A Markov Modulated Multi-server Queue with Negative Customers – The MM CPP/GE/c/L G-Queue

Ram Chakka* Peter G. Harrison†

Abstract

We obtain the queue length probability distribution at equilibrium for a multi-server queue with generalised exponential service time distribution and either finite or infinite waiting room. This system is modulated by a continuous time Markov phase process. In each phase, the arrivals are a superposition of a positive and a negative arrival stream, each of which is a compound Poisson process with phase dependent parameters, i.e. a Poisson point process with bulk arrivals having geometrically distributed batch size. Such a queueing system is well suited to B-ISDN/ATM networks since it can account for both burstiness and correlation in traffic. The result is exact and is derived using the method of spectral expansion applied to the two dimensional (queue length by phase) Markov process that describes the dynamics of the system. Several variants of the system are considered, applicable to different modelling situations, such as server breakdowns, cell losses and load balancing.

We also consider the departure process and derive its batch size distribution and the Laplace transform of the interdeparture time probability density function. From this, a recurrence formula is obtained for its moments. The analysis therefore provides the basis of a building block for modelling networks of switching nodes in terms of their internal arrival processes.

*Department of Computer Science, Norfolk State University, Norfolk, VA 23504, USA. Email: ram_chakka@hotmail.com

†Department of Computing, Imperial College, 180 Queen's Gate, London, SW7 2BZ. Tel: +44 171 594 8363 Email: pgh@doc.ic.ac.uk

1 Introduction

Various models have been proposed for describing the traffic that arises in today's telecommunications systems, such as ATM and the Internet, which often exhibits burstiness – i.e. batches of transmission units (e.g. packets) arrive together – and correlation between interarrival times. These models include the compound Poisson process (CPP) in which the interarrival times are assumed to have generalised exponential (GE) probability distribution [19], the Markov modulated Poisson process (MMPP) and self-similar traffic models such as Fractional Brownian Motion (FBM) [22, 27]. A CPP traffic model often gives a good representation of burstiness, e.g. [9, 1], but not of the auto-correlations observed in much real traffic. Conversely, the MMPP models can capture auto-correlation but not burstiness, e.g. [21, 5, 8]. The self-similar models such as FBM can account for both auto-correlation and burstiness, but they are analytically intractable in a queueing context.

We introduce a new queueing/traffic system, the Markov Modulated CPP/GE/c/L G-queue. This is a multi-server queue with GE service times and with both positive and negative arrival streams, each of which is a CPP, i.e. a Poisson point process with batch arrivals of geometrically distributed size. In other words, interarrival times are also GE random variables. All three GE distributions (for positive and negative customer interarrival times and for service time) are modulated by a continuous time Markov phase process. Negative customers remove (positive) customers in the queue and have been used to model random neural networks, task termination in speculative parallelism, faulty components in manufacturing systems and server breakdowns [13, 11, 6, 7, 15]. The name G-queue has been adopted for queues with negative customers in acknowledgement of Gelenbe who first introduced them [10]. This queueing model can account for burstiness and correlation, but in addition the negative customers can represent additional behaviours such as breakdowns, killing signals, cell losses and load balancing.

In the next section we define the MM CPP/GE/c/L G-queue and model it as a continuous time Markov chain in section 3, where we also derive its equilibrium Kolmogorov equations and transform these into a form suitable for solution by the method of spectral expansion [24]. The steady state joint probability distribution for the queue length and phase is then obtained by this method in section 4. Obviously the MMCPP/GE/c/L, MMCPP/MMGE/c/L and the MMPP/M/c/L [5] queues, with only positive customers, are special cases of this more general queueing model. The matrix geometric method [26] can also be used, instead of spectral expansion, and we consider this alternative with its relative merits in section 5. In section 6, an alternate killing discipline by negative customers is considered where customers actually being served are immune and cannot be removed by a negative arrival. This killing discipline is appropriate for modelling load balancing where a customer in service would never be moved to a less utilised server. At the opposite extreme, the killing discipline in which customers are removed from the head of the queue – one in service being the first to go – is considered in section 7. This discipline is appropriate for modelling server breakdowns where a customer in service is lost and maybe also part of the queue. The departure process of the queue is determined in section 8 as the Laplace transform of the probability density function of interdeparture time, whence follow recurrence formulae for its moments.

The paper concludes in section 9 with a discussion of the implications of our results on the modelling of ATM and IP neworks. Certainly, the analysis provides the basis of a building block for networks of routers described in terms of their internal arrival processes. Each node in the model can represent both burstiness and correlation and its parameters can be matched to observations made in real traffic. In addition, the modulation of the *whole* node facilitates the modelling of different router characteristics for different types of traffic in the following sense. The rates of the different traffic streams can be modelled as the rates of different phases in the modulating Markov chain. Changes in traffic type are represented by changes in phase, given by the generators of this Markov chain. This is one way in which traffic becomes correlated. However, in our new queue, any change in behaviour of the server that may occur when the traffic type changes can be *synchronised* since the service time distribution is controlled by the same Markov chain.

At the same time, other features can be accounted for by the negative customers, which traffic may also exhibit (synchronised) correlation. In particular, we expect to model breakdowns and cell losses that arise indirectly from buffer overflow. An initial loss can be described simply when we use a finite capacity queue, although (as we shall see) at greater computational cost. However, when one cell of a message is lost, the rest may be of no use and need to be removed. This deletion can be modelled by negative customers, viewed as "killing signals". In fact, it may be a realistic approximation to use negative customers to model initial cell loss too in an infinite queue model, avoiding the computational overhead involved in finite capacity queues.

