AN ANALYSIS OF A MODIFIED M/G/1 QUEUE USING A MARTINGALE TECHNIQUE

Matthew Roughan

Journal of Applied Probability, vol.33, March 1996

Abstract: We consider a variation of the M/G/1 queue in which, when the system contains more than k customers, it switches from its initial general service distribution to a different general service distribution until the server is cleared, whereupon it switches back to the original service distribution. Using a technique of Baccelli and Makowski (1985,1989) we define a martingale with respect to an embedded process and from this arrive at a relationship between the process and a modified Markov renewal process. Using this an analysis of the stationary behaviour of the queue is possible.

Keywords: M/G/1 queue, threshold, martingales, Doob's optional sampling theorem, queueing theory.

1 Introduction

We shall consider a single-server queueing process in which the arrivals are Poisson with rate λ, and the service times are distributed according to one or the other of two general probability distribution functions $A(\cdot)$ and $B(\cdot)$. We define the number of customers in the system at time t to be $X(t)$.

If we consider the queue to start with $X(0) = 0$, initially the service times are determined by $A(\cdot)$. This continues until a stopping time $\tau_1(0)$. We call this period phase 1. At time $\tau_1(0)$, the service-time distribution switches to $B(\cdot)$. The service-times follow this distribution until a second stopping time $\tau_2(0)$ when there is a return to services governed by $A(\cdot)$. We call this second period phase 2.

For each time t the stopping time $\tau_2(t)$ occurs at the first time $\tau > t$, when $X(\tau) = 0$, immediately after a departure from the server. The other stopping time $\tau_1(t)$ can be any stopping time τ, which occurs immediately after a departure, such that $t \leq \tau \leq \tau_2(t)$ almost surely.

Thus the two types of service alternate. The motivating situation is when the times $\tau_1(t)$ correspond to the first time after t when the system has more than a certain number of customers in it at the end of a service, or the system becomes empty. This could correspond to a system where the server serves at a given rate until the system becomes overfull, at which point it changes its service strategy until it has cleared the queue, whereupon it changes back to its old strategy. For example, under sufficient load the server can request extra, possibly expensive, resources in order to complete its task more quickly. It releases these extra resources when the system becomes empty.

As is the case with the usual M/G/1 queue we consider the embedded, discrete-time process formed when one observes the number of customers in the system after departures. In this case this embedded process does not form a Markov chain, as with the standard M/G/1 case, without the additional complexity of supplementary variables. We follow the approach of Baccelli and Makowski (1989) in defining an exponential martingale with respect to the embedded process, and from this we establish a relationship between the forward recurrence times in a type of modified discrete-time Markov renewal process, and the queue length. This is the fundamental relationship of this paper and is expressed in Theorem 4.1. From analysis of the Markov renewal process, the limiting probability generating function of the system size may be expressed in terms of that of the state sojourn times of the Markov renewal process. These may be calculated using the martingale once again, in conjunction with conventional probabilistic arguments.

2 Preliminaries

We assume throughout that all of the random variables and stochastic elements of this paper are defined on some underlying probability space $(\mathcal{P}, \mathcal{F}, \Omega)$ and we denote the indicator function of an event A in $\mathcal{F}$ by $I(A)$.

Following Baccelli and Makowski (1985,1989) we consider the embedded process (X_n), where X_n is the number of customers in the system as seen by the nth departing customer. Note this is not a Markov chain as it depends on the history of the process and not just upon the current state. The number of arrivals during the nth service is given by A_n or B_n if the system is in phase 1 or 2 respectively. (A_n) and (B_n) are two sequences of independent, identically distributed random variables. We assume that $X_0 = 0$ a.s., that is, there is a departure at time 0 which leaves the system empty. We can then define the two generating functions $a(z)$ and $b(z)$ by

$$a(z) = E\left[z^{A_1}\right], \qquad b(z) = E\left[z^{B_1}\right].$$

Both $a(z)$ and $b(z)$ can be written as Laplace-Stieltjes transforms of their respective probability distribution functions,

$$a(z) = A^*(\lambda(1-z)), \qquad b(z) = B^*(\lambda(1-z)).$$

If we take the mean number of arrivals during a service of type A or type B to be ρ_a or ρ_b respectively, then $a'(1) = \rho_a$ and $b'(1) = \rho_b$. We also define two sequences of **N**-valued stopping times (S_n) and (T_n) by

$$\begin{aligned} S_0 &= 0, \\ S_i &= \begin{cases} \inf\{m > S_{i-1} | X_m = 0\}, & \text{if the set is non-empty,} \\ \infty, & \text{otherwise,} \end{cases} \\ T_i &= \begin{cases} \inf\{S_i < m \leq S_{i+1} | X_m \geq k\}, & \text{if the set is non-empty,} \\ S_{i+1}, & \text{otherwise.} \end{cases} \end{aligned}$$

We can define the sequences of events (C_n) and (D_n) respectively by

$$\begin{aligned} C_n &= \bigcup_{i=0}^{\infty} \{S_i \leq n < T_i\}, \\ D_n &= \bigcup_{i=0}^{\infty} \{T_i \leq n < S_{i+1}\}. \end{aligned}$$

