On the distribution of the number of customers in the symmetric $M/G/1$ queue

Denis Denisov* Artëm Sapozhnikov†

31 March 2006

We consider an $M/G/1$ queue with symmetric service discipline. The class of symmetric service disciplines contains, in particular, the preemptive last-come-first-served discipline and the processor-sharing discipline. It has been conjectured in Kella, Zwart and Boxma [1] that the marginal distribution of the queue length at any time is identical for all symmetric disciplines if the queue starts empty. In this paper we show that this conjecture is true if service requirements have an Erlang distribution. We also show by a counterexample, involving the hyperexponential distribution, that the conjecture is generally not true.

Keywords: SYMMETRIC QUEUE, TIME-DEPENDENT ANALYSIS, INSENSITIVITY, PROCESSOR-SHARING QUEUE, LAST COME FIRST SERVED QUEUE

1 Introduction

In this paper we consider the M/G/1 queue with the symmetric service discipline which is defined as follows. Customers arrive according to a Poisson process with rate λ and have independent and identically distributed service times $\{B_i\}_{i\geq 1}$. Let $p_i^{(n)}$ be a sequence of positive numbers such that for each n, $p_1^{(n)} + p_2^{(n)} + \cdots + p_n^{(n)} = 1$. If there are n customers in the queue then the customer in position i gets a fraction $p_i^{(n)}$ of the service rate. If a new customer arrives at the queue with n customers he moves into position i with probability $p_i^{(n+1)}$; customers in positions $i, i+1, \ldots, n$ move to positions $i+1, i+2, \ldots, n+1$.

The symmetric queueing discipline has been introduced by Kelly [2]. It contains both two important disciplines: the preemptive Last Come First Served (LCFS) discipline and Processor Sharing (PS) discipline. It is proved in Section 3.3 of [2] that for the symmetric M/G/1 queue the distribution of the queue length in *steady state* is geometric with probability of success $1-\rho$, where ρ is the traffic intensity. In particular, it is insensitive to the service discipline and depends only on the mean of the service times .

Recently, [1] has studied *time-dependent*, rather than steady-state, properties of the queue length process $\{Q_t, t \geq 0\}$ of the symmetric M/G/1 queue. In particular, it has been shown that if $Q_0 = 0$, then at any moment of time the M/G/1 LCFS queue and PS queue coincide in distribution, i.e. $Q_t^{PS} =_D Q_t^{LCFS}$, for any fixed $t \geq 0$. Also, it has been shown that if $\tau(q)$ is an independent, exponentially distributed random variable, then $Q_{\tau(q)}^{LCFS}$ has a geometric distribution. It has been conjectured in [1] that Q_t has the same distribution for any M/G/1

*EURANDOM, Eindhoven, The Netherlands; denisov@eurandom.tue.nl

†Department of Mathematics / Boole Centre for Research in Informatics, University College Cork, Cork, Ireland; as2@proba.ucc.ie

symmetric queue. In this paper we show that this conjecture is true if the service requirements have an Erlang distribution (see Theorem 2.2 below). However, in general this conjecture does not hold and we show this by a simple counterexample (see Section 4 below).

Let $\beta(s) = \mathbf{E}e^{-sB_1}$ be the Laplace-Stieltjes transform (LST) of the service distribution B_1 and define the net input process $Y(t) = \sum_{i=1}^{N(t)} B_i - t$. This is a Lévy process with exponent $\varphi(s) = s - \lambda(1 - \beta(s))$, that is $\mathbf{E}e^{-sX(t)} = e^{t\varphi(s)}$. Let $s^* = \inf\{s : \varphi(s) > 0\}$. Since $\varphi(s)$ is continuous and strictly increasing on $[s^*, \infty)$, it has an inverse, which we denote by $\kappa(q), q \geq 0$. In [1, 3] the following result is proved for LCFS and PS queues.

Proposition 1.1. *Let $\tau(q)$ be an independent exponentially distributed random variable with rate $q > 0$. If $Q_0 = 0$, then*

$$\mathbf{P}(Q_{\tau(q)} = n) = \left(1 - \frac{q}{\kappa(q)}\right)^n \frac{q}{\kappa(q)}.$$

The paper is organized as follows. In Section 2 we give the result in case the service requirements have an Erlang distribution (Theorem 2.1). We prove Theorem 2.1 in Section 3. We give a counterexample in Section 4.

