5 Examples of $M/M/1$ type models

In this chapter we present some simple variations on the $M/M/1$ system; we will first summarize some of results for the $M/M/1$ system.

5.1 The $M/M/1$ system

In the $M/M/1$ system customers arrive according to a Poisson process and the service times of the customers are independent and identically exponentially distributed. This system can be described by a Markov process with states i, where i is simply the number of customers in the system. The generator Q of this Markov proces is given by

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \dots \\ 0 & \mu & -(\lambda+\mu) & \mu & 0 & \dots \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{1}$$

where λ is the arrival rate and μ the service rate (with $\lambda < \mu$). The corresponding transition-rate diagram of the $M/M/1$ model is shown in figure 1.

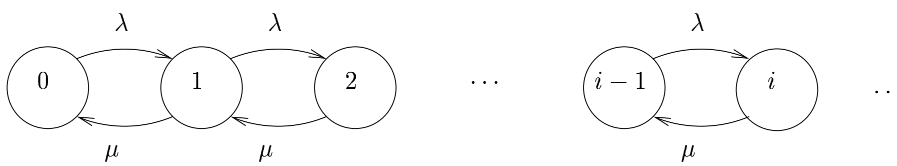

Figure 1: Transition-rate diagram for the $M/M/1$ model

Let p_i denote the (equilibrium) probability of state i, $i \geq 0$. From the transition-rate diagram it is easy to derive the equilibrium equations for the state probabilities p_i yielding

$$\begin{aligned} p_0\lambda &= p_1\mu, \\ p_i(\lambda+\mu) &= p_{i-1}\lambda + p_{i+1}\mu, \qquad i = 1, 2 \dots, \end{aligned}$$

or by rearranging all terms at the same side of the equation,

$$-p_0\lambda + p_1\mu = 0, \tag{2}$$

$$p_{i-1}\lambda - p_i(\lambda+\mu) + p_{i+1}\mu = 0, \qquad i = 1, 2 \dots \tag{3}$$

Together with the normalization equation, this set of equations has a (unique) geometric solution

$$p_i = (1-\rho)\rho^i, \qquad i = 0, 1, 2, \dots, \tag{4}$$

where $\rho = \lambda/\mu$.

An important feature of the system above is that transitions are restricted to neighboring states only, i.e., from state i to state $i-1$ or from state i to $i+1$. In the following sections we will consider models that share this feature, but in these models the simple state i is replaced by a set of states referred to as *level* i and the equilibrium distribution is a matrix generalization of (4); i.e., ρ will be replaced by a *rate matrix* R.

The excess probabilities for the waiting time W in the $M/M/1$ system may be computed by conditioning on the state at arrival. Given that there are i customers in the system at arrival, the waiting time is Erlang-i distributed with mean i/μ. By PASTA, the probability of finding i customers at arrival is p_i. Hence, we get

$$\begin{aligned} P(W>t) &= \sum_{i=1}^{\infty}(1-\rho)\rho^i\sum_{j=0}^{i-1}\frac{(\mu t)^j}{j!}e^{-\mu t} = \sum_{j=0}^{\infty}\frac{(\mu t)^j}{j!}e^{-\mu t}\sum_{i=j+1}^{\infty}(1-\rho)\rho^i \\ &= \sum_{j=0}^{\infty}\frac{(\mu t)^j}{j!}e^{-\mu t}\rho^{j+1} = \rho e^{-\mu(1-\rho)t}, \qquad t\geq 0. \end{aligned} \tag{5}$$

Thus the excess probabilities of the waiting time are exponential; for the models in the following sections we will generalize this result to mixtures of exponentials.

5.2 Machine with setup times

Let us consider a machine processing jobs in order of arrival. Jobs arrive according to a Poisson stream with rate λ and the processing times are exponential with mean $1/\mu$. For stability we assume that $\rho=\lambda/\mu<1$. The machine is turned off when the system is empty and it is turned on again when a new job arrives. But turning on the machine requires a setup time, which is exponential with mean $1/\theta$. We are interested in the effect of the setup time on the production lead time.

This model can be respresented as a Markov process with states (i,j) where i is the number of jobs in the system and j indicates the state of the machine: $j=0$ means that the machine is off or in the setup phase, $j=1$ means that it is on (i.e., ready to process jobs). The transition-rate diagram is displayed in figure 2. It looks similar to figure 1, except that each state i has been replaced by the set of states $\{(i,0),(i,1)\}$. This set of states is called level i. Transitions are now restricted to neigboring levels.