The question remains as to how to combine the new queues together into a network. One possibility is to use the properties we derive for the departure processes of the queues. These can be combined (from different queues) and split according to the specified network topology to yield a MMCPP input stream for each queue, in general requiring an iterative procedure in a network with cycles. The parameterisation of each constituent queue would be costly without further approximation. For example, the number of phases in a union of MMCPPs is the product of the individual numbers of phases. Hence an approximation based on matching the moments and/or autocorrelations of constituent processes with those of a MMCPP with fixed dimension would seem appropriate. An alternative methodology is suggested in the conclusion of this paper.

2 System description

2.1 Modulation

The arrival and service processes are modulated by a continuous time, irreducible Markov phase process with N states. Let Q be the generator matrix of this process, given by

$$Q = \begin{bmatrix} -q_1 & q_{1,2} & \cdots & q_{1,N} \\ q_{2,1} & -q_2 & \cdots & q_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N,1} & q_{N,2} & \cdots & -q_N \end{bmatrix} ,$$

where $q_{i,k} (i \neq k)$ is the instantaneous transition rate from phase i to phase k, and

$$q_i \;=\; \sum_{j=1}^{N} q_{i,j} \quad ; \quad q_{i,i} = 0 \qquad (i = 1, \ldots, N)$$

Let $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ be the vector of equilibrium probabilities of the modulating phases. Then, $\mathbf{r}$ is uniquely determined by the equations:

$$\mathbf{r}Q = 0 \quad ; \quad \mathbf{r}\mathbf{e}_N = 1 \, .$$

where $\mathbf{e}_N$ stands for the column vector with N elements, each of which is unity.

2.2 The arrival process

The arrival process is the superposition of two CPP arrival streams in each of the modulating phases. One of these CPP processes is of (positive) customers and the other of negative customers. The parameters of the GE inter-arrival time distribution for the positive customers in phase i are (σ_i, θ_i), and (ρ_i, δ_i) are those of the negative customers. That is, the inter-arrival time probability distribution function is $1-(1-\theta_i)e^{-\sigma_i t}$, in phase i, for the positive customers and $1-(1-\delta_i)e^{-\rho_i t}$ for the negative customers. Thus, the arrival *point*-processes are Poisson, with batches arriving at each point having geometric size distribution. Specifically, the probability that a batch is of size s is $(1-\theta_i)\theta_i^{s-1}$, in phase i, for the positive customers, and $(1-\delta_i)\delta_i^{s-1}$ for the negative customers.

The overall average arrival rates of the positive customers ($\overline{\sigma}$) and negative customers ($\overline{\rho}$) can then be determined as,

$$\overline{\sigma} = \sum_{i=1}^{N} \frac{r_i \sigma_i}{1-\theta_i} \qquad ; \qquad \overline{\rho} = \sum_{i=1}^{N} \frac{r_i \rho_i}{1-\delta_i} \ .$$

2.3 The GE multi-server

The service facility has c homogeneous servers, each with GE-distributed service times with parameters (μ_i, ϕ_i) in phase i. The service discipline is FCFS and each server serves at most one positive customer at any given time. Negative customers neither wait in the queue, nor are served. The operation of the GE server is similar to that described for the CPP arrival processes above. However, the batch size associated with a service completion is bounded by one more than the number of customers waiting to commence service at the departure instant. For queues of length $c \le j < L+1$ (including any customers in service), the maximum batch size at a departure instant is $j-c+1$, only one server being able to complete a service period at any one instant under the assumption of exponentially distributed batch-service times. Thus, the probability that a departing batch has size s is $(1-\phi_i)\phi_i^{s-1}$ for $1 \le s \le j-c$ and ϕ_i^{j-c} for $s = j-c+1$. In particular, when $j = c$, the departing batch has size 1 with probability one, and this is also the case for all $1 \le j \le c$ since each customer is already engaged by a server and there are then no customers waiting to commence service.

It is assumed that the first positive customer in a batch arriving at an instant when the queue length is less than c (so that at least one server is free) *never* skips service, i.e. always has an exponentially distributed service time.

2.4 Negative customers

A negative customer removes a positive customer in the queue (or being served) according to a specified *killing discipline*. We consider here two variants of the RCE killing discipline (removal of the customer from the end of the queue), where the most recent positive arrivals are removed [16] first. The first variant, RCE-*inimmune servicing*, removes the most recent positive arrival regardless of whether it is in service or waiting; a negative arrival has no effect only when it encounters an empty queue and all servers idle. The second variant does *not* allow a customer actually in service to be removed: a negative customer that arrives when there are no positive customers waiting to start service has no effect. We say that customers in service are immune from killing and that the service itself is RCE-*immune servicing*.

The first variant is that of the traditional negative customer, suited to the modelling of killing signals in speculative parallelism, for example. It can also be used to model cell losses caused by the arrival of a corrupted cell or one encountering a full buffer, when the preceding cells of a packet would be discarded [15]. The second variant is a modification suitable for the modelling of load balancing where work is transferred from overloaded queues but never work that is actually in progress.

When a batch of negative customers of size $l\,(1 \leq l < j)$ arrives, in the first variant (customers in service not immune), l positive customers are removed leaving the remaining $j-l$ positive customers in the system. If $l \geq j$, then all the j positive customers are removed, leaving the system empty. For the immune case, the second variant, when a batch of negative customers of size $l\,(1 \leq l < j-c)$ arrives, l positive customers are removed from the end of the queue leaving the remaining $j-l$ positive customers in the system. If $l \geq j-c \geq 1$, then $j-c$ positive customers are removed, leaving none waiting to commence service (queue length equal to c). If $j \leq c$, the negative arrivals have no effect.

