Applied Probability Trust (January 18, 2022)

SIMPLE ANALYTICAL SOLUTIONS FOR THE M^b/E$_k$/1/m, E$_k$/M^b/1/m AND RELATED QUEUES

BENNY VAN HOUDT,* *University of Antwerp*

Abstract

In this paper we revisit some classical queueing systems such as the M^b/E$_k$/1/m and E$_k$/M^b/1/m queue for which fast numerical and recursive methods exist to study their main performance measures.

We present simple explicit results for the loss probability and queue length distribution of these queueing systems as well as for some related queues such as the M^b/D/1/m, the D/M^b/1/m queue and fluid versions thereof. In order to establish these results we first present a simple analytical solution for the invariant measure of the M/E$_k$/1 queue that appears to be new.

Keywords: Queueing systems; analytical solution; loss probability; bulk service; bulk arrivals; Erlang-k distribution

2020 Mathematics Subject Classification: Primary 60K25

Secondary 68M20

1. Introduction

Queueing theory is a well established discipline in the field of applied probability with applications in various areas [1, 2, 4, 6, 7]. While simple closed form results for various finite capacity queueing systems have been derived, these systems typically have an underlying Markov chain with a birth-death structure.

In this paper we revisit a number of elementary finite capacity queueing systems which do not have an underlying birth-death structure and for which, to the best of our knowledge, no simple closed form results have appeared in the literature. More specifically, we present a simple analytical solution for the M^b/E$_k$/1/m, the E$_k$/M^b/1/m queue and some closely related queues. In an M^b/E$_k$/1/m queue arrivals occur in batches of size b according to a Poisson process, a job requires an Erlang-k service

* Postal address: Middelheimlaan 1, B2020 Antwerp, Belgium

* Email address: benny.vanhoudt@uantwerpen.be

time, a single server processes jobs in FCFS order and at most m jobs can be present in the queue. An $\mathrm{E}_k/\mathrm{M}^b/1/\mathrm{m}$ also offers room for up to m jobs, but jobs arrive according to a renewal process with Erlang-k inter-arrival times and are processed in batches (of size b), where the service time of a batch is exponential.

The contributions of the paper are the following:

1. We first present a closed form expression for the stationary measure of a simple class of Markov chains in Section 2, which immediately yields an expression for the queue length distribution of the $\mathrm{M}/\mathrm{E}_k/1$ queue, that appears to be new.

2. We prove a number of properties of this stationary measure in Section 3.

3. Sections 4 and 5 then present the analytical solution for the $\mathrm{M}^b/\mathrm{E}_k/1/\mathrm{m}$ and the $\mathrm{E}_k/\mathrm{M}^b/1/\mathrm{m}$ queue, respectively.

4. By letting k tend to infinity, we obtain simple closed form results for the $\mathrm{M}^b/\mathrm{D}/1/\mathrm{m}$ and $\mathrm{D}/\mathrm{M}^b/1/\mathrm{m}$ queue in Section 6. Setting $b = 1$ for the $\mathrm{M}^b/\mathrm{D}/1/\mathrm{m}$ queue yields the analytical solution of the M/D/1/m queue presented in [3].

5. Finally, in Section 7 we let b tend to infinity which yields closed form results for some fluid queues with jumps.

To establish the results in Sections 4 and 5 we make use of a general truncation result for the $\mathrm{M}^X/\mathrm{GI}/1/\mathrm{m}$ and $\mathrm{GI}/\mathrm{M}^Y/1/\mathrm{m}$ queue established by Miyazawa in [10]. It should however be noted that in a number of cases the use of this truncation result can be replaced by a censoring argument. We further note that the analytical results presented in the paper can also be used to numerically compute the loss probability and queue length distribution of the $\mathrm{M}^b/\mathrm{E}_k/1/\mathrm{m}$ and the $\mathrm{E}_k/\mathrm{M}^b/1/\mathrm{m}$ queue in $\mathrm{O}(km)$ time.

Finite capacity queueing systems are typically studied using numerical methods [8, 12, 13]. Such numerical methods can also give rise to closed form results in some particular cases (e.g., the M/PH/1/m queue [12]), but these are expressed in matrix form and therefore do not immediately yield a simple analytical solution.

The work presented in this paper was motivated by some recent work on load-balancing in large-scale systems. More precisely, in [5] a number of load balancing policies are studied that have bounded queue lengths in the large-scale limit. In case of the resource pooling policy (initially introduced in [14]) the maximum queue length of this load balancing policy when subject to phase-type distributed job sizes, can be determined via the loss probability in an M/PH/1/m queue. As the loss probability in an M/PH/1/m queue is minimized over all phase-type distributions of order k by an Erlang-k distribution (due to [10]), the loss probability in an M/E$_k$/1/m queue can be used to establish a lower bound on the maximum queue length of the resource pooling policy over all order k phase-type distributions.

2. A simple Markov chain

We define a continuous time Markov chain $(X_t^{(\ell,x)})_{t\geq 0}$ on the state space $\{0,1,\ldots\}$ such that $X_t^{(\ell,x)}$ decreases by one at rate $q\ell$ and increases by ℓ at rate qx. Note that this Markov chain is positive recurrent for $x<1$, transient for $x>1$ and null-recurrent when $x=1$. Further any stationary measure of this chain is independent of q.

Let $(p_0^{(\ell)}(x), p_1^{(\ell)}(x), p_2^{(\ell)}(x),\ldots)$ be any stationary measure of $(X_t^{(\ell,x)})_{t\geq 0}$. The first ℓ balance equations imply that $\ell p_1^{(\ell)}(x) = x p_0^{(\ell)}(x)$ and $\ell p_n^{(\ell)}(x) = (\ell+x)p_{n-1}^{(\ell)}(x)$ for $n=2,\ldots,\ell$, meaning

$$p_n^{(\ell)}(x) = p_0^{(\ell)}(x)\frac{x}{\ell}\left(1+\frac{x}{\ell}\right)^{n-1},$$

for $n=1,\ldots,\ell$. From the remaining balance equations we see

$$p_n^{(\ell)}(x) = \left(1+\frac{x}{\ell}\right)p_{n-1}^{(\ell)}(x) - \frac{x}{\ell}p_{n-1-\ell}^{(\ell)}(x), \tag{1}$$

for $n>\ell$. Therefore given $p_0^{(\ell)}(x)$ the chain has a unique stationary measure.