Let $p(i,j)$ denote the equilibrium probability of state (i,j), $i\geq 0$, $j=0,1$; clearly $p(0,1)=0$. From the transition-rate diagram we obtain by equating the flow out of a state and the flow into that state the following set of equilibrium equations,

$$p(0,0)\lambda = p(1,1)\mu, \tag{6}$$
$$p(1,0)(\lambda+\theta) = p(0,0)\lambda, \tag{7}$$
$$p(1,1)(\lambda+\mu) = p(1,0)\theta+p(2,1)\mu, \tag{8}$$
$$p(i,0)(\lambda+\theta) = p(i-1,0)\lambda, \qquad i=2,3,\ldots \tag{9}$$
$$p(i,1)(\lambda+\mu) = p(i,0)\theta+p(i+1,1)\mu+p(i-1,1)\lambda, \qquad i=2,3,\ldots \tag{10}$$

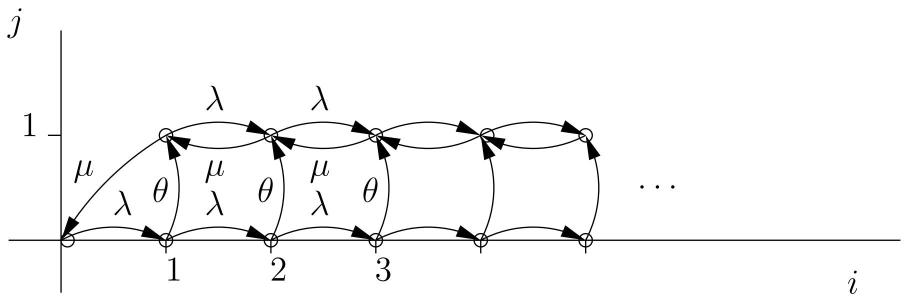

Figure 2: Transition-rate diagram for the $M/M/1$ model with setup times

The structure of the equations (9)-(10) is closely related to the similar set of equations (3). This becomes more striking by rewriting (9)-(10) in vector-matrix notation:

$$p_{i-1}A_0 + p_iA_1 + p_{i+1}A_2 = 0, \qquad i > 1, \tag{11}$$

where $p_i = (p(i,0), p(i,1))$ and

$$A_0 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} -(\lambda+\theta) & \theta \\ 0 & -(\lambda+\mu) \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}.$$

Obviously, if we can determine the equilibrium probabilities $p(i,j)$, then we can also compute the mean number of jobs in the system, and by Little's law, the mean production lead time. We now present two methods to determine the equilibrium probabilities. The first one is known as the *matrix-geometric method*, the other one is referred to as the *spectral expansion method*; see, e.g. [3, 1, 2]. Let us start with the matrix-geometric approach.

We first simplify the equilibrium equations (11) by eliminating the vector p_{i+1}. By equating the flow from level i to level $i+1$ to the flow from level $i+1$ to i (this is known as the *balance principle*) we obtain

$$(p(i,0) + p(i,1))\lambda = p(i+1,1)\mu$$

or in vector-matrix notation

$$p_iA_3 = p_{i+1}A_2$$

where

$$A_3 = \begin{pmatrix} 0 & \lambda \\ 0 & \lambda \end{pmatrix}.$$

Substituting this equation into (11) yields

$$p_{i-1}A_0 + p_i(A_1 + A_3) = 0, \qquad i > 1,$$

or

$$p_i = -p_{i-1}A_0(A_1+A_3)^{-1} = p_{i-1}R, \tag{12}$$

where

$$R = -A_0(A_1 + A_3)^{-1} = \begin{pmatrix} \lambda/(\lambda+\theta) & \lambda/\mu \\ 0 & \lambda/\mu \end{pmatrix}.$$

Iterating (12) leads to the matrix-geometric solution

$$p_i = p_1 R^{i-1}, \qquad i \geq 1. \tag{13}$$

Hence it is very similar to the solution for the $M/M/1$ model given by (cf. (4))

$$p_i = p_1 \rho^{i-1}, \qquad i \geq 1.$$

Finally, $p(0,0)$ and p_1 follow from the equations (6)-(8) and the normalization equation

$$1 = \sum_{i,j} p(i,j) = p(0,0) + p_1(I - R)^{-1}e,$$

where I is the identity matrix and e the column vector of ones. From (13) we obtain for $E(L)$, the mean number of jobs in the system,

$$E(L) = \sum_{i=1}^{\infty} i p_i e = \sum_{i=1}^{\infty} i p_1 R^{i-1} e = p_1(I - R)^{-2}e.$$

Finally, application of Little's law yields the mean production lead time, $E(S)$.

We now demonstrate the spectral expansion method. This method first seeks solutions of the equations (11) of the simple form

$$p_i = y \cdot x^{i-1}, \qquad i = 1, 2, \ldots,$$

where $y = (y(0), y(1)) \neq 0$ and $|x| < 1$. The latter is required, since we want to be able to normalize the solution afterwards. Substitution of this form into (11) and dividing by common powers of x gives

$$y\left(A_0 + xA_1 + x^2 A_2\right) = 0.$$

Hence, the desired values of x are the roots inside the unit circle of the determinantal equation

$$\det(A_0 + xA_1 + x^2 A_2) = 0. \tag{14}$$

In this case we have

$$\det(A_0 + xA_1 + x^2 A_2) = (\lambda - (\lambda + \theta)x)(\mu x - \lambda)(x - 1).$$