We also consider a further killing discipline – removal of the customer from the head of the queue (RCH) – as a third variant in section 7. Here, customers are removed from the head of the queue, i.e. the earliest arrivals go first. This discipline is suitable for modelling server breakdowns, as already noted.

2.5 The queueing capacity

L is the queueing capacity, in all phases, including the customers in service, if any. L can be finite or infinite. We assume, when the number of customers is j and the arriving batch size of positive customers is greater than $L - j$ (assuming finite L), that only $L - j$ customers are taken in and the rest are rejected.

2.6 Condition for stability

When L is finite, the system is ergodic since the representing Markov process is irreducible. Otherwise, when $L = \infty$, the overall average departure rate increases with the queue length and its maximum (that is, when $J(t) \to \infty$) can be determined as,

$$\overline{\mu} = c \sum_{i=1}^{N} \frac{r_i \mu_i}{1 - \phi_i}. \tag{1}$$

Hence, we conjecture the necessary and sufficient condition for the existence of steady state probabilities is

$$\overline{\sigma} < \overline{\rho} + \overline{\mu} \, . \tag{2}$$

which is equivalent to,

$$\begin{aligned} \mathbf{r} \left[\sum_{s=1}^{\infty} s \Theta^{s-1} (E - \Theta) \Sigma \right] \mathbf{e}_N \quad < \quad & \mathbf{r} \left[\sum_{s=1}^{\infty} s \Phi^{s-1} (E - \Phi) C \right] \mathbf{e}_N \\ & + \quad \mathbf{r} \left[\sum_{s=1}^{\infty} s \Delta^{s-1} (E - \Delta) R \right] \mathbf{e}_N \end{aligned} \tag{3}$$

where, $\Theta, \Phi, \Delta, R, E$ are defined in section 3. This stability condition is valid for the RCE variants 1 and 2, not for the RCH discipline, which is dealt with in section 7.

3 The steady state balance equations

The state of the system at any time t can be specified completely by two integer-valued random variables, $I(t)$ and $J(t)$. $I(t)$ varies from 1 to N, representing the phase of the modulating Markov chain, and $0 \leq J(t) < L + 1$ represents the number of positive customers in the system at time t, including any in service. The system is now modelled by a continuous time discrete state Markov process, $\overline{Y}$ (Y if L is infinite), on a rectangular lattice strip. Let $I(t)$, the phase, vary in the horizontal direction and $J(t)$, the queue length or *level*, in the vertical direction. We denote the steady state probabilities, when they exist, by $\{p_{i,j} \mid 1 \leq i \leq N, 0 \leq j < L+1\}$, where $p_{i,j} = \lim_{t\to\infty} \text{Prob}(I(t) = i, J(t) = j)$, and we write $\mathbf{v}_j = (p_{1,j}, \ldots, p_{N,j})$.

In this section, only the first variant[1] of the RCE killing discipline is considered; other variants are considered separately in later sections. Initially we assume L is finite and later extend the analysis to the case of the process Y. The process $\overline{Y}$ evolves with the following instantaneous transition rates:

(a) $q_{i,k}$ – purely lateral transition rate – from state (i, j) to state (k, j), for all $j \geq 0$ and $1 \leq i, k \leq N$ $(i \neq k)$, caused by a phase transition in the Markov chain governing the arrival phase process;

(b) $B_{i,j,j+s}$ – s-step upward transition rate – from state (i, j) to state $(i, j+s)$, for all phases i, caused by a new batch arrival of size s positive customers. For a given j, s is bounded when L is finite and unbounded when L is infinite.

The s-step downward transition rate in phase i, denoted by $C_{i,j+s,j}$ $(0 \leq j \leq L-1\,,\ 1 \leq s \leq L-j)$, is more involved and is split into the following cases:

(c) $C_{i,j+s,j}$ – s-step downward transition rate above the threshold – from state $(i, j+s)$ to state (i, j), for $c \leq j \leq L-1\,;\ 1 \leq s \leq L-j$ and all phases i, caused by either a batch service completion of size s or a batch arrival of negative customers of size s;

(d) $C_{i,c-1+s,c-1}$ – bounded s-step downward transition rate through the threshold – from state $(i, c-1+s)$ to state $(i, c-1)$, for $c \geq 2$ and for all phases i, caused by a batch arrival of negative customers of size s or a batch service completion of size $s\,(1 \leq s \leq L-c+1)$;

(e) $C_{i,s,0}$ – bounded s-step downward transition rate to the empty queue – from state (i, s) to state $(i, 0)$, for all phases i, $s \geq 1$ and $c = 1$, caused by a batch service completion of size s or a batch arrival of negative customers of size $\geq s$;

[1]This steady state solution is applicable for the more general case in which customers in the queue can be killed by negative customers in any order.

(f) $C_{i,j+1,j}$ – bounded 1-step downward transition rate below the threshold – from state $(i, j+1)$ to state (i, j), for all phases i and $1 \leq j \leq c-2$, $c > 1$, caused by a single service completion or the arrival of a single negative customer;

(g) $C_{i,j+s,j}$ – bounded s-step downward transition rate through the threshold caused by negative arrivals – from state $(i, j+s)$ to state (i, j), for all phases i and $1 \leq j \leq c-2$, $c \geq 3$, $s \geq 2$, caused by the batch arrival of negative customers of size $s \geq 2$ (mutually exclusive with case (d));

(h) $C_{i,s,0}$ – bounded s-step downward transition rate to the empty queue caused by negative arrivals – from state (i, s) to state $(i, 0)$, for all phases i and $c \geq 2$, $s \geq 2$, caused by a batch arrival of negative customers of size $\geq s$;

(i) $C_{i,1,0}$ – bounded 1-step downward transition rate to the empty queue – from state $(i, 1)$ to state $(i, 0)$, for all phases i and $c \geq 2$, caused by a single departure or a batch arrival of negative customers of size ≥ 1.

