# 7 $G/M/1$ type models

In this chapter we consider $G/M/1$ type models, i.e., generalizations of the ordinary $G/M/1$ queue, and we state some of the main results; for a detailed exposition of the analysis of $G/M/1$ type models the reader is referred to [4]. In the next section we first treat the $G/M/1$ queue.

## 7.1 The $G/M/1$ system

In the $G/M/1$ queue customers arrive one by one with interarrival times identically and independently distributed according to an arbitrary distribution function $F_A(\cdot)$ with density $f_A(\cdot)$. The mean interarrival time is equal to $1/\lambda$. The service times are exponentially distributed with mean $1/\mu$. For stability we again require that the occupation rate $\rho = \lambda/\mu$ is less than one. The state of the $G/M/1$ queue at time $t$ can be described by the pair $(i, x)$ where $i$ denotes the number of customers in the system and $x$ the residual interarrival time. So we need a complicated two-dimensional state description. But the state description is much easier at special points in time. If we look at the system just before arrival instants, then the state description can be simplified to $i$ only, because $x = 0$ just before an arrival. Below we are going to study this *Markov chain embedded on arrival instants*. To specify the transition probabilities of this Markov chain we first introduce the probabilities $a_n$ defined as the probability that exactly $n$ customers are served during an interarrival time (assuming there are at least $n$ customers present at the start of the interarrival time). By conditioning on the length of the interarrival time it follows that

$$a_n = \int_{t=0}^{\infty} \frac{(\mu t)^n}{n!} e^{-\mu t} f_A(t) d(t), \qquad n = 0, 1, 2, \ldots \tag{1}$$

Further let $b_n$ denote the probability that *more than* $n$ customers are served during an interarrival time, so

$$b_n = \sum_{k>n} a_k.$$

Then the transition probability matrix $P$ takes the form

$$P = \begin{pmatrix} b_0 & a_0 & 0 & 0 & 0 & \cdots \\ b_1 & a_1 & a_0 & 0 & 0 & \cdots \\ b_2 & a_2 & a_1 & a_0 & 0 & \cdots \\ b_3 & a_3 & a_2 & a_1 & a_0 & \cdots \\ b_4 & a_4 & a_3 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

The equilibrium probabilities $p_i$ satisfy the equilibrium equations

$$\begin{aligned} p_0 &= p_0 b_0 + p_1 b_1 + p_2 b_2 + \cdots \\ &= \sum_{n=0}^{\infty} p_n b_n \qquad (2) \\ p_i &= p_{i-1} a_0 + p_i a_1 + p_{i+1} a_2 + \cdots \\ &= \sum_{n=0}^{\infty} p_{i-1+n} a_n, \qquad i = 1, 2, \ldots \qquad (3) \end{aligned}$$

To solve the equilibrium equations we try to find solutions of the form

$$p_i = \sigma^i, \quad i = 0, 1, 2, \ldots \qquad (4)$$

Substitution of this form into equation (3) and dividing by the common power $\sigma^{i-1}$ yields

$$\sigma = \sum_{n=0}^{\infty} \sigma^n a_n \,.$$

Of course we know that $a_n$ is given by (1). Hence we have

$$\begin{aligned} \sigma &= \sum_{n=0}^{\infty} \sigma^n \int_{t=0}^{\infty} \frac{(\mu t)^n}{n!} e^{-\mu t} f_A(t) dt \\ &= \int_{t=0}^{\infty} e^{-(\mu - \mu\sigma)t} f_A(t) dt. \end{aligned}$$

The last integral can be recognised as the Laplace-Stieltjes transform of the interarrival time. Thus we arrive at the following equation

$$\sigma = \widetilde{A}(\mu - \mu\sigma), \qquad (5)$$

where

$$\widetilde{A}(s) = \int_{t=0}^{\infty} e^{-st} f_A(t) dt.$$

Clearly, $\sigma = 1$ is a root of equation (5), since $\widetilde{A}(0) = 1$. But this root is not useful, because we must be able to normalize the solution of the equilibrium equations. It can be shown that as long as $\rho < 1$ equation (5) has a unique root $\sigma$ in the range $0 < \sigma < 1$, and this is the root which we seek. Note that the remaining equilibrium equation (2) is also satisfied by (4), since the equilibrium equations are dependent. We finally have to normalize solution (4) yielding the geometric form (cf. (37) for the $E_r/M/1$ queue in chapter 5)

$$p_i = (1 - \sigma)\sigma^i, \qquad i = 0, 1, 2, \ldots,$$

In the following section we introduce a model, in continuous time, for which the generator $Q$ has the same transition strucure as the transition probability matrix $P$ for the $G/M/1$ queue. But in this model the simple state $i$ is replaced by a set of states (referred to as level $i$). Its equilbrium distribution will have a matrix-geometric form (or a sum of geometric terms).