The next theorem presents a simple closed form expression for $p_n^{(\ell)}(x)$:

Theorem 1. *The unique stationary measure $(p_0^{(\ell)}(x), p_1^{(\ell)}(x), p_2^{(\ell)}(x),\ldots)$ with $p_0^{(\ell)}(x) = 1$ of the Markov chain $(X_t^{(\ell,x)})_{t\geq 0}$ can be expressed as*

$$p_n^{(\ell)}(x) = c_n^{(\ell)}(x) - c_{n-1}^{(\ell)}(x),$$

with $c_0^{(\ell)}(x)=1$,

$$c_n^{(\ell)}(x) = \sum_{i=0}^{\lfloor \frac{n}{\ell+1}\rfloor}(-1)^i\left(\frac{x}{\ell}\right)^i\binom{n-i\ell}{i}\left(1+\frac{x}{\ell}\right)^{n-i(\ell+1)}, \tag{2}$$

for $n>0$ and $c_n^{(\ell)}(x)=0$ for $n<0$.

Proof. The proof is by direct verification of the balance equations (it is also possible to derive this expression using the generating function approach). For $n \leq \ell$ we have $c_n^{(\ell)}(x) = (1+x/\ell)^n$ and therefore $p_n^{(\ell)}(x) = (1+x/\ell)^{n-1}x/\ell$ as required. For $n > \ell$, we verify (1), by showing that $c_n^{(\ell)}(x) + xc_{n-(\ell+1)}^{(\ell)}(x)/\ell = (1+x/\ell)c_{n-1}^{(\ell)}(x)$. Now,

$$\begin{aligned}\frac{x}{\ell}c_{n-(\ell+1)}^{(\ell)}(x) &= \sum_{i=0}^{\lfloor \frac{n}{\ell+1} \rfloor -1} (-1)^i \left(\frac{x}{\ell}\right)^{i+1} \binom{n-(i+1)\ell-1}{i} \left(1+\frac{x}{\ell}\right)^{n-(i+1)(\ell+1)} \\ &= -\sum_{i=1}^{\lfloor \frac{n}{\ell+1} \rfloor} (-1)^i \left(\frac{x}{\ell}\right)^{i} \binom{n-i\ell-1}{i-1} \left(1+\frac{x}{\ell}\right)^{n-i(\ell+1)}.\end{aligned}$$

Therefore,

$$\begin{aligned}c_n^{(\ell)}(x) + \frac{x}{\ell}c_{n-(\ell+1)}^{(\ell)}(x) &= \left(1+\frac{x}{\ell}\right)^n \\ &\quad + \sum_{i=1}^{\lfloor \frac{n}{\ell+1} \rfloor} (-1)^i \left(\frac{x}{\ell}\right)^{i} \left[\binom{n-i\ell}{i} - \binom{n-i\ell-1}{i-1}\right] \left(1+\frac{x}{\ell}\right)^{n-i(\ell+1)} \\ &= \left(1+\frac{x}{\ell}\right) \sum_{i=0}^{\lfloor \frac{n-1}{\ell+1} \rfloor} (-1)^i \left(\frac{x}{\ell}\right)^{i} \binom{n-1-i\ell}{i} \left(1+\frac{x}{\ell}\right)^{n-1-i(\ell+1)} \\ &= \left(1+\frac{x}{\ell}\right) c_{n-1}^{(\ell)}(x), \end{aligned} \tag{3}$$

where the second equality holds as both binomial coefficients are equal to one when n is a multiple of $(\ell+1)$ and $i = n/(\ell+1)$. □

Remarks:

1. The sequence $c_n^{(\ell)}(x)$ is increasing in n as $p_n^{(\ell)}(x)$ is a stationary measure, meaning $p_n^{(\ell)}(x) \geq 0$. Moreover, given x and ℓ, we can numerically compute the values of $c_1^{(\ell)}(x), c_2^{(\ell)}(x) \ldots, c_n^{(\ell)}(x)$ in $O(n)$ time by noting that $c_n(x) = (1+x/\ell)^n$ for $n \leq \ell$ and using (3) for $n > \ell$.

2. As the stationary measure of an irreducible transient or null-recurrent Markov chain always sums to infinity, this result implies that $c_n^{(k)}(x)$ tends to infinity for $x \geq 1$ as n tends to infinity. A direct proof of this is also presented in Section 3.

Corollary 2.1. *The queue length distribution $Q_\lambda^{(b,k)}$ in an $M^b/E_k/1$ queue with mean* *job size equal to $1/b$ and batch arrival rate $\lambda < 1$ is given by*

$$\mathbb{P}[Q_\lambda^{(b,k)} \leq j] = (1-\lambda) c_{jk}^{(bk)}(\lambda)$$
$$= (1-\lambda) \sum_{i=0}^{\lfloor \frac{jk}{bk+1} \rfloor} (-1)^i \left(\frac{\lambda}{bk}\right)^i \binom{(j-ib)k}{i} \left(1+\frac{\lambda}{bk}\right)^{(j-ib)k-i}.$$

Proof. If we keep track of the number of phases present in an $\mathrm{M}^b/\mathrm{E}_k/1$ queue with batch arrival rate λ and mean job size $1/b$, we obtain the Markov chain $(X_t^{(bk,\lambda)})_{t\geq 0}$ with $q = 1$. The result therefore follows from Theorem 1 as the $\mathrm{M}^b/\mathrm{E}_k/1$ queue with load λ is empty with probability $1-\lambda$. □

Setting $b = 1$ in the above result appears to yield a new expression for the queue length distribution of the $\mathrm{M}/\mathrm{E}_k/1$ queue. The queue length distribution of the $\mathrm{M}/\mathrm{E}_k/1$ queue is typically expressed in terms of the k roots of a degree k polynomial (that are computed numerically), see [1]. A more involved explicit expression for the queue length distribution in the $\mathrm{M}/\mathrm{E}_k/1$ queue that does not require the computation of any roots can be found in [9, p167].

