1}{16}$ sec,
- Type 2: $E[T_2] = \frac{150}{4800} = \frac{1}{32}$ sec.

As a result the parameters of the exponential distributions are $\mu_1 = 16$ and $\mu_2 = 32$, respectively. A random arriving packet is of Type 1 or Type 2 with probability 0.5. We apply the conditional expectation [3]:

$$E[T] = \frac{1}{2}E[T_1] + \frac{1}{2}E[T_2] = \frac{3}{64}$$

Similarly, we calculate the mean square:

$$E[T^2] = \frac{1}{2}E[T_1^2] + \frac{1}{2}E[T_2^2] = \frac{1}{2}\frac{2}{\mu_1^2} + \frac{1}{2}\frac{2}{\mu_2^2} = 16^{-2} + 32^{-2} = \frac{5}{4} \cdot 16^{-2}.$$

The variance of T is derived from $Var[T] = E[T^2] - (E[T])^2$. Then we compute the standard deviation σ_T as $\sigma_T = \sqrt{Var[T]}$, and finally the coefficient of variation is given as: $c_T = \frac{\sigma_T}{E[T]}$.

[3] This is similar to exercise 1.2

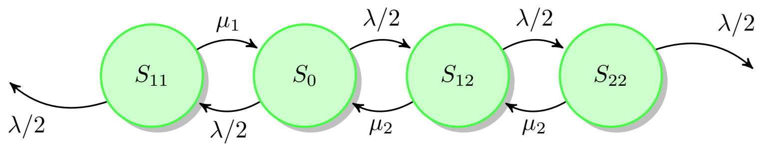


Figure 1: State Diagram for Exercise 3.4

b) For this part of the exercise, we need to draw the Markov Chain (Fig. 1) and solve it in the **steady state**. The state space must be defined in such a way that we can guarantee that all transitions – from state to state – have an exponential rate. We choose here to define such a Markov chain with 4 states:
State 0; Empty buffer.
State 11; 1 packet of Type 1.
State 21; 1 packet of Type 2.
State 22; 2 packets of Type 2.
Then we solve the balance equations in the local form:

$$\mu_2 P_{22} = \lambda/2 P_{21}$$
$$\mu_2 P_{21} = \lambda/2 P_0$$
$$\mu_1 P_{11} = \lambda/2 P_0$$
$$P_0 + P_{11} + P_{21} + P_{22} = 1 (\text{norm. equation})$$

Solution:

$$P_0 = 0.670, P_{21} = 0.105, P_{22} = 0.016, P_{11} = 0.209.$$

An accepted message of Type 1 can only arrive at state 0, otherwise it is rejected. So its the average service time will be $E[T_1]$.
An accepted message of Type 2 can arrive at states 0 and 21, otherwise it is rejected. Then, the average service time will be $(E[T_2]P_0 + 2E[T_2]P_{(21)})/(P_0 + P_{21})$. [4]

c) The loss probabilities are equal to the probabilities of the system being in BLOCKING states, for each of the two packet types. We underline that this is always true for *homogeneous* Markov chains, that is, Markov chains where the arrival rates do not depend on the system state.

2.4 Exercise 3.5

Consider a Markovian system with discouraged job arrivals. Jobs arrive to a server in a Poisson fashion, with an intensity of one job per 7 seconds. The jobs observe the queue. They do NOT join the queue with probability l_k if they observe k jobs in the queue. $l_k = k/4$ if $k < 4$, or 0, otherwise. The service time is exponentially distributed with mean time of 6 seconds.

a) Determine the mean number of customers in the system, and
b) the number of jobs served in 100 seconds.

Solution:

[4] The occurrence of acceptance reduces the sample space to two states only. Then the probabilities are normalized.

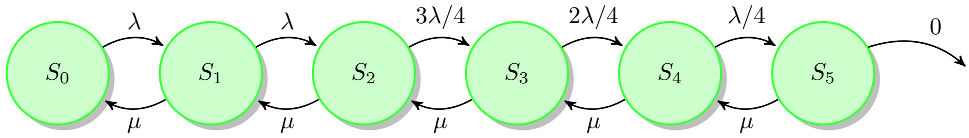


Figure 2: State diagram for Exercise 3.5

a) This is a simple model but requires careful design. After building the correct state diagram, the solution is found, based on the LOCAL balance equations.

We have a system with 6 states. *State space*: S_k : k jobs in the system. The system diagram is shown in Fig. 2).

Balance Equation System:

$$\begin{aligned}
&\lambda P_0 = \mu P_1 \\
&\lambda P_1 = \mu P_2 \\
&3\lambda/4 P_2 = \mu P_3 \\
&\lambda/2 P_2 = \mu P_4 \\
&\lambda/4 P_2 = \mu P_5 \\
&\textstyle\sum_{k=1}^{5} P_k = 1
\end{aligned}$$

Solution: $P_0 \approx 0.3$, and the remaining probabilities are computed based on P_0 and the equations above. After determining the state probabilities, we derive the average number of customers in the system through

$$E[N] = \sum_{k=0}^{5} k \cdot P_k$$

We find $E[N] \approx 1.43$.

b) We have, here, a system with different arrival rates in each state. These systems are defined as *non-homogeneous*. However, the service rate is constant. The server is busy with probability $(1-P_0)$. When it is busy, it serves jobs. The service rate is $\mu = 1/6sec^{-1}$. As a result, the server can serve $100 \cdot \mu \cdot (1-P_0)$ jobs in 100 seconds on AVERAGE!

2.5 Exercise 3.6

Consider a network node that can serve 1 and store 2 packets altogether. Packets arrive to the node according to a Poisson process. Serving a packet involves two independent sequentially performed tasks: the ERROR CHECK and the packet TRANSMISSION to the output link. Each task requires an exponentially distributed time with an average of 30msec. Give, that we observe that the node is empty in 60% of the time, what is the average time spend in the node for one packet?

Solution: As always, we need to construct the state diagram is such a way

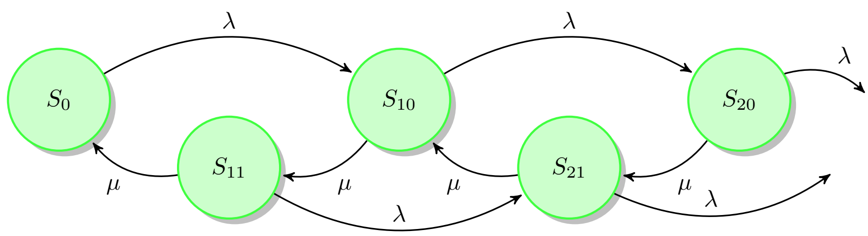


Figure 3: State Diagram for Exercise 3.6

that all transitions rates are guaranteed to be exponential. The selected state space:

S_0: Empty network node,
S_{11}: One packet under transmission,
S_{10}: One packet under error-check,
S_{20}: One packet under error-check and one buffered,
S_{21}: One packet under transmission and one buffered,

The state diagram is shown in Fig. 3. We can form the *global* balance equations parameterized by λ. Then we apply information that is given: $P_0 = 0.6$; This extra information enables the solution of the system of equations, and leads to the calculation of λ:

$$\begin{aligned}
&\lambda P_0 = \mu P_{11} \\
&(\lambda + \mu)P_{11} = \mu P_{10} \\
&(\lambda + \mu)P_{10} = \lambda P_0 + \mu P_{21} \\
&\mu P_{21} = \lambda P_{11} + \mu P_{22} \\
&P_{11} + P_{10} + P_{21} + P_{22} = 1 - P_0 = 0.4
\end{aligned}$$

Solution: $P_{10} \approx 0.1636, P_{11} \approx 0.1337, P_{20} \approx 0.0365, P_{21} \approx 0.0663, \lambda \approx 7.63$. For the calculation of the total average *system* time for a packet, we apply Little's formula.

$$\overline{N} = \lambda_{eff} \cdot E[T_{sys}] \to E[T_{sys}] = \frac{\overline{N}}{\lambda_{eff}} = \frac{1 \cdot (P_{10} + P_{11}) + 2 \cdot (P_{21} + P_{22})}{\lambda_{eff}}.$$

We always apply the effective arrival rate at Little's formula, because the formula needs the actual average arrival rate at the system, excluding possible drops. Here, the effective rate is not equal to λ, since we have packet drops. However, since the arrival rate for this system does not change with time, the effective arrival rate is simply:

$$\lambda_{eff} = \lambda \cdot (P_0 + P_{10} + P_{11}).$$

3 Chapter 4 – Queuing Systems

3.1 Exercise 4.1

Packets arrive to a communication node with a single output link according to a Poisson Process. Give the Kendall notation for the following cases:

1. *the packet lengths are exponentially distributed, the buffer capacity at the node is infinite*
2. *the packet length is fixed, the buffer can store n packets*
3. *the packet length is L with probability p_L amd l with probability p_l and there is no buffer in the node*

Solution: Kendall Notation

1. Arrival Process
2. Service Time
3. Number of Servers
4. Number of Total Positions (servers and queues)
5. Population

The Poisson arrivals (M) and the Single server (1) are fixed: M/?/1/?/?