C_n and D_n are the events that the system, at departure n, is in phase 1 or phase 2 respectively. Clearly $I(C_n) = 1 - I(D_n)$ for all n. From these definitions

$$X_{n+1} = X_n - I(X_n \neq 0) + A_{n+1} I(C_n) + B_{n+1} I(D_n). \tag{1}$$

We also define the increasing sequence of sigma-algebras $(\mathcal{F}_n)$ by

$$\mathcal{F}_n = \sigma\{X_m | \, 0 \leq m \leq n\},$$

and take

$$\mathcal{F}_\infty = \bigcup_{n=0}^{\infty} \mathcal{F}_n.$$

2.1 The martingale

We can now define a martingale $(M_n(z))$ with filtration $(\mathcal{F}_n)$ by

$$\begin{aligned} M_0(z) &= z^{X_0}, \\ M_n(z) &= z^{X_n} \prod_{k=0}^{n-1} \frac{z^{I(X_k \neq 0)}}{a(z) I(C_k) + b(z) I(D_k)}, \end{aligned}$$

for $z \in [0,1]$. $M_n(z)$ is a $\mathcal{F}_n$-measurable, non-negative, real-valued random variable. To demonstrate that it is a martingale we use (1), the fact that X_m, A_m, B_m, C_m and D_m are all $\mathcal{F}_n$-measurable for $m \leq n$ and that A_{n+1} and B_{n+1} are independent to show that

$$\begin{aligned}
E\left[M_{n+1}(z)|\mathcal{F}_n\right] &= E\left[z^{X_{n+1}} \prod_{k=0}^{n} \frac{z^{I(X_k \neq 0)}}{a(z)I(C_k)+b(z)I(D_k)} \right| \left. \mathcal{F}_n \right] \\
&= E\left[z^{A_{n+1}I(C_n)+B_{n+1}I(D_n)} \right| \left. \mathcal{F}_n \right] z^{X_n - I(X_n \neq 0)} \prod_{k=0}^{n} \frac{z^{I(X_k \neq 0)}}{a(z)I(C_k)+b(z)I(D_k)} \\
&= z^{X_n}\left[a(z)I(C_n)+b(z)I(D_n)\right] z^{-I(X_n \neq 0)} \prod_{k=0}^{n} \frac{z^{I(X_k \neq 0)}}{a(z)I(C_k)+b(z)I(D_k)} \\
&= z^{X_n} \prod_{k=0}^{n-1} \frac{z^{I(X_k \neq 0)}}{a(z)I(C_k)+b(z)I(D_k)} \\
&= M_n(z) \qquad\qquad\qquad\qquad \text{a.s.}
\end{aligned}$$

As the martingale is non-negative, $E\left[|M_n|\right] = E\left[M_n\right] < \infty$, for $z \in [0,1]$.

2.2 Stopping times

We define two sequences of stopping times as follows

$$\begin{aligned}
\tau_1(n) &= \begin{cases} n, & \text{if } X_n \in \text{phase } 2 \text{ and } n < \infty, \\ \inf\{m > n | (X_m > k) \text{ or } (X_m = 0)\}, & \text{if } X_n \in \text{phase } 1, n < \infty, \text{ set non-empty}, \\ \infty, & \text{otherwise}, \end{cases} \\
\tau_2(n) &= \begin{cases} \inf\{m > n | X_m = 0\}, & \text{if } n < \infty \text{ and the set is non-empty}, \\ \infty, & \text{otherwise}, \end{cases}
\end{aligned}$$

where $k \in \mathbf{N}$ is the threshold. Also we may define four sequences of times $(\nu_1(n))$, $(\nu_2(n))$, $(\mu_1(n))$ and $(\mu_2(n))$ in the same way but it is more instructive to define them as follows

$$\begin{aligned}
\nu_1(n) &= \tau_1(n) - n, \\
\nu_2(n) &= \tau_2(n) - n, \\
\mu_1(n) &= \begin{cases} 0, & X_n = 0, \\ \nu_1(n), & X_n \neq 0, \end{cases} \\
\mu_2(n) &= \begin{cases} 0, & X_n = 0, \\ \nu_2(n), & X_n \neq 0. \end{cases}
\end{aligned}$$

The only difference between the μs and νs occurs when $X_n = 0$. We shall later use $\mu_1(n)$ and $\mu_2(n) - \mu_1(n)$ as forward recurrence times in the Markov renewal process while $\nu_1(0)$ and $\nu_2(0) - \nu_1(0)$ will be used as sojourn times.

3 Use of the optional sampling theorem

Lemma 3.1 *For the definitions above, $a(z) > z,\ \forall z \in [0,1)$, if and only if $a'(1) \leq 1$. Similarly for $b(z)$.*

Proof: This follows from Takács' lemma (Takács, 1962, page 46).

Lemma 3.2 *For any a.s. finite stopping time* γ *and* $\lambda, \mu_a > 0$

$$
\begin{array}{lrcll}
(i) & \tau_1(\gamma) & < & \infty & a.s., \\
(ii) & E[\tau_1(\gamma)] & < & \infty, & \\
(iii) & E\left[\omega^{\tau_1(\gamma)}\right] & < & \infty, &
\end{array}
$$

where $\omega \in [0, \alpha]$, $\alpha = \sup_{z\in[0,1]} (z/a(z))$.

Proof: (i) and (ii) follow directly from Williams (1991, page 101). The proof of (iii) is as follows. From (i) and Lemma 3.3 it is sufficient to consider $E\left[\omega^{\tau_1(0)}\right]$. Using the theorem of total expectation we see

$$
E\left[\omega^{\tau_1(0)}\right] = \omega\, p\{\tau_1(0) = 1\} + \sum_{j=1}^{k}\sum_{i=2}^{\infty} \omega^i\, p\{\tau_1(0) = i | X_{i-1} = j\} p\{X_{i-1} = j\}. \tag{2}
$$

We take $a_i = p\{A_1 = i\}$. Now the values of $p\{\tau_1(0) = i|X_{i-1} = j\}$ are shown in Table 1 in Section 5. Thus substituting in (2) we arrive at

$$
\begin{aligned}
E\left[\omega^{\tau_1(0)}\right] &= \omega \left\{ \left(a_0 + \sum_{l=k+1}^{\infty} a_l\right)\left(1 + \sum_{i=1}^{\infty}\omega^i p\{X_i = 1\}\right) \right. \\
&\qquad \left. + \sum_{j=2}^{k}\left(\sum_{l=k+1}^{\infty} a_{l-j+1}\right)\sum_{i=1}^{\infty}\omega^i p\{X_i = j\} \right\} \\
&\le \omega \sum_{j=1}^{k}\left(\sum_{l=k+1}^{\infty} a_{l-j+1}\right)\sum_{i=1}^{\infty}\omega^i p\{X_i = j\}.
\end{aligned}
$$