3. Properties of $c_n^{(\ell)}(x)$

We now study $c_n^{(\ell)}(x)$ in some more detail. More specifically, we show that $c_n^{(\ell)}(x)$ is a degree n polynomial in x and provide two simple closed form expressions for its coefficients. To prove one of the expressions, we rely on the following well-known equality:

Proposition 3.1. (j-th difference formula of a degree j polynomial.) *Let $j \geq 0$ be an integer and y, Δ real numbers, then*

$$\sum_{i=0}^{j} (-1)^i \binom{j}{i} Q(y+i\Delta) = (-1)^j a_j \Delta^j j!, \tag{4}$$

where $Q(x) = \sum_{q=0}^{j} a_q x^q$ is any degree j polynomial.

Theorem 2. *Let $c_n^{(\ell)}(x)$ be defined by (2), then $c_n^{(\ell)}(x)$ is a degree n polynomial in x. Let $[x^s]c_n^{(\ell)}(x)$ denote the coefficient of x^s of the polynomial $c_n^{(\ell)}(x)$, then for $0 \leq s \leq n$*

we have

$$[x^s]c_n^{(\ell)}(x) = \frac{1}{\ell^s}\sum_{i=0}^{\min(s,\lfloor (n-s)/\ell\rfloor)} (-1)^i \binom{s}{i}\binom{n-i\ell}{s} \tag{5}$$

$$= 1 - \frac{1}{\ell^s}\sum_{i=\lfloor n/\ell\rfloor+1}^{s} (-1)^{i+s}\binom{s}{i}\binom{s+i\ell-n-1}{s}, \tag{6}$$

where the second expression implies that $[x^s]c_n^{(\ell)}(x) = 1$ *for* $s \leq \lfloor n/\ell \rfloor$.

Proof. Expanding $(1+x/\ell)^{n-i(\ell+1)}$ in (2) yields

$$\begin{aligned}
c_n^{(\ell)}(x) &= \sum_{i=0}^{\lfloor \frac{n}{\ell+1}\rfloor} (-1)^i \binom{n-i\ell}{i} \sum_{s=0}^{n-i(\ell+1)} \binom{n-i(\ell+1)}{s}\left(\frac{x}{\ell}\right)^{s+i}.\\
&= \sum_{i=0}^{\lfloor \frac{n}{\ell+1}\rfloor} (-1)^i \binom{n-i\ell}{i} \sum_{s=i}^{n-i\ell} \binom{n-i(\ell+1)}{s-i}\frac{x^s}{\ell^s}.\\
&= \sum_{s=0}^{n} \frac{x^s}{\ell^s} \sum_{i=0}^{\min(s,\lfloor (n-s)/\ell\rfloor)} (-1)^i \binom{n-i\ell}{i}\binom{n-i(\ell+1)}{s-i}.
\end{aligned}$$

The expression in (5) now follows by applying the *cancellation identity* which states that

$$\binom{t}{r}\binom{t-r}{a} = \binom{t}{r+a}\binom{r+a}{r},$$

with $t = n - i\ell$, $r = i$ and $a = s - i$.

To obtain (6) we first note that

$$\frac{1}{\ell^s}\sum_{i=0}^{s}(-1)^i\binom{s}{i}\binom{n-i\ell}{s} = \frac{1}{\ell^s}\sum_{i=0}^{s}(-1)^i\binom{s}{i}\frac{Q(n-i\ell)}{s!} = 1,$$

due to (4) with $j = s$, $y = n$, $\Delta = -\ell$ and $Q(x) = \prod_{q=0}^{s-1}(x-q)$. Therefore (5) implies that

$$[x^s]c_n^{(\ell)}(x) = 1 - \frac{1}{\ell^s}\sum_{i=\lfloor (n-s)/\ell\rfloor+1}^{s} (-1)^i\binom{s}{i}\binom{n-i\ell}{s}.$$

Note that s exceeds $n - i\ell$ when $i \geq \lfloor (n-s)/\ell \rfloor + 1$. Hence, $\binom{n-i\ell}{s}$ is zero when $n - i\ell \geq 0$, that is, for $i \leq \lfloor n/\ell \rfloor$ and it suffices to sum i from $\lfloor n/\ell \rfloor + 1$ to s. □

Theorem 3. *The values* $c_n^{(\ell)}(x)$ *obey the following simple recursion:*

$$c_n^{(\ell)}(x) = 1 + \frac{x}{\ell}\sum_{j=1}^{\min(n,\ell)} c_{n-j}^{(\ell)}(x), \tag{7}$$

for $n \geq 1$. Further,

$$1 = [x^0]c_n^{(\ell)}(x) \geq [x^1]c_n^{(\ell)}(x) \geq \ldots \geq [x^n]c_n^{(\ell)}(x) = \ell^{-n} \geq 0, \tag{8}$$

$c_n^{(\ell)}(x)$ is convex and increasing in x on $[0,\infty)$ and $c_{j\ell}^{(\ell)}(x) \geq \sum_{i=0}^{j} x^i$.

Proof. For $n \leq \ell$ we have $c_n^{(\ell)}(x) = (1 + x/\ell)^n$, which implies that the recursion holds for $n \leq \ell$. For $n > \ell$, we have by (3) that

$$c_n^{(\ell)}(x) - c_{n-1}^{(\ell)}(x) = \frac{x}{\ell}\left(c_{n-1}^{(\ell)}(x) - c_{n-(\ell+1)}^{(\ell)}(x)\right).$$

Plugging this into

$$c_n^{(\ell)}(x) - c_\ell^{(\ell)}(x) = \sum_{i=\ell+1}^{n}\left(c_i^{(\ell)}(x) - c_{i-1}^{(\ell)}(x)\right),$$

yields

$$\begin{aligned} c_n^{(\ell)}(x) - c_\ell^{(\ell)}(x) &= \frac{x}{\ell}\sum_{j=1}^{\ell} c_{n-j}^{(\ell)}(x) - \frac{x}{\ell}\sum_{i=0}^{\ell-1} c_i^{(\ell)}(x) \\ &= \frac{x}{\ell}\sum_{j=1}^{\ell} c_{n-j}^{(\ell)}(x) - (c_\ell^{(\ell)}(x) - 1) \end{aligned}$$

which proves the recursion.