The $B_{\cdot,\cdot,\cdot}$ and $C_{\cdot,\cdot,\cdot}$ terms respectfully denote batch arrival rates and batch departure rates (caused by either service completions or negative arrivals). They are therefore defined as follows, consistent with the above cases:

$$
\begin{aligned}
B_{i,j,j+s} &= (1-\theta_i)\theta_i^{s-1}\sigma_i \qquad (\forall i\ ;\ 0 \leq j \leq L-2\ ;\ j < j+s < L) \\
B_{i,j-s,j} &= (1-\theta_i)\theta_i^{s-1}\sigma_i \qquad (\forall i\ ;\ 0 \leq j-s \leq L-2\ ;\ j-s < j < L) \\
B_{i,j,L} &= \sum_{s=L-j}^{\infty} (1-\theta_i)\theta_i^{s-1}\sigma_i = \theta_i^{L-j-1}\sigma_i \qquad (\forall i\ ;\ j \leq L-1) \\
C_{i,j+s,j} &= (1-\phi_i)\phi_i^{s-1}c\mu_i + (1-\delta_i)\delta_i^{s-1}\rho_i \\
&\qquad\qquad (\forall i\ ;\ c \leq j \leq L-1\ ;\ 1 \leq s \leq L-j) \\
&= \phi_i^{s-1}c\mu_i + (1-\delta_i)\delta_i^{s-1}\rho_i \\
&\qquad\qquad (\forall i\ ;\ c > 1\ ;\ j = c-1\ ;\ 1 \leq s \leq L-c+1) \\
&= \phi_i^{s-1}\mu_i + \delta_i^{s-1}\rho_i \\
&\qquad\qquad (\forall i\ ;\ c = 1\ ;\ j = c-1 = 0\ ;\ 1 \leq s \leq L) \\
&= (j+1)\mu_i + (1-\delta_i)\rho_i \qquad (\forall i\ ;\ 1 \leq j \leq c-2\ ;\ s = 1) \\
&= (1-\delta_i)\delta_i^{s-1}\rho_i \qquad (\forall i\ ;\ 1 \leq j \leq c-2\ ;\ 2 \leq s \leq L-j) \\
&= \delta_i^{s-1}\rho_i \qquad (\forall i\ ;\ c > 1\ ;\ j = 0\ ;\ 2 \leq s \leq L)
\end{aligned}
$$

$$= \quad \mu_i + \rho_i \qquad (\forall i \; ; \; j = 0 \; ; \; s = 1)$$

For a conciser notation, we define the following matrices (where E is the unit $N \times N$ matrix):

$$\begin{aligned}
B_{j-s,j} &= \text{Diag}\left[(1-\theta_1)\theta_1^{s-1}\sigma_1, (1-\theta_2)\theta_2^{s-1}\sigma_2, \ldots, (1-\theta_N)\theta_N^{s-1}\sigma_N\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad (j < L \; ; \; s = 1, 2, \ldots, j) \\
&= \text{Diag}\left[\theta_1^{s-1}\sigma_1, \theta_2^{s-1}\sigma_2, \ldots, \theta_N^{s-1}\sigma_N\right] \qquad (j = L \; ; \; s = 1, 2, \ldots, L) \\
B_s &= B_{j-s,j} \qquad (j < L) \\
B &= B_1 = (E - \Theta)\Sigma \\
\Sigma &= \text{Diag}\left[\sigma_1, \sigma_2, \ldots, \sigma_N\right] \\
\Theta &= \text{Diag}\left[\theta_1, \theta_2, \ldots, \theta_N\right] \\
R &= \text{Diag}\left[\rho_1, \rho_2, \ldots, \rho_N\right] \\
\Delta &= \text{Diag}\left[\delta_1, \delta_2, \ldots, \delta_N\right] \\
M &= \text{Diag}\left[\mu_1, \mu_2, \ldots, \mu_N\right] \\
\Phi &= \text{Diag}\left[\phi_1, \phi_2, \ldots, \phi_N\right] \\
C_j &= jM \qquad (1 \le j \le c) \\
&= cM = C \qquad (j \ge c) \\
C_{j+s,j} &= \text{Diag}\left[C_{1,j+s,j}, C_{2,j+s,j}, \ldots, C_{N,j+s,j}\right]
\end{aligned}$$

Then, we get,

$$\begin{aligned}
B_s &= \Theta^{s-1}B = \Theta^{s-1}(E - \Theta)\Sigma \\
\sum_{s=1}^{\infty} B_s &= \Sigma \\
B_{L-s,L} &= \Theta^{s-1}\Sigma \\
C_{j+s,j} &= C(E - \Phi)\Phi^{s-1} + R(E - \Delta)\Delta^{s-1} \\
&\qquad (c \le j \le L-1 \; ; \; s = 1, 2, \ldots, L-j) \\
&= C\Phi^{s-1} + R(E - \Delta)\Delta^{s-1} \qquad (c > 1 \; ; \; j = c-1 \; ; \; 1 \le s \le L-c+1) \\
&= C\Phi^{s-1} + R\Delta^{s-1} \qquad (c = 1 \; ; \; j = c-1 = 0 \; ; 1 \le s \le L) \\
&= C_{j+1} + R(E - \Delta) \qquad (1 \le j \le c-2 \; ; \; s = 1) \\
&= R(E - \Delta)\Delta^{s-1} \qquad (1 \le j \le c-2 \; ; \; 2 \le s \le L-j) \\
&= R\Delta^{s-1} \qquad (c > 1 \; ; \; j = 0 \; ; \; 2 \le s \le L) \\
&= C_1 + R \qquad (j = 0 \; ; \; s = 1)
\end{aligned}$$

