The M/M/1 queue with gated random order of service

The $M/M/1$ queue with gated random order of service

Jacques Resing
Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Ronald Rietman
Philips Research Laboratories
Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Abstract

We analyse an $M/M/1$ queueing model with gated random order of service discipline. In this service discipline there is a waiting room, in which arriving customers are collected, and a service queue. Each time the service queue becomes empty, all customers in the waiting room are put instantaneously and in random order in the service queue. We find the joint stationary distribution of the number of customers in waiting room and service queue. Furthermore, we obtain the bivariate Laplace–Stieltjes transform of the joint distribution of the sojourn times of a customer in the waiting room and the service queue.

Keywords: $M/M/1$ queue, gated random order of service, two-dimensional Markov process, compensation method, sojourn times.

1 Introduction

In this paper we consider an $M/M/1$ queueing model with gated random order of service discipline. In this service discipline customers are first gathered in an unordered waiting room before they are put in random order in an ordered service queue at the moments that this latter queue becomes empty.

The model is motivated by a situation encountered in multi-access communication in cable networks. Cable networks are currently being upgraded to support bidirectional data transport. The system is thus extended with an "upstream" channel to complement the "downstream" channel that is already present. This upstream channel is shared among many stations so that contention resolution is essential for data transport. An efficient way to carry out the upstream data transport is via a request-grant mechanism. Stations request data slots in contention with other stations via contention trees. After a successful request, data transfer follows in reserved slots, not in contention with other stations.

There are two versions of the contention resolution mechanism via contention trees: the free access variant and the blocked access variant (see Mathys and Flajolet [8]). Essential features of the blocked access variant are

- requests competing in the same tree leave the tree in random order;
- new requests arriving when a tree is in progress have to wait until the current tree is resolved before they can be part of a tree itself.

Exactly these two features lead us to the study of the queueing model with gated random order of service discipline. Here, customers in the service queue represent the requests currently competing in the tree. Customers in the waiting room represent the requests waiting until the current tree is resolved. Recently, motivated by the same application, also a machine-repair model with gated random order of service discipline at the repair facility has been studied by Boxma, Denteneer and Resing [5].

For the $M/M/1$ queue with gated random order of service, we are particularly interested in the joint stationary distribution of the sojourn times of a customer in the waiting room and the service queue. In order to find this distribution, we will first study the two-dimensional Markov process describing the number of customers in the waiting room and in the service queue. The stationary distribution of this two-dimensional Markov process is found using a compensation method for queueing problems developed by Adan [1].

A queueing model with two stages of waiting has been studied before by Ali and Neuts [3] (see also Boxma and Cohen [4] for a related model). The essential difference between the models in [3, 4] and our model is that in [3, 4] the transfer of customers from the waiting room to the service queue takes some random transfer time $T > 0$, while in our model this transfer time is equal to 0. Furthermore, also the feature that customers, after transfer, are put in a random order in the service queue is not captured in the models in [3, 4].

The rest of the paper is organized in the following way. In section 2 we describe in detail the model under consideration. Next, section 3 is devoted to the derivation of the joint stationary distribution of the number of customers in the waiting room and in the service queue. In section 4 we present an asymptotic result for the stationary distribution of the number of customers in the waiting room, given that the total number of customers in the waiting room and service queue together equals N. Finally, in section 5 we present results for the joint stationary distribution of the sojourn times of customers in the waiting room and the service queue.

2 Model description

Customers arrive according to a Poisson process with rate λ at a single server system. The service times of the customers are exponentially distributed with parameter μ. We assume that $\rho := \lambda/\mu < 1$. The waiting area in front of the server consists of two parts: an unordered waiting room and an ordered service queue, both of infinite capacity. Upon arrival a customer first enters the waiting room. Each time the service queue becomes empty, all customers present in the waiting room at that moment are transferred instantaneously from the waiting room to the service queue. They are put in random order into this queue and the order in which they are put into the queue determines the order in which they are taken into service later on. If there are no customers present in the waiting room at the moment that the service queue becomes empty, the server waits for the next customer to arrive, transfers this customer immediately to the service queue and starts serving this customer. It follows that the service queue cannot be empty unless the waiting room is empty. The service discipline that we obtain in this way, customers first waiting "behind a gate" in the waiting room and after that put into random order in the service queue, will be called *gated random order of service.*