1. M/M/1, as the buffer is infinite
2. M/D/1/n+1, as the service is deterministic and the buffer is n
3. M/G/1/1, as the service is general and there is no buffer

3.2 Exercise 4.2

Give the Kendall notation for the following systems. Telephone calls arrive to a PBX with C output links. The calls arrive as Poisson process and the call holding times are exponentially distributed.

1. *Calls arriving when all the output links are busy are blocked*
2. *Up to c calls can wait when all the output links are blocked*

Solution:

1. M/M/C/C
2. M/M/C/C+c

3.3 Exercise 4.3

Why is it not a good idea to have a G/G/10/12/5 System? Solution: 10 servers for 5 users!

4 Exercise 4.4

Which system provides the best performance, an M/M/3/300/100 or an M/M/3/100/100?

Solution: They have the same performance, since the users fit to both queues. Of course, the first system wastes buffer positions!

4.1 Exercise 4.5

A PBX was installed to handle the voice traffic generated by 300 employees in an office. Each employee on average makes 2 calls per hour with an average call duration of 4.5 minutes The PBX has 90 outgoing links.

1. *What is the offered load to the PBX?*
2. *What is the utilization of the outgoing links? Assume that calls arriving when all the links are busy are queued up.*

Solution: Offered Load $\rho \to \lambda\overline{T} = 300 \cdot \frac{2}{60} \cdot 4.5 = 45$ Erlang. Generally, the actual load is not the offered load.

Link Utilization:

$$\frac{\text{actual load}}{\#\text{ servers}}$$

The existence of an infinite queue here means that no load is dropped, or that the offered load is the actual load. So,

$$\text{utilization} = \frac{\text{offered load}}{90} = \frac{45}{90} = 0.5$$

.

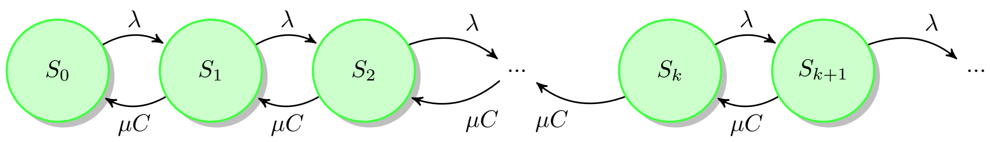


Figure 4: System diagram for the M/M/1 chain of exercise 5.1

5 Chapter 5 – M/M/1 Systems

5.1 Exercise 5.1

In a computer network a link has a transmission rate of C bit/s. Messages arrive to this link in a Poisson fashion with rate λ messages per second. Assume that the messages have exponentially distributed length with a mean of $1/\mu$ bits and the messages are queued in a FCFS fashion if the link is busy.
a) Determine the minimum required C for given λ and μ such that the average system time (service time + waiting time) is less than a given time T_0.

Solution: System Description

- Single communication link: C bits per second
- Poisson arrivals: λ messages per second
- Exponential Service times: $E[T] = E[X]/C = 1/(\mu C)$, so the *exponential rate* is μC.
- First Come First Served policy
- Infinite Queue[5]

This is a typical M/M/1 System. We see the system diagram in Fig. 4. We first derive the state distribution (steady-state) of this system through the solution of the balance equations. We *define* $\rho = \lambda/(\mu C)$. For a no-loss system, ρ is the OFFERED and, at the same time, the ACTUAL load.

$$\lambda P_0 = (\mu C)P_1 \rightarrow P_1 = \rho P_0$$
$$\lambda P_1 = (\mu C)P_2 \rightarrow P_2 = \rho P_1 = \rho^2 P_0$$
$$\lambda P_2 = (\mu C)P_3 \rightarrow P_3 = \rho P_2 = \rho^3 P_0$$

....

$$\lambda P_k = (\mu C)P_{k+1} \rightarrow P_{k+1} = \rho P_k = \rho^k P_0$$

....

Then, we calculate the P_0 through the normalization equation:

$$\sum_{k=0}^{\infty} P_k = 1 \rightarrow \sum_{k=0}^{\infty} \rho^k P_0 = 1 \rightarrow P_0 \sum_{k=0}^{\infty} \rho^k = 1 \rightarrow P_0 \cdot \frac{1}{1-\rho} = 1 \rightarrow P_0 = 1 - \rho.$$

[5] If no buffer capacity is mentioned, we always assume that this is infinite.

Finally, the state distribution is given as

$$P_k = (1-\rho)\rho^k.$$

We, now, derive the average number of messages in the system, using the state distribution:

$$\overline{N} = \sum_{k=0}^{\infty} kP_k = \sum_{k=0}^{\infty} k(1-\rho)\rho^k = (1-\rho)\rho \sum_{k=0}^{\infty} k\rho^{k-1} =$$
$$= (1-\rho)\rho \sum_{k=0}^{\infty} \frac{d\rho^k}{d\rho} = (1-\rho)\rho \frac{d\left(\sum_{k=0}^{\infty}\rho^k\right)}{d\rho} = (1-\rho)\rho \frac{d(1/(1-\rho))}{d\rho} = \frac{\rho}{1-\rho}.$$

In order to solve the first question we can use the LITTLE's formula:

$$\overline{N} = \lambda_{\text{eff}} E[T_{total}] \to E[T_{\text{total}}] = \frac{\overline{N}}{\lambda} = \frac{\rho/(1-\rho)}{\lambda} = \frac{\lambda/(\mu C)/(1-\lambda/(\mu C))}{\lambda},$$

since $\lambda_{eff} = \lambda$, so, finally,

$$E[T_{total}] = \frac{1}{(\mu C) - \lambda}.$$

The minimum required C is determined by:

$$\frac{1}{\mu C - \lambda} \leq T_0 \to \mu C - \lambda \geq T_0^{-1} \to C \geq \frac{\lambda + T_0^{-1}}{\mu}.$$

5.2 Exercise 5.5

Consider a queuing system with a single server. The arrival events can be modeled with Poisson distribution, but two customers arrive at the system at each arrival event. Each customer requires an exponentially distributed service time.

1. Draw the state diagram
2. Determine p_k using local balance equations
3. Let $P(z) = \sum_{k=0}^{\infty} z^k p_k$. Calculate $P(z)$ for the system. Note, that $P(z)$ must be finite for $|z| < 1$, and we know $P(1) = 1$.
4. Calculate the mean number of customers in the system with the help of $P(z)$ and compare it with the one of the M/M/1 system.

Solution: The system can be described by an M/M/1 model, since there is a single server, the service times are exponential service and the arrival process is Poisson. We must notice, however, that this Poisson Process models arrival events, but the events consist of two customer arrivals. (The departure events are still one-by-one, though.)

As always, for a Markovian System we must guarantee that all transitions are exponential. We define the usual state space: S_k : k customers in the

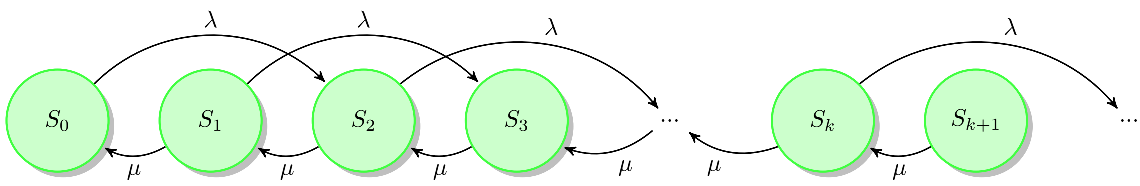


Figure 5: System diagram for the M/M/1 chain of exercise 5.5

system. Then, the state diagram is straightforward. Special care must be taken on determining the transitions and rates from state to state.

$$\textbf{Departure rate} = \mu$$
$$\textbf{Arrival Event rate} = \lambda$$

Clearly, the average customer arrival rate is 2λ and is NOT Poisson! What IS Poisson is the group arrival rate. We also DEFINE $\rho = \frac{\lambda}{\mu}$. This is neither the offered nor the actual load. We just use ρ to define this fraction.
The system diagram is given in Fig. 5.
Local Balance Equations:

$$\lambda P_0 = \mu P_1$$
$$\lambda P_{k-2} + \lambda P_{k-1} = \mu P_k, \quad k \geq 2$$

We can go ahead and solve them numerically. Alternatively, we can use the ZT methodology, since we only want to compute the average number of customers.