Hence, we find two roots, namely

$$x_1 = \frac{\lambda}{\lambda + \theta}, \qquad x_2 = \frac{\lambda}{\mu}.$$

For $i = 1, 2$, let $y = y_i$ be a nonnull solution of

$$y\left(A_0 + x_i A_1 + x_i^2 A_2\right) = 0.$$

The final step of the spectral expansion method is to linearly combine the two simple solutions to also satify the boundary equations (6)-(8); note here that the equilibrium equations (9)-(10) are linear. So we set

$$p_i = c_1 y_1 x_1^{i-1} + c_2 y_2 x_2^{i-1}, \qquad i = 1, 2, \ldots \tag{15}$$

where the coefficients c_1 and c_2 and $p(0,0)$ follow from the boundary equations (6)-(8) and the normalization equation

$$1 = p(0,0) + \frac{c_1 y_1 e}{1 - x_1} + \frac{c_2 y_2 e}{1 - x_2}.$$

Using representation (15) we obtain

$$E(L) = \sum_{i=1}^{\infty} i p_i e = \frac{c_1 y_1 e}{(1 - x_1)^2} + \frac{c_2 y_2 e}{(1 - x_2)^2},$$

and, again, application of Little's law produces the mean production lead time.

The two methods presented above are closely related: x_1 and x_2 are the eigenvalues of the rate matrix R and y_1 and y_2 are the corresponding eigenvectors.

It is also possible to derive the distribution of the production lead time. By conditioning on the state seen on arrival and using PASTA we get

$$\begin{aligned} P(S > t) &= \sum_{i=0}^{\infty} P(T + B_1 + \cdots + B_{i+1} > t) p(i,0) \\ &+ \sum_{i=1}^{\infty} P(B_1 + \cdots + B_{i+1} > t) p(i,1), \end{aligned}$$

where T is an exponential (residual) setup time with mean $1/\theta$ and $B_1, B_2, \ldots$ are independent exponential processing times with mean $1/\mu$ (and independent of the setup time T). Substituting the expressions (15) for $p(i,j)$ and using that a geometric sum of exponential random variables is again exponential (verify!), we get, after some algebra,

$$P(S > t) = \frac{1}{\mu(1-\rho) - \theta} \left[\mu(1-\rho) e^{-\theta t} - \theta e^{-\mu(1-\rho)t}\right].$$

Hence, the density of S is given by

$$f_S(t) = -\frac{d}{dt} P(S > t) = \frac{\mu(1-\rho)\theta}{\mu(1-\rho) - \theta} \left[e^{-\theta t} - e^{-\mu(1-\rho)t}\right]. \tag{16}$$

From (16) it follows that

$$\widetilde{S}(s) = E(e^{-sS}) = \int_{t=0}^{\infty} e^{-st} f_S(t)dt = \frac{\theta}{\theta + s} \cdot \frac{\mu(1-\rho)}{\mu(1-\rho) + s},$$

which implies that S is the sum of two independent exponentials, one with parameter θ and the other with parameter $\mu(1-\rho)$.

Remark 5.1 The mean number of jobs in the system, $E(L)$, and the mean production lead time, $E(S)$, can also be determined by combining the PASTA property and Little's law. Based on PASTA we know that the average number of jobs in the system seen by an arriving job equals $E(L)$, and each of them (also the one being processed) has a (residual) processing time with mean $1/\mu$. With probability $1-\rho$ the machine is not in operation on arrival, so that the job also has to wait for the setup phase with mean $1/\theta$. Further, the job has to wait for its own processing time. Hence

$$E(S) = (1-\rho)\frac{1}{\theta} + E(L)\frac{1}{\mu} + \frac{1}{\mu},$$

and together with Little's law

$$E(L) = \lambda E(S),$$

we find

$$E(S) = \frac{1/\mu}{1-\rho} + \frac{1}{\theta}.$$

The first term at the right-hand side is the mean production lead time in the system without setup times (i.e., the machine is always on). The second term is the mean setup time. Clearly, the mean setup time is exactly the extra mean delay caused by turning off the machine when there is no work. In fact, it can be shown (by using, e.g., a sample path argument, or see (16)) that the extra delay is an exponential time with mean $1/\theta$.

5.3 Unreliable machine

In this section we consider an unreliable machine processing jobs. The machine breaks down at random instants, whether it is processing or not. Thus the machine is subject to so-called *time-dependent* breakdowns (as opposed to *operational dependent* breakdowns, which can only occur when the machine is processing a job; see also remark 5.3). As soon as the machine has been repaired, processing resumes at the point where it was interrupted. To obtain some insight in the effects of the breakdowns we study the following simple model.

Jobs arrive according to a Poisson stream with rate λ. The processing times are exponential with mean $1/\mu$. The time between two breakdowns is exponentially distributed with mean $1/\eta$. The repair time is also exponentially distributed with mean $1/\theta$.