2 Queue with Erlang distributed service requirements

In this section we study symmetric queues for which customers arrive according to a Poisson process with intensity parameter λ and their service requirements B_n have Erlang distribution with parameters N and μ i.e., $B_n = B_{n,1} + \ldots + B_{n,N}$ for independent $B_{n,j}$ exponentially distributed with parameter μ. We prove the following theorem.

Theorem 2.1. *Let $Q_0 = 0$. Then, for any $t \geq 0$, the distribution of Q_t does not depend on $\{p_i^{(n)},\ 1 \leq i \leq n,\ n \geq 1\}$.*

In particular, Proposition 1.1 holds for symmetric Erlang queues.

We consider a Markov process X_t on a state space $\mathcal{X} \cup \{0\}$, where

$$\mathcal{X} = \{(x_1, \ldots, x_l),\ \ l \geq 1,\ x_i \in \{1, \ldots, N\}\}. \tag{2.1}$$

In the definition above $(x_1, \ldots, x_l)$ corresponds to a queue with l customers in which the i-th customer is in the x_i-th service stage. For any vector $(x_1, \ldots, x_l)$, we denote its length as $|(x_1, \ldots, x_l)|$. Note that $Q_t = |X_t|$ is the queue length at time t.

We note that X_t is a Markov jump process. The time it spends in state 0 before it jumps to a different state has an exponential distribution with parameter λ. The time it spends in any other state before it jumps to a different state has an exponential distribution with parameter $\lambda + \mu$. We want to prove independence of the distribution of the queue length Q_t of a symmetric queue from a service discipline (i.e. independence with respect to $\{p_i^{(m)},\ 1 \leq i \leq m,\ m \geq 1\}$). It is well known that for work conserving queues $\mathbf{P}(Q_t = 0)$ does not depend on the service discipline. Therefore we can omit the time X_t spends at 0 by adding one customer in the queue at each time it becomes empty. It means that we consider a modified Markov process which jumps from N to 1 with the same probability as X_t jumps from the state (N) to (0). From now on we are going to work only with the modified process. Therefore we also denote it X_t. The new process is defined on $\mathcal{X}$. The time it spends in any state before it jumps to a different state has an exponential distribution with parameter λ.

Let $\{\xi_i\}$ be a sequence of independent $\text{Exp}(\lambda + \mu)$ random variables. It corresponds to the times between subsequent jumps of X_t. Let $N(t) = \max\{i\ :\ \sum_{j=1}^{i} \xi_j \leq t\}$ be the number of jumps on $(0,t]$. Then

$$\mathbf{P}(Q_t = i) = \sum_{n=0}^{\infty} \mathbf{P}(Q_t = i \mid N(t) = n)\mathbf{P}(N(t) = n) = \sum_{n=0}^{\infty} \mathbf{P}(|Y_{n+1}| = i)\mathbf{P}(N(t) = n),$$

where Y_n is an embedded Markov chain corresponding to X_t.

It is sufficient to prove that, for any $n \geq 1$ and $i \geq 1$, $\mathbf{P}(|Y_n| = i)$ does not depend on the service discipline. We prove a more general result. We introduce subsets of $\mathcal{X}$. For $k \geq 1$, let

$$\mathcal{U}_k = \left\{ (x_1, \ldots, x_l) \in \mathcal{X} \ : \ \sum_{i=1}^{l} x_i = k \right\}. \tag{2.2}$$

Remark 1.

$$|\mathcal{U}_k| = |\mathcal{U}_{k-1}| + \ldots + |\mathcal{U}_{(k-N)^+}|,$$

in particular, $|\mathcal{U}_k| = 2^{k-1}$ for $k \leq N$.

We prove the following theorem.