Note that the Markov process Y (or $\overline{Y}$) is neither of the M/G/1 type nor of the GI/M/1 type. Hence, existing standard solution methods, such as the spectral expansion method (SEM) or the matrix geometric method (MGM), cannot be directly applied for solving the balance equations. These methods for two-dimensional Markov chains require *bounded* vertical transitions. Because of the (unbounded) nature of the geometric batch sizes, we do not have this property. In the case of finite L, the vertical dimension of the whole process is finite, of course, but for non-trivial problems L is so large as to make the above methods intractable.

3.1 Transformed balance equations

The key result of section 3 is the transformation of the raw balance equations (with unbounded vertical transitions) into a set of equations in the form of a bounded recurrence relation amenable to SEM and MGM, together with a set of boundary conditions that yield the unique steady state solution.

Proposition 1 *The balance equations of the* Markov modulated CPP/GE/c/L queue *with positive and negative customers (MM CPP/GE/c/L G-queue), with RCE-inimmune servicing, may be transformed into the following equations (for $L-3>c$)*[2]*:*

(i) For $c+1 \leq j \leq L-3$,

$$\mathbf{v}_{j-1}Q_0 + \mathbf{v}_jQ_1 + \mathbf{v}_{j+1}Q_2 + \mathbf{v}_{j+2}Q_3 = 0 \; .$$

(ii) For $j = L, L-1, L-2$,

$$\sum_{s=1}^{L} \mathbf{v}_{L-s}\Theta^{s-1}\Sigma + \mathbf{v}_L(Q-C-R) = 0 \; ;$$

$$\begin{aligned} \mathbf{v}_{L-2}\,[\Sigma - (Q-C-R)\Theta] \\ +\mathbf{v}_{L-1}\,[Q - \Sigma - C(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Delta)] \\ +\mathbf{v}_L\,[C(E-\Phi)(E-\Theta\Phi) + R(E-\Delta)(E-\Theta\Delta)] &= 0 \end{aligned}$$

[2]If $L \leq c+3$, then the resulting Markov process with $(L+1)N$ states can be solved quite easily using a direct method [29].

$$
\begin{aligned}
\mathbf{v}_{L-3}\left[\Sigma-(Q-C-R)\Theta\right] & \\
+\mathbf{v}_{L-2}\left[Q(E+\Theta\Phi)-\Sigma(E+\Phi)-C(E+\Theta)\right. & \\
\left.-R(E+\Theta\Phi+\Theta-\Theta\Delta)\right] & \\
+\mathbf{v}_{L-1}\left[-Q\Phi+\Sigma\Phi+C+R(E+(\Phi-\Delta)(E+\Theta-\Theta\Delta))\right] & \\
+\mathbf{v}_{L}\left[R(E-\Delta)(E-\Theta\Delta)(\Delta-\Phi)\right] & = 0
\end{aligned}
$$

(iii) For $j=c, c-1, c-2$ $(c \geq 2)$,

$$
\begin{aligned}
\mathbf{v}_{c-1}\left[\Sigma-(Q-C_{c-1}-R)\Theta\right] & \\
+\mathbf{v}_{c}\left[Q-\Sigma-C(E+\Theta)-R(E+\Theta-\Theta\Delta)\right] & \\
+\sum_{s=1}^{L-c}\mathbf{v}_{c+s}\left[C\Phi^{s-1}(E-\Phi-\Theta\Phi)+R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] & = 0 \\
(for\ c \geq 2) &
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}_{c-2}\left[\Sigma-(Q-C_{c-2}-R)\Theta\right] & \\
+\mathbf{v}_{c-1}\left[Q-\Sigma-C_{c-1}(E+\Theta)-R(E+\Theta-\Theta\Delta)\right] & \\
+\sum_{s=1}^{L-c+1}\mathbf{v}_{c-1+s}\left[C\Phi^{s-1}+R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] & = 0 \\
(for\ c \geq 3) &
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}_{c-3}\left[\Sigma-(Q-C_{c-3}-R)\Theta\right] & \\
+\mathbf{v}_{c-2}\left[Q-\Sigma-C_{c-2}(E+\Theta)-R(E+\Theta-\Theta\Delta)\right] & \\
+\mathbf{v}_{c-1}\left[C_{c-1}+R(E-\Delta)(E-\Theta\Delta)\right] & \\
+\sum_{s=2}^{L-c+2}\mathbf{v}_{c-2+s}\left[R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] & = 0 \\
(for\ c \geq 4) &
\end{aligned}
$$

(iv) For $2 \leq j \leq c-3$,

$$
\mathbf{v}_{j-1}Q_{0,j}+\mathbf{v}_{j}Q_{1,j}+\mathbf{v}_{j+1}Q_{2,j}+\mathbf{v}_{j+2}Q_{3,j}=0\ .
$$