**Remark 7.1** There is a simple probabilistic argument why the equilibrium probabilities $p_i$ are of the form (4). The ratio $p_{i+1}/p_i$ is the expected number of visits to state $i+1$ inbetween two successive visits to state $i$. The structure of $P$ implies that $p_{i+1}/p_i$ is the same for all $i$, i.e., this ratio does not depend on $i$ (Why?).

## 7.2 The $G/M/1$ type Model

We consider a Markov process, the state space of which consists of the *boundary states* $(0, j)$ where $j$ ranges from 0 to $n$, and a semi infinite strip of states $(i, j)$ where $i$ ranges from 1 to $\infty$ and $j$ from 0 to $m$. The states are ordered lexicographically, that is, $(0,0), (0,1), \ldots, (0,n), (1,0), \ldots, (1,m), (2,0), \ldots, (2,m), \ldots$. The set of boundary states $\{(0,0), (0,1), \ldots, (0,m)\}$ will be called *level* 0, and the set of states $\{(i,0), (i,1), \ldots, (i,n)\}$, $i \geq 1$, will be called *level* $i$. We partition the state space according to these levels, and for this partitioning we assume that the generator $Q$ is of the form

$$Q = \begin{pmatrix} B_{00} & B_{01} & 0 & 0 & 0 & \cdots \\ B_{10} & B_{11} & A_0 & 0 & 0 & \cdots \\ B_{20} & A_2 & A_1 & A_0 & 0 & \cdots \\ B_{30} & A_3 & A_2 & A_1 & A_0 & \cdots \\ B_{40} & A_4 & A_3 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix},$$

where the matrix $B_{00}$ is of dimension $(n+1) \times (n+1)$, $B_{0,1}$ of dimension $(n+1) \times (m+1)$, the matrices $B_{i0}$, $i \geq 1$, of dimension $(m+1) \times (n+1)$, and $B_{11}$ and $A_i$, $i \geq 0$, are square matrices of dimension $m+1$. Let

$$A = \sum_{i=0}^{\infty} A_i.$$

Note that $A$ is a generator; it describes the behavior of the Markov process $Q$ in the (vertical) $j$-direction only. We assume that the Markov process $Q$ is irreducible and that the generator $A$ has exactly one communicating class. For the stability of $Q$ we have the same result as theorem 6.3: the Markov process $Q$ is ergodic if and only if

$$\pi A_0 e < \pi \sum_{i=2}^{\infty} (i-1) A_i e,$$

where $e$ is the column vector of ones and $\pi = (\pi_0, \pi_1, \ldots, \pi_m)$ is the equilibrium distribution of the Markov process with generator $A$; so

$$\pi A = 0, \qquad \pi e = 1.$$

In the sequel we will assume that the Markov process $Q$ is ergodic. Thus the equilibrium probabilities $p(i,j)$ exist. Let $p_i$ denote the vector of equilibrium probabilities of level $i$, so

$$p_0 = (p(0,0), p(0,1), \ldots, p(0,n)), \qquad p_i = (p(i,0), p(0,1), \ldots, p(0,m)), \qquad i = 1, 2, \ldots.$$

These probability vectors $p_i$ satisfy the equilibrium equations

$$p_0 B_{00} + p_1 B_{10} + \sum_{n=2}^{\infty} p_n B_{n0} = 0, \tag{6}$$

$$p_0 B_{01} + p_1 B_{11} + \sum_{n=2}^{\infty} p_n A_2 = 0, \tag{7}$$

$$\sum_{n=0}^{\infty} p_{i-1+n} A_n = 0, \qquad i = 2, 3, \ldots, \tag{8}$$

and, of course, the normalization equation,

$$\sum_{i=0}^{\infty} p_i e = 1.$$

In the following section we describe the matrix-geometric results, which are very similar to the ones in section 6.2.