The number of customers in the waiting room at time t will be denoted by $X_1(t)$ and the number of customers in the service queue (including the customer in service) at time t will be denoted by $X_2(t)$. Clearly, the stochastic process $\{(X_1(t), X_2(t)) : t \geq 0\}$ is a two-dimensional Markov process. The next section is devoted to the determination of the steady-state probabilities

$$\pi(k,n) = \lim_{t\to\infty} P(X_1(t) = k, X_2(t) = n),$$

of this Markov process. In the sequel we denote with (X_1, X_2) a pair of random variables with joint distribution given by the probabilities $\pi(k,n)$. Remark that $X_1 + X_2$ has the same distribution as the stationary number of customers in an ordinary $M/M/1$ queue with FCFS service, i.e.,

$$P(X_1 + X_2 = j) = (1-\rho)\rho^j, \qquad j = 0, 1, 2, \ldots. \tag{1}$$

3 The steady-state probabilities $\pi(k,n)$

The balance equations for the steady-state probabilities $\pi(k,n)$ are given by

$$\rho\pi(0,0) = \pi(0,1), \tag{2}$$
$$(\rho+1)\pi(0,1) = \rho\pi(0,0) + \pi(1,1) + \pi(0,2), \tag{3}$$
$$(\rho+1)\pi(0,n) = \pi(n,1) + \pi(0,n+1), \qquad n \geq 2, \tag{4}$$
$$(\rho+1)\pi(k,n) = \rho\pi(k-1,n) + \pi(k,n+1), \qquad k \geq 1, n \geq 1. \tag{5}$$

The next theorem states that the probabilities $\pi(k,n)$ are given by an infinite sum of product forms.

Theorem 1 *The unique probability distribution which solves the balance equations (2)–(5) is given by*

$$\pi(0,0) = 1-\rho, \tag{6}$$

$$\pi(k,n) = \sum_{m=1}^{\infty} c_m s_m^k s_{m-1}^{n-1}, \quad k \geq 0,\, n \geq 1, \tag{7}$$

where

$$s_m = \frac{\rho(1-\rho^m)}{1-\rho^{m+1}}, \quad m \geq 0, \tag{8}$$

and

$$c_m = (1-\rho)^3 \frac{\rho^m}{(1-\rho^m)(1-\rho^{m+1})}, \quad m \geq 1. \tag{9}$$

Proof of Theorem 1:
Clearly, we have $\pi(0,0) = 1-\rho$. Furthermore, substitution of (2) in (3) gives

$$\rho\pi(0,1) = \pi(1,1) + \pi(0,2). \tag{10}$$

Hence, we are looking for probabilities $\pi(k,n), k \geq 0, n \geq 1$, satisfying (4), (5) and (10) and such that

$$\sum_{k=0}^{\infty}\sum_{n=1}^{\infty} \pi(k,n) = \rho. \tag{11}$$

To find these probabilities, we use a compensation method for queueing problems developed by Adan [1] (see also Adan, Wessels and Zijm [2]). The method attempts to solve the balance equations by a linear combination of product forms. This is achieved by first determining a basis of product-form solutions satisfying (5). Subsequently this basis is used to construct a linear combination that also satisfies (4). The basis of product-form solutions contains uncountably many elements. Therefore a procedure is needed to select the appropriate elements. This procedure is based on a compensation argument: after introducing the first term, countably many terms are added to compensate for the error made in the equations (4). Finally, remark that once (4) and (5) are satisfied, we automatically satisfy the last balance equation (10) due to the fact that the balance equations are dependent.

So, first we seek solutions of the form $\alpha^k\beta^{n-1}$ satisfying the equations (5) for all k and n. Substitution of this product form into (5) and division by common powers yields

$$\alpha = g(\beta) \qquad \text{with} \qquad g(\beta) := \frac{\rho}{1+\rho-\beta}. \tag{12}$$

Because later in the analysis the solution has to be normalized, the factors α and β are required to satisfy $|\alpha| < 1$ and $|\beta| < 1$. The points on the curve (12) inside this region determine a continuum of product forms satisfying equations (5).