This system can be described by a Markov process with states (i, j) where i is the number of jobs in the system and j indicates the state of the machine; the machine is up if $j = 1$, it is down and in repair if $j = 0$. The transition-rate diagram of this system is shown in figure 3. It again looks similar to figure 1, except that each state i has been replaced by the set of states $\{(i,0),(i,1)\}$.

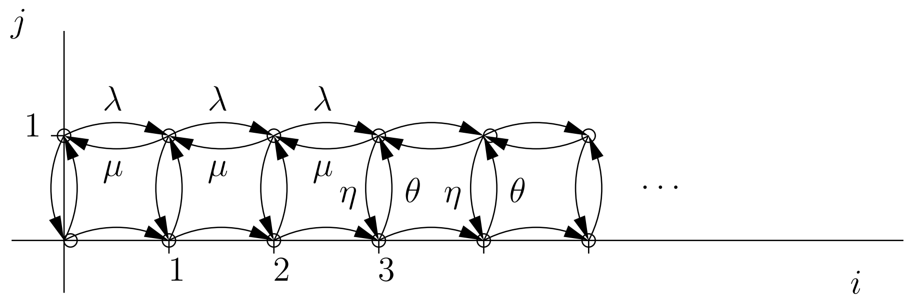

Figure 3: Transition-rate diagram for the $M/M/1$ model with time-dependent breakdowns

Let ρ_U denote the fraction of time the machine is up, so

$$\rho_U = \frac{1/\eta}{1/\eta + 1/\theta}.$$

Then, for stability, we have to require that

$$\frac{\lambda}{\mu} < \rho_U. \tag{17}$$

Let $p(i,j)$ denote the equilibrium probability of state (i,j). From the transition-rate diagram we get the following balance equations for the states $(0,0)$ and $(0,1)$,

$$\begin{aligned} p(0,0)(\lambda+\theta) &= p(0,1)\eta, & (18)\\ p(0,1)(\lambda+\eta) &= p(0,0)\theta + p(1,1)\mu, & (19)\end{aligned}$$

and for all states (i,j) with $i \geq 1$,

$$\begin{aligned} p(i,0)(\lambda+\theta) &= p(i-1,0)\lambda + p(i,1)\eta, \qquad i = 1,2,\ldots & (20)\\ p(i,1)(\lambda+\eta+\mu) &= p(i,0)\theta + p(i+1,1)\mu + p(i-1,1)\lambda, \quad i = 1,2,\ldots & (21)\end{aligned}$$

In vector-matrix notation these equations can be written as (cf. (11))

$$p_{i-1}A_0 + p_i A_1 + p_{i+1}A_2 = 0, \qquad i \geq 1,$$

where $p_i = (p(i,0), p(i,1))$ and

$$A_0 = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} -(\lambda+\theta) & \theta \\ \eta & -(\lambda+\mu+\eta) \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 \\ 0 & \mu \end{pmatrix}.$$

Similarly as for the $M/M/1$ model with setup times we can show that the solution to these equations has a matrix-geometric form

$$p_i = p_0 R^i, \qquad i \geq 0,$$

with

$$R = \frac{\lambda}{\mu}\begin{pmatrix} (\eta+\mu)/(\lambda+\theta) & 1 \\ \eta/(\lambda+\theta) & 1 \end{pmatrix},$$

or the following spectral expansion form

$$p_i = c_1 y_1 x_1^i + c_2 y_2 x_2^i, \qquad i = 0, 1, 2, \ldots$$

with x_1 and x_2 being the roots of

$$\mu(\lambda+\theta)x^2 - \lambda(\lambda+\mu+\eta+\theta)x + \lambda^2 = 0.$$

Based on these expressions for the equilibrium probabilities $p(i,j)$ it is easy to find closed-form expressions for the mean number of jobs in the system, $E(L)$, and the mean production lead time, $E(S)$. And, with some more effort, it can be shown that the distribution of the production lead time is a mixture of two exponentials.

In table 1 we compare the impact of frequent and small breakdowns on the mean production leadtime, with infrequent and long breakdowns. The mean processing time is 1 hour ($\mu = 1$). The average number of jobs that arrives during a week (40 hours) is 32, so $\lambda = 0.8$ jobs per hour. In each example, $\rho_U = 0.9$. The rate η at which the machine breaks down is varied from (on average) every 10 minutes till once a week. In the former case the mean repair time is 1.1 minute, in the latter case it is more dramatic, namely nearly half a day (4.4 hours). The results indicate that it is better to have frequent and small breakdowns than infrequent and long breakdowns. Note that as η and θ both tend to infinity such that $\eta/\theta = 1/9$, then $E(S)$ tends to 10, which is the mean sojourn time in an $M/M/1$ with arrival rate 0.8 and service rate 0.9.