Using induction on n, this recurrence together with (5) implies (8). Hence $c_n^{(\ell)}(x)$ is a degree n polynomial in x with non-negative coefficients and therefore all its derivatives with respect to x are non-negative on $[0,\infty)$. The lower bound on $c_{j\ell}^{(\ell)}(x)$ follows by noting that $[x^s]c_{j\ell}^{(\ell)}(x) = 1$, for $s = 0,\ldots,j$ due to Theorem 2, while $[x^s]c_{j\ell}^{(\ell)}(x) \geq 0$ for $s > j$. □

4. The $M^b/E_k/1/m$ queue

We now derive results for the $M^b/E_k/1/m$ queue with batch arrival rate λ by relying on the following Theorem by Miyazawa. We assume without loss of generality that the mean (Erlang distributed) service time of a job equals $1/b$, such that the load equals λ (which may exceed 1).

Theorem 4. (Theorem 2.1 in [10].) *Let $(\psi_0, \psi_1, \ldots)$, with $\psi_0 = 1$, be the unique stationary measure of the Markov chain $(Y_t)_{t\geq 0}$ that keeps track of the number of* *jobs in an $M^X/G/1$ queue immediately after a service completion. The queue length distribution in an $M^X/G/1/m$ queue immediately after a service completion can be written as*

$$\mathbb{P}[Q^{M^X/G/1/m}(service) = j] = \frac{\psi_j}{\sum_{i=0}^{m-1}\psi_i} \tag{9}$$

for $j = 0, \ldots, m-1$.

Miyazawa also derived a formula for the loss probability in an $M^X/G/1/m$ queue using the above result. This formula is however incorrect in case of batch arrivals (as the probability that the server is busy should be $1 - p_0^a m_H$ instead of $1 - p_0^a$). It is easy to see that the correct formula for the loss probability in case of batch arrivals is given by

$$\mathbb{P}_{loss}^{M^X/G/1/m} = 1 - \frac{\sum_{i=0}^{m-1}\psi_i}{m_H + \lambda\sum_{i=0}^{m-1}\psi_i} = 1 - \frac{1}{\rho}\left(1 - \frac{m_H}{m_H + \lambda\sum_{i=0}^{m-1}\psi_i}\right), \tag{10}$$

where m_H is the mean batch size (see also [Equation (3.7)][11]). Note that the same argument to prove Theorem 4.1 in [10] remains valid with this corrected loss probability formula.

We are now in a position to express the queue length distribution and loss probability of the $M^b/E_k/1/m$ queue in terms of the polynomials $c_n^{(kb)}(\lambda)$.

Theorem 5. *The loss probability $\mathbb{P}_{loss}^{\lambda}(k,b,m)$ in the $M^b/E_k/1/m$ queue is given by*

$$\mathbb{P}_{loss}^{\lambda}(k,b,m) = 1 - \frac{1}{\lambda}\left(1 - \frac{b}{b + \lambda\sum_{i=0}^{m-1}\psi_i^{(k,b)}(\lambda)}\right), \tag{11}$$

with

$$\psi_i^{(k,b)}(\lambda) = \frac{kb}{\lambda}\left(c_{ik+1}^{(kb)}(\lambda) - c_{ik}^{(kb)}(\lambda)\right) = c_{ik}^{(kb)}(\lambda) - c_{(i-b)k}^{(kb)}(\lambda), \tag{12}$$

and

$$\sum_{i=0}^{j}\psi_i^{(k,b)}(\lambda) = \sum_{i=0}^{\min(j,b-1)} c_{(j-i)k}^{(kb)}(\lambda), \tag{13}$$

for $j \geq 0$, where $c_n^{(\ell)}(x)$ is given by (2). The queue length distribution $Q_\lambda^{(k,b,m)}$ can be expressed as

$$\mathbb{P}[Q_\lambda^{(k,b,m)} \leq j] = \frac{b c_{jk}^{(kb)}(\lambda)}{b + \lambda\sum_{i=0}^{m-1}\psi_i^{(k,b)}(\lambda)}$$

for $j = 0, \ldots, m-1$.

Proof. If we keep track of the number of phases present in the $\mathrm{M}^b/\mathrm{E}_k/1$ queue, we obtain the Markov chain $(X_t^{(\ell,x)})_{t\geq 0}$ with $x=\lambda$, $\ell=kb$ and $q=1$. Theorem 1 therefore implies that $p_n^{(kb)}(\lambda)=c_n^{(kb)}(\lambda)-c_{n-1}^{(kb)}(\lambda)$ is a stationary measure for this chain.

The queue length becomes i after a service completion from state $ik+1$. Therefore the measure $\psi_i^{(k,b)}(\lambda)$ used in Theorem 4 is proportional to $p_{ik+1}^{(kb)}(\lambda)$ and is given by

$$\psi_i^{(k,b)}(\lambda)=\frac{kb}{\lambda}(c_{ik+1}^{(kb)}(\lambda)-c_{ik}^{(kb)}(\lambda)), \tag{14}$$

where the factor kb/λ is such that $\psi_0^{(k,b)}(\lambda)=1$. From (7) we note that

$$c_{ik+1}^{(kb)}(\lambda)-c_{ik}^{(kb)}(\lambda)=\frac{\lambda}{kb}(c_{ik}^{(kb)}(\lambda)-c_{(i-b)k}^{(kb)}(\lambda)), \tag{15}$$

for $i\geq b$ and

$$c_{ik+1}^{(kb)}(\lambda)-c_{ik}^{(kb)}(\lambda)=\frac{\lambda}{kb}c_{ik}^{(kb)}(\lambda), \tag{16}$$

for $i<b$. These equalities can be combined with (14) to yield (12). It is now easy to verify that

$$\sum_{i=0}^{j}\psi_i^{(k,b)}(\lambda)=\sum_{i=0}^{\min(j,b-1)}c_{(j-i)k}^{(kb)}(\lambda). \tag{17}$$

Further, by Theorem 4 and (2.9) in [10] we have

$$\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival)=j]=\frac{\psi_j^{(k,b)}(\lambda)}{b+\lambda\sum_{i=0}^{m-1}\psi_i^{(k,b)}(\lambda)},$$

for $j=0,\ldots,m-1$. Note that this is the queue length distribution just prior to a job arrival. In case of batch arrivals, the $i-1$ jobs preceding the i-th job of a batch are also considered as being part of the queue just before the arrival of the i-th job.