Next, we construct a linear combination of these product forms, which also satisfies the equations (4). We start with initial term $c_1\alpha_1^k\beta_1^{n-1}$, where $\beta_1 = 0$, $\alpha_1 = g(\beta_1) = \rho/(1+\rho)$ and c_1 is some constant. If we rewrite equations (4) in the form

$$(\rho+1)\pi(0,n) - \pi(0,n+1) = \pi(n,1), \qquad n \geq 2, \tag{13}$$

and plug in the initial term $c_1\alpha_1^k\beta_1^{n-1}$ into (13), we obtain 0 on the left-hand side and $c_1\alpha_1^n$ on the right-hand side. Hence, we conclude that $c_1\alpha_1^k\beta_1^{n-1}$ does not satisfy equations (13) and to compensate for this error we have to add a second term so that we also obtain $c_1\alpha_1^n$ on the left-hand side. This means that we add an extra term $c_2\alpha_2^k\beta_2^{n-1}$ with (α_2, β_2) on the curve (12) such that for all $n \geq 2$,

$$c_2(\rho+1)\beta_2^{n-1} - c_2\beta_2^n = c_1\alpha_1^n.$$

This immediately gives

$$\beta_2 = \alpha_1, \quad \alpha_2 = g(\beta_2), \quad c_2 = c_1\frac{\alpha_1}{\rho+1-\alpha_1}.$$

Of course, by adding the term $c_2\alpha_2^k\beta_2^{n-1}$, we also obtain an extra term $c_2\alpha_2^n$ on the right-hand side of (13). Hence, we are going to add a third term $c_3\alpha_3^k\beta_3^{n-1}$ to compensate for this term $c_2\alpha_2^n$. Similarly as before, it follows immediately that

$$\beta_3 = \alpha_2, \quad \alpha_3 = g(\beta_3), \quad c_3 = c_2\frac{\alpha_2}{\rho+1-\alpha_2}.$$

Continuing in this way, we add extra terms $c_m\alpha_m^k\beta_m^{n-1}$ where α_m, β_m and c_m are defined iteratively by

$$\beta_m = \alpha_{m-1}, \quad \alpha_m = g(\beta_m), \quad c_m = c_{m-1}\frac{\alpha_{m-1}}{\rho+1-\alpha_{m-1}}.$$

If we now can prove that

(a) the error terms $c_m\alpha_m^n$ go to zero for $m \to \infty$,

(b) $\sum_{m=1}^{\infty} c_m\alpha_m^k\beta_m^{n-1}$ converges,

then we have that

$$\pi(k,n) = \sum_{m=1}^{\infty} c_m\alpha_m^k\beta_m^{n-1}$$

satisfies both equations (4) and (5). Both (a) and (b) follow straightforwardly from the fact that the iterates α_m, β_m and c_m are explicitly given by

$$\alpha_m = (\overbrace{g\circ\cdots\circ g}^{m\text{ times}})(0) = \frac{\rho(1-\rho^m)}{1-\rho^{m+1}},$$

$$\beta_m = (\overbrace{g\circ\cdots\circ g}^{m-1\text{ times}})(0) = \frac{\rho(1-\rho^{m-1})}{1-\rho^m},$$

$$c_m = c_1 \cdot \prod_{k=1}^{m-1}\frac{\alpha_k}{\rho+1-\alpha_k} = c_1\rho^{m-1}\cdot\frac{1-\rho}{1-\rho^m}\cdot\frac{1-\rho^2}{1-\rho^{m+1}}.$$

Finally, the remaining constant c_1 follows from the normalization equation (11). It turns out that

$$c_1 = \frac{\rho(1-\rho)^2}{1-\rho^2}.$$

The result of the theorem now follows.

Remark 1 From Theorem 1 we see that $P(X_1 + X_2 = 0) = 1 - \rho$ and for $j > 0$

$$\begin{aligned} P(X_1 + X_2 = j) &= \sum_{k=0}^{j-1} \pi(k, j-k) = \sum_{m=1}^{\infty} c_m \frac{s_m^j - s_{m-1}^j}{s_m - s_{m-1}} \\ &= \sum_{m=1}^{\infty} (1-\rho)(s_m^j - s_{m-1}^j) = (1-\rho)\rho^j, \end{aligned}$$

which was already mentioned in (1).