(v) For $j=1,0$

$$
\begin{aligned}
\mathbf{v}_{0}\left[\Sigma-Q\Theta\right]+\mathbf{v}_{1}\left[Q-\Sigma-(C_{1}+R)(E+\Theta)\right] & \\
+\sum_{s=1}^{L-1}\mathbf{v}_{1+s}\left[C_{1+s,1}-C_{1+s,0}\Theta\right] & = 0
\end{aligned}
$$

$$\mathbf{v}_0\left[Q-\Sigma\right]+\mathbf{v}_1\left[C_1+R\right]+\sum_{s=2}^{L}\mathbf{v}_s\left[C\Phi^{s-1}(1-I_{c-1>0})+R\Delta^{s-1}\right]=0$$

(vi) Normalisation

$$\sum_{j=0}^{L}\mathbf{v}_j\mathbf{e}_N=1\ .$$

where,

$$\begin{aligned}
Q_{0,j} &= \Sigma-(Q-C_{j-1}-R)\Theta \\
Q_{1,j} &= Q(E+\Theta\Delta)-\Sigma(E+\Delta)-C_j(E+\Theta+\Theta\Delta)-R(E+\Theta) \\
Q_{2,j} &= -Q\Delta+\Sigma\Delta+C_{j+1}(E+\Delta+\Theta\Delta)+R \\
Q_{3,j} &= -C_{j+2}\Delta \\
Q_0 &= \Sigma-(Q-C-R)\Theta \\
Q_1 &= Q(E+\Theta\Phi+\Theta\Delta)-\Sigma(E+\Delta+\Phi)-C(E+\Theta+\Theta\Delta) \\
&\quad -R(E+\Theta+\Theta\Phi) \\
Q_2 &= -Q(\Phi+\Delta+\Theta\Phi\Delta)+\Sigma(\Phi+\Delta+\Delta\Phi)+C(E+\Delta+\Theta\Delta) \\
&\quad +R(E+\Phi+\Theta\Phi) \\
Q_3 &= Q\Phi\Delta-\Sigma\Phi\Delta-C\Delta-R\Phi \\
C_{1+s,0} &= C\Phi^s+R\Delta^s \qquad (c=1) \\
&= R\Delta^s \qquad (c>1) \\
C_{1+s,1} &= C(E-\Phi)\Phi^{s-1}+R(E-\Delta)\Delta^{s-1} \qquad (c=1) \\
&= C\Phi^{s-1}+R(E-\Delta)\Delta^{s-1} \qquad (c=2) \\
&= C_2+R(E-\Delta) \qquad (c\geq 3\ ;\ s=1) \\
&= R(E-\Delta)\Delta^{s-1} \qquad (c\geq 3\ ;\ s\geq 2) \\
I_{k>0} &= 1 \qquad (k>0) \\
&= 0 \qquad (k\leq 0)
\end{aligned}$$

Proof

The balance equations are:

$$\sum_{s=1}^{L}\mathbf{v}_{L-s}B_{L-s,L}+\mathbf{v}_L(Q-C-R)=0 \qquad \text{for the } L^{th} \text{ row,} \tag{4}$$

$$\sum_{s=1}^{j} \mathbf{v}_{j-s} B_s + \mathbf{v}_j \left[Q - \Sigma - C_j - R I_{j>0} \right] + \sum_{s=1}^{L-j} \mathbf{v}_{j+s} C_{j+s,j} = 0 \qquad (5)$$

$$(0 \leq j \leq L-1)$$

Substituting $B_{L-s,L} = \Theta^{s-1}\Sigma$, the balance equation for the L^{th} row becomes,

$$\sum_{s=1}^{L} \mathbf{v}_{L-s} \Theta^{s-1} \Sigma + \mathbf{v}_L (Q - C - R) = 0 \; . \qquad (6)$$

Substituting $B_s = \Theta^{s-1}(E - \Theta)\Sigma$, the balance equations for the levels j and $j-1$, $(0 \leq j-1, j \leq L-1)$ respectively are:

$$\sum_{s=1}^{j} \mathbf{v}_{j-s} \Theta^{s-1} (E - \Theta) \Sigma + \mathbf{v}_j \left[Q - \Sigma - C_j - R I_{j>0} \right] + \sum_{s=1}^{L-j} \mathbf{v}_{j+s} C_{j+s,j} = 0 \qquad (7)$$

$$(0 \leq j \leq L-1)$$

$$\sum_{s=1}^{j-1} \mathbf{v}_{j-1-s} \Theta^{s-1} (E - \Theta) \Sigma + \mathbf{v}_{j-1} \left[Q - \Sigma - C_{j-1} - R I_{j-1>0} \right] \qquad (8)$$

$$+ \sum_{s=1}^{L-j+1} \mathbf{v}_{j-1+s} C_{j-1+s,j-1} \;\; = \;\; 0$$

$$(0 \leq j-1 \leq L-1 \text{ or } 1 \leq j \leq L)$$

These equations have a similar structure due to the geometric distribution of batch size; many terms in equation 8, if multiplied by Θ, also appear in equation 7. Hence we modify the balance equation for level j as follows: post-multiply (8) by Θ and subtract the resulting equation from (7) to get the modified equation for level j $(1 \leq j \leq L-1)$ as:

$$\mathbf{v}_{j-1} \left[\Sigma - (Q - C_{j-1} - R I_{j-1>0}) \Theta \right] + \mathbf{v}_j \left[Q - \Sigma - C_j - R I_{j>0} - C_{j,j-1} \Theta \right]$$

$$+ \sum_{s=1}^{L-j} \mathbf{v}_{j+s} \left[C_{j+s,j} - C_{j+s,j-1} \Theta \right] = 0 \qquad (1 \leq j \leq L-1) \quad (9)$$

This same idea, similar to the method of summing geometric series used in high school, is applied repeatedly for each of the geometric distributions corresponding to positive arrivals (just done above), service completions and negative arrivals. Positive arrivals are unconstrained below the capacity L, but the threshold and empty queue states place restrictions on the negative arrival and service completion processes. We therefore consider five cases separately.