We consider the parametric local balance equation:

$$\mu P_k = \lambda P_{k-1} + \lambda P_{k-2} \rightarrow$$
$$\rightarrow \sum_{k=2}^{\infty} z^k \mu P_k = \sum_{k=2}^{\infty} z^k (\lambda P_{k-1} + \lambda P_{k-2})$$
$$\rightarrow \mu(P(z) - zP_1 - P_0) = \sum_{k=2}^{\infty} \lambda z^k P_{k-1} + \sum_{k=2}^{\infty} \lambda z^k P_{k-2}$$
$$\rightarrow \mu(P(z) - zP_1 - P_0) = \lambda z \sum_{k=2}^{\infty} \lambda z^{k-1} P_{k-1} + \lambda z^2 \sum_{k=2}^{\infty} \lambda z^{k-2} P_{k-2}$$
$$\rightarrow \mu(P(z) - zP_1 - P_0) = \lambda z (P(z) - P_0) + \lambda z^2 P(z)$$

We solve the equation with respect to $P(z)$

$$P(z) = \frac{\mu P_0 + \mu z P_1 - \lambda z P_0}{\mu - \lambda z - \lambda z^2} = \frac{P_0 + zP_1 - \rho z P_0}{1 - \rho z - \rho z^2}. \tag{6}$$

We need to apply two conditions that HOLD, in order to determine the unknown terms above. The first condition comes from the balance equation that we did not consider. We replace $P_1 = \rho P_0$ in (6), and obtain:

$$P(z) = \frac{P_0}{1 - \rho z - \rho z^2}. \tag{7}$$

The second condition comes from the NORMALIZATION in the probability or in the Z-domain:

$$\sum_{k=0}^{\infty} P_k = 1, \quad \text{or,} \quad P(z=1) = 1.$$

Replacing that in (7) we obtain $P_0 = 1 - 2\rho$, so finally

$$P(z) = \frac{1-2\rho}{1-\rho z-\rho z^2} \tag{8}$$

Finally, we need to compute the mean number of customers. We have

$$\overline{N} = \left[\frac{dP(z)}{dz}\right]_{z=1}.$$

Proof:

$$\left[\frac{dP(z)}{dz}\right]_{z=1} = \left[\frac{d\sum_{k=0}^{\infty} z^k P_k}{dz}\right]_{z=1} = \left[\sum_{k=0}^{\infty} k z^{k-1} P_k\right]_{z=1} = \sum_{k=0}^{\infty} k P_k = \overline{N}.$$

So, this is what we will do. We differentiate the derived ZT in (8):

$$\frac{dP(z)}{dz} = \frac{(-1)(1-2\rho)(-\rho-2\rho z)}{(1-\rho z-\rho z^2)^2}$$

Replacing $z = 1$ we obtain

$$\overline{N} = \frac{3\rho}{1-2\rho} = \frac{3\lambda}{\mu-2\lambda}.$$

The typical M/M/1 system with the same average customer arrival rate (2λ) and service rate (μ) has $\overline{N}_{M/M/1} = \frac{\rho}{1-\rho}$, where ρ is its offered load, and is equal to $\rho = 2\lambda/\mu$. So, finally,

$$\overline{N}_{M/M/1} = \frac{2\lambda}{\mu-2\lambda}$$

so it is different, and, actually, less. Why?

5.3 Exercise 5.6

A queuing system has one server and infinite queuing capacity. The number of customers in the system can be modeled as a birth-death process with $\lambda_k = \lambda$ and $\mu_k = k\mu$, $k = 0, 1, 2, \ldots$ thus, the server increases the speed of the service with the number of customers in the queue. Calculate the average number of customers in the system as a function of $\rho = \lambda/\mu$.

Solution: The system is an M/M/1 queue, since it has infinite buffer, 1 server, and Markovian arrival and departure process. However, as we can see, it is not a typical M/M/1 case, as the service rates depend on the current system state. The system diagram is shown in Fig. 6. We need to solve the system of balance equations:

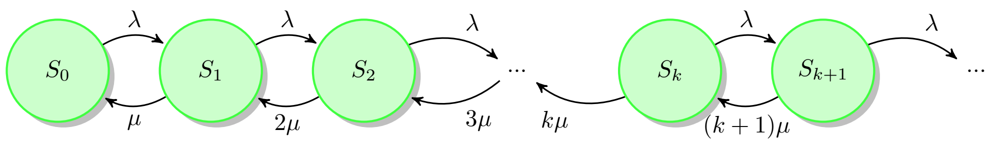


Figure 6: System diagram for the M/M/1 chain of exercise 5.6

$\lambda P_0 = \mu P_1 \rightarrow P_1 = \rho P_0$
$\lambda P_1 = 2\mu P_2 \rightarrow P_2 = \frac{1}{2}\rho P_1 = \frac{1}{2}\rho^2 P_0$
$\lambda P_2 = 3\mu P_3 \rightarrow P_3 = \frac{1}{3}\rho P_2 = \frac{1}{2\cdot 3}\rho^3 P_0$
....
$\lambda P_{k-1} = k\mu P_k \rightarrow P_k = \frac{1}{k}\rho P_{k-1} = ... = \frac{1}{k!}\rho^k P_0$
....
$\sum_{k=0}^{\infty} P_k = 1$ (normalization)

From the last general equation and the normalization equation we obtain the state distribution:

$$\sum_{k=0}^{\infty} \frac{\rho^k}{k!} P_0 = 1 \rightarrow P_0 e^{\rho} = 1 \rightarrow P_0 = e^{-\rho}.$$

so finally, for each k

$$P_k = \frac{\rho^k}{k!} e^{-\rho}$$

so the state distribution is POISSON! Then, we can calculate the average number of customers from the state distribution

$$\overline{N} = \sum_{k=0}^{\infty} k P_k = \rho$$

or simply say that the average is ρ, from the Poisson distribution.
From LITTLE we can, also, calculate the average system time

$$E[T_{total}] = \frac{\overline{N}}{\lambda} = \frac{1}{\mu}.$$

This means that the arriving customers only stay in the system for an average time equal to the service time![6]

5.4 Exercise 5.7

Customers arrive to a single server system in groups of 1,2,3 and 4 customers. The number of customers per group is i.i.d. There are in total 4 places in the

[6]This is equivalent to the case where there is no queue and each customer is served in parallel with the others, so actually this system is equivalent to an M/M/∞ system!

system. If a group of customers does not fit into the system, none of the members of the group joins the queue. 10% of the customers arrive in a group of 1, 20% of the customers arrive in groups of 2, 30% in a group of 3 and 40% in a group of 4 customers. The average number of arriving customers is 75 customers per hour, the interarrival time between groups is exponentially distributed. The service time is exponentially distributed with a mean of 0.5 minutes.

1. *Give the Kendall notation of the system and draw the state transition diagram.*
2. *Calculate the average number of customers in the queue and the mean waiting time per customer.*
3. *Calculate the probability that the system is full and the probability that a customer arriving in a group of k customers can not join the queue.*
4. *Calculate the probability that an arriving customer in general can not join the queue and the probability that an arriving group of customers can not join the queue.*
5. *What is the average waiting time for a customer arriving in a group of 3 customers?*

Solution: This is a very interesting problem that reveals the problems when the arriving process is complex and not straightforward so it must be derived.

First, we give the Kendall notation of the system. We have:

- Exponential GROUP inter-arrival times, so the arrival time will be Markovian.[7]
- The service times are exponentially distributed
- The system has a single server
- The total capacity is 4

Consequently, the Kendall notation is M/M/1/4.
We must draw the state transition diagram. We consider the typical state space where S_k means "k customers in the system". As a result the system has 5 states in total. The service rates are always the same, with

$$\mu = \frac{1}{E[T_s]} = \frac{1}{0.5/60} = 120h^{-1}.$$

The difficulty lies in deriving the arrival transition rates. We are given that the number of customers per group is i.i.d. We assume, naturally, that the GROUP arrival process is HOMOGENEOUS, that is, groups arrive in each state with the same rate! Also, since the inter-arrival times between GROUP arrivals are exponential we conclude that the GROUP arrivals is a Poisson process.

[7] It is our task to find an appropriate state space where the event arrival process is Poisson.

We are given that the number of customers per group is an i.i.d. process, but we are NOT given the distribution. Let

$$q_1, \quad q_2, \quad q_3, \quad q_4$$

denote the probabilities that a random arriving group contains 1,2,3,4 customers respectively.

Let λ_G denote the Poisson **group** arrival rate. Then, the individual rates for each groups is ALSO a Poisson process, based on the *Poisson split* property, with rates

$$\lambda_G q_1, \quad \lambda_G q_2, \quad \lambda_G q_3, \quad \lambda_G q_4.$$

For any $i = 1, 2, 3, 4$,

$$\lambda_G \cdot q_i$$

defines the (average) rate of arrivals for group's of type i, and, consequently,

$$\lambda_G \cdot q_i \cdot i$$

defines the (average) rate of arrivals of customers belonging to group of type i.

Based on the given data from the exercise regarding the ratio of customers arriving in any of the groups, we obtain the following equations:

$$\lambda_G q_1 \cdot 1 = 10\% \cdot 75$$

$$\lambda_G q_2 \cdot 2 = 20\% \cdot 75$$

$$\lambda_G q_3 \cdot 3 = 30\% \cdot 75$$

$$\lambda_G q_4 \cdot 4 = 40\% \cdot 75$$

From the above it is clear that $q_1 = q_2 = q_3 = q_4 \rightarrow q_i = \frac{1}{4}, \quad \forall i = 1, 2, 3, 4$. Finally, using any of the above equations we compute the group arrival rate:

$$\lambda_G = 30 \text{ groups / hour}$$

We can now complete the state diagram (Fig. 7). Then, we solve the LOCAL balance equations, to define the state probabilities.