Remark 2 From Theorem 1 we see that the marginal distributions of X_1 and X_2 are given by

$$P(X_1 = k) = \begin{cases} 1 - \rho + \displaystyle\sum_{m=1}^{\infty} (1-\rho)\rho^m (1 - s_m), & \text{if } k = 0, \\ \displaystyle\sum_{m=1}^{\infty} (1-\rho)\rho^m (1 - s_m) s_m^k, & \text{if } k > 0, \end{cases}$$

and

$$P(X_2 = n) = \begin{cases} 1 - \rho, & \text{if } n = 0, \\ \displaystyle\sum_{m=1}^{\infty} (1-\rho)\rho^m (1 - s_{m-1}) s_{m-1}^{n-1}, & \text{if } n > 0. \end{cases}$$

Remark 3 An alternative proof of Theorem 1 can be given by using generating functions. If we introduce

$$Q(x, y) = \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} \pi(k, n) x^k y^{n-1},$$

we obtain from (4), (5) and (10) using straightforward calculations

$$\left[\rho(1-x) + \left(1 - \frac{1}{y}\right)\right] Q(x, y) = \left(1 - \frac{1}{y}\right) Q(0, 0) + \frac{1}{y} \left(Q(y, 0) - Q(x, 0)\right). \tag{14}$$

Next, if we define $f(x) = [1 + \rho(1-x)]^{-1}$ and substitute $y = f(x)$ in (14), the left-hand side of (14) equals zero. Using that $Q(x, y)$ is bounded for $|x| \leq 1, |y| \leq 1$, we conclude that for $y = f(x)$ also the right-hand side of (14) has to be zero, and hence that

$$Q(x, 0) = Q(f(x), 0) + (f(x) - 1) Q(0, 0).$$

Iteration of this equation yields an infinite-sum expression for $Q(x, 0)$. After substitution of this expresssion in (14), we obtain an expression for $Q(x, y)$, from which after some calculations the results of Theorem 1 can be obtained.

4 Division of customers over waiting room and service queue

In Theorem 1 in the previous section we give the steady-state probabilities $\pi(k,n)$ as an infinite sum of product forms. Although these expressions converge quite rapidly if ρ is not very close to 1, they are not very practical if one wants to compare, for instance, the values of $\pi(0,100)$, $\pi(50,50)$ and $\pi(99,1)$. In this section we use the results from the previous section to obtain more knowledge about the division of customers over the waiting room and the service queue when the total number of customers is large. We derive formulas that converge rapidly for ρ close to 1. These formulas reveal a remarkable oscillatory behaviour that is more pronounced for small ρ.

Let $P_N(k)$ be the probability that the waiting room contains k customers, given that the total number of customers in the system equals N. The probability $P_N(k)$ is given by

$$P_N(k) = \frac{\pi(k, N-k)}{(1-\rho)\rho^N}. \tag{15}$$

Since $\sum_{k=0}^{N-1} P_N(k) = 1$, $P_N(k)$ will be of the order of $1/N$. If $NP_N(k)$ with $k/N = \xi$ fixed would have a limit as $N \to \infty$, we would not even need all results from the previous section to obtain this limit. More specifically, if we assume that

$$\lim_{N\to\infty} NP_N([N\xi]) = f(\xi) \tag{16}$$

with normalization

$$\int_0^1 f(\xi)\, d\xi = 1 \tag{17}$$

then the balance equations (5) imply that for large N

$$(\rho+1)f(\xi) = \left(1-\frac{1}{N}\right)^{-1} f\left(\left(1-\frac{1}{N}\right)^{-1}\left(\xi-\frac{1}{N}\right)\right) + \rho\left(1+\frac{1}{N}\right)^{-1} f\left(\xi\left(1+\frac{1}{N}\right)^{-1}\right). \tag{18}$$

Taylor expanding and equating terms of order $1/N$ gives the differential equation

$$(1-\rho)f(\xi) - (1-\xi(1-\rho))f'(\xi) = 0 \tag{19}$$

which has the properly normalized solution

$$f(\xi) = \frac{1-\rho}{\ln(1/\rho)(1-\xi(1-\rho))}. \tag{20}$$

On the other hand, balance equations (4) require that

$$(\rho+1)f(0)=\rho\left(1+\frac{1}{N}\right)^{-1}\left[f\left(\left(1+\frac{1}{N}\right)^{-1}\right)+f(0)\right]. \tag{21}$$

Taylor expansion in $1/N$ gives to lowest order $f(0)=\rho f(1)$, which is satisfied by the function f from (20), and to first order $f(1)+f'(1)=-f(0)$, which is not satisfied by the function f from (20), as that f satisfies $f(1)+f'(1)=f(0)/\rho^2$ instead. So the function f cannot satisfy all relevant balance equations. It follows that $NP_N([N\xi])$ does not have a limit as $N\to\infty$. This is expressed in the following theorem.