Case 1: $c+1 \le j \le L-3$

Consider the balance equations for the levels, $c+1 \le j \le L-1$. For this range, we have

$$C_{j-1} = C \; ; \; C_j = C$$

$$\begin{aligned} C_{j,j-1} &= C(E-\Phi) + R(E-\Delta) \\ &\quad (\text{for } c \le j-1 \le L-1\ , \text{ hence for } c+1 \le j \le L-1) \end{aligned}$$

$$\begin{aligned} C_{j+s,j} &= C(E-\Phi)\Phi^{s-1} + R(E-\Delta)\Delta^{s-1} \\ &\quad (\text{for } c \le j \le L-1\ , \text{ hence for } c+1 \le j \le L-1) \end{aligned}$$

$$\begin{aligned} C_{j+s,j-1} &= C(E-\Phi)\Phi^{s} + R(E-\Delta)\Delta^{s} \\ &\quad (\text{for } c \le j-1 \le L-1\ , \text{ hence for } c+1 \le j \le L-1) \end{aligned}$$

Substituting the above in (9), the balance equation for level j becomes,

$$\begin{aligned} &\mathbf{v}_{j-1}\left[\Sigma - (Q-C-R)\Theta\right] \\ &+\mathbf{v}_j\left[Q - \Sigma - C(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Delta)\right] \\ &+\sum_{s=1}^{L-j}\mathbf{v}_{j+s}\left[C(E-\Phi)\Phi^{s-1}(E-\Theta\Phi) + R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] = 0 \\ &(c+1 \le j \le L-1) \end{aligned} \tag{10}$$

Using the above, the balance equation for the level $j+1$ ($c+1 \le j+1 \le L-1$) is obtained by replacing j by $j+1$ in (10), as

$$\begin{aligned} &\mathbf{v}_{j}\left[\Sigma - (Q-C-R)\Theta\right] \\ &+\mathbf{v}_{j+1}\left[Q - \Sigma - C(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Delta)\right] \\ &+\sum_{s=1}^{L-j-1}\mathbf{v}_{j+1+s}\left[C(E-\Phi)\Phi^{s-1}(E-\Theta\Phi) + R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] = 0 \\ &(c+1 \le j+1 \le L-1 \text{ or } c \le j \le L-2) \end{aligned} \tag{11}$$

As above with equation (7) and Θ, we further modify the balance equation (10) for level j by subtracting from it equation (11) post-multiplied by Φ, to get

$$\begin{aligned}
\mathbf{v}_{j-1}\left[\Sigma-(Q-C-R)\Theta\right] & \qquad (12)\\
+\mathbf{v}_j\left[Q(E+\Theta\Phi)-\Sigma(E+\Phi)-C(E+\Theta)-R(E+\Theta+\Theta\Phi-\Theta\Delta)\right] &\\
+\mathbf{v}_{j+1}\left[-Q\Phi+\Sigma\Phi+C+R(E+(\Phi-\Delta)(E+\Theta-\Theta\Delta))\right] &\\
+\sum_{s=2}^{L-j}\mathbf{v}_{j+s}\left[R(E-\Delta)\Delta^{s-2}(E-\Theta\Delta)(\Delta-\Phi)\right] &= 0\\
(c+1\le j\le L-2) &
\end{aligned}$$

Using the above equation, the modified equation pertaining to the level $j+1\,(c+1\le j+1\le L-2)$ is obtained by replacing j by $j+1$ in (12), giving

$$\begin{aligned}
\mathbf{v}_{j}\left[\Sigma-(Q-C-R)\Theta\right] & \qquad (13)\\
+\mathbf{v}_{j+1}\left[Q(E+\Theta\Phi)-\Sigma(E+\Phi)-C(E+\Theta)-R(E+\Theta+\Theta\Phi-\Theta\Delta)\right] &\\
+\mathbf{v}_{j+2}\left[-Q\Phi+\Sigma\Phi+C+R(E+(\Phi-\Delta)(E+\Theta-\Theta\Delta))\right] &\\
+\sum_{s=2}^{L-j-1}\mathbf{v}_{j+1+s}\left[R(E-\Delta)\Delta^{s-2}(E-\Theta\Delta)(\Delta-\Phi)\right] &= 0\\
(\text{for } c+1\le j+1\le L-2 \text{ or } c\le j\le L-3) &
\end{aligned}$$

Further, subtracting equation (13) post-multiplied by Δ from equation (12), we get

$$\mathbf{v}_{j-1}Q_0+\mathbf{v}_jQ_1+\mathbf{v}_{j+1}Q_2+\mathbf{v}_{j+2}Q_3=0 \quad (c+1\le j\le L-3) \qquad (14)$$

where,

$$\begin{aligned}
Q_0 &= \Sigma-(Q-C-R)\Theta\\
Q_1 &= Q(E+\Theta\Phi+\Theta\Delta)-\Sigma(E+\Delta+\Phi)-C(E+\Theta+\Theta\Delta)\\
&\quad -R(E+\Theta+\Theta\Phi)\\
Q_2 &= -Q(\Phi+\Delta+\Theta\Phi\Delta)+\Sigma(\Phi+\Delta+\Phi\Delta)+C(E+\Delta+\Theta\Delta)\\
&\quad +R(E+\Phi+\Theta\Phi)\\
Q_3 &= Q\Phi\Delta-\Sigma\Phi\Delta-C\Delta-R\Phi
\end{aligned}$$

Notice that the coefficient matrices Q_0, Q_1, Q_2, Q_3 are *independent* of j; we say this range of levels j constitutes the *repeating case.*

The other (non-repeating) cases, corresponding to balance equations for level j near to the upper bound L, near to the threshold c, below the threshold and near to 0 (empty queue), are similar and considered in the Appendix.