In order to find $\sum_{i=1}^{\infty}\omega^i p\{X_i = j\}$, we define the vector

$$
\mathbf{v}^i = (p\{X_i = 1\}, p\{X_i = 2\}, \cdots, p\{X_i = k\}),
$$

the sub-stochastic probability transfer matrix

$$
\mathbf{P_k} = \begin{pmatrix}
a_1 & a_2 & a_3 & \cdots & a_{k-1} & a_k \\
a_0 & a_1 & a_2 & \cdots & a_{k-2} & a_{k-1} \\
0 & a_0 & a_1 & \cdots & a_{k-3} & a_{k-2} \\
 & & & \vdots & & \\
0 & 0 & 0 & \cdots & a_0 & a_1
\end{pmatrix}, \tag{3}
$$

and $\mathbf{v}^1 = (a_1, a_2, \ldots, a_k)$, the vector of initial probabilities given a transition from $X_0 = 0$. Then

$$
\mathbf{v}^i = \mathbf{v}^1 {\mathbf{P_k}}^{i-1}.
$$

We seek conditions under which

$$
\sum_{i=1}^{\infty} \omega^i {\mathbf{P_k}}^i
$$

converges. From Householder (1964, page 54) this series converges for $|\omega|\rho(\mathbf{P_k}) < 1$, where $\rho(\mathbf{P_k})$ is the spectral radius of matrix $\mathbf{P_k}$. Given this condition the series converges to $(\mathbf{I}-\omega\mathbf{P_k})^{-1}$. From Barnett and Storey (1970, Theorem 2-9-3) we see that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for any matrix norm $\|\mathbf{A}\|$. From this the series converges if $|\omega|\|\mathbf{P_k}\| < 1$.

We define $\|\mathbf{A}\|_z$ for $z \in (0,1)$ by

$$\|\mathbf{A}\|_z = \max_{i=1,..,k}\left[\sum_{j=1}^{k}|a_{ij}|z^{j-i}\right],$$

for a matrix $\mathbf{A} = (a_{ij})$. This satisfies properties (I),(II),(III) and (IV) of Householder (page 41) and is consistent with the vector norm

$$\|\mathbf{x}\|_z = \max_{i=1,..k}|x_i|z^{1-i}.$$

Now

$$\begin{aligned}\|\mathbf{P_k}\|_z &= \max\left\{\sum_{j=1}^{k}a_jz^{j-1},\ \sum_{j=0}^{k-1}a_jz^{j-1}\right\}\\ &< \sum_{j=0}^{k}a_jz^{j-1}\\ &< \frac{a(z)}{z},\end{aligned}$$

so that

$$\frac{1}{\|\mathbf{P_k}\|_z} > \frac{z}{a(z)}.$$

Thus there exists a $z_0 \in [0,1)$ such that

$$\omega \leq \sup_{z\in[0,1)}\frac{z}{a(z)} < \frac{1}{\|\mathbf{P_k}\|_{z_0}}.$$

Thus the series converges for the desired range and hence the result.

Lemma 3.3 *For any a.s. finite stopping time γ and $\rho_b < 1$*

$$\tau_2(\gamma) < \infty \ \ a.s.$$

Proof: (i) If the system is in phase 1 we use Lemma 3.2 (i) which implies that the system gets to phase 2 in a finite amount of time with probability one.
(ii) Once the system is in phase 2 it behaves as an ordinary M/G/1 queue with mean service time $1/\mu_b$, until reaching state 0 again. Thus the normal stability conditions for the M/G/1 queue provide the result.

Theorem 3.1 *If $\rho_a \in (0,1)$ and $\rho_b \in [0,1)$, the stopping time $\tau_2(\gamma)$ is regular for the martingale $\{M_n(z), n = 0,1,\ldots\}$ and for $z \in [0,1]$ the relationship below holds*

$$E\left[I(\gamma<\infty,\tau_2(\gamma)<\infty)\left(\frac{z}{a(z)}\right)^{\nu_1(\gamma)}\left(\frac{z}{b(z)}\right)^{\nu_2(\gamma)-\nu_1(\gamma)}\middle|\mathcal{F}_\gamma\right] \ = \ I(\gamma<\infty)z^{X_\gamma}z^{I(X_\gamma=0)} \ \ a.s.$$

Proof: This is omitted as it follows directly from a similar proof in Baccelli and Makowski (1985). The case with $\rho_a = 0$ is excluded as the system is then trivial.

Theorem 3.2 *The results of Theorem 3.1 can be extended to the case $\rho_a \in (0, \infty)$.*