Theorem 2 *The large N behaviour of $NP_N(k)$, with k/N kept fixed at ξ, is*

$$\begin{aligned}NP_N(k)\sim{}&\frac{(1-\rho)}{\ln(1/\rho)(1-\xi(1-\rho))}\\&\times\sum_{q=-\infty}^{\infty}\Gamma\left(1-\frac{2\pi iq}{\ln(1/\rho)}\right)e^{2\pi iq\ln[N(1-\rho)(1-\xi(1-\rho))]/\ln(1/\rho)}.\end{aligned} \tag{22}$$

Proof of Theorem 2:
Inserting the explicit formulas for $\pi(k,N-k)$ gives

$$\begin{aligned}P_N(k)&=\sum_{m=1}^{\infty}\frac{\rho^m(1-\rho)^2}{(1-\rho^m)(1-\rho^{m+2})}\left(\frac{(1-\rho^{m+1})^2}{(1-\rho^m)(1-\rho^{m+2})}\right)^k\left(\frac{1-\rho^m}{1-\rho^{m+1}}\right)^N\\&=\sum_{m=1}^{\infty}\frac{\rho^m(1-\rho)^2}{(1-\rho^m)(1-\rho^{m+2})}\left[\left(1+\frac{\rho^m(1-\rho)^2}{(1-\rho^m)(1-\rho^{m+2})}\right)^\xi\left(1-\frac{\rho^m(1-\rho)}{1-\rho^{m+1}}\right)\right]^N\end{aligned} \tag{23}$$

for $k=0,\ldots,N-2$, $\xi:=k/N$ (so $0\leq\xi<1$) and

$$\begin{aligned}P_N(N-1)&=\sum_{m=1}^{\infty}\rho^{m-1}\left(\frac{1-\rho}{1-\rho^m}\right)^2\left(\frac{1-\rho^m}{1-\rho^{m+1}}\right)^N\\&=(1-\rho)^2\sum_{m=1}^{\infty}\frac{\rho^{m-1}}{(1-\rho^m)^2}\left[1-\frac{\rho^m(1-\rho)}{1-\rho^{m+1}}\right]^N.\end{aligned} \tag{24}$$

The large N behaviour can be obtained using techniques employed in the appendices of Janssen and de Jong [7]. When both N and k become large, with $\xi=k/N$ fixed, only terms where the factor between square brackets in (23) and (24) is close to 1 survive, so only the large m terms for which ρ^m is small contribute to the sum. Taking logarithms, Taylor expanding in ρ^m around $\rho^m=0$, keeping only the first term, and exponentiating result in

$$P_N(k)\sim(1-\rho)^2\sum_{m=1}^{\infty}\rho^m\exp\left[-N\rho^m(1-\rho)(1-\xi(1-\rho))\right], \tag{25}$$

$$P_N(N-1)\sim(1-\rho)^2\sum_{m=1}^{\infty}\rho^{m-1}\exp\left[-N\rho^m(1-\rho)\right]. \tag{26}$$

Defining $K = N(1-\rho)(1-\xi(1-\rho))$, multiplying and dividing by K, we obtain

$$P_N(k) \sim \frac{(1-\rho)^2}{K} \sum_{m=1}^{\infty} K\rho^m e^{-K\rho^m}. \tag{27}$$

This expression is also valid (for large N) for $P_N(N-1)$ (with $\xi = 1$ and $K = N(1-\rho)\rho$), since

$$\sum_{m=1}^{\infty} \rho^{m-1} e^{-K\rho^{m-1}} = e^{-K} + \sum_{m=1}^{\infty} \rho^m e^{-K\rho^m} \tag{28}$$

so the error made decreases exponentially in K. In fact now the summation range of m can be extended to *all* integers with negligible error, since for $m \leq 0$ and $0 < \rho < 1$ it holds that $\rho^m \geq 1$, so the summand decreases very rapidly as K increases. Only the large m terms ($m \to +\infty$) contribute to the sum in the large K limit. Defining $t = -\ln(\rho) > 0$ and $s = \ln(K)$, the sum becomes

$$\sum_{m=1}^{\infty} K\rho^m e^{-K\rho^m} \sim \sum_{m=-\infty}^{\infty} K\rho^m e^{-K\rho^m} = \sum_{m=-\infty}^{\infty} e^{s-mt-e^{s-mt}}. \tag{29}$$