Theorem 2.2. *Let Y_n be the Markov chain defined above. For $k \geq 1$ and $n \geq 1$, let*

$$P(k,n) = \mathbf{P}(Y_n \in \mathcal{U}_k). \tag{2.3}$$

Then

1. $P(k,n)$ do not depend on $\{p_i^{(m)},\ 1 \leq i \leq m,\ m \geq 1\}$, and, moreover, for any $(x_1, \ldots, x_l) \in \mathcal{U}_k$,

$$\mathbf{P}(Y_n = (x_1, \ldots, x_l)) = \left(\frac{\lambda}{\lambda + \mu}\right)^{l-1} \left(\frac{\mu}{\lambda + \mu}\right)^{k-l} P(k,n). \tag{2.4}$$

2. $P(k,n)$ satisfies the following recursion:

$$P(k,n) = P(k-1, n-1) + \frac{\lambda}{\lambda + \mu} \left(\frac{\mu}{\lambda + \mu}\right)^{N} P(k+N, n-1). \tag{2.5}$$

Remark 2. The recursion (2.5) simply means that the Markov chain jumps to $\mathcal{U}_k$ from $\mathcal{U}_{k-1}$ or from a subset of $\mathcal{U}_{k+N}$ which consists of vectors such that at least one of the components is N.

3 Proof of Theorem 2.2

It is clear that $P(k,n) = 0$ for $k > n$. We prove the result by induction.

The result holds for $n = 1$. Indeed, $P(k,1) = \delta_k(1)$.

We assume that $P(k,n)$ does not depend on $\{p_i^{(m)},\ 1 \leq i \leq m,\ m \geq 1\}$ for any k, and, for any $(x_1, \ldots, x_l) \in \mathcal{U}_k$, (2.4) holds. We show that the result holds for Y_{n+1}.

We fix any state $(x_1, \ldots, x_l) \in \mathcal{U}_k$.

$$\begin{aligned} \mathbf{P}\left(Y_{n+1} = (x_1, \ldots, x_l)\right) &= \sum_{(y_1, \ldots, y_m) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_m) \mapsto (x_1, \ldots, x_l)\right) \mathbf{P}\left(Y_n = (y_1, \ldots, y_m)\right) \\ &+ \sum_{(y_1, \ldots, y_m) \in \mathcal{U}_{k+N}} \mathbf{P}\left((y_1, \ldots, y_m) \mapsto (x_1, \ldots, x_l)\right) \mathbf{P}\left(Y_n = (y_1, \ldots, y_m)\right), \end{aligned}$$

where $a \mapsto b$ stands for a transition from a to b in one step. We write the last two summands as

$$\mathbf{P}\left(Y_{n+1} = (x_1, \ldots, x_l)\right) = \Sigma_1 + \Sigma_2. \tag{3.1}$$

We evaluate Σ_1 and Σ_2 separately.

$$\begin{aligned}
\Sigma_1 &= \sum_{(y_1,\ldots,y_{l-1}) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_{l-1}) \mapsto (x_1, \ldots, x_l)\right) \mathbf{P}\left(Y_n = (y_1, \ldots, y_{l-1})\right) \\
&+ \sum_{(y_1,\ldots,y_l) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_l) \mapsto (x_1, \ldots, x_l)\right) \mathbf{P}\left(Y_n = (y_1, \ldots, y_l)\right) \\
&= \left(\frac{\lambda}{\lambda+\mu}\right)^{l-2} \left(\frac{\mu}{\lambda+\mu}\right)^{k-l+1} P(k-1,n) \sum_{(y_1,\ldots,y_{l-1}) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_{l-1}) \mapsto (x_1, \ldots, x_l)\right) \\
&+ \left(\frac{\lambda}{\lambda+\mu}\right)^{l-1} \left(\frac{\mu}{\lambda+\mu}\right)^{k-l} P(k-1,n) \sum_{(y_1,\ldots,y_l) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_l) \mapsto (x_1, \ldots, x_l)\right).
\end{aligned}$$