## 7.3 The matrix-geometric method

Provided the Markov process $Q$ is ergodic, the equilibrium probability vectors $p_i$ are given by the matrix-geometric form

$$p_i = p_1 R^{i-1}, \qquad i = 1, 2, \ldots, \tag{9}$$

where the matrix $R$ is the *minimal nonnegative solution* of the matrix equation

$$\sum_{n=0}^{\infty} R^n A_n = 0. \tag{10}$$

The matrix $R$ has spectral radius less than one (so $I - R$ is invertable). Note that, if $R$ satisfies (10), then it is easily seen that the matrix-geometric form (9) for $p_i$ indeed satisfies the equilibrium eqations for the levels $i > 1$; substitution of (9) into the left-hand side of (8) yields $p_{i-1} \sum_{n=0}^{\infty} R^n A_n$, which vanishes if $R$ satisfies (10). The boundary equations for $p_0$ and $p_1$ are exactly the same as for the $M/M/1$ type model, treated in section 6.2. Hence, in comparison with the $M/M/1$ results, the only difference is that the matrix-quadratic equation for $R$ is replaced by equation (10); this of course complicates the computation of $R$. Equation (10) can be rewritten as

$$R = -(A_0 + \sum_{n=2}^{\infty} R^n A_n) A_1^{-1}.$$

To solve this equation we first have to *truncate the infinite sum* at $N$, say, and then compute an approximation for $R$ by successive substitutions, i.e.,

$$R_{k+1} = -(A_0 + \sum_{n=2}^{N} R_k^n A_n) A_1^{-1}, \qquad k = 0, 1, 2, \ldots$$

starting with $R_0 = 0$. The larger $N$, the better the resulting approximation for $R$, but also the higher the computational effort to compute this approximation.

We finally mention that the rate matrix $R$ has the same probabilistic interpretation as described in section 6.2.

## 7.4 Spectral expansion method

Along the same lines as in section 6.3 it can be shown that the equilbrium probability vectors $p_i$ can be expressed as

$$p_i = \sum_{j=0}^{m} c_j y_j x^{i-1}, \qquad i = 1, 2, \ldots$$

where $x_0, x_1, \ldots, x_m$ are the roots inside the unit circle of

$$\det(\sum_{n=0}^{\infty} x^n A_n) = 0. \tag{11}$$

The vector $y_j$, $j = 0, 1, \ldots, m$, is a nonnul solution of

$$y \sum_{n=0}^{\infty} x^n A_n = 0.$$

The difficulties, however, with this approach are (i) to prove that equation (11) has indeed $m+1$ (different) roots $x$ with $|x| < 1$, and (ii) the computation of these roots. In the next chapter we will consider a special class of $G/M/1$ models, for which these difficulties can be resolved.

## 7.5 Example: The $G/PH/1$ queue

In this section we will study a single-server queue with phase-type service times and arbitrarily distributed interarrival times. The interarrival time distribution is denoted by $F_A(\cdot)$ with density $f_A(\cdot)$ and mean $1/\lambda$. The service times have a mixed Erlang-$r$ distribution with scale parameter $\mu$. This means that with probability $q_n$ the service time is the sum of $n$ exponential phases, each with the same parameter $\mu$, $n = 1, 2, \ldots, r$. The phase representation of this distribution is shown in figure 1.

The system behavior will be analyzed at arrival instants. The state on arrival instants can be described by the pair $(i, j)$, where $i$ is the number of customers in the system and $j$ the number of remaining service phases of the customer in service just before an arrival. This *two-dimensional* description leads to a $G/M/1$ type model, as studied in this chapter; see [4, 2, 3, 5, 6] for efficient algorithms for the computation of the matrix-geometric solution.