Viewed as a function of s, the rightmost expression is periodic with period t, so it can be written as a Fourier series:

$$\sum_{m=-\infty}^{\infty} e^{s-mt-e^{s-mt}} = \sum_{q=-\infty}^{\infty} e^{2\pi i q s/t} a_q(t), \tag{30}$$

with

$$\begin{aligned} a_q(t) &= \frac{1}{t}\int_0^t e^{-2\pi i q s/t} \sum_{m=-\infty}^{\infty} e^{s-mt-e^{s-mt}}\, ds \\ &= \frac{1}{t}\int_{-\infty}^{\infty} e^{-2\pi i q s/t} e^{s-e^s}\, ds \\ &= \frac{1}{t}\int_0^{\infty} y^{-2\pi i q/t} e^{-y}\, dy \\ &= \frac{1}{t}\Gamma\left(1 - \frac{2\pi i q}{t}\right). \end{aligned} \tag{31}$$

Substituting this expression for $a_q(t)$ into equation (30) and making use of equations (29) and (27), the theorem follows.

Because of the well-known reflection property of the Gamma function,

$$|\Gamma(1-iy)|^2 = \frac{\pi y}{\sinh(\pi y)}, \tag{32}$$

the amplitudes $|a_q(t)|$ of the Fourier coefficients decrease exponentially with exponential factor $\exp(-\pi^2|q|/t)$, so for small ρ, i.e., large t, more terms are needed to obtain accurate numerical answers. For ρ close to 1, i.e., small t, only a few terms are needed. The amplitudes of the oscillatory terms in the right-hand side of equation (22) depend on $|q|$, ρ and ξ, they are independent of N. The phases depend on all these variables.

5 Sojourn time distribution

Let S_1 and S_2 be the stationary sojourn time of a customer in the waiting room and the service queue, respectively. We are interested in the bivariate Laplace-Stieltjes transform of these stationary sojourn times.

Theorem 3 *The bivariate Laplace–Stieltjes transform* $\phi(u,v) = E(e^{-uS_1 - vS_2})$ *is given by*

$$\phi(u,v) = (1-\rho)\frac{\mu}{\mu+v} - \frac{\mu^2}{uv}\sum_{m=1}^{\infty}\frac{c_m}{s_m}\ln\left(1 - s_m \cdot \frac{u}{u+\mu(1-s_{m-1})}\cdot\frac{v}{v+\mu(1-s_m)}\right). \tag{33}$$

Proof of Theorem 3:
According to the PASTA property, an arriving customer sees with probability $\pi(k,n)$ the system in state (k,n). If $(k,n) = (0,0)$ the customer is immediately taken into service. If $(k,n) \neq (0,0)$, the customer first has to wait in the waiting room while the n customers in the service queue are processed. If during this waiting time r additional customers arrive, our customer will then be in the ℓ^{th} ($\ell = 1,\ldots,k+1+r$) position in the service queue, with probability $(k+1+r)^{-1}$. This gives

$$\phi(u,v) = \pi(0,0)\frac{\mu}{\mu+v} + \sum_{k=0}^{\infty}\sum_{n=1}^{\infty}\pi(k,n)\cdot \int_0^{\infty}\frac{(\mu x)^{n-1}}{(n-1)!}\mu e^{-\mu x}e^{-ux}\sum_{r=0}^{\infty}\frac{(\lambda x)^r}{r!}e^{-\lambda x}\frac{1}{k+1+r}\sum_{\ell=1}^{k+1+r}\left(\frac{\mu}{\mu+v}\right)^{\ell}dx.$$

Now, using the formulas

$$\sum_{\ell=1}^{k+1+r}\left(\frac{\mu}{\mu+v}\right)^{\ell} = \frac{\mu}{v}\cdot\left[1-\left(\frac{\mu}{\mu+v}\right)^{k+r+1}\right], \tag{34}$$

$$\sum_{n=1}^{\infty}\pi(k,n)\frac{(\mu x)^{n-1}}{(n-1)!} = \sum_{m=1}^{\infty}c_m s_m^k e^{\mu x s_{m-1}}, \tag{35}$$