As the arrivals occur in batches of size b, we have

$$\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival)=j]=\sum_{i=0}^{\min(j,b-1)}\frac{1}{b}\mathbb{P}[Q_\lambda^{(k,b,m)}=j-i].$$

This implies that

$$\begin{aligned}\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival)=j]-\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival)=j-1]\\ =\frac{1}{b}(\mathbb{P}[Q_\lambda^{(k,b,m)}=j]-\mathbb{P}[Q_\lambda^{(k,b,m)}=j-b]1[j\geq b]).\end{aligned}$$

Reordering yields the recursion

$$\mathbb{P}[Q_\lambda^{(k,b,m)} = j] = b\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival) = j] - b\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival) = j-1] + \mathbb{P}[Q_\lambda^{(k,b,m)} = j-b]1[j \geq b].$$

Hence,

$$\begin{aligned}\mathbb{P}[Q_\lambda^{(k,b,m)} = j] &= b\sum_{i=0}^{\lfloor j/b \rfloor} \Big(\mathbb{P}[Q_\lambda^{(k,b,m)}(arrival) = j-ib] \\ &\quad - \mathbb{P}[Q_\lambda^{(k,b,m)}(arrival) = j-ib-1]\Big) \\ &= \frac{b\sum_{i=0}^{\lfloor j/b \rfloor}(\psi_{j-ib}^{(k,b)}(\lambda) - \psi_{j-ib-1}^{(k,b)}(\lambda))}{b + \lambda\sum_{i=0}^{m-1}\psi_i^{(k,b)}(\lambda)},\end{aligned}$$

with $\psi_{-1}^{(k,b)}(\lambda) = 0$. Finally, (12) implies that

$$c_{jk}^{(kb)}(\lambda) = \psi_j^{(k,b)}(\lambda) + c_{jk-bk}^{(kb)}(\lambda),$$

which can be used to show that

$$c_{jk}^{(kb)}(\lambda) = \sum_{i=0}^{\lfloor j/b \rfloor} \psi_{j-ib}^{(k,b)}(\lambda),$$

and similarly as $\psi_{-1}^{(k,b)}(\lambda) = 0$

$$c_{(j-1)k}^{(kb)}(\lambda) = \sum_{i=0}^{\lfloor j/b \rfloor} \psi_{j-ib-1}^{(k,b)}(\lambda).$$

Therefore

$$\mathbb{P}[Q_\lambda^{(k,b,m)} = j] = \frac{b(c_{jk}^{(kb)}(\lambda) - c_{(j-1)k}^{(kb)}(\lambda))}{b + \lambda\sum_{i=0}^{m-1}\psi_i^{(k,b)}(\lambda)},$$

□

We now focus on two special cases: the M/E$_k$/1/m queue and the M^b/M/1/m queue:

Corollary 4.1. *The loss probability $\mathbb{P}_{loss}^\lambda(k,1,m)$ in an M/E$_k$/1/m queue is given by*

$$\mathbb{P}_{loss}^\lambda(k,m) = 1 - \frac{1}{\lambda}\left(1 - \frac{1}{1+\lambda c_{(m-1)k}^{(k)}(\lambda)}\right), \tag{18}$$

where $c^{(k)}_{(m-1)k}(\lambda)$ is given by (2). The queue length distribution can be expressed as

$$\mathbb{P}[Q^{(k,1,m)}_\lambda \le j] = \frac{c^{(k)}_{jk}(\lambda)}{1+\lambda c^{(k)}_{(m-1)k}(\lambda)},$$

for $j = 0, \ldots, m-1$.

Proof. Setting $b = 1$ implies that $\sum_{i=0}^{j} \psi_i^{(k,1)}(\lambda) = c^{(k)}_{jk}(\lambda)$. □

Corollary 4.2. *The loss probability $\mathbb{P}_{loss}(1, b, m)$ in an $M^b/M/1/m$ queue is given by*

$$\mathbb{P}^\lambda_{loss}(1, b, m) = 1 - \frac{1}{\lambda}\left(1 - \frac{1}{c^{(b)}_m(\lambda)}\right), \tag{19}$$

where $c^{(b)}_m(\lambda)$ is given by (2). The queue length distribution can be expressed as

$$\mathbb{P}[Q^{(1,b,m)}_\lambda \le j] = \frac{c^{(b)}_j(\lambda)}{c^{(b)}_m(\lambda)},$$

for $j = 0, \ldots, m$.

Proof. Setting $k = 1$ immediately implies that $\psi_i^{(1,b)}(\lambda) = b(c^{(b)}_{i+1}(\lambda) - c^{(b)}_i(\lambda))/\lambda$ and therefore $\sum_{i=0}^{j} \psi_i^{(1,b)}(\lambda) = b(c^{(b)}_{j+1}(\lambda) - 1)/\lambda$. □

5. The $E_k/M^b/1/m$ queue

Consider the $E_k/M^b/1/m$ queue with arrival rate λ and a mean batch service rate $1/b$, such that the load equals λ. To express the queue length distribution and loss probability we rely on the following duality result by Miyazawa:

Theorem 6. (Theorem 3.1 in [10].) *Let $(\phi_0, \phi_1, \ldots)$, with $\phi_0 = 1$, be the unique stationary measure of the Markov chain $(Z_t)_{t\ge 0}$ that keeps track of the number of jobs in a* **modified** *$M^X/G/1$ queue immediately after a service completion where the modification is such that the batch size equals one when the queue is empty. Let λ be the batch arrival rate and G the service time distribution.*

The queue length distribution in a $GI/M^X/1/m$ queue with inter-arrival time distribution G and batch service rate λ just prior to an arrival can be written as

$$\mathbb{P}[Q^{GI/M^X/1/m}(arrival) = j] = \frac{\phi_{m-j}}{\sum_{i=0}^{m}\phi_i} \tag{20}$$

for $j = 0, \ldots, m$.