Since

$$\sum_{(y_1,\ldots,y_{l-1}) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_{l-1}) \mapsto (x_1, \ldots, x_l)\right) = \frac{\lambda}{\lambda+\mu} \sum_{i\,:\,x_i=1} p_i^{(l)} \tag{3.2}$$

and

$$\sum_{(y_1,\ldots,y_l) \in \mathcal{U}_{k-1}} \mathbf{P}\left((y_1, \ldots, y_l) \mapsto (x_1, \ldots, x_l)\right) = \frac{\mu}{\lambda+\mu} \sum_{i\,:\,x_i \neq 1} p_i^{(l)}, \tag{3.3}$$

we obtain

$$\Sigma_1 = \left(\frac{\lambda}{\lambda+\mu}\right)^{l-1} \left(\frac{\mu}{\lambda+\mu}\right)^{k-l} P(k-1,n). \tag{3.4}$$

Similarly, we compute

$$\begin{aligned}
\Sigma_2 &= \sum_{(y_1,\ldots,y_{l+1}) \in \mathcal{U}_{k+N}} \mathbf{P}\left((y_1, \ldots, y_{l+1}) \mapsto (x_1, \ldots, x_l)\right) \mathbf{P}\left(Y_n = (y_1, \ldots, y_{l+1})\right) \\
&= \left(\frac{\lambda}{\lambda+\mu}\right)^{(l+1)-1} \left(\frac{\mu}{\lambda+\mu}\right)^{k+N-(l+1)} P(k+N,n) \\
&\cdot \sum_{(y_1,\ldots,y_{l+1}) \in \mathcal{U}_{k+N}} \mathbf{P}\left((y_1, \ldots, y_{l+1}) \mapsto (x_1, \ldots, x_l)\right).
\end{aligned}$$

Since

$$\sum_{(y_1,\ldots,y_{l+1}) \in \mathcal{U}_{k+N}} \mathbf{P}\left((y_1, \ldots, y_{l+1}) \mapsto (x_1, \ldots, x_l)\right) = \sum_{i=1}^{l+1} \frac{p_i^{(l+1)}\mu}{\lambda+\mu} = \frac{\mu}{\lambda+\mu},$$

we obtain

$$\Sigma_2 = \left(\frac{\lambda}{\lambda+\mu}\right)^{l-1} \left(\frac{\mu}{\lambda+\mu}\right)^{k-l} \left\{ \frac{\lambda}{\lambda+\mu} \left(\frac{\mu}{\lambda+\mu}\right)^{N} P(k+N,n) \right\}. \tag{3.5}$$

The result now follows from (3.1), (3.4) and (3.5).

Q.E.D.

Theorem 2.1 follows from Theorem 2.2 and reduction to the analysis of the embedded Markov chain which resulted from the uniformization procedure described in Section 2.

Remark 3. The property of the symmetric queue was essentially used in (3.2)–(3.4).

4 Counterexample

Once Theorem 2.1 is proved for the symmetric queues with Erlang distributed service requirements, it is natural to ask if it still holds when service requirements have the phase type distribution. If it were true, a classical approximation procedure (see e.g. [2, Lemma 3.9]) would give a result for the symmetric queues with general service requirements. Unfortunately the answer is no. In this section we give an example of a symmetric queue for which Theorem 2.1 does not hold. Let, as before, customers arrive in the queue according to a Poisson process with intensity parameter λ, and the service requirements are independent and identically distributed with the density function

$$\frac{1}{2}\mu_1 e^{-\mu_1 x} + \frac{1}{2}\mu_2 e^{-\mu_2 x}, \qquad x \geq 0.$$

Then, the LST of service time B_1 is equal to $\beta(s) = \frac{1}{2}(\frac{\mu_1}{\mu_1+s} + \frac{\mu_2}{\mu_2+s})$. This system could be considered as a model with customers of two types: customers of both types arrive according to independent Poisson processes with intensity parameter $\lambda/2$ and their service requirements are independent and exponentially distributed with parameters μ_1 and μ_2 respectively. We consider a symmetric queue with the following service discipline:

$$p_1^{(1)} = 1, \;\; p_1^{(2)} = p, p_2^{(2)} = q = 1-p, \;\; p_i^{(n)} = \delta_{n,i}, \text{ for } 1 \leq i \leq n, \;\; n \geq 3.$$

Note that the case of $q = 1$ corresponds to the LCFS discipline.