Proof: When $z < a(z)$ the proof above applies and so we will only consider the case when $z \geq a(z)$. From Lemma 3.3, $\tau_2(n)$ is almost surely finite. We use Neveu (1975, IV-**3**-16) to show the stopping time $\tau_2(0)$ is regular for the martingale and hence $\tau_2(n)$ is regular. Condition (1) of this proposition,

$$\int_{\{\tau_2(0)<\infty\}} |M_{\tau_2(0)}(z)| dP < \infty,$$

is automatically satisfied for our martingale. Condition (2),

$$\lim_{n\to\infty} \int_{\{\tau_2(0)>n\}} |M_n(z)| dP = 0,$$

can be seen to be satisfied as follows. The martingale is non-negative so we start with

$$\begin{aligned} |M_n(z)| I(\tau_2(0) > n) &= I(\tau_2(0) > n) z^{X_n} \prod_{k=0}^{n-1} \frac{z^{I(X_k \neq 0)}}{a(z) I(C_k) + b(z) I(D_k)} \\ &\leq I(\tau_2(0) > n) \left(\frac{z}{a(z)}\right)^{\tau_1(0)\wedge n} \left(\frac{z}{b(z)}\right)^{(n-\tau_1(0))\vee 0} \quad a.s., \end{aligned}$$

on $z \in [0, 1)$. Lemma 3.1 implies $z/b(z) < 1$ for $z \in [0, 1)$ and $\rho_b < 1$ so we get

$$I(\tau_2(0) > n) \left(\frac{z}{a(z)}\right)^{\tau_1(0)\wedge n} \left(\frac{z}{b(z)}\right)^{(n-\tau_1(0))\vee 0} \leq I(\tau_2(0) > n) \left(\frac{z}{a(z)}\right)^{\tau_1(0)},$$

which tends to zero almost surely as n tends to infinity. Also

$$|M_n(z)| I(\tau_2(0) > n) \leq \left(\frac{z}{a(z)}\right)^{\tau_1(0)} \quad a.s.,$$

which has finite expectation (by Lemma 3.2). So by the dominated convergence theorem

$$E\left[M_n(z) I(\tau_2(0) > n)\right] \to 0,$$

as $n \to \infty$. Neveu (1975, proposition IV-**3**-12) implies that Doob's optional sampling theorem can be applied, whence the result.

4 The relationship

As outlined in the introduction we wish to establish a relationship between the queue length and a modified discrete-time Markov renewal process, the modification being of the type described in Nakagawa and Osaki (1976) in which not all of the states are renewal states. That is, the epochs of transition to certain states are not renewal points. The case in question is the simplest non-trivial version of the Markov renewal process of type I in Nakagawa and Osaki (1976) in which there is

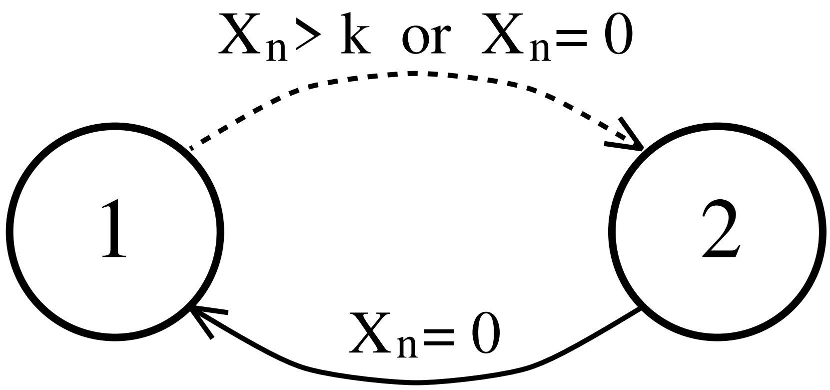

Figure 1: The Markov renewal process (state 2 is the non-renewal state).

only one renewal state and one non-renewal state (see Figure **??**). The transition times correspond to the stopping times $\{S_i, T_i\}$ as defined in Section 2, that is, transitions from $1 \to 2$ occur at the times $\{T_i\}$ while transitions from $2 \to 1$ occur at the times $\{S_i\}$. Clearly a transition from $1 \to 2$ may occur at the same time as a transition from $2 \to 1$ when $T_i = S_{i+1}$. In this case we treat it as though the transitions occur in the order $1 \to 2 \to 1$ with zero time spent in state 2.

In this process $S_{i+1} - T_i$ is not independent of $\mathcal{F}_{T_i}$, however $S_{i+1} - S_i$ is independent of $\mathcal{F}_{S_i}$. Recurrence follows from Lemma 3.3. Thus the process is of the type of Markov renewal processes described above.

In this context $\mu_1(\gamma)$ and $\mu_2(\gamma) - \mu_1(\gamma)$ are forward recurrence times, that is, the amount of time from stopping time γ spent in states 1 or 2, respectively, prior to the transition from state 2 to 1 which is associated with the renewal. Also, given that the process starts with a renewal at time 0, the times $\nu_1(0)$ and $\nu_2(0) - \nu_1(0)$ are sojourn times for the Markov renewal process. The connection between the forward recurrence times and the sojourn times in this type of Markov renewal process can be derived by an extension of the Key renewal theorem. A proof of this can be found in Roughan (1994).

The following result establishes the fundamental relationship between the probability generating function for the queueing process and that of the forward recurrence times of the Markov renewal process.

Theorem 4.1 *For* $n < \infty$, $\rho_a > 0$, $\rho_b < 1$ *and* $z \in [0, 1)$,

$$E\left[z^{X_n}\right] = E\left[\left(\frac{z}{a(z)}\right)^{\mu_1(n)}\left(\frac{z}{b(z)}\right)^{\mu_2(n)-\mu_1(n)}\right].$$

Proof: We use Theorem 3.2 with $\gamma = n$ where $n < \infty$ a.s. Multiplying by the indicator function $I(X_n \neq 0)$, noting that $I(\cdot)^2 = I(\cdot)$ and that $I(S)I(S^c) = 0$ and also using the $\mathcal{F}_n$-measurability of $I(X_n \neq 0)$, we get

$$E\left[I(X_n \neq 0)I(\nu_2(n) < \infty)\left(\frac{z}{a(z)}\right)^{\nu_1(n)}\left(\frac{z}{b(z)}\right)^{\nu_2(n)-\nu_1(n)}\middle|\mathcal{F}_n\right] = I(X_n \neq 0)z^{X_n} \;\; a.s.$$