We can compute BOTH the average number of customers in the system, and the average number of customers in the queue:

$$\overline{N}_{queue} = 1 \cdot P_2 + 2 \cdot P_3 + 3 \cdot P_4$$
$$\overline{N} = 1 \cdot P_1 + 2 \cdot P_2 + 3 \cdot P_3 + 4 \cdot P_4$$

For the average waiting time, we could apply the LITTLE result. For that we need the effective customer arrival rate, which is different from the nominal, since there are losses in the system.

It is important to notice again that the system is HOMOGENEOUS in group arrivals but not in customer arrivals.

We have

$$\lambda_{eff} = P_0 \lambda_G \cdot (q_1 + 2q_2 + 3q_3 + 4q_4) + P_1 \lambda_G \cdot (q_1 + 2q_2 + 3q_3) + P_2 \lambda_G \cdot (q_1 + 2q_2) + P_3 \lambda_G \cdot (q_1).$$

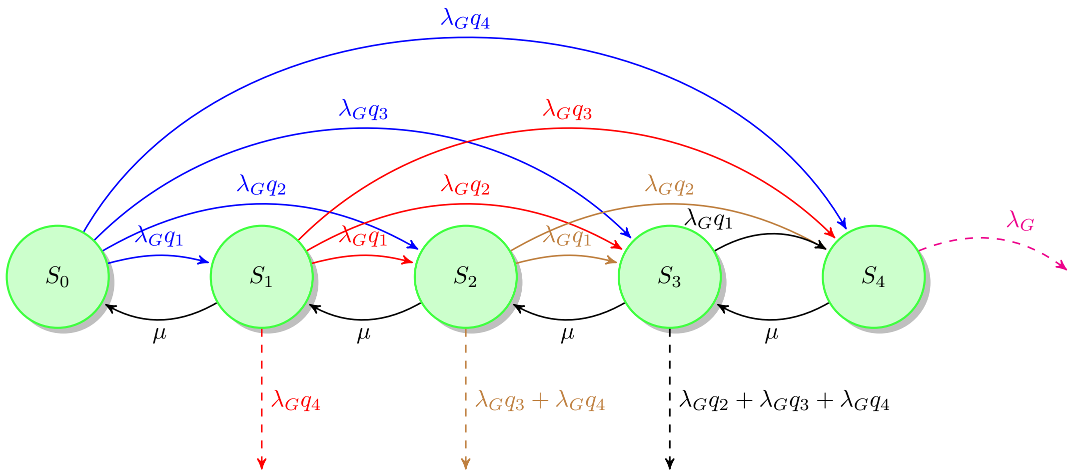


Figure 7: System diagram for the M/M/1 chain of exercise 5.7

Then, from LITTLE we compute first the system time, $\overline{N} = \lambda_{eff} E[T]$, and then the average waiting time will be $\overline{W} = E[T] - E[T_s]$.

The probability that the system is full (as seen by an independent observer) is simply P_4.

The probability that a customer of a group k does not join the queue is, actually, the probability that the whole particular group does not join the queue. Since the system is homogeneous in group arrivals (a random group SEES state distribution),

$$\begin{aligned}
\Pr(\text{a random group 1 is blocked}) &= P_4 \\
\Pr(\text{a random group 2 is blocked}) &= P_4 + P_3 \\
\Pr(\text{a random group 3 is blocked}) &= P_4 + P_3 + P_2 \\
\Pr(\text{a random group 4 is blocked}) &= P_4 + P_3 + P_2 + P_1
\end{aligned}$$

An arriving customer in general, belongs to groups 1,2,3,4 with probabilities 10%, 20%, 30% and 40%. So given these probabilities, he follows the group blocking probabilities:

$$\Pr(\text{a random customer is blocked}) =$$
$$= 40\% \cdot (P_1 + P_2 + P_3 + P_4) + 30\% \cdot (P_2 + P_3 + P_4) + 20\%2 \cdot (P_3 + P_4) + 10\% \cdot P_4$$

An arriving group of customers is blocked with probability

Pr(a random group is blocked) =
$= \sum_{i=1}^{4} \Pr\{\text{a random group has } i \text{ customers}\} \cdot \Pr\{\text{a random group } i \text{ is blocked}\} =$
$= P_4 + P_3 \cdot 3/4 + P_2 \cdot 1/2 + P_1 \cdot 1/4.$

A customer that arrives in group of 3 customers MEANS that the arrived group sees either state 0 or state 1, otherwise there is no mean of WAITING time, since the group is rejected. The arrivals are homogeneous in groups, so the groups see state 0, 1 with probabilities P_0, P_1, respectively. So

$$\overline{W}_3 = \frac{P_0}{P_0 + P_1} \cdot W_3^0 + \frac{P_1}{P_0 + P_1} \cdot W_3^1$$

where the two waiting times are

$$W_3^0 = \frac{1}{3}(0.5 + 1 + 0), \qquad W_3^1 = \frac{1}{3}(0.5 + 1 + 1.5)$$

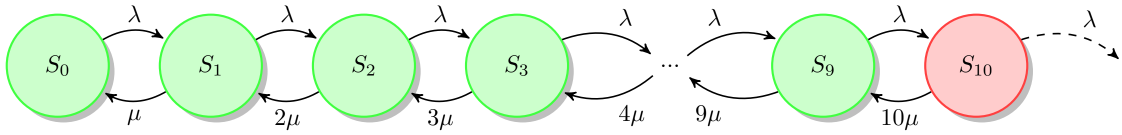


Figure 8: System diagram for exercise 6.5

6 Chapter 6 – M/M/m/m Loss Systems

Abstract

The importance for this tutorial is to introduce the students into the loss systems, and, to underline the difference between the concepts of call-blocking and time-blocking probabilities, and to understand under which conditions these probabilities are identical. The selected exercises introduce, also, the Erlang tables, which is an important tool for easy calculations for the blocking probabilities.

6.1 Exercise 6.5

A telephone switch has 10 output lines and a large number of incoming lines. Upon arrival a call on the input line is assigned an output line if such line is available – otherwise the call is blocked and lost. The output line remains assigned to the call for its entire duration which is of exponentially distributed length. Assume that 180 calls / hour arrive in Poisson fashion whereas the mean call duration is 110 seconds.

1. *Determine the blocking probability.*
2. *How many calls are rejected per hour?*
3. *What is the average load per server (in Erlang)?*
4. *What is the maximum arrival rate at which a blocking probability of (at most) 2% can be guaranteed?*

Solution: This is an introductory problem for the Markovian Loss Systems. We start with the system mode and the Kendall Notation:

- Poisson arrivals with rate $\lambda = 180$ calls per hour
- Exponential service times with rate $\mu = \frac{1}{E[T]} = \frac{3600}{110}$
- 10 Servers
- No buffer

The Kendall Notation is: **M/M/10/10**.

We draw the system diagram (Fig. 8) and, based on that, we build the balance equations to compute the state probabilities. A state, S, defines the number of active calls in the system. We define

$$\rho = \frac{\lambda}{\mu}$$

which, here, is also the *offered load* of the system; of course this is NOT the actual load, since the system may drop calls when it is blocked. We have:

$$\lambda P_0 = \mu P_1 \to P_1 = \rho P_0$$
$$\lambda P_1 = 2\mu P_2 \to P_2 = \tfrac{\rho}{2} P_1 = \tfrac{\rho^2}{2} P_0$$
$$\lambda P_2 = 3\mu P_3 \to P_3 = \tfrac{\rho}{3} P_2 = \tfrac{\rho^3}{3 \cdot 2} P_0$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\lambda P_{k-1} = k\mu P_k \to P_k = \tfrac{\rho}{k} P_{k-1} = \tfrac{\rho^k}{k!} P_0$$

We use the normalization equation to compute the P_0:

$$\sum_{k=0}^{n} P_k = 1 \to \sum_{k=0}^{n} \frac{\rho^k}{k!} P_0 = 1 \to P_0 = \frac{1}{\sum_{k=0}^{n} \frac{\rho^k}{k!}}.$$

where, here, $n = 10$. The blocking probability is the probability that the system can not accept new calls, and this is equal to the probability of the system being in state S_n (S_{10}). This probability is, also, the *TIME BLOCKING*, i.e. the percentage of time the system can not accept new calls, and it is the one seen by an independent observer.

$$P_{block} = P_n = \frac{\frac{\rho^n}{n!}}{\sum_{k=0}^{n} \frac{\rho^k}{k!}}$$

General rule: In Markov chains where the arrival rates do not depend on the system state, the time blocking is also **equal** to the probability of a random event being blocked. The latter is defined as *CALL BLOCKING* probability.

You can compute the P_{block} from the equation above, but you could also do it by using the Erlang tables.