Lemma 5.1. *Let* $(\phi_0^{(k,b)}(x), \phi_1^{(k,b)}(x), \ldots)$, *with* $\phi_0^{(k,b)}(x) = 1$, *be the unique stationary measure of the Markov chain* $(Z_t^{(k,b,x)})_{t\geq 0}$ *that keeps track of the number of jobs in a* **modified** $M^b/E_k/1$ *queue with load* x *immediately after a service completion where the modification is such that the batch size equals one when the queue is empty, then*

$$\phi_i^{(k,b)}(x) = c_{ik}^{(kb)}(x) - c_{(i-1)k}^{(kb)}(x), \tag{21}$$

for $i \geq 0$.

Proof. The stationary measure of the $\mathrm{M}^X/\mathrm{G}/1$ queue without the modification is given by (12). The expression for the complementary generating functions in (2.7) in [10] implies that for $0 < z < z_0$ for some z_0 sufficiently small

$$\sum_{i=0}^{\infty} \phi_i^{(k,b)}(x) z^i = \frac{1-z}{1-z^b} \sum_{i=0}^{\infty} \psi_i^{(k,b)}(x) z^i = \sum_{i=0}^{\infty} \psi_i^{(k,b)}(x) z^i (1-z) \sum_{j=0}^{\infty} z^{bi},$$

which yields

$$\phi_i^{(k,b)}(x) = \sum_{j=0}^{\lfloor i/b \rfloor} \psi_{i-jb}^{(k,b)}(x) - \sum_{j=0}^{\lfloor (i-1)/b \rfloor} \psi_{i-jb-1}^{(k,b)}(x).$$

Combined with (12) this proves the statement. □

Lemma 5.2. *The polynomials defined by* (2) *obey the following equality*

$$\sum_{i=0}^{k-1} c_{jk+i}^{(kb)}(x) = \frac{kb}{x} \sum_{i=0}^{\lfloor j/b \rfloor} (c_{(j+1)k-ibk}^{(kb)}(x) - c_{jk-ibk}^{(kb)}(x)), \tag{22}$$

for $k, b > 0$ *and* $j \geq 0$.

Proof. By (3) we have

$$c_{n-i\ell+1}^{(\ell)}(x) - c_{n-i\ell}^{(\ell)}(x) = \frac{x}{\ell}(c_{n-i\ell}^{(\ell)}(x) - c_{n-(i+1)\ell}^{(\ell)}(x)),$$

for $n - i\ell + 1 > 0$. Summing over i then yields

$$c_n^{(\ell)}(x) = \frac{\ell}{x} \sum_{i=0}^{\lfloor n/\ell \rfloor} (c_{n-i\ell+1}^{(\ell)}(x) - c_{n-i\ell}^{(\ell)}(x)),$$

as $c^{(\ell)}_{n-\ell-\lfloor n/\ell\rfloor\ell}(x) = 0$. Setting $\ell = kb$ and $n = jk$ we therefore have

$$\begin{aligned}\sum_{s=0}^{k-1} c^{(kb)}_{jk+s}(x) &= \frac{kb}{x}\sum_{s=0}^{k-1}\sum_{i=0}^{\lfloor (jk+s)/kb\rfloor}(c^{(kb)}_{jk+s-ikb+1}(x) - c^{(kb)}_{jk+s-ikb}(x)) \\ &= \frac{kb}{x}\sum_{i=0}^{\lfloor j/b\rfloor}\sum_{s=0}^{k-1}(c^{(kb)}_{jk+s-ikb+1}(x) - c^{(kb)}_{jk+s-ikb}(x)) \\ &= \frac{kb}{x}\sum_{i=0}^{\lfloor j/b\rfloor}(c^{(kb)}_{jk-ikb+k}(x) - c^{(kb)}_{jk-ikb}(x)),\end{aligned}$$

where the second equality holds as $\lfloor (jk+s)/kb\rfloor = \lfloor j/b\rfloor$ for $s \in \{0,\dots,k-1\}$. □

Theorem 7. *The loss probability $\tilde{\mathbb{P}}^{\lambda}_{loss}(k,b,m)$ in the $E_k/M^b/1/m$ queue with load λ is given by*

$$\tilde{\mathbb{P}}^{\lambda}_{loss}(k,b,m) = \frac{1}{c^{(kb)}_{mk}(1/\lambda)}. \tag{23}$$

The queue length distribution is given by

$$\mathbb{P}[\tilde{Q}^{(k,b,m)}_{\lambda} \geq m-j] = \frac{\sum_{i=0}^{k-1} c^{(kb)}_{jk+i}(1/\lambda)}{kc^{(kb)}_{mk}(1/\lambda)},$$

for $j = 0,\dots,m-1$.

Proof. The loss formula follows by combining Theorem 6 with Lemma 5.1 where we note that the $M^b/E_k/1$ queue has load $1/\lambda$ if the dual $E_k/M^b/1/m$ queue has load λ.

By Theorem 6, (2.9) in [10] and Lemma 5.1 we have for $j = 0,\dots,m-1$:

$$\mathbb{P}[\tilde{Q}^{(k,b,m)}_{\lambda}(service) = j] = \frac{c^{(kb)}_{(m-j)k}(1/\lambda) - c^{(kb)}_{(m-j-1)k}(1/\lambda)}{c^{(kb)}_{mk}(1/\lambda) - 1},$$

where $\tilde{Q}^{(k,b,m)}_{\lambda}(service)$ represents the queue length immediately after a job completion in an $E_k/M^b/1/m$ queue, where the $b-i$ jobs after the i-th job of a batch are still counted as present in the queue when the i-th job completes service.