and

$$\int_0^{\infty} x^r e^{-\gamma x}dx = r!\gamma^{-r-1}, \tag{36}$$

we obtain after straightforward calculations

$$\phi(u,v) = (1-\rho)\frac{\mu}{\mu+v} + \frac{\mu^2}{v}\sum_{m=1}^{\infty}\frac{c_m}{\gamma_m}\sum_{k=0}^{\infty}\sum_{r=0}^{\infty}\frac{1}{k+r+1}s_m^k\left(\frac{\lambda}{\gamma_m}\right)^r\left[1-\left(\frac{\mu}{\mu+v}\right)^{k+r+1}\right], \tag{37}$$

where γ_m is defined as

$$\gamma_m = \lambda + \mu(1 - s_{m-1}) + u. \tag{38}$$

Finally, the sum over k and r in (37) can be evaluated using

$$\sum_{k=0}^{\infty} \sum_{r=0}^{\infty} \frac{1}{k+r+1} x^k y^r = \frac{\ln(1-x) - \ln(1-y)}{y-x}. \tag{39}$$

This results in

$$\begin{aligned}
\phi(u,v) &= (1-\rho)\frac{\mu}{\mu+v} + \frac{\mu^2}{v} \sum_{m=1}^{\infty} \frac{c_m}{\lambda - \gamma_m s_m} \ln\left(\frac{(1-s_m)(1 - \frac{\lambda}{\gamma_m} \cdot \frac{\mu}{\mu+v})}{(1 - \frac{\lambda}{\gamma_m})(1 - s_m \cdot \frac{\mu}{\mu+v})} \right) \\
&= (1-\rho)\frac{\mu}{\mu+v} - \frac{\mu^2}{uv} \sum_{m=1}^{\infty} \frac{c_m}{s_m} \ln\left(1 - s_m \cdot \frac{u}{u + \mu(1-s_{m-1})} \cdot \frac{v}{v + \mu(1-s_m)} \right)
\end{aligned}$$

and hence the theorem follows.

From (33), after substitution of $v = 0$ and $u = 0$, respectively, and using

$$-\ln(1-x) = \sum_{k=1}^{\infty} \frac{x^k}{k}, \tag{40}$$

we obtain

$$\begin{aligned}
E(e^{-uS_1}) &= (1-\rho) + \sum_{m=1}^{\infty} \frac{c_m}{(1-s_m)(1-s_{m-1})} \cdot \frac{\mu(1-s_{m-1})}{u + \mu(1-s_{m-1})} \\
&= (1-\rho) + \sum_{m=1}^{\infty} (1-\rho)\rho^m \cdot \frac{\mu(1-s_{m-1})}{u + \mu(1-s_{m-1})} \\
&= (1-\rho) + \rho \sum_{m=0}^{\infty} (1-\rho)\rho^m \cdot \frac{\mu(1-s_m)}{u + \mu(1-s_m)}
\end{aligned} \tag{41}$$

and

$$\begin{aligned}
E(e^{-vS_2}) &= (1-\rho)\frac{\mu}{\mu+v} + \sum_{m=1}^{\infty} \frac{c_m}{(1-s_m)(1-s_{m-1})} \cdot \frac{\mu(1-s_m)}{v + \mu(1-s_m)} \\
&= (1-\rho)\frac{\mu}{\mu+v} + \sum_{m=1}^{\infty} (1-\rho)\rho^m \cdot \frac{\mu(1-s_m)}{v + \mu(1-s_m)} \\
&= \sum_{m=0}^{\infty} (1-\rho)\rho^m \cdot \frac{\mu(1-s_m)}{v + \mu(1-s_m)}.
\end{aligned} \tag{42}$$

Remark 4 Both the sojourn time in the waiting room, given that this sojourn time is greater than zero, and the sojourn time in the service queue are equal to the same geometric mixture of exponentials.

Let us now concentrate on finding the first two moments of S_1, S_2 and $S = S_1 + S_2$. The moments of S_2 follow from (42),

$$E(S_2) = \sum_{m=0}^{\infty}(1-\rho)\rho^m \frac{1}{\mu(1-s_m)} = \frac{1}{\mu}\sum_{m=0}^{\infty}\rho^m(1-\rho^{m+1}) = \frac{1}{\mu}\cdot\frac{1}{1-\rho^2}, \tag{43}$$

$$\begin{aligned} E(S_2^2) &= \sum_{m=0}^{\infty}(1-\rho)\rho^m \frac{2}{\mu^2(1-s_m)^2} = \frac{2}{\mu^2(1-\rho)}\sum_{m=0}^{\infty}\rho^m(1-\rho^{m+1})^2 \\ &= \frac{1}{\mu^2}\cdot\frac{2(1+\rho^2)}{(1-\rho^2)(1-\rho^3)}. \end{aligned} \tag{44}$$