Taking expectations gives

$$E\left[I(X_n \neq 0)I(\nu_2(n) < \infty)\left(\frac{z}{a(z)}\right)^{\nu_1(n)}\left(\frac{z}{b(z)}\right)^{\nu_2(n)-\nu_1(n)}\right] \;=\; E\left[I(X_n \neq 0)z^{X_n}\right].$$

The term $I(\nu_2(n) < \infty)$ can be removed, as Lemma 3.2 (iii) and the fact that $z/b(z) < 1$ entails that the contribution to the expectation made when $\nu_2(n) = \infty$ is zero. Note that

$$\begin{array}{lcl} [X_n = 0] & \iff & [\mu_1(n) = 0,\ \mu_2(n) = 0], \\ [X_n \neq 0] & \iff & [\mu_1(n) = \nu_1(n),\ \mu_2(n) = \nu_2(n)]. \end{array}$$

From this we arrive at

$$\begin{aligned} p\{\mu_1(n) = 0, \mu_2(n) = 0\} \;\;+\;\; & E\left[I(\mu_1(n), \mu_2(n) \text{ not both } 0)\left(\frac{z}{a(z)}\right)^{\mu_1(n)}\left(\frac{z}{b(z)}\right)^{\mu_2(n)-\mu_1(n)}\right] \\ = \;\; & p\{X_n = 0\} + E\left[I(X_n \neq 0)z^{X_n}\right], \end{aligned}$$

and hence

$$E\left[z^{X_n}\right] \;=\; E\left[\left(\frac{z}{a(z)}\right)^{\mu_1(n)}\left(\frac{z}{b(z)}\right)^{\mu_2(n)-\mu_1(n)}\right].$$

Remark: If we take $Q_n^*(x, y)$ to be the probability generating function for the forward recurrence times in the Markov renewal process, the previous theorem implies that

$$E\left[z^{X_n}\right] = Q_n^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right). \tag{4}$$

Theorem 4.2 *For $\rho_a > 0$, $\rho_b < 1$ and $z \in [0, 1)$ and given that $X_0 = 0$,*

$$E\left[\left(\frac{z}{a(z)}\right)^{\nu_1(0)}\left(\frac{z}{b(z)}\right)^{\nu_2(0)-\nu_1(0)}\right] \;=\; z.$$

Proof: As in the proof of Theorem 4.1 we multiply by an indicator function, in this case $I(X_n = 0)$, and then take expectations to derive

$$E\left[I(X_n = 0)\left(\frac{z}{a(z)}\right)^{\nu_1(n)}\left(\frac{z}{b(z)}\right)^{\nu_2(n)-\nu_1(n)}\right] \;=\; E\left[z\, I(X_n = 0)\right],$$

which when we assume $X_0 = 0$ gives

$$E\left[\left(\frac{z}{a(z)}\right)^{\nu_1(0)}\left(\frac{z}{b(z)}\right)^{\nu_2(0)-\nu_1(0)}\right] \;=\; z.$$

Remark: This provides a relationship for the number of customers served in each phase during the busy period. Unlike the unaltered M/G/1 queue, discussed in Baccelli and Makowski (1989), this does not uniquely determine the probability generating function for the number of customers served. However, it does uniquely determine this probability generating function in conjunction with the result for $E\left[\left(\frac{z}{b(z)}\right)^{\nu_2(0)-\nu_1(0)}\right]$, which is described later in Theorems 5.1 and 5.2. As noted before that the times $\nu_1(0)$ and $\nu_2(0)-\nu_1(0)$ can also be considered to be sojourn times in the Markov renewal process so that if $F^*(x,y)$ is the probability generating function for the sojourn times then

$$F^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right) = z. \tag{5}$$

4.1 The limit as $n \to \infty$.

Theorem 4.3 *For the definitions above and $\rho_a > 0$, $\rho_b < 1$ we have*

$$\lim_{n\to\infty} E\left[z^{X_n}\right] = \frac{1}{m}\left[\frac{F^*(1, z/b(z)) - F^*(z/a(z), z/b(z))}{1 - z/a(z)} + \frac{1 - F^*(1, z/b(z))}{1 - z/b(z)}\right],$$

where m is a normalising constant.

Proof: If we take the probability generating function for the forward recurrence times from time n in the Markov renewal process to be $Q_n^*(x,y)$ then from equation (4),

$$E\left[z^{X_n}\right] = Q_n^*\left(\frac{z}{a(z)}, \frac{z}{b(z)}\right)$$

on $z \in [0,1)$. From Roughan (1994) we have the result

$$\lim_{n\to\infty} Q_n^*(x,y) = \frac{1}{m}\left[\frac{F^*(1,y) - F^*(x,y)}{1-x} + \frac{1 - F^*(1,y)}{1-y}\right],$$

where m is a normalising constant and $F^*(x,y)$ is the probability generating function for the sojourn times in the modified Markov renewal process. This is true for all x and y such that the left-hand side converges. For $x \in [0,\alpha]$ and $y \in [0,1)$, $x^{\mu_1(n)}y^{\mu_2(n)-\mu_1(n)}$ is bounded above by $x^{\mu_1(n)}$ and Lemma 3.2 (iii) implies that $E\left[x^{\mu_1(n)}\right] < \infty$. Thus from the dominated convergence theorem, $Q_n^*(x,y)$ converges for all $x \in [0,\alpha]$ and $y \in [0,1)$. Hence the result follows for all $\rho_a > 0$ and $\rho_b \in [0,1)$.

Corollary 4.1 *For $\rho_a > 0$, $\rho_b \in [0,1)$ and $z \in [0,1)$ we have*

$$E\left[z^{X}\right] = \frac{1}{m}\left[\frac{F^*(1, z/b(z)) - z}{1 - z/a(z)} + \frac{1 - F^*(1, z/b(z))}{1 - z/b(z)}\right],$$

where $X(t) \to X$ in distribution as $t \to \infty$.