Alternatively the state on arrival instants can be described by the *one-dimensional* states $i$ where $i$ is the total number of uncompleted service phases in the system. In doing

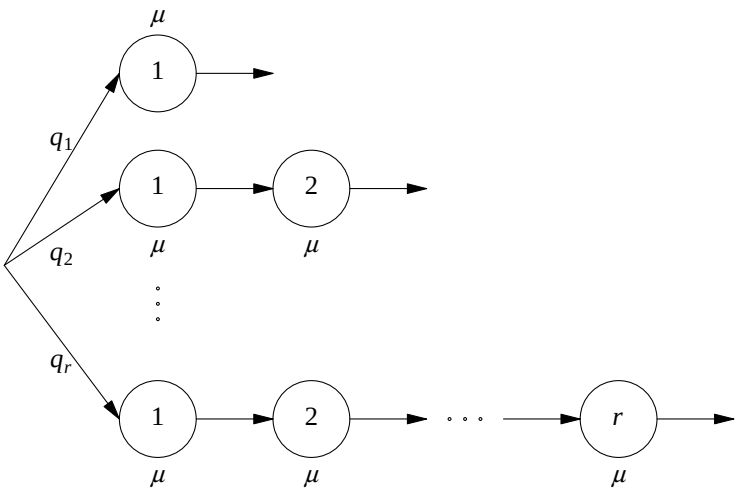


Figure 1: Phase representation of the mixed Erlang service time distribution

so, we loose part of the information, since we cannot determine the number of customers in the system from this description (except when the service times have a pure Erlang distribution). But information on the number of uncompleted service phases is all that is needed to determine the waiting time. As we will see, the analysis of this one-dimensional Markov chain is very similar to the analysis of the embedded Markov chain for the $G/M/1$ system in section 7.1.

The transition probability $p_{ij}$ from state $i \geq 0$ to state $j > 0$ is given by

$$p_{ij} = \begin{cases} q_1 a_{i+1-j} + q_2 a_{i+2-j} + \cdots + q_r a_{i+r-j} & i \geq 0\,, \;\; 0 < j \leq i + r, \\ 0 & i \geq 0\,, \;\; j > i + r, \end{cases}$$

where $a_n$ is defined as the probability that $n$ service stages are completed during an inter-arrival time, so

$$a_n = \int_0^\infty \frac{(\mu t)^n}{n!} e^{-\mu t} f_A(t) dt\,, \qquad n \geq 0\,.$$

Note that the Markov chain on arrival instants is irreducible and aperiodic. Henceforth it will be assumed that the offered load $\rho$, defined by

$$\rho = \lambda \left( q_1 \cdot \frac{1}{\mu} + q_2 \cdot \frac{2}{\mu} + \cdots + q_r \cdot \frac{r}{\mu} \right),$$

is less than 1. Then the equilibrium probabilities $p_i$ of finding $i$ customers on arrival exist. For $i > 0$ the equilibrium equations are given by

$$p_i = q_1 \sum_{n=0}^{\infty} p_{i-1+n} a_n + q_2 \sum_{n=0}^{\infty} p_{i-2+n} a_n + \cdots + q_r \sum_{n=0}^{\infty} p_{i-r+n} a_n, \tag{12}$$

where by convention

$$p_{-1} = p_{-2} = \cdots = p_{1-r} = 0. \tag{13}$$

Note that we do not pay attention to the equilibrium equation in state 0; since the equilibrium equations are dependent, this one will be satisfied automatically once we have satisfied (12). To solve the equilibrium equations (12) we try to find $r$ basis solutions of the form

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

Substitution of this form into (12) and division by $\sigma^{i-1}$ yields

$$\sigma^r = \left(q_1\sigma^{r-1} + q_2\sigma^{r-2} + \cdots + q_r\right) \sum_{n=0}^{\infty} a_n \sigma^n,$$

and thus, by substituting the expression for $b_k$, we find the following equation for $\sigma$,

$$\sigma^r = \left(q_1\sigma^{r-1} + q_2\sigma^{r-2} + \cdots + q_r\right) E(e^{-\mu(1-\sigma)A}), \tag{14}$$

where the generic random variable $A$ has distribution $F_A(\cdot)$. Clearly, only solutions with $|\sigma| < 1$ are useful. By using Rouché's Theorem it can be shown that equation (14) has exactly $r$ roots inside the unit circle (cf. [1]). We assume that these roots are all different and label them $\sigma_1, \sigma_2, \ldots, \sigma_r$. Now we take the linear combination

$$p_i = \sum_{k=1}^{r} c_k (1 - \sigma_k)\sigma_k^i.$$

For any choice of the coefficients $c_k$, this linear combination satisfies (12); it remains to determine the coefficients $c_k$ such that the convention (13) is satisfied. Substitution of this linear combination into (13) yields