η	θ	$E(S)$
6	54	10.02
3	27	10.03
1	9	10.1
0.125	1.125	10.8
0.0625	0.5625	11.6
0.025	0.225	14

Table 1: The mean production leadtime $E(S)$ as a function of the break-down rate η for fixed $\rho_U = 0.9$

Remark 5.2 The PASTA property and Little's law can be used to determine $E(L)$ and $E(S)$ directly, i.e. without knowledge of the detailed probabilities $p(i,j)$. An arriving job finds on average $E(L)$ jobs in the system and each of them has an exponential processing time with mean $1/\mu$. But the processing times are interrupted by random breakdowns; when the machine starts processing a job, then after an exponential time with mean $1/(\mu+\eta)$ the machine either finishes the job (with probability $\mu/(\mu+\eta)$) or it breaks down (with probability $\eta/(\mu+\eta)$). Hence, if $E(G)$ denotes the mean processing time *including* breakdowns, we get

$$E(G) = \frac{1}{\mu+\eta} + \frac{\mu}{\mu+\eta}\cdot 0 + \frac{\eta}{\mu+\eta}\cdot\left(\frac{1}{\theta} + E(G)\right),$$

so

$$E(G) = \frac{1}{\mu} + \frac{\eta}{\mu}\cdot\frac{1}{\theta} = \frac{1}{\mu\rho_U}\,.$$

Further, with probability $1-\rho_U$ the machine is already down on arrival, in which case our job has an extra mean delay of $1/\theta$. Summarizing we have

$$E(S) = (E(L)+1)E(G) + (1-\rho_U)\frac{1}{\theta} = (E(L)+1)\frac{1}{\mu\rho_U} + (1-\rho_U)\frac{1}{\theta}\,.$$

Then, with Little's law stating that $E(L) = \lambda E(S)$, we immediately obtain

$$E(S) = \frac{1/(\mu\rho_U) + (1-\rho_U)/\theta}{1-\lambda/(\mu\rho_U)}\,.$$

Remark 5.3 In case breakdowns can only occur while the machine is processing a job (so-called operational dependent breakdowns), we have to slightly adapt the model; state $(0,0)$ is not possible since the machine cannot go down while it is idle, see figure 4. The analysis of this model proceeds along exactly the same lines as the model with time-dependent breakdowns.

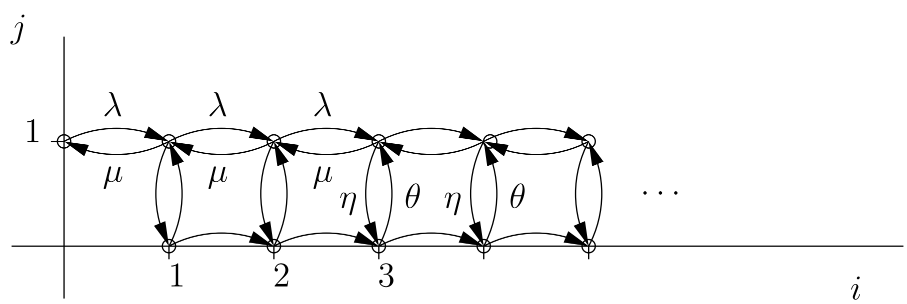

Figure 4: Transition-rate diagram for the $M/M/1$ model with operational dependent breakdowns

5.4 The $M/E_r/1$ model

We consider a single-server queue. Customers arrive according to a Poisson process with rate λ and they are served in order of arrival. The service times are Erlang-r distributed with mean r/μ. For stability we require that the occupation rate

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

is less than one. This system can be described by a Markov process with states (i, j) where i is the number of customers waiting in the queue and j is the remaining number of service phases of the customer in service. The transition-rate diagram is shown in figure 5.

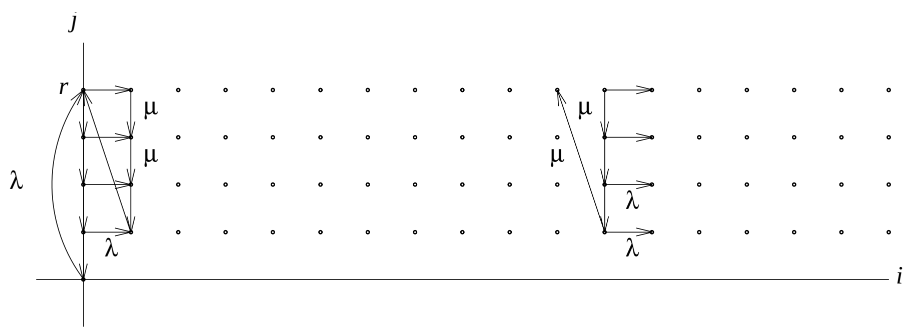

Figure 5: Transition-rate diagram for the $M/E_r/1$ model

Let $p(i, j)$ denote the equilibrium probability of state (i, j). From the transition-rate diagram we get the following balance equations for the states (i, j) with $i \geq 1$,

$$\begin{aligned} p(i,j)(\lambda+\mu) &= p(i-1,j)\lambda + p(i,j+1)\mu, \qquad j = 1,\ldots,r-1, &(22)\\ p(i,r)(\lambda+\mu) &= p(i-1,r)\lambda + p(i+1,1)\mu, &(23)\end{aligned}$$

or in vector-matrix notation

$$p_{i-1}A_0 + p_iA_1 + p_{i+1}A_2 = 0, \qquad i \geq 1, \tag{24}$$

where $p_i = (p(i,1),\ldots,p(i,r))$ and

$$A_0 = \begin{pmatrix} \lambda & 0 & \\ 0 & \ddots & 0 \\ & 0 & \lambda \end{pmatrix}, \quad A_1 = \begin{pmatrix} -(\lambda+\mu) & 0 & & \\ \mu & \ddots & 0 & \\ 0 & \ddots & \ddots & 0 \\ & 0 & \mu & -(\lambda+\mu) \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & \cdots & 0 & \mu \\ \vdots & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & & 0 \end{pmatrix}.$$