- Servers: $n = 10$
- Offered Load: $\rho = \frac{\lambda}{\mu} = \frac{180}{\frac{3600}{110}} = 5.5$

We search in the tables for the solution. $E_{10}(5.5) \approx 0.029$.

The rate of rejected calls is ALWAYS given by: $\lambda \cdot P_{call\_block}$. However, as stated above, here,

$$P_{call\_block} = P_{block}$$

since the arrival process is independent of the system state[8]. We get:

$$\lambda_{rejected} = \lambda P_{block} \approx 180 \cdot 0.029 = 5.27.$$

[8]or, as we say, alternatively, the arrival process is homogeneous.

To find the average load per server, we first need to find the total ACTUAL load. The actual load is the nominal (offered) load minus the rejected load:

$$\lambda_{eff} = \lambda - \lambda_{rejected} = \lambda(1 - P_1 0) = 5.5 \cdot (1 - 0.029)$$

For the average load per server, we divide the above with the number of severs.[9]

As we see the blocking probability is above the target of 2%. So we must decrease the offered load, keeping the number of servers to 10. We check in the tables the maximum offered load that guarantees the blocking probability below the target. Then, from that we find the nominal arrival rate.

7 Exercise 6.3

We consider two types of call arrivals to a cell in a mobile telephone network: new calls that originate in a cell and calls that are handed over from neighboring cells. It is desirable to give preference to handover calls over new calls. For this reason, some of the channels in the cell are reserved for handover calls, while the rest of the channels are available to both types of calls. For the questions below, assume the following: Channels in the cell are held for two minutes on average, with exponential distribution. All calls arrive according to a Poisson process with rate $\lambda_{nc} = 125$ calls per hour for new calls, and $\lambda_{ho} = 50$ calls per hour for handover calls. The cell has a capacity of 10 channels; each call occupies 1 channel.

1. *Draw a state diagram of the channel occupancy in a cell when 2 channels are used exclusively for handover calls.*
2. *Calculate the blocking probability in the cell if no channels are reserved for handover calls. What is the average number of channels used?*
3. *Find the minimum number of channels reserved for handover calls so that their blocking probability is below 1 percent. What is the blocking probability for the new calls in this case?*

Solution:

This is an interesting exercise that models a Mobile network, like GSM, UTRAN etc. We have a system which prioritizes some calls (handover) over some others (new).

We draw the state transition diagram for the case of 2 reserved channels for handover calls (Fig. 9)

- Poisson arrivals – $\lambda = \lambda_{nc} + \lambda_{ho}$ for states 0,1,2,...,7, $\lambda = \lambda_{ho}$ for states 8,9. This is a nce application of the Poisson SPLIT property.
- Exponential Service times – $\mu = \frac{1}{E[t]} = \frac{1}{2/60} = 30$.
- 10 servers, no buffer

[9] Check also exercise 4.5.

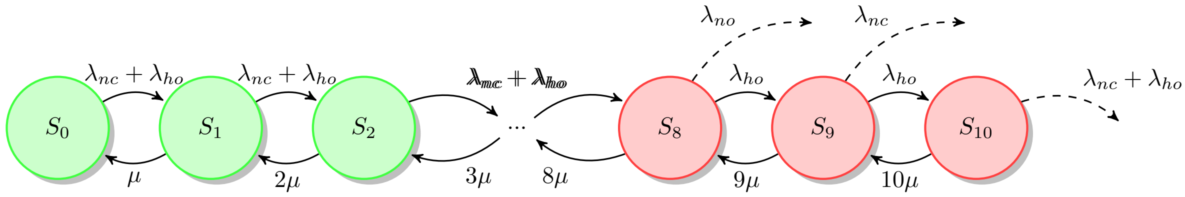


Figure 9: State diagram for exercise 6.3. Case of 2 reserved channels for handover calls. Notice that the system accepts only handover arrivals when only 2 channels remain vacant. In state 10, both handover and new calls are dropped.

We can draw and solve the balance equations for this system, and calculate, for example, the average number of active calls in the cell. The offered load in the system is

$$\lambda_{total} \cdot E[T] = (\lambda_{nc} + \lambda_{ho}) \cdot E[T].$$

What is the actual load of this system?

We consider, now, the case where no channels are reserved for handover calls. The Kendall notation is then M/M/10/10. So, the system resembles a standard M/M/10/10 case!

- Offered Load: $\rho = (\lambda_{nc} + \lambda_h o)E[T] = 175 \cdot \frac{2}{60} = \frac{35}{6}$.
- Servers: $n = 10$
- Blocking Probability: $E_{10}(\frac{35}{6}) = ...$

Consequently, the effective, or actual load of the system is $\lambda_{eff} = 175 \cdot (1 - E_{10}(35/6)) = ...$

We can apply the state probabilities to derive the average number of calls in the cell. However, since we know the effective load, and the average system time, we can also apply the LITTLE formula

$$\overline{N} = \lambda_{eff} \cdot E[T] = ...$$

For the third question we need to go step by step. First, we do the calculations with one reserved channel. The diagram in shown in Fig. 10 We have the result:

$$P_{block} = P_{10} = \frac{\frac{\lambda_{ho}}{10\mu}\frac{\lambda_{tot}^9}{9!\mu^9}}{1 + \frac{\lambda_{tot}}{\mu} + \frac{\lambda_{tot}^2}{2\mu^2} + ... + \frac{\lambda_{tot}^9}{9!\mu^9} + \frac{\lambda_{ho}\lambda_{tot}^9}{10!\mu^{10}}} = \frac{\frac{\lambda_{ho}}{10\mu}}{\frac{1}{E_9(\rho)} + \frac{\lambda_{ho}}{10\mu}}$$

If we replace the number we find a blocking probability of 1.1%, which is above the target. We now do the calculations for two reserved channels (Fig. 11):

$$P_{block} = P_{10} = \frac{\frac{\lambda_{ho}^2}{10\mu 9\mu}\frac{\lambda_{tot}^8}{8!\mu^8}}{1 + \frac{\lambda_{tot}}{\mu} + \frac{\lambda_{tot}^2}{2\mu^2} + ... + \frac{\lambda_{tot}^8}{8!\mu^8} + \frac{\lambda_{ho}\lambda_{tot}^8}{9!\mu^9} + \frac{\lambda_{ho}^2\lambda_{tot}^8}{10!\mu^{10}}} = \frac{\frac{\lambda_{ho}^2}{10\cdot 9\mu^2}}{\frac{1}{E_8(\rho)} + \frac{\lambda_{ho}}{9\mu} + \frac{\lambda_{ho}^2}{10\cdot 9\mu^2}}$$

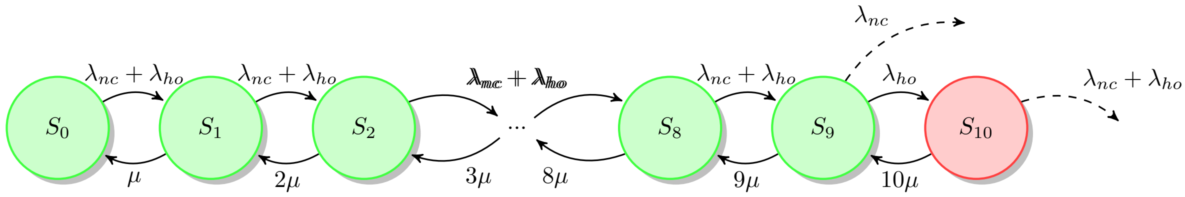


Figure 10: State diagram for exercise 6.3c. Case of 1 reserved channel for handover calls. Notice that the system accepts only handover arrivals when only one channel remains vacant. State S_9 is now green, to show that it is a non-blocked state for handover calls.

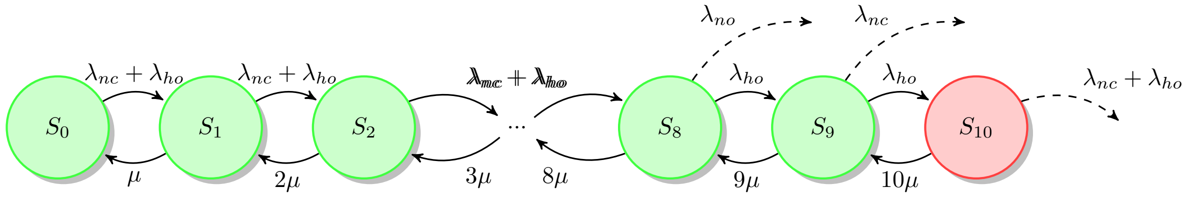


Figure 11: State diagram for exercise 6.3c. Case of 2 reserved channel for handover calls. Notice that the system accepts only handover arrivals when only one channel remains vacant. State S_9 is now green, to show that it is a non-blocked state for handover calls.

We do again the calculations and find a probability of 0.2%, which is acceptable.

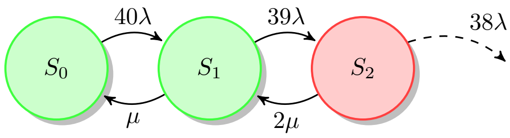
Finally, the blocking probability for the new calls is $P_8 + P_9 + P_{10}$, since the system is time homogeneous and the last two channels are reserved for the handovers.