As the batch service completions occur at rate $1/b$, we have

$$\mathbb{P}[\tilde{Q}^{(k,b,m)}_{\lambda}(service) = m-j] = \frac{\mathbb{P}[m-j+1 \leq \tilde{Q}^{(k,b,m)}_{\lambda} \leq \min(m-j+b,m)]}{b\lambda(1-\tilde{P}^{\lambda}_{loss}(k,b,m))},$$

for $j = 1, \ldots, m$. Combining the previous two equations with the expression for $\tilde{P}^{\lambda}_{loss}(k,b,m)$ shows that

$$\mathbb{P}[m-j+1 \leq \tilde{Q}_\lambda^{(k,b,m)} \leq \min(m-j+b,m)] = \frac{\lambda b(c_{jk}^{(kb)}(1/\lambda) - c_{(j-1)k}^{(kb)}(1/\lambda))}{c_{mk}^{(kb)}(1/\lambda)},$$

for $j = 0, \ldots, m-1$. The probability $\mathbb{P}[\tilde{Q}_\lambda^{(k,b,m)} \geq m-j]$ for can therefore be expressed as

$$\begin{aligned}
&\mathbb{P}[\tilde{Q}_\lambda^{(k,b,m)} \geq m-j] \\
&\qquad = \sum_{i=0}^{\lfloor j/b \rfloor} \mathbb{P}[m-j-ib \leq \tilde{Q}_\lambda^{(k,b,m)} \leq \min(m-j-ib+b-1,m)] \\
&\qquad = \frac{\lambda b}{c_{mk}^{(kb)}(1/\lambda)} \sum_{i=0}^{\lfloor j/b \rfloor} (c_{((j+1)-ib)k}^{(kb)}(1/\lambda) - c_{(j-ib)k}^{(kb)}(1/\lambda)),
\end{aligned}$$

for $j = 0, \ldots, m-1$. The proof now completes by applying Lemma 5.2 with $x = 1/\lambda$. □

We now focus on two special cases: the $E_k/M/1/m$ queue and the $M/M^b/1/m$ queue:

Corollary 5.1. *The loss probability $\tilde{\mathbb{P}}^{\lambda}_{loss}(k,1,m)$ in an $E_k/M/1/m$ queue is given by $1/c_{mk}^{(k)}(1/\lambda)$, where $c_n^{(\ell)}(x)$ is given by (2). The queue length distribution can be expressed as*

$$\mathbb{P}[\tilde{Q}_\lambda^{(k,1,m)} \geq m-j] = \frac{\lambda(c_{(j+1)k}^{(k)}(1/\lambda) - 1)}{c_{mk}^{(k)}(1/\lambda)},$$

for $j = 0, \ldots, m-1$.

Proof. Setting $b = 1$ implies that

$$\mathbb{P}[\tilde{Q}_\lambda^{(k,1,m)} \geq m-j] = \frac{\sum_{i=0}^{k-1} c_{jk+i}^{(k)}(1/\lambda)}{k c_{mk}^{(k)}(1/\lambda)} = \frac{\lambda(c_{(j+1)k}^{(k)}(1/\lambda) - 1)}{c_{mk}^{(k)}(1/\lambda)},$$

due to (3) with $\ell = k$ and $x = 1/\lambda$. □

Corollary 5.2. *The loss probability $\tilde{\mathbb{P}}_{loss}(1,b,m)$ in an $M/M^b/1/m$ queue is given by $1/c_m^{(b)}(1/\lambda)$, where $c_n^{(\ell)}(x)$ is given by (2). The queue length distribution can be expressed as*

$$\mathbb{P}[\tilde{Q}_\lambda^{(1,b,m)} \geq m-j] = \frac{c_j^{(b)}(1/\lambda)}{c_m^{(b)}(1/\lambda)},$$

for $j = 0, \ldots, m$.

Proof. Setting $k = 1$ in Theorem 7 yields the result. □

It is worth noting that the loss probability of the E_k/M/1/m queue and the M/M^k/1/mk queue are identical, even though their associated finite state Markov chains are not.

6. The M^b/D/1/m and D/M^b/1/m queue

We now present closed form results for the M^b/D/1/m and D/M^b/1/m queue. Setting $b = 1$ in the M^b/D/1/m queue yields the results in [3] (as $\xi_j(\lambda)$ corresponds to b_j in [3]).

Theorem 8. *The loss probability $\mathbb{P}^{\lambda}_{loss}(\infty, b, m)$ in the M^b/D/1/m queue is given by*

$$\mathbb{P}^{\lambda}_{loss}(\infty, b, m) = 1 - \frac{1}{\lambda}\left(1 - \frac{b}{b + \lambda \sum_{i=0}^{\min(b-1,m-1)} \xi_{(m-1-i)/b}(\lambda)}\right), \tag{24}$$

with

$$\xi_y(x) = \sum_{i=0}^{\lfloor y \rfloor} \frac{(-1)^i x^i}{i!}(y-i)^i e^{x(y-i)}, \tag{25}$$

for $y \geq 0$ real. The queue length distribution $Q^{(\infty,b,m)}_{\lambda}$ can be expressed as

$$\mathbb{P}[Q^{(\infty,b,m)}_{\lambda} \leq j] = \frac{b\xi_{j/b}(\lambda)}{b + \lambda \sum_{i=0}^{\min(m-1,b-1)} \xi_{(m-1-i)/b}(\lambda)},$$

for $j = 0, \ldots, m-1$.