Proof: Cooper (1972, page 154), implies that for this system in equilibrium, the arrivals see the same distribution as that the departures leave. Thus as Poisson arrivals see time averages (PASTA, Wolff, 1989), we can see that

$$\lim_{n\to\infty} E\left[z^{X_n}\right] = \lim_{t\to\infty} E\left[z^{X(t)}\right].$$

From the bounded convergence theorem

$$\lim_{t\to\infty} E\left[z^{X(t)}\right] = E\left[\lim_{t\to\infty} z^{X(t)}\right].$$

Relationship (5) states $F^*(z/a(z), z/b(z)) = z$ and hence the result.

5 Calculating $F^*(1, y)$

From Corollary 4.1 we can see we need to calculate $F^*\left(1, z/b(z)\right)$ in order to calculate the solution. Again we resort to the martingale arguments of Section 3.

Theorem 5.1 *Given the previous definitions, $\rho_a > 0$, $\rho_b < 1$, for all $z \in [0, 1]$*

$$F^*\left(1, \frac{z}{b(z)}\right) = E\left[z^{X_{\tau_1(0)}}\right].$$

Proof: We use Theorem 3.2 again, choosing $\gamma = \tau_1(0)$. We have

$$E\left[I(\nu(\tau_1(0)) < \infty)\left(\frac{z}{a(z)}\right)^{\nu_1(\tau_1(0))}\left(\frac{z}{b(z)}\right)^{\nu_2(\tau_1(0))-\nu_1(\tau_1(0))}\middle|\mathcal{F}_{\tau_1(0)}\right] \quad = \quad z^{X_{\tau_1(0)}} z^{I(X_{\tau_1(0)}=0)} \quad a.s.$$

Following the procedure of Theorems 4.1 and 4.2 we derive

$$E\left[\left(\frac{z}{a(z)}\right)^{\mu_1(\tau_1(0))}\left(\frac{z}{b(z)}\right)^{\mu_2(\tau_1(0))-\mu_1(\tau_1(0))}\right] \quad = \quad E\left[z^{X_{\tau_1(0)}}\right].$$

Now $\mu_1(\tau_1(0)) = 0$ and $\mu_2(\tau_1(0)) = \nu_2(0) - \nu_1(0)$, so

$$E\left[\left(\frac{z}{b(z)}\right)^{\nu_2(0)-\nu_1(0)}\right] \quad = \quad E\left[z^{X_{\tau_1(0)}}\right],$$

and therefore

$$F^*\left(1, \frac{z}{b(z)}\right) = E\left[z^{X_{\tau_1(0)}}\right].$$

Theorem 5.2 *Given the previous definitions, $k \geq 1$, $\rho_a > 0$, $\rho_b < 1$ and $z \in [0, 1)$,*

$$E\left[z^{X_{\tau_1(0)}}\right] = \frac{a(z)}{z}\left(1 - \frac{z}{a(z)}\right)\left(\mathbf{e_1}(\mathbf{I} - \mathbf{P_k})^{-1}\mathbf{z}^t\right) + z,$$

where $\mathbf{P_k}$ is the sub-stochastic matrix defined by (3), $\mathbf{I}$ is the k x k identity matrix, $\mathbf{z} = (z, z^2, \ldots, z^k)$ and $\mathbf{e_1} = (1, 0, \ldots, 0)$.

Proof: We consider $E\left[z^{X_{\tau_1(0)}}\right]$. Using the theorem of total expectation we get

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] \quad &= \quad E\left[z^{X_1}|\tau_1(0) = 1\right]p\{\tau_1(0) = 1\} \\ &+ \sum_{j=1}^{k}\sum_{i=2}^{\infty} E\left[z^{X_i}|\tau_1(0) = i, X_{i-1} = j\right]p\{\tau_1(0) = i|X_{i-1} = j\}p\{X_{i-1} = j\}, \quad (6) \end{aligned}$$

as $X_0 = 0$. The values of

$$E\left[z^{X_i}|\tau_1(0) = i, X_{i-1} = j\right]p\{\tau_1(0) = i|X_{i-1} = j\} \tag{7}$$

are shown in Table 1. Thus, substituting the values of (7) in (6) we arrive at the equation

		$E\left[z^{X_{\tau_1(0)}} \middle\| \tau_1(0)=i, X_{i-1}=j\right]$	$p_{\{\tau_1(0)=i \| X_{i-1}=j\}}$	the product (7)
$i=1$	$j=0$	$\frac{a_0+\sum_{l=k+1}^{\infty} a_l z^l}{a_0+\sum_{l=k+1}^{\infty} a_l}$	$a_0+\sum_{l=k+1}^{\infty} a_l$	$a(z)-\sum_{l=1}^{k} a_l z^l$
$i>1$	$j=0$	0	0	0
$i>1$	$j=1$	$\frac{a_0+\sum_{l=k+1}^{\infty} a_l z^l}{a_0+\sum_{l=k+1}^{\infty} a_l}$	$a_0+\sum_{l=k+1}^{\infty} a_l$	$a(z)-\sum_{l=1}^{k} a_l z^l$
$i>1$	$1<j\leq k$	$\frac{\sum_{l=k+1}^{\infty} a_{l-j+1} z^l}{\sum_{l=k+1}^{\infty} a_{l-j+1}}$	$\sum_{l=k+1}^{\infty} a_{l-j+1}$	$a(z)z^{j-1}-\sum_{l=0}^{k-j+1} a_l z^{l+j-1}$

Table 1:

$$E\left[z^{X_{\tau_1(0)}}\right] = \left(a(z)-\sum_{l=1}^{k} a_l z^l\right)+\sum_{j=1}^{k}\left(a(z)z^{j-1}-\sum_{l=(j-1)\vee 1}^{k} a_{l-j+1}z^l\right)\sum_{i=1}^{\infty} p\{X_i=j\}.$$