We first determine the probabilities $p(i, j)$ by using the matrix-geometric approach. Let level i denote the set of states $\{(i, 1), \ldots, (i, r)\}$. By balancing the flow between level i and level $i + 1$ we get

$$(p(i,1) + \cdots + p(i,r))\lambda = p(i+1,1)\mu$$

or

$$p_i A_3 = p_{i+1} A_2, \tag{25}$$

where

$$A_3 = \begin{pmatrix} 0 & \cdots & 0 & \lambda \\ \vdots & & 0 & \lambda \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & \lambda \end{pmatrix}.$$

To eliminate p_{i+1} we substitute equation (25) into (24) yielding

$$p_{i-1} A_0 + p_i (A_1 + A_3) = 0.$$

Hence

$$p_i = p_{i-1} R,$$

where

$$R = -A_0 (A_1 + A_3)^{-1}.$$

Note that $A_1 + A_3$ is invertable, since it is a transient (or leak) generator. Iterating the above equation yields

$$p_i = p_0 R^i, \qquad i = 0, 1, 2, \ldots$$

Finally the probabilities $p(0,0)$ and p_0 follow from the equilibrium equations for the states $(0,0), \ldots, (0,r)$ and the normalization equation.

To apply the spectral expansion method we substitute the simple form

$$p(i,j) = y(j) x^i, \qquad i \geq 0, 1 \leq j \leq r,$$

into the equilibrium equations (22)-(23), yielding

$$y(j) x(\lambda + \mu) = y(j)\lambda + y(j+1) x\mu, \qquad j = 1, \ldots, r-1, \tag{26}$$

$$y(r) x(\lambda + \mu) = y(r)\lambda + y(1) x^2 \mu. \tag{27}$$

Hence

$$\frac{y(j+1)}{y(j)} = \frac{x(\lambda + \mu) - \lambda}{x\mu} = \text{constant} \equiv y,$$

so

$$y(j) = y^j, \qquad j = 1, \ldots, r.$$

Substituting this into (26)-(27) gives

$$\begin{aligned} x(\lambda + \mu) &= \lambda + yx\mu, \\ x(\lambda + \mu) &= \lambda + \frac{x^2 \mu}{y^{r-1}}. \end{aligned}$$

This set of equations is equivalent to

$$
\begin{aligned}
x &= y^r, & (28)\\
y^r(\lambda+\mu) &= \lambda + y^{r+1}\mu. & (29)
\end{aligned}
$$

It can be shown that equation (29) has exactly r different (possibly complex) roots with $|y| < 1$; label these roots $y_1, \ldots, y_r$. Thus we find r *basis solutions* of the form

$$
p(i,j) = y_k^j x_k^i, \qquad k = 1, \ldots, r,
$$

where $x_k = y_k^r$. The next step is to take a linear combination of these basis solutions; so we set

$$
p(i,j) = \sum_{k=1}^{r} c_k y_k^j x_k^i, \qquad i = 0,1,2,\ldots, j = 1,\ldots,r, \tag{30}
$$

and determine the coefficients $c_1, \ldots, c_r$ and $p(0,0)$ such that the equilibrium equations for the states $(0,j)$, $0 \le j \le r$ and the normalization equation are satisfied.

From the probabilities $p(i,j)$ we can also compute the distribution of the amount of work in the system, expressed in terms of uncompleted service phases. Let p_n denote the probability that the number of uncompleted service phases in the system is equal to n, $n = 0,1,2,\ldots$. In state (i,j) the number of uncompleted service phases is $n = i \cdot r + j$, and thus, from (30) and (28),

$$
p_n = \sum_{k=1}^{r} c_k y_k^n = \sum_{k=1}^{r} d_k (1-y_k) y_k^n, \qquad n = 0,1,2,\ldots,
$$

where $d_k = c_k/(1-y_k)$.

The excess probabilities for the waiting time W may be computed in exactly the same way as for the $M/M/1$ model. By conditioning on the number of uncompleted service phases in the system just before an arrival and using PASTA, we obtain (cf. (5))

$$
\begin{aligned}
P(W > t) &= \sum_{n=1}^{\infty} p_n \sum_{j=0}^{n-1} \frac{(\mu t)^j}{j!} e^{-\mu t}\\
&= \sum_{k=1}^{r} d_k \sum_{n=1}^{\infty} (1-y_k) y_k^n \sum_{j=0}^{n-1} \frac{(\mu t)^j}{j!} e^{-\mu t}\\
&= \sum_{k=1}^{r} d_k y_k e^{-\mu(1-y_k)t}, \qquad t \ge 0.
\end{aligned}
$$

Remark 5.4 The vector $(y_k, y_k^2, \ldots, y_k^r)$ is the row eigenvector of the rate matrix R for eigenvalue x_k, $k = 1, \ldots, r$.