7.1 Exercise 8 [Collection of Exam Problems]

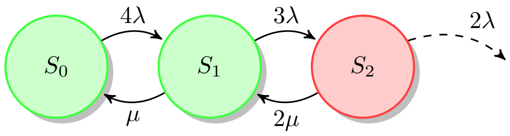
Consider two M/M/2 loss systems, A and B. A has 40 subscribers; to B there are 4 subscribers connected. Each subscriber generates on average 1 call per minute. The mean service time is 6 seconds.

1. *Compare the call blocking probability for A and B*
2. *Compare the mean blocking time for A and B*
3. *Compare the mean time without blocking for A and B*
4. *How many additional servers do you have to provide to system A to achieve a time blocking that is not higher than that of system B?*

Solution: There is a clear difference between the two system, and that is, the user population. In Fig. 12 we depict the two system diagrams. Both systems have finite population, i.e. the user arrival rate depends on the system state.



(a)



(b)

Figure 12: State diagrams for the A and B systems of exercise 8.

However, the first system (A) could be approximated with a typical M/M/2/2 system, since the population size (40) is high compared to the number of servers (2). In other words, $40\lambda \approx 39\lambda \approx 38\lambda$. We can not say the same for system B.

System A The offered load is

$$\rho = \frac{40\lambda}{\mu} = 40\lambda \cdot E[T] = 40 \cdot 1 \cdot \frac{6}{60} = 4 \text{ Erl.}$$

We can either draw and solve the balance equations, or use the Erlang Tables to compute the TIME blocking probability (P_2). In the second case, we obtain:

$$E_m(\rho) = E_2(4) = 0.615.$$

We notice that the TIME Blocking probability is equal to the CALL blocking probability since the arrival intensity is constant and does not depend on the state (we assumed typical M/M/2/2). So, we obtain:

$$P_{block} = P_2 = 0.615$$

The mean blocking time (T_b) is the average time the system CONTINUOUSLY remains in state S_2, i.e. from the time it arrives at S_2 until it departs to S_1. The transition rate ($S_2 \rightarrow S_1$) is 2μ, so the average time the system stays at S_2 is $E[T_b] = 1/2\mu = 3$sec.

The **mean time without blocking** (T_n) is the average time between a **departure** from state S_2 and an **arrival** at state S_2. It is **not** the percentage of time without blocking! That would be, simply, $P_0 + P_1$.

We must calculate $E[T_n]$ based on $P_b, E[T_b]$. We define a **cycle** of the system as a set of adjacent blocking and non-blocking periods. Consider that we observe the system for a large period of cycles, say M. Then it holds:

$$P_b = \lim_{M\to\infty} \frac{\sum_{i=1}^{M} T_b^i}{\sum_{i=1}^{M} T_b^i + T_n^i}, \tag{9}$$

where T_b^i is the duration of the blocking period at cycle i and T_n^i is the duration of the non-blocking period at cycle i. We rewrite the above as

$$P_b = \lim_{M\to\infty} \frac{\frac{1}{M}\sum_{i=1}^{M} T_b^i}{\frac{1}{M}\sum_{i=1}^{M} T_b^i + T_n^i} = \frac{\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} T_b^i}{\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} T_b^i + T_n^i} \tag{10}$$

and finally:

$$P_b = \frac{\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} T_b^i}{\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} T_b^i + \lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} T_n^i} = \frac{E[T_b]}{E[T_b] + E[T_n]}. \tag{11}$$

From the above we can calculate $E[T_n]$ with respect to P_b and $E[T_b]$.

The solution for system B is similar to the one for system A, EXCEPT that we can not use the approximation from the Erlang tables, so we need to derive the steady-state solution by solving the balance equations.

We calculate the mean blocking time and the mean time without block in the same way.

Note, however, that the call blocking probability is different, now, since this chain is non-homogeneous, as the arrival rate is state-dependent! In particular:

$$P_{\text{call block}} = \frac{2\lambda P_2}{4\lambda P_0 + 3\lambda P_1 + 2\lambda P_2}$$

8 Tutorial 7 – M/M/m systems

8.1 Exercise 7.2

The performance of a system with one processor and another with two processors will be compared. Let the inter-arrival times of jobs be exponentially distributed with parameter λ. We consider, first, the system with one processor. The service time of the jobs is exponentially distributed with a mean of 0.5sec.

1. *For an average response time of 2.5sec (the total time for a job in the system) how many jobs per second can be handled?*
2. *For an increase of λ with 10% how much will the response time increase?*
3. *Calculate the average waiting time, the average number of customers in the server and the utilization of the server. What is the probability of the server being empty?*

Let us now compare this system with a system of two cheaper processors, each with a mean service time of 1 sec.

1. *How many jobs can now be handled per second with a mean response time of 2.5 sec?*
2. *For an increase of λ with 10% how much will the response time increase?*
3. *Calculate the average waiting time, the average number of customers in the server and the utilization of the server. What is the probability of both of the servers being busy?*

Solution: We consider first the system with one processor. Clearly, since we are not given any information regarding the buffer data, and we are also told that there is waiting time, we can safely conclude that the buffer is **infinite**.

As a result we are dealing with a typical M/M/1 system, and we will derive the answers from the M/M/1 formulas. The average number of jobs in the M/M/1 system is given as

$$\overline{N} = \frac{\rho}{1-\rho} = \frac{\lambda/2}{1-\lambda/2} = \frac{\lambda}{2-\lambda}$$

since $\mu = \frac{1}{E[T]} = \frac{1}{0.5} = 2$. Applying the LITTLE's formula, we get

$$\overline{T}_{\text{system}} = \frac{\overline{N}}{\lambda} = \frac{1}{2-\lambda} \to \lambda = 1.6$$

since $\overline{T}_{\text{system}} = 2.5$. Assume now that we increase the arrival rate by 10%. The new rate will be $\hat{\lambda} = 1.1 \cdot 1.6 = 1.76$. This will change the offered load of the system to $\hat{\rho} = \frac{\hat{\lambda}}{\mu} = \frac{1.76}{2} = 0.88$. Then the response time will be given in a similar way:

$$\hat{T}_{\text{system}} = \frac{1}{\mu - \hat{\lambda}} = \frac{1}{2-1.76} = 4.1666\text{sec}$$

So the increase is $\frac{4.166-2.5}{2.5} = 66.66\%$

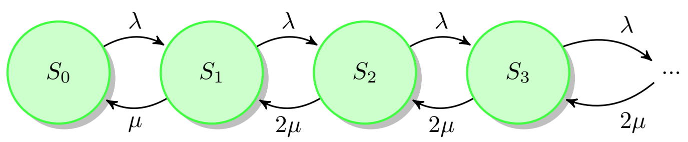


Figure 13: System diagram for the second model of exercise 7.2

For the initial ($\lambda = 1.6$) system the average waiting time is $\overline{W} = \overline{T}_{\text{system}} - E[T] = \overline{T}_{\text{system}} - \frac{1}{\mu} = 2\text{sec}$, the mean number of customers in the system is $\overline{N} = \rho/(1-\rho) = 0.8/0.2 = 4$, and the probability of empty server is $1 - \rho = 0.2$. The utilization is of course ρ and is equal to the mean number of customers in the server.

We consider now the second system. This one has 2 processors with $\mu = 1$, so half of the service rate. The offered load for this system is $\rho = \lambda/1 = \lambda$. The state diagram is given in Fig 13. We could use the formulas for the M/M/2 system. However, since we only have 2 servers, we can solve it analytically with the Balance Equations.

$$\begin{aligned}
&\lambda P_0 = \mu P_1 \to P_1 = \rho P_0 \\
&\lambda P_1 = 2\mu P_2 \to P_2 = \tfrac{1}{2}\rho P_1 = \tfrac{\rho^2}{2} P_0 \\
&\lambda P_2 = 2\mu P_3 \to P_3 = \tfrac{1}{2}\rho P_2 = \tfrac{\rho^3}{4} P_0 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\lambda P_{k-1} = \mu P_k \to P_k = \tfrac{\rho^k}{2^{k-1}} P_0
\end{aligned}$$

From the above equation set and the normalization equation we derive

$$P_0 \left(1 + \rho \sum_{k=1}^{\infty} \frac{\rho^{k-1}}{2^{k-1}} \right) = 1 \to P_0 = \frac{2-\rho}{2+\rho}.$$

Then, we calculate the mean number of customers in the system, from the state distribution

$$\overline{N} = \sum_{k=1}^{\infty} k P_k = \sum_{k=1}^{\infty} k \frac{\rho^k}{2^{k-1}} \frac{2-\rho}{2+\rho} = \ldots = \frac{4\rho}{(2+\rho)(2-\rho)} = \frac{4\lambda}{(2+\lambda)(2-\lambda)}.$$

Then, based on LITTLE we compute the average system time:

$$\overline{T}_{\text{system}} = \frac{\overline{N}}{\lambda} = \frac{4}{(2+\lambda)(2-\lambda)} \to \lambda = \sqrt{\frac{6}{2.5}} \approx 1.54919$$

This is the maximum allowed arrival rate, in order to guarantee the target response time of 2.5 seconds. As we can see, it is lower than 1.6, meaning that the second system is worse than the first, so for the same performance objective it can serve less incoming load.