In order to find the $\sum_{i=1}^{\infty} p\{X_i=j\}$, we define $\mathbf{v}^i$, $\mathbf{P_k}$ and $\mathbf{v}^1$ as in the proof of Lemma 3.2. Then

$$\mathbf{v}^i = \mathbf{v}^1 \mathbf{P_k}^{i-1}.$$

Using standard probabilistic arguments we can see that

$$\sum_{i=1}^{\infty} \mathbf{v}^i = \mathbf{v}^1(\mathbf{I}-\mathbf{P_k})^{-1},$$

and since $\mathbf{v}^1 = \mathbf{e_1}\mathbf{P_k}$ we have

$$\mathbf{v}^1(\mathbf{I}-\mathbf{P_k})^{-1} = -\mathbf{e_1}+\mathbf{e_1}\left(\mathbf{I}-\mathbf{P_k}\right)^{-1}.$$

Taking $\mathbf{g}(z)=(g_1(z), g_2(z), \cdots, g_k(z))$, where

$$g_j(z) = \left(a(z)z^{j-1}-\sum_{l=(j-1)\vee 1}^{k} a_{l-j+1}z^l\right),$$

we get

$$\begin{aligned} E\left[z^{X_{\tau_1(0)}}\right] &= g_1(z)+\sum_{i=1}^{\infty}\left(\sum_{j=1}^{k} p\{X_i=j\}g_j(z)\right) \\ &= \mathbf{e_1}\mathbf{g}^t(z)+\sum_{i=1}^{\infty}\mathbf{v}^i\mathbf{g}^t(z) \\ &= (\mathbf{e_1}-\mathbf{e_1}+\mathbf{e_1}\left(\mathbf{I}-\mathbf{P_k}\right)^{-1})\mathbf{g}^t(z) \\ &= \mathbf{e_1}\left(\mathbf{I}-\mathbf{P_k}\right)^{-1}\mathbf{g}^t(z), \end{aligned}$$

where, $\mathbf{g}^t(z)$ denotes the transpose of the row-vector $\mathbf{g}(z)$. We can simplify $g_j(z)$ and hence $\mathbf{g}(z)$ by taking $\mathbf{z}=(z, z^2, \ldots, z^k)$, to derive

$$\mathbf{g}^t(z) = \frac{a(z)}{z}\mathbf{z}^t-\mathbf{P_k}\mathbf{z}^t,$$

and therefore we can write $E\left[z^{X_{\tau_1(0)}}\right]$ as

$$\begin{aligned}
E\left[z^{X_{\tau_1(0)}}\right] &= \mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\left(\frac{a(z)}{z}\mathbf{z}-\mathbf{P_k}\mathbf{z}^t\right) \\
&= \frac{a(z)}{z}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)-\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{P_k}\mathbf{z}^t\right) \\
&= \frac{a(z)}{z}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)+\left(\mathbf{e_1}\mathbf{I}\mathbf{z}^t\right)-\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right) \\
&= \frac{a(z)}{z}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)+z-\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right) \\
&= \frac{a(z)}{z}\left(1-\frac{z}{a(z)}\right)\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)+z.
\end{aligned}$$

6 The equilibrium solution

The solution can now be found.

Theorem 6.1 *With the previous definitions, $\rho_a > 0$, $\rho_b \in [0,1)$, $k \geq 1$ and $z \in [0,1)$ the equilibrium solution is*

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z)+\{b(z)-a(z)\}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)}{b(z)-z}\right],$$

where $m(1-\rho_b) = 1+\{\rho_a-\rho_b\}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{1}^t\right)$ and $\mathbf{1} = (1,1,1,\ldots,1)$.

Proof: Substituting in the results of Corollary 4.1 and Theorem 5.2 gives

$$\begin{aligned}
E\left[z^X\right] &= \frac{1}{m}\left[\frac{F^*\left(1,\frac{z}{b(z)}\right)-z}{1-\frac{z}{a(z)}}+\frac{1-F^*\left(1,\frac{z}{b(z)}\right)}{1-\frac{z}{b(z)}}\right] \\
&= \frac{1}{m}\left[\frac{1-z}{1-\frac{z}{b(z)}}+\frac{F^*\left(1,\frac{z}{b(z)}\right)-z}{1-\frac{z}{a(z)}}-\frac{F^*\left(1,\frac{z}{b(z)}\right)-z}{1-\frac{z}{b(z)}}\right] \\
&= \frac{1}{m}\left[\frac{b(z)(1-z)}{b(z)-z}+\frac{\left\{\frac{z}{a(z)}-\frac{z}{b(z)}\right\}\left[F^*\left(1,\frac{z}{b(z)}\right)-z\right]}{(1-\frac{z}{a(z)})(1-\frac{z}{b(z)})}\right] \\
&= \frac{1}{m}\left[\frac{b(z)(1-z)+\{b(z)-a(z)\}\left(\mathbf{e_1}(\mathbf{I}-\mathbf{P_k})^{-1}\mathbf{z}^t\right)}{b(z)-z}\right].
\end{aligned}$$

In this theorem we treat m as a normalising constant. It can be interpreted as the mean number of services during a busy period. Either of these facts may be used to derive its enunciated value.
Remark: The probability generating function for the stationary M/G/1 queue is simply

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z)}{b(z)-z}\right], \tag{8}$$

where $m = 1/(1-\rho)$. This suggests that the result of Theorem 6.1 is merely this solution plus a correcting term derived from the modified behaviour during phase 1. This is related to the fact that this modification is equivalent to an extended modification of the boundary conditions of the queue.