Furthermore, the moments of S_1 follow directly from Remark 4,

$$E(S_1) = \rho E(S_2) = \frac{1}{\mu}\cdot\frac{\rho}{1-\rho^2}, \tag{45}$$

$$E(S_1^2) = \rho E(S_2^2) = \frac{1}{\mu^2}\cdot\frac{2\rho(1+\rho^2)}{(1-\rho^2)(1-\rho^3)}. \tag{46}$$

Adding (43) and (45), we obtain

$$E(S) = \frac{1}{\mu}\left(\frac{\rho}{1-\rho^2} + \frac{1}{1-\rho^2}\right) = \frac{1}{\mu}\cdot\frac{1}{1-\rho}, \tag{47}$$

which is of course equal to the mean sojourn time in the standard $M/M/1$ queue with FCFS service. In order to find $E(S^2)$, we first calculate $E(S_1 S_2)$ by using (33) and (40). We obtain

$$\begin{aligned} E(S_1 S_2) &= \frac{\partial^2}{\partial u\,\partial v}\phi(u,v)|_{u=v=0} \\ &= \frac{1}{\mu^2}\sum_{m=1}^{\infty}\frac{c_m(1+\frac{1}{2}s_m)}{(1-s_{m-1})^2(1-s_m)^2} \\ &= \frac{1}{\mu^2}\frac{\rho(2+3\rho+\rho^3)}{2(1-\rho^2)(1-\rho^3)}, \end{aligned} \tag{48}$$

where the second equality follows from a Taylor series expansion of $\phi(u,v)$. Combination of (44), (46) and (48) now gives

$$E(S^2) = \frac{1}{\mu^2}\frac{2+4\rho+5\rho^2+2\rho^3+\rho^4}{(1-\rho^2)(1-\rho^3)}, \tag{49}$$

and so the variance of the total sojourn time is given by

$$\sigma^2(S) = \frac{1}{\mu^2}\cdot\frac{1}{(1-\rho)^2}\left(1+\frac{\rho^2(1+\rho^2)}{(1+\rho)(1+\rho+\rho^2)}\right). \tag{50}$$

In comparison, the variance of the sojourn time in the model with FCFS service discipline and in the model with random order of service are given by (see, e.g., Cohen [6])

$$\sigma^2_{FCFS}(S) = \frac{1}{\mu^2} \cdot \frac{1}{(1-\rho)^2}, \tag{51}$$

$$\sigma^2_{ROS}(S) = \frac{1}{\mu^2} \cdot \frac{1}{(1-\rho)^2} \left(1 + \frac{2\rho^2}{2-\rho}\right). \tag{52}$$

From (50), (51) and (52) we see that the variance in the model with gated random order of service discipline is larger than in the model with FCFS service discipline but smaller than in the model with random order of service.

Acknowledgements

We thank Onno Boxma and Dee Denteneer for stimulating discussions.

References

[1] Adan, I.J.B.F. (1991) A compensation approach for queueing problems, Ph.D. thesis, Eindhoven University of Technology.

[2] Adan, I.J.B.F., Wessels, J, and Zijm, W.H.M. (1993). A compensation approach for two-dimensional Markov processes. *Adv. Appl. Probab.* **25**, 783-817.

[3] Ali, O.M.E. and Neuts, M.F. (1984) A service system with two stages of waiting and feedback of customers. *J. Appl. Probab.* **21**, 404-413.

[4] Boxma, O.J. and Cohen, J.W. (1991) The $M/G/1$ queue with permanent customers. *IEEE J. Sel. Areas Commun.* **9**, 179-184.

[5] Boxma, O.J., Denteneer, D. and Resing, J.A.C. (2002) Some models for contention resolution in cable networks. In: Proceedings Networking 2002, Pisa, Italy.

[6] Cohen, J.W. (1982) *The Single Server Queue*, North-Holland, Amsterdam.

[7] Janssen, A.J.E.M. and de Jong, M.J.M. (2000) Analysis of contention tree algorithms. *IEEE Trans. Inform. Theory* **46**, 2163-2172.

[8] Mathys, P. and Flajolet, Ph. (1985). Q-ary collision resolution algorithms in random-access systems with free or blocked channel access. *IEEE Trans. Inform. Theory* **31**, 217-243.