We increase, now the λ by 10%, so $\hat{\lambda} = 1.704$ [of the original = 1.54919 load]. The average system time will, then be

$$\hat{T}_{\text{system}} = \frac{4}{(2+1.704)(2-1.704)} = \ldots = 3.65\text{sec}$$

So the increase is $\frac{3.65-2.5}{2.5} = ...$

Note that for a fair comparison with the previous system we must compute the response time for $\lambda = 1.76$. Calculations give:

$$\hat{T}_{\text{system}} = \frac{4}{(2+1.76)(2-1.76)} = ... = 4.4326\text{sec}$$

which is larger than the 4.1666 of the previous system, as expected.

The average waiting time will be given through the series:

$$\overline{W} = \sum_{k=2}^{\infty} P_k = 1 - P_0 - P_1 = ...$$

but, simply, can be given as $\overline{T}_{\text{system}} - \frac{1}{\mu} = 1.5$ sec.

The average number of customers in the servers (together) is: $\overline{N}_{\text{server}} = P_1 + 2 \cdot (1 - P_0 - P_1) = ...$

Utilization of an ARBITRARY server: $U = 0 \cdot P_0 + \frac{1}{2} P_1 + 1 \cdot (1 - P_0 - P_1)$, since at state S_1 a server is utilized with probability 1/2. Alternatively, we can divide the offered load, which is also the actual load, by the number of servers.

The probability that both servers are busy is: $1 - P_0 - P_1$.

8.2 Exercise 7.6

Consider a pure delay system where customers arrive according to a Poisson process with intensity $\lambda = 3$. The service time is exponentially distributed with mean value 1/3. The queuing discipline is FCFS. There are two servers in the system, and one of them is always available. The other one starts service when the queue length would become two (so that it immediately becomes one). If there are no more customers in the queue, the server which becomes idle first is closed (and stays closed until the queue length becomes two again). Let us denote the state of the system with (i,j) where i is the total number of customers in the system and j is the number of open servers. Give the Kendall notation of the system and draw the system diagram.

Solution: In this problem the state space is given. S_{ij} denotes the state where the total number of customers [in the server(s) and in the queue] is i and the number of active servers is j. Given the state space we draw the system diagram in Fig. 14. The system has 2 servers, infinite queue, Poisson arrivals and exponential service times, so the Kendall notation is: M/M/2, although it is not a typical case, as you see in the diagram.

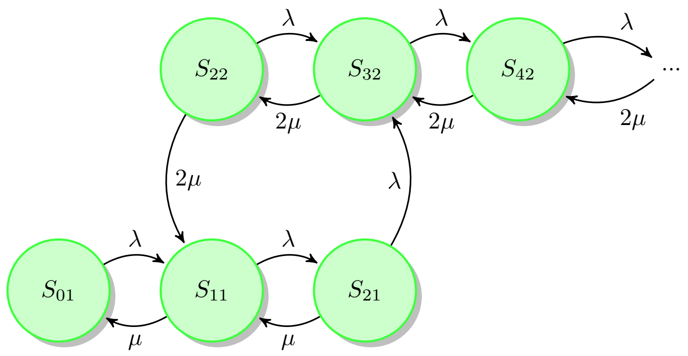


Figure 14: System diagram for exercise 7.6. Notice that the system transits from S_{21} to S_{32}, since this new arrival would make the queue length become 2, so the second server is activated, and the queue length remains one, while both servers, now, serve a customer.

9 Tutorial 8 – M/M/m systems with limited number of customers

10 Exercise 6.6

There are K computers in an office, and a single repairman. Each computer breaks down after an exponentially distributed time with parameter α. The repair takes an exponentially distributed amount of time with parameter β. Only one computer is being repaired at a time, computers break down independently from the repair process, and repair times and lifetimes of the computers are independent.

1. *What is the probability that i computers are working at time t?*
2. *What is the average failure rate (i.e. the average number of computers that fail per time unit)?*
3. *What is the percentage of time a repairman is busy?*
4. *What is the percentage of time when all computers are out-of-order (broken)?*
5. *How many computers should we have if we would like to have K computers to work on average?*

Solution:

By reading the first question of this problem we already realize what the desired *State Space* could be. We try with the obvious selection:

$$\text{State} \quad S_i: \quad i \text{ computers working}, \quad i = 0, 1, 2, ..., K$$

The K computers in the system is the *population size* of our model. And it is, clearly, **finite**. The single repairman represents the *servers* in our model (1). Each computer needs an exponentially distributed time for repair, so the service rates in our model are exponential. Finally, each computer breaks-down after an exponentially distributed time (after its repair) so the arrival process in the system is, also, Markovian.

To summarize:

- A break-down of a computer is an "arrival"
- A repair of a computer is a "departure" or a "service"

Based on the above, we depict the diagram in Fig. 10. We denote $\gamma = \beta/\alpha$. We draw the Balance Equations:

$$\begin{aligned}
&\beta P_0 = \alpha P_1 \rightarrow P_1 = \tfrac{\beta}{\alpha} P_0 = \gamma P_0, \\
&\beta P_1 = 2\alpha P_2 \rightarrow P_2 = \tfrac{\beta}{2\alpha} P_1 = \tfrac{1}{2}\gamma P_1 = \tfrac{1}{2}\gamma^2 P_0, \\
&\beta P_1 = 3\alpha P_2 \rightarrow P_2 = \tfrac{\beta}{3\alpha} P_1 = \tfrac{1}{3}\gamma P_1 = \tfrac{1}{2\cdot 3}\gamma^3 P_0, \\
&\text{...} \\
&\beta P_{K-1} = K\alpha P_K \rightarrow P_K = \tfrac{\beta}{K\alpha} P_{K-1} = ... = \tfrac{1}{K!}\gamma^K P_0.
\end{aligned} \qquad (12)$$

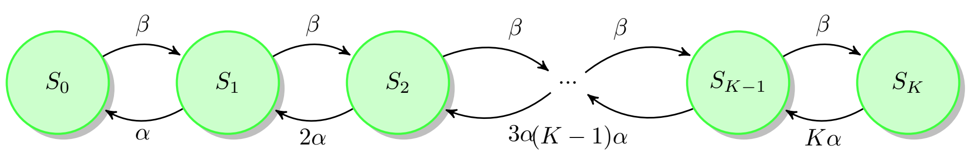


Figure 15: System diagram for exercise 6.6

From the above it is clear that the probability of an arbitrary state S_i is

$$P_i = \frac{\gamma^i}{i!} P_0, \quad i = 0, 1, 2, ..., K \tag{13}$$

From the above and the normalization equation we derive the P_0:

$$\sum_{j=1}^{K} P_j = 1 \rightarrow \sum_{j=1}^{K} \frac{1}{j!} \gamma^j P_0 = 1 \rightarrow P_0 = \frac{1}{\sum_{j=1}^{K} \frac{1}{j!} \gamma^j}. \tag{14}$$

P_0 gives the probability that all computers are out of order. In addition, the probability that i computers are working (in the steady-state) is given by:

$$P_i = \frac{\gamma^i}{i! \sum_{j=1}^{K} \frac{\gamma^j}{j!}}. \tag{15}$$

The repairman is busy in all states, except state S_K, where all computers are working properly. So the percentage of time the repairman is busy, will be:

$$P_{\text{rep. busy}} = 1 - P_K. \tag{16}$$

Next, we want to calculate the average failure rate ($\lambda_{\text{fail rate}}$). We notice that the failure rate **depends** on the state of the system. For example, the failure rate in state S_K is $K\alpha$, while, in case S_0 is it zero! We use the conditional expectation calculation to compute it:

$$\lambda_{\text{fail rate}} = \sum_{j=1}^{K} j \cdot \alpha P_j = \alpha \sum_{j=1}^{K} j \cdot P_j, \tag{17}$$

where P_j is given above. The average number of working computers will be calculated based on the state probability distribution, that is:

$$\overline{N} = \sum_{i=1}^{K} i \cdot P_i = \sum_{i=1}^{K} i \cdot \frac{\gamma^i}{i! \sum_{j=1}^{K} \frac{\gamma^j}{j!}} \tag{18}$$

We would like to have a system with K computers working (on average). Assume the required number of computers is M. Then M can be found as the lowest integer that satisfies the inequality:

$$\overline{N} \geq K \quad \rightarrow \quad \sum_{i=1}^{M} i \cdot \frac{\gamma^i}{i! \sum_{j=1}^{M} \frac{\gamma^j}{j!}} \geq K. \tag{19}$$

We can find the require M by trial-and-error.
Extra question: What is the probability that a broken computer can not be repaired immediately?
This probability can be found by taking the *ratio* of the rate of computer break-downs that can not be served immediately, over the total average rate of computer break-downs:

$$\lambda_W = \frac{\sum_{j=1}^{K-1} \alpha j P_j}{\sum_{j=1}^{K} \alpha j P_j} \tag{20}$$

11 Exercise 8.6

In a kitchen dormitory corridor there are two hobs for cooking and 3 places on the sofa. There are 8 students living in the corridor, each of them goes on average every $1/\alpha$ hours to the kitchen to cook (if he is not cooking already), the inter-arrival time is exponentially distributed. If on arrival the 2 hobs are occupied, the student looks for a place on the sofa. If the sofa is occupied as well, the student goes back to his room and tries again at a later time. Students spend an exponentially distributed amount of time cooking with mean $1/\beta$. $\alpha = 0.5$ hours, $\beta = 1$ hours.