In table 2 we list for varying values of ρ and r the mean waiting time and some waiting time probabilities. The squared coefficient of variation of the service time is denoted by c_B^2. We see that the variation in the service times is important to the behavior of the system. Less variation in the service times leads to smaller waiting times.

ρ	r	c_B^2	$E(W)$		$P(W > t)$		
				t	5	10	20
0.8	1	1	4		0.29	0.11	0.02
	2	0.5	3		0.21	0.05	0.00
	4	0.25	2.5		0.16	0.03	0.00
	10	0.1	2.2		0.12	0.02	0.00
0.9	1	1	9		0.55	0.33	0.12
	2	0.5	6.75		0.46	0.24	0.06
	4	0.25	5.625		0.41	0.18	0.04
	10	0.1	4.95		0.36	0.14	0.02

Table 2: Performance characteristics for the $M/E_r/1$ with mean service time equal to 1

Remark 5.5 The mean production lead time $E(W)$ and the mean number of jobs waiting in the queue, $E(L^q)$, can also be determined by PASTA and Little. An arriving customer has to wait for the customers in the queue and, if the server is busy, for the one in service. According to the PASTA property, the mean number of customers waiting in the queue is equal to $E(L^q)$ and the probability that the server is busy on arrival is equal to ρ, i.e. the fraction of time the server is busy. Hence,

$$E(W) = E(L^q)\frac{r}{\mu} + \rho E(R), \tag{31}$$

where $E(R)$ denotes the mean *residual service time* of the customer in service. If the server is busy on arrival, then with probability $1/r$ he is busy with the first phase of the service time, also with probability $1/r$ he is busy with the second phase, and so on. So the mean residual service time $E(R)$ is equal to

$$\begin{aligned} E(R) &= \frac{1}{r}\cdot\frac{r}{\mu} + \frac{1}{r}\cdot\frac{r-1}{\mu} + \cdots + \frac{1}{r}\cdot\frac{1}{\mu} \\ &= \frac{r+1}{2}\cdot\frac{1}{\mu}. \end{aligned}$$

Substitution of this expression into (31) yields

$$E(W) = E(L^q)\frac{r}{\mu} + \rho\cdot\frac{r+1}{2}\cdot\frac{1}{\mu}.$$

Together with Little's law, stating that

$$E(L^q) = \lambda E(W)$$

we find

$$E(W) = \frac{\rho}{1-\rho}\cdot\frac{r+1}{2}\cdot\frac{1}{\mu}.$$

5.5 The $E_r/M/1$ model

In this section we consider a single-server queue with exponential service times with mean $1/\mu$. The arrival process is not Poisson. The interarrival times are Erlang-r distributed with mean r/λ; i.e., the time between two arrivals is a sum of r independent exponential phases, each with mean $1/\lambda$. For stability we assume that the occupation rate

$$\rho = \frac{\lambda}{r} \cdot \frac{1}{\mu}$$

is less than one. The states of the Markov process describing this system are the pairs (i, j), where i denotes the number of customers in the system and j the phase of the arrival process; i.e., $j = r$ means that already $r-1$ phases of the interarrival have been completed, so there is only one phase to go before the next arrival. The transition-rate diagram is depicted in figure 6.

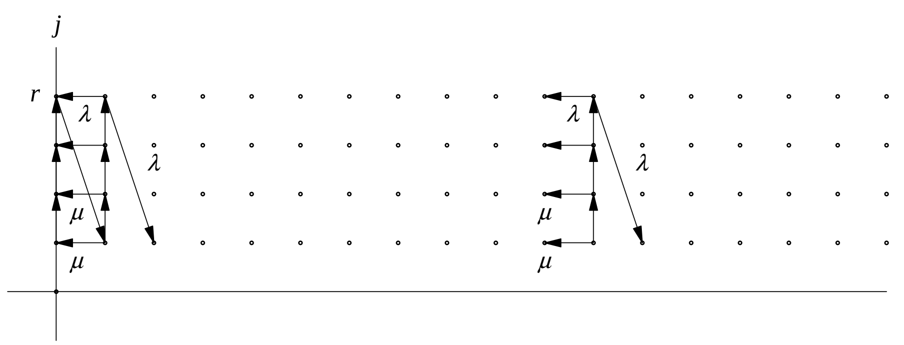

Figure 6: Transition-rate diagram for the $E_r/M/1$ model

Let us denote the state probabilities by $p(i, j)$. The equilibrium equations for the states (i, j) with $i \geq 1$ are formulated below.