Proof. The result follows from Theorem 5 by letting k tend to infinity and noting that $\xi_{j/b}(\lambda) = \lim_{k\to\infty} c^{(kb)}_{jk}(\lambda)$. □

Theorem 9. *The loss probability $\tilde{\mathbb{P}}^{\lambda}_{loss}(\infty, b, m)$ in the D/M^b/1/m queue is given by*

$$\tilde{\mathbb{P}}^{\lambda}_{loss}(\infty, b, m) = \frac{1}{\xi_{m/b}(1/\lambda)}, \tag{26}$$

where $\xi_y(x)$ is given by (25). The queue length distribution $\tilde{Q}^{(\infty,b,m)}_{\lambda}$ can be expressed as

$$\mathbb{P}[\tilde{Q}^{(\infty,b,m)}_{\lambda} \geq m - j] = \frac{\lambda b \sum_{i=0}^{\lfloor j/b \rfloor} (\xi_{(j+1)/b-i}(1/\lambda) - \xi_{j/b-i}(1/\lambda))}{\xi^{(b)}_{m}(1/\lambda)},$$

for $j = 0, \ldots, m-1$.

Proof. The result follows from Theorem 7 and Lemma 5.2 by letting k tend to infinity. □

7. Some finite capacity fluid queues

In this section we present results for the following two fluid queues. In both queues we have a buffer with capacity m. In the first queueing system fluid drains from the queue at rate 1 and arrivals occur at rate λ that instantaneously add one unit of fluid to the buffer unless the fluid in the buffer exceeds $m-1$, in which case the fluid level becomes m. Note that this fluid queue corresponds to the work present in an M/D/1 queue with bounded workload and partial acceptance.

The second queueing system is the dual of the first, where the fluid level increases at rate λ and Poisson arrivals at rate 1 instantaneously remove one unit of fluid from the buffer or empty the buffer if the fluid is below 1. We denote these queueing systems as the fluid M/D/1 and fluid D/M/1 queue, respectively.

Corollary 7.1. *The fraction of lost work in the fluid M/D/1 is given by*

$$\lim_{b\to\infty} \mathbb{P}^{\lambda}_{loss}(1,b,bm) = 1 - \frac{1}{\lambda}\left(1 - \frac{1}{\xi_m(\lambda)}\right),$$

where $\xi_y(x)$ is defined in (25)*. The probability that the fluid level is at most $y \in [0,m]$ is given by $\xi_y(\lambda)/\xi_m(\lambda)$.*

Proof. The fluid M/D/1 queue is equivalent to the M^b/M/1/bm queue if we let b tend to infinity and renormalize the buffer size by b, therefore the result follows from Corollary 4.2 and the fact that $\xi_y(x) = \lim_{b\to\infty} c^{(b)}_{by}(x)$. □

Corollary 7.2. *The fraction of lost fluid in the fluid D/M/1 is given by*

$$\lim_{b\to\infty} \tilde{\mathbb{P}}^{\lambda}_{loss}(1,b,bm) = \frac{1}{\xi_m(1/\lambda)},$$

where $\xi_y(x)$ is defined in (25)*. The probability that the fluid level exceeds $m-y$ with $y \in [0,m]$ is given by $\xi_y(1/\lambda)/\xi_m(1/\lambda)$.*

Proof. The fluid D/M/1 queue is equivalent to the M/M^b/1/bm queue if we let b tend to infinity and renormalize the buffer size by b, therefore the result follows from Corollary 5.2 and the fact that $\xi_y(x) = \lim_{b\to\infty} c^{(b)}_{by}(x)$. □

Note that the above results are also valid if m is not an integer.

8. Conclusions

In this paper we presented simple analytical expressions for the $\mathrm{M}^b/\mathrm{E}_k/1/\mathrm{m}$ and $\mathrm{E}_k/\mathrm{M}^b/1/\mathrm{m}$ queue as well as for some closely related queueing systems. We demonstrated that both the loss probability and queue length distribution can be expressed in terms of the polynomials $c_n^\ell(x)$ given by (2).

We further note that these polynomials can also be used to obtain analytical expressions for other queueing systems not considered in this paper such as the $\mathrm{M}/\mathrm{M}^b/1/\mathrm{m}$ queue where the server is interrupted if the number of jobs in the system is below b. It may also be possible to extend some of the results in this paper to multi-server systems.

Funding information

This work was supported by the FWO-Flanders project G033119N.

References

[1] Adan, I. and Resing, J. (2001). *Queueing Theory: Ivo Adan and Jacques Resing.* Eindhoven University of Technology. Department of Mathematics and Computing Science.

[2] Asmussen, S. (2008). *Applied probability and queues* vol. 51. Springer Science & Business Media.

[3] Brun, O. and Garcia, J.-M. (2000). Analytical solution of finite capacity M/D/1 queues. *Journal of Applied Probability* **37,** 1092–1098.

[4] Gross, D. and Harris, C. (1974). *Fundamentals of Queueing Theory.* John Wiley and Sons, New York.

[5] Hellemans, T., Kielanski, G. and Van Houdt, B. (2022). Performance of load balancers with bounded maximum queue length in case of non-exponential job sizes. *arXiv preprint arXiv.org/abs/2201.03905*.

[6] Kleinrock, L. (1975). *Queueing Systems Vol. I.* Wiley, New York.

[7] Kleinrock, L. (1976). *Queueing Systems Vol. II.* Wiley, New York.

[8] Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods and stochastic modeling.* SIAM, Philadelphia.

[9] Medhi, J. (2002). *Stochastic models in queueing theory.* Elsevier.

[10] Miyazawa, M. (1990). Complementary generating functions for the MX/GI/1/k and GI/MY/1/k queues and their application to the comparison of loss probabilities. *Journal of Applied Probability* **27,** 684–692.

[11] Miyazawa, M. and Shanthikumar, J. G. (1991). Monotonicity of the loss probabilities of single server finite queues with respect to convex order of arrival or service processes. *Probability in the Engineering and Informational Sciences* **5,** 43–52.

[12] Neuts, M. (1981). *Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach.* John Hopkins University Press.

[13] Neuts, M. (1989). *Structured Stochastic Matrices of M/G/1 type and their applications*. Marcel Dekker, Inc., New York and Basel.

[14] Tsitsiklis, J. N. and Xu, K. (2012). On the power of (even a little) resource pooling. *Stochastic Systems* **2,** 1–66.