$$c_1(1-\sigma_1)\tau_1^i + c_2(1-\sigma_2)\tau_2^i + \cdots + c_r(1-\sigma_r)\tau_r^i = 0, \qquad i = 1, 2, \ldots, r-1,$$

where $\tau_k = 1/\sigma_k$. These equations are of a *VanderMonde-type* and therefore, they can be solved explicitly using Cramer's rule. Then we get

$$c_k = \frac{C}{\prod_{j=1}^{r}(1-\tau_j)} \frac{\prod_{j\neq k}(1-\tau_j)}{\prod_{j\neq k}(\tau_k - \tau_j)}, \qquad k = 1, \ldots, r,$$

for some constant $C$. This constant follows from the normalization equation, which, by using Lagrange's interpolation formula, leads to

$$C = \prod_{j=1}^{r}(1-\tau_j).$$

Our findings are summarized in the following theorem.

**Theorem 7.2** *For all* $i = 0, 1, 2, \ldots$,

$$p_i = \sum_{k=1}^{r} c_k (1 - \sigma_k)\sigma_k^i,$$

*where $\sigma_1, \ldots, \sigma_r$ are the roots with $|\sigma| < 1$ of equation (14) and (with $\tau_j = 1/\sigma_j$)*

$$c_k = \frac{\prod_{j \neq k}(1 - \tau_j)}{\prod_{j \neq k}(\tau_k - \tau_j)}, \qquad k = 1, \ldots, r\,.$$

The arrival probabilities $p_i$ are of the same form as the one for the standard $G/M/1$ queue (i.e., a sum of geometric distributions). Thus the waiting time distribution can also be found in the same way as for the $G/M/1$, yielding

$$P(W > t) = \sum_{k=1}^{r} c_k \sigma_k e^{-\mu(1-\sigma_k)t}, \qquad t \geq 0. \tag{15}$$

Based on (15) it is easy to find expressions for the moments of the (conditional) waiting time. Hence the problem of finding the waiting time distribution has been reduced to that of finding the roots $\sigma_k$ of (14).

In the special case of (pure) Erlang-$r$ service time the roots $\sigma_k$ can be found very efficiently. Then (14) simplifies to

$$\sigma^r = E(e^{-\mu(1-\sigma)A}).$$

The idea is to reduce this equation for $r$ roots to $r$ equations for a single root, by raising both sides of the above equation to the power $1/r$. This leads to

$$\sigma = \phi F(\sigma)\,, \tag{16}$$

where $\phi$ satisfies $\phi^r = 1$ and

$$F(\sigma) = \sqrt[r]{E(e^{-\mu(1-\sigma)A})}\,.$$

Thus $\phi$ can be selected from the $r$ unity roots $e^{2\pi i m/r}$, $m = 0, 1, \ldots, r-1$. For each choice of $\phi$ equation (16) is a fixed point equation. We can try to find the root of (16) with $|\sigma| < 1$ by using the iteration scheme

$$\sigma^{(k+1)} = \phi F(\sigma^{(k)})\,, \qquad k = 0, 1, \ldots$$

starting with $\sigma^{(0)} = 0$. For certain classes of interarrival time distributions it can be shown that, indeed, the sequence $\sigma^{(0)}, \sigma^{(1)}, \ldots$ converges to the desired root; see [1] for more details. These classes include deterministic, shifted exponential, gamma, mixed Erlang and hyper-exponential distributions.

# References


[1] I.J.B.F. Adan, Y. Zhao, *Analyzing $GI/E_r/1$ queues*. Operations Research Letters, 19 (1996), pp. 183–190.

[2] W.K. Grassman, *The $GI/PH/1$ queue: a method to find the transition matrix*. Infor., 20 (1982), pp. 144–156.

[3] D.M. Lucantoni, *Efficient algorithms for solving the non-linear matrix equations arising in phase type queues*. Stochastic Models, 1 (1985), pp. 29–51.

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

[5] V. Ramaswami, D.M. Lucantoni, *Moments of the stationary waiting time in the $GI/PH/1$ queue*. J. Appl. Prob., 25 (1988), pp. 636–641.

[6] V. Ramaswami, G. Latouche, *An experimental evaluation of the matrix-geometric method for the $GI/PH/1$ queue.* Stochastic Models, 5 (1989), pp. 629–667.