Let $\tau(\alpha)$ be an independent random variable exponentially distributed with parameter $\alpha > 0$. We show that for the symmetric queue introduced above $\overline{P}_q \overset{\text{def}}{=} \mathbf{P}(Q_{\tau(\alpha)} \geq 2)$ does depend on q. It is sufficient to show that $\overline{P}_q$ is different for $q = 1$ and $q = 1/2$. For $q = 1$, it is known [1] that $\overline{P}_1 = \left(1 - \frac{\alpha}{\kappa(\alpha)}\right)^2$, where $\kappa(\alpha)$ is the inverse function for $\varphi(s) = s - \lambda(1 - \beta(s))$.

Let

$$\gamma = \frac{\mu_1}{\mu_1 + \mu_2},$$

$$\pi_1 = \pi_1(\alpha) = \frac{\mu_1}{\mu_1 + \kappa(\alpha)}, \quad \pi_2 = \pi_2(\alpha) = \frac{\mu_2}{\mu_2 + \kappa(\alpha)}, \quad \pi_{1,2} = \pi_{1,2}(\alpha) = \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 + 2\kappa(\alpha)}.$$

We show that, for $q = 1/2$,

$$\overline{P} := \overline{P}_{1/2} = \left(1 - \frac{\alpha}{\kappa(\alpha)}\right)\left\{\frac{\lambda}{\lambda+\alpha} - \frac{\alpha}{\lambda+\alpha}\frac{R}{\frac{\lambda}{\lambda+\alpha}(\pi_1+\pi_2) - 4 - R}\right\}, \tag{4.1}$$

where

$$R = \frac{\lambda}{\lambda+\alpha}\left(\frac{\lambda}{\lambda+\alpha}\left(\pi_1\pi_2 + \pi_{1,2}(\gamma\pi_1 + (1-\gamma)\pi_2)\right) - (\pi_1 + \pi_2 + 2\pi_{1,2})\right).$$

It can be shown analytically that the above expressions for $\overline{P}_1$ and $\overline{P}_{1/2}$ are different for different values of q. But it is much easier to verify it numerically. For $\lambda = 1$, $\mu_1 = 1$, $\mu_2 = 10$ and $\alpha = 1$, we have $\kappa(\alpha) = 1.346215241$ and

$$\overline{P}_1 = 0.06613987328 \neq \overline{P}_{1/2} = 0.05720076818.$$

Now we prove (4.1). We denote the two types of customers as a and b. Then (a) stays for the queue with a single customer of type a, (b) stays for the queue with a single customer of type b. A use of the total probability formula and memoryless property of the exponential distribution gives

$$\overline{P} = \mathbf{P}(Q_{\tau(\alpha)} \geq 2 \mid \tau(\alpha) \leq \tau_{\text{bp}})\mathbf{P}(Q_{\tau(\alpha)} \neq 0) = \mathbf{P}(Q_{\tau(\alpha)} \geq 2 \mid \tau(\alpha) \leq \tau_{\text{bp}})\left(1 - \frac{\alpha}{\kappa(\alpha)}\right),$$

where τ_{bp} is the first busy period. We denote $\widetilde{P} = \mathbf{P}(Q_{\tau(\alpha)} \geq 2 \mid \tau(\alpha) \leq \tau_{\text{bp}})$. Therefore, as in Section 2, it is sufficient to consider a queue for which the state 0 is deleted, and which jumps with intensity λ from the state (a) to (a,a) or (a,b) with probabilities $1/2$, and from the state (b) to (a,b) or (b,b) with probabilities $1/2$. Note that the states (a,b) and (b,a) are indistinguishable, since $p = q = 1/2$.

Let T_n be the n-th return to $\{(a),(b)\}$, $T_0 = 0$. The time the queue spends in the set $\{(a),(b)\}$ from the time T_n is exponentially distributed with parameter λ. We denote it ξ_n. Hence

$$\widetilde{P} = \sum_{n=1}^{\infty} \mathbf{P}(T_{n-1} + \xi_n < \tau(\alpha) < T_n) = \sum_{n=1}^{\infty} \left(\mathbf{E}e^{-\alpha T_{n-1}} \frac{\lambda}{\lambda+\alpha} - \mathbf{E}e^{-\alpha T_n} \right) \tag{4.2}$$

$$= \frac{\lambda}{\lambda+\alpha} - \frac{\alpha}{\lambda+\alpha} \sum_{n=1}^{\infty} \mathbf{E}e^{-\alpha T_n}. \tag{4.3}$$