7 Limiting cases

There are several special cases of this system which have been examined in detail in the literature. The simplest is the M/G/1 queue. The probability generating function for the equilibrium number in the M/G/1 queueing system is expressed in (8). This is the solution we expect to get from the following special cases. If $a(z) = b(z)$ the solution is immediate. If $k \to \infty$ and $\rho_a < 1$ we also expect this solution. As $k \to \infty$, $F^*(1, y)$, which is the generation function for the time spent in phase 2, approaches 1. This is because the probability that zero time is spent in the second phase approaches one. Hence

$$E[z^X] = \frac{1}{m}\left[\frac{1-z}{1-\frac{z}{a(z)}}\right] = \frac{1}{m}\frac{a(z)(1-z)}{a(z)-z}.$$

The case when $k = 0$ includes such examples as the M/G/1 queue with server vacations or a warmup time. These are situations in which the service-time distribution is different for a customer arriving at an empty system. The solution to these systems is given in Yeo (1962, Theorem 3) as

$$E\left[z^X\right] = \frac{1}{m}\left\{\frac{b(z)-za(z)}{b(z)-z}\right\},$$

where m is given by

$$m = \frac{\rho_b - 1 - \rho_a}{\rho_b - 1}.$$

In this case, using our technique, Theorem 5.1 gives $F^*(1, y)$ as

$$F^*\left(1, \frac{z}{b(z)}\right) = E\left[z^{X_{\tau'(n)}}\right].$$

The right-hand side of this equation is simply $a(z)$. Once $F^*(1, y)$ is known we can write the solution using Corollary 4.1 as

$$\begin{aligned} E[z^X] &= \frac{1}{m}\left[\frac{(a(z)-z)(1-\frac{z}{b(z)}) + (1-a(z))(1-\frac{z}{a(z)})}{(1-\frac{z}{a(z)})(1-\frac{z}{b(z)})}\right] \\ &= \frac{1}{m}\left[\frac{b(z)-za(z)}{b(z)-z}\right]. \end{aligned}$$

8 Conclusion

Conventionally the type of queueing system considered in this paper could be subsumed under the general block-matrix methodology of the M/G/1 type (see Neuts (1989)). While versatile and able to deal with a number of extensions of this sort of problem, the block-matrix approach is better suited to numerical computation rather than the derivation of analytical formulae such as we derive here. Of course the generating function expressed in Theorem 6.1 requires the inversion of a kxk matrix $(\mathbf{I} - \mathbf{P_k})$. However this matrix is already in lower Hessenburg form (see Golub and van Loan (1983)) which makes it much less computationally expensive to calculate its inverse.

The block-matrix methodology would typically include the phase in the current state of the process, making the state of the embedded process $(X_n, phase)$. Thence the probability transition matrix would be partitioned into sub-blocks so as to obtain a repeated structure such as

$$\mathbf{P} = \begin{pmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ A_0 & A_1 & A_2 & A_3 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ & & & \vdots & \end{pmatrix}.$$

This can then be used to create a numerical algorithm to deduce the equilibrium probabilities of the system. In this current system it is difficult to see how A_k and B_k could be less than kxk matrices making the computations involved in the algorithm of at least the same order of difficulty as the matrix inversion required for the solution herein.

The method of this paper can be extended in a number of ways. For instance throughout this paper we have swapped to the second service distribution when the system had more than k customers in it. Any stopping time $\tau_1(t)$ which occurs at the end of a service and which is less than $\tau_2(t)$ and greater than t is acceptable. Thus it could occur after a number of services in the busy period, or at a completely random time. Given a threshold for which Lemma 3.2 holds, Theorems 3.1-2, 4.1-3, 5.1 and Corollary 4.1 all remain unchanged. Only Theorem 5.2 must be modified to suit the altered threshold. Extensions to multiple threshold problems can also be made.

Baccelli and Makowski have used their technique to produce results on the transient behaviour of the M/G/1 queue. The analysis of this paper can also be extended to deal with transients, though the formulae involved are very complicated and of questionable utility.

References

[1] BACCELLI, F. AND MAKOWSKI, A.M. (1985) Direct martingale arguments for stability: the M/GI/1 case. *Systems Control Letters*, **6**, 181–186.

[2] BACCELLI, F. AND MAKOWSKI, A.M. (1989) Dynamic, transient and stationary behaviour of the M/GI/1 queue via martingales. *Annals of Probability*, **17**(4), 1691–1699.

[3] BARNETT, S. AND STOREY, S. (1970) *Matrix Methods in Stabililty.* Thomas Nelson and Sons Ltd.

[4] COOPER, R.B. (1972) *Introduction to Queueing Theory*. The Macmillian Company.

[5] GOLUB, G.H. AND VAN LOAN, C.F. (1983) *Matrix Computations.* North Oxford Academic.

[6] HOUSEHOLDER, A.S. (1964) *The Theory of Matrices in Numerical Analysis.* Blaisdell Publishing Company.

[7] NAKAGAWA, T. AND OSAKI, S. (1976) Markov renewal processes with some non-regeneration points and their applications to reliability theory. *Microelectronics and Reliabliity*, **15**, 633-636.

[8] NEVEU, J. (1975) *Discrete-Parameter Martingales*. North-Holland Publishing Company.

[9] NEUTS, M.F. (1989) *Structured Stochastic Matrices of M/G/1 Type and their Applications*. Marcel Dekker.

[10] ROUGHAN, M. (1994) *An Application of Martingales to Queueing Theory.* PhD. thesis, Department of Applied Mathematics, University of Adelaide.

[11] TAKÁCS, L. (1962) *Introduction to the Theory of Queues*. Oxford University Press.

[12] WILLIAMS, D. (1991) *Probability with Martingales*. Cambridge University Press.

[13] WOLFF, R.W. (1989) *Stochastic Modelling and the Theory of Queues*. Prentice Hall International.

[14] YEO, G.F. (1962) Single server queue with modified service mechanisms, *Journal of Operations Research*, **1**, 499-507.