1. *Draw the state transition diagram*
2. *Calculate the mean waiting time of the students*
3. *Calculate the ratio of time the kitchen is completely full, e.g. a student arriving has to go back to his room.*
4. *Calculate the probability that a student finds the kitchen completely full*
5. *Calculate the probability that a student has to wait more than 2 hours (supposing that he can sit down in the kitchen)*

Solution:

This is a complex problem that discusses all important matters in the Markovian systems with finite population size. It is, perhaps, easy to choose the state space; it could, clearly, be the number of places occupied by students at some point in time, either in the kitchen counter (hobs) or the kitchen sofa:

$$\text{State Space:} \quad S_k, \quad k \text{ positions occupied}, \quad k = 0, 1, ..., 5. \tag{21}$$

The population of the system is the number of students (8). The number of servers is the number of hobs (2) and the number of queuing positions is the number of places on the sofa (3). The inter-arrival times between student arrivals at the kitchen is exponentially distributed and the rate depends on the remaining number of students that are still at their rooms (not already cooking or waiting at the sofa). The service times are the cooking times and are also exponentially distributed. Consequently, the system is Markovian, with Kendall notation: M/M/2/5/8. The system diagram is depicted in Fig. 16. We define:

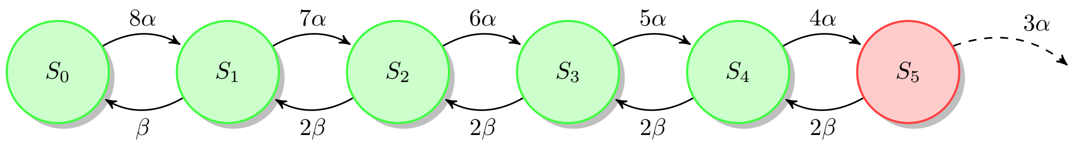


Figure 16: System diagram for exercise 8.6

$$\rho = \frac{\alpha}{\beta} = \frac{1}{2}.$$

Notice that is **not** the offered load at the system. We draw the balance equations:

$$\begin{aligned}
&8\alpha P_0 = \beta P_1 \to P_1 = 8\rho P_0, \\
&7\alpha P_1 = 2\beta P_2 \to P_2 = \tfrac{7}{2}\rho P_1 = 28\rho^2 P_0, \\
&6\alpha P_2 = 2\beta P_3 \to P_3 = 3\rho P_2 = 84\rho^3 P_0, \\
&5\alpha P_3 = 2\beta P_4 \to P_4 = \tfrac{5}{2}\rho P_3 = 210\rho^4 P_0, \\
&4\alpha P_4 = 2\beta P_5 \to P_5 = 2\rho P_2 = 420\rho^5 P_0
\end{aligned} \tag{22}$$

Applying the normalization equation:

$$\sum_{k=0}^{5} P_k = 1, \tag{23}$$

we calculate the values of the state probabilities.

Through the state probabilities we calculate, first, the percentage of time the kitchen is completely full, simply:

$$P_{\text{full}} = P_5. \tag{24}$$

Now, since the system has finite population, the arrival rates depend on the state, and, so the probability that a random student finds the system (kitchen) full is NOT equal to P_5. The **average student arrival rate** is calculated using the conditional expectation and taking into account the different arrival rates in each state:

$$\overline{\lambda} = \sum_{k=0}^{5} \lambda_k \cdot P_k, \tag{25}$$

where λ_k is the arrival rate at state S_k. Based on the diagram we obtain:

$$\overline{\lambda} = 8\alpha P_0 + 7\alpha P_1 + 6\alpha P_2 + 5\alpha P_3 + 4\alpha P_4 + 3\alpha P_3. \tag{26}$$

The **average student block rate** is the average rate of students that are blocked, i.e. find the kitchen completely full. Since they are only blocked at state S_5 we obtain:

$$\overline{\lambda}_{\text{block}} = 3\alpha P_5. \tag{27}$$

The ratio between the two average arrival rates gives the percentage of students that are blocked, or, equivalently, the probability that an arbitrary student is blocked (P_{blocked}):

$$P_{\text{blocked}} = \frac{\overline{\lambda}_{\text{block}}}{\overline{\lambda}} = \frac{3P_5}{8P_0 + 7P_1 + 6P_2 + 5P_3 + 4P_4 + 3P_3} \tag{28}$$

Clearly, the **effective** arrival rate of the system is the total average arrival rate minus the blocked arrival rate:

$$\lambda_{\text{eff}} = \overline{\lambda} - \overline{\lambda}_{\text{block}} = 8\alpha P_0 + 7\alpha P_1 + 6\alpha P_2 + 5\alpha P_3 + 4\alpha P_4. \tag{29}$$

We can find the mean waiting time of the student with the help of LITTLE's formula. First, we need to compute the average number of students in the kitchen:

$$\overline{N} = \sum_{i=0}^{5} kP_k = P_1 + 2P_2 + 3P_3 + 4P_4 + 5P_5. \tag{30}$$

Then, using LITTLE we can find the average SYSTEM time for a student:

$$E[T_{\text{system}}] = \frac{\overline{N}}{\lambda_{\text{eff}}}. \tag{31}$$

Then the average waiting time will be equal to the average system time minus the average service (cooking) time:

$$\overline{W} = E[T_{\text{system}}] - \frac{1}{\beta}. \tag{32}$$

There is another way to do this; that of considering the arrivals at each state and the waiting time a student experiences given the state of arrival. Here, we must reject the blocked arrivals cause they have no waiting time.

The last question is a bit challenging. We are asked to consider ONLY those students that are not rejected. So we must consider students that arrive at states $S_k, \quad k = 0, ...4$. Generally, we use the total probability theorem, and obtain:

$$\Pr\{W > 2h\} = \sum_{k=0}^{4} \Pr\{W > 2h | \text{Student sees state } S_k\} \cdot \Pr\{\text{Student sees state } S_k\} \tag{33}$$

1. If the student observes state S_0 or S_1 there is not waiting time. (Cooks immediately)
2. If the student observes state S_2 the student waits for an exponential amount of time with parameter 2β.
3. If the student observes state S_3 the student waits for a sum of two exponential amounts of time, each with parameter 2β.

4. If the student observes state S_4 the student waits for a sum of three exponential amounts of time, each with parameter 2β.

One option is to realize that the sum of exponential variables is an Erlang-distributed variable, and use the Erlang distribution formulas to calculate the $\Pr\{W > 2h|.....\}$.

There is however, a better option. We can realize that the times between departures are exponentially distributed variables with rate 2β, considering fully loaded kitchen counter. So, the number of departures within some time interval will be a Poisson random variable! This is the duality between the exponential and the Poisson distributions!

So, for example, if the student observes S_4 he will want for more than T=2h, if less than 3 departures occur during these two hours, and the number of these departures is Poisson $(2\beta \cdot T)$:

$$
\begin{aligned}
\Pr\{W > T|\text{Student sees state } S_4\} &= \left(\tfrac{(2\beta T)^0}{0!} + \tfrac{(2\beta T)^1}{1!} + \tfrac{(2\beta T)^2}{2!}\right) e^{-2\beta T} \\
\Pr\{W > T|\text{Student sees state } S_3\} &= \left(\tfrac{(2\beta T)^0}{0!} + \tfrac{(2\beta T)^1}{1!}\right) e^{-2\beta T} \\
\Pr\{W > T|\text{Student sees state } S_2\} &= \left(\tfrac{(2\beta T)^0}{0!} + \right) e^{-2\beta T}
\end{aligned} \tag{34}
$$

The last thing to calculate the probabilities that a random student observes a particular state. Using the same reasoning as in the calculation of the blocked students, we get:

$$
\begin{aligned}
\Pr\{\text{Student sees state } S_2\} &= \tfrac{6\alpha P_2}{\lambda_{\text{eff}}} \\
\Pr\{\text{Student sees state } S_2\} &= \tfrac{5\alpha P_3}{\lambda_{\text{eff}}} \\
\Pr\{\text{Student sees state } S_2\} &= \tfrac{4\alpha P_4}{\lambda_{\text{eff}}}
\end{aligned} \tag{35}
$$

Notice that we used λ_{eff} because we exclude those students that arrive at state S_5 and are blocked (because they have no waiting time).