$$\begin{aligned} p(i,1)(\lambda+\mu) &= p(i-1,r)\lambda + p(i+1,1)\mu, && (32)\\ p(i,j)(\lambda+\mu) &= p(i,j-1)\lambda + p(i+1,j)\mu, \qquad j = 2,\dots,r. && (33) \end{aligned}$$

In these equations we now substitute

$$p(i,j) = y(j)x^i, \qquad i = 1,2,\dots, j = 1,\dots,r;$$

also for $i = 0$ and $j = r$ (so $p(0, r) = y(r)$). This leads to

$$\begin{aligned} y(1)x(\lambda+\mu) &= y(r)\lambda + y(1)x^2\mu,\\ y(j)(\lambda+\mu) &= y(j-1)\lambda + y(j)x\mu, \qquad j = 2,\dots,r. \end{aligned}$$

Hence

$$\frac{y(j)}{y(j-1)} = \frac{\lambda}{\lambda+\mu-x\mu} = \text{constant} \equiv y,$$

so

$$y(j) = y^j, \qquad j = 1, \ldots, r,$$

where y satisfies

$$\begin{aligned} x(\lambda+\mu) &= y^{r-1}\lambda + x^2\mu, \\ y(\lambda+\mu) &= \lambda + yx\mu. \end{aligned}$$

This gives that $x = y^r$ and

$$x = \left(\frac{\lambda}{\lambda+\mu-\mu x}\right)^r. \tag{34}$$

Let $y_1 = \sqrt[r]{x_1}$. Then we eventually find

$$p(i,j) = c_1 y_1^j x_1^i, \qquad i = 1, 2, \ldots, j = 1, \ldots, r, \tag{35}$$

and this form is also valid for $p(0,r)$. The coefficient c_1 and the boundary probabilities $p(0,1), \ldots, p(0,r-1)$ follow from the balance equations for the states $(0,1), \ldots, (0,r)$ and the normalization equation.

Solution (35) may also be written in matrix-geometric form; it is easily verified that

$$p_i = (p(i,1), \ldots, p(i,r)) = p_1 R^{i-1}, \qquad i = 1, 2, \ldots,$$

where

$$R = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \\ y_1 & y_1^2 & \cdots & y_1^r \end{pmatrix}. \tag{36}$$

To determine the excess probabilities of the waiting time we *cannot use* PASTA, since the arrival process is not Poisson. Let L^a denote the number of customers in the system just before an arrival. Then we have

$$P(L^a = n) = \frac{p(n,r)\lambda}{\sum_{i=0}^{\infty} p(i,r)\lambda} = (1-x_1)x_1^n, \qquad n = 0, 1, 2, \ldots, \tag{37}$$

and hence (cf. (5)),

$$\begin{aligned} P(W > t) &= \sum_{n=1}^{\infty} P(L^a = n) \sum_{j=0}^{n-1} \frac{(\mu t)^j}{j!} e^{-\mu t} \\ &= \sum_{n=1}^{\infty} (1-x_1)x_1^n \sum_{j=0}^{n-1} \frac{(\mu t)^j}{j!} e^{-\mu t} \\ &= x_1 e^{-\mu(1-x_1)t}, \qquad t \geq 0. \end{aligned}$$

In table 3 we list for varying values of ρ and r the mean waiting time $E(W)$ and $P(W > 1/\mu)$. The squared coefficient of variation of the interarrival time is denoted by

c_A^2. We see that the variation in the interarrival times is important to the behavior of the system. Less variation in the interarrival times leads to smaller waiting times. Comparison of the mean waiting time $E(W)$ in tables 2 and 3 also suggests that variation in interarrival times has a stronger effect on waiting times than variation in service times.

ρ	r	c_A^2	$E(W)$	$P(W > 1/\mu)$
0.8	1	1	4	0.65
	2	0.5	2.84	0.57
	4	0.25	2.27	0.51
0.9	1	1	9	0.81
	2	0.5	6.59	0.76
	4	0.25	5.38	0.72

Table 3: Performance characteristics for the $E_r/M/1$ with mean service time equal to 1

Remark 5.6 Equation (34) can be rewritten as

$$x = \widetilde{A}(\mu - \mu x),$$

where $\widetilde{A}(s)$ is the Laplace-Stieltjes transform of the Erlang-r distribution with scale parameter λ; i.e.

$$\widetilde{A}(s) = \left(\frac{\lambda}{\lambda + s}\right)^r.$$

As we will see later on, this equation is important in the analysis of the $G/M/1$ queue.

References

[1] G. Latouche and V. Ramaswami (1999), *Introduction to Matrix Analytic Methods in Stochastic Modeling*. SIAM.

[2] I. Mitrani and D. Mitra, A spectral expansion method for random walks on semi-infinite strips, in: R. Beauwens and P. de Groen (eds.), *Iterative methods in linear algebra*. North-Holland, Amsterdam (1992), 141–149.

[3] M.F. Neuts (1981), *Matrix-geometric solutions in stochastic models*. The John Hopkins University Press, Baltimore.