Note that

$$\mathbf{E}e^{-\alpha T_n} = \left(\frac{\lambda}{\lambda+\alpha}\right)^n \mathbf{E}e^{-\alpha \widetilde{T}_n},$$

where $\widetilde{T}_n$ is the total time the queue spends outside the set $\{(a),(b)\}$ up to the time T_n. Conditioned that the queue starts from the state (a) or (b) we denote $\widetilde{T}_n$ as $\widetilde{T}_n(a)$ or $\widetilde{T}_n(b)$ respectively. A lengthy but straightforward computation gives a recursion for the Laplace-Stieltjes transforms of $\widetilde{T}_n(a)$ and $\widetilde{T}_n(b)$:

$$\mathbf{E}e^{-\alpha \widetilde{T}_n(a)} = \frac{1}{2}\left(\pi_1 + (1-\gamma)\pi_{1,2}\right) \mathbf{E}e^{-\alpha \widetilde{T}_{n-1}(a)} + \frac{1}{2}\gamma\pi_{1,2}\mathbf{E}e^{-\alpha \widetilde{T}_{n-1}(b)}, \tag{4.4}$$

and

$$\mathbf{E}e^{-\alpha \widetilde{T}_n(b)} = \frac{1}{2}(1-\gamma)\pi_{1,2}\mathbf{E}e^{-\alpha \widetilde{T}_{n-1}(a)} + \frac{1}{2}\left(\pi_2 + \gamma\pi_{1,2}\right) \mathbf{E}e^{-\alpha \widetilde{T}_{n-1}(b)}. \tag{4.5}$$

Let

$$S(a) = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\lambda+\alpha}\right)^n \mathbf{E}e^{-\alpha \widetilde{T}_n(a)}, \quad S(b) = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\lambda+\alpha}\right)^n \mathbf{E}e^{-\alpha \widetilde{T}_n(b)}. \tag{4.6}$$

Then

$$S = \sum_{n=1}^{\infty} \mathbf{E}e^{-\alpha \widetilde{T}_n} = \frac{1}{2}\left(S(a) + S(b)\right). \tag{4.7}$$

From (4.4) and (4.5) we obtain a system of equation for $S(a)$ and $S(b)$

$$2S(a) = \frac{\lambda}{\lambda+\alpha}\left(\pi_1 + \pi_{1,2}\right) + \frac{\lambda}{\lambda+\alpha}\left\{\left(\pi_1 + (1-\gamma)\pi_{1,2}\right) S(a) + \gamma\pi_{1,2}S(b)\right\}, \tag{4.8}$$

$$2S(b) = \frac{\lambda}{\lambda+\alpha}\left(\pi_2 + \pi_{1,2}\right) + \frac{\lambda}{\lambda+\alpha}\left\{(1-\gamma)\pi_{1,2}S(a) + \left(\pi_2 + \gamma\pi_{1,2}\right) S(b)\right\}, \tag{4.9}$$

which results in (4.1).

Acknowledgment. We would like to thank Onno Boxma, Neil O'Connell and Bert Zwart for drawing attention to this problem. We are also grateful to Onno Boxma and Bert Zwart for many useful comments and suggestions.

This work was done while Artëm Sapozhnikov was visiting EURANDOM. He would like to thank EURANDOM for the hospitality. The research of Denis Denisov was supported by the Dutch BSIK project (*BRICKS*) and the EURO-NGI project. The research of Artëm Sapozhnikov was supported by Science Foundation Ireland Grant No. SFI04/RP1/I512.

References

[1] Kella, O., Zwart, B. and Boxma, O. (2005) Some time-dependent properties of symmetric $M/G/1$ queues. *J.Appl.Prob.* **42**, 223–234.

[2] Kelly, F.P. (1979). *Reversibility and Stochastic Networks.* John Wiley, Chichester.

[3] Kitaev, M.Yu. (1993). The $M/G/1$ processor-sharing model: transient behavior. *Queueing Systems* **14**, 239–273.