Finally, since the sum of the steady state probabilities of all the states is 1, we have the normalising equation

$$\sum_{j=0}^{L} \mathbf{v}_j \mathbf{e}_N = 1 \tag{15}$$

3.2 Illustrative examples

For illustration, we now consider three special cases when the problem reduces to well known queues:

1. The M/M/1 G-queue with no modulation and no batches;
2. The MMCPP/M/1 queue with no negative customers and no batches;
3. The M/M/1 G-queue with batch arrivals

Here, we derive the transformed balance equations only but will observe that they are the same as the usual raw balance equations in examples 1 and 2. The steady state probabilities will be verified after the description of our SEM algorithm in the next section. Since we will only be considering the levels above the threshold (the repeating case), all servers are busy and we take $c = 1$ in each example without loss of generality. We do not consider the boundary at L either – equivalently we assume $L = \infty$.

3.2.1 M/M/1 G-queue

In this example, there is no modulation, so $Q = 0, N = 1$, there are no batches of more than one customer, so $\Theta = (\theta_1) = \Phi = (\phi_1) = \Delta = (\delta_1) = 0$ and $c = 1$. Proposition 1 then yields (abbreviating σ_1, μ_1, ρ_1 by σ, μ, ρ respectively):

$$\begin{aligned} Q_0 &= \sigma \\ Q_1 &= -(\sigma + \mu + \rho) \\ Q_2 &= \mu + \rho \\ Q_3 &= 0 \end{aligned}$$

This gives the balance equations

$$p_{j-1}\sigma - p_j(\sigma + \mu + \rho) + p_{j+1}(\mu + \rho) = 0$$

for $j \geq 2$. These are precisely the balance equations of the M/M/1 G-queue for states $j > 0$. In this example (only), we also look at the balance equation corresponding to the empty queue. Case (v) of proposition 1 gives:

$$-p_0\sigma + p_1(\mu + \rho) = 0$$

again as expected. Thus we observe that the negative customers effectively just increase the rate of the server to $\mu + \rho$ as far as the equilibrium queue length probabilities are concerned.

3.2.2 MMCPP/M/1 queue

This time we have modulation, so $Q \neq 0, N > 1$, but again there are no batches of more than one customer, so $\Theta = \Phi = \Delta = 0$, and $c = 1$. Proposition 1 then yields:

$$\begin{aligned} Q_0 &= \Sigma \\ Q_1 &= Q - \Sigma - C - R \\ Q_2 &= C + R \\ Q_3 &= 0 \end{aligned}$$

This gives the balance equation

$$\mathbf{v}_{j-1}\Sigma + \mathbf{v}_j(Q - \Sigma - C - R) + \mathbf{v}_{j+1}(C + R) = 0$$

for $j \geq 2$. Again, this is well known for the MMCPP/M/1 queue.

3.2.3 M/M/1 G-queue with batch arrivals

This example is the same as the first except that $\theta_1 \neq 0$ or $\delta_1 \neq 0$. We obtain, dropping the suffix $_1$ as above,

$$\begin{aligned} Q_0 &= \sigma + (\mu + \rho)\theta \\ Q_1 &= -\sigma(1 + \delta) - \mu(1 + \theta + \theta\delta) - \rho(1 + \theta) \\ Q_2 &= \sigma\delta + \mu(1 + \delta + \theta\delta) + \rho \\ Q_3 &= -\mu\delta \end{aligned}$$

Note that $Q_3 \neq 0$, a necessary condition for which is that there must be negative arrivals with non-unit batch size, i.e. $\delta > 0$. The equations resulting from proposition 1 are certainly not the balance equations of the original problem, because of the unbounded transitions in the queue length, and little more can be said until we investigate their solution in the next section.

4 Spectral expansion solution of the balance equations

The set of equations (14) concerning the levels $c+1$ to $L-3$ have coefficient matrices Q_0, Q_1, Q_2, Q_3 that are independent of j and hence have an efficient solution by the matrix geometric method [26] or the spectral expansion method [24, 4, 23]. We opt for the latter and consider the former in section 5. This is not to say that either method is inherently superior in any way. Indeed, we shall see that the two are closely related.

Define the matrix polynomials $Z(\lambda)$ and $\overline{Z}(\xi)$ as,

$$\begin{aligned} Z(\lambda) &= Q_0 + Q_1\lambda + Q_2\lambda^2 + Q_3\lambda^3 \,, &(16)\\ \overline{Z}(\xi) &= Q_3 + Q_2\xi + Q_1\xi^2 + Q_0\xi^3 \,. &(17) \end{aligned}$$

Then, the spectral expansion solution for $\mathbf{v}_j$ $(c \le j \le L-1)$ is given by,

$$\mathbf{v}_j = \sum_{k=1}^{N} a_k \boldsymbol{\psi}_k \lambda_k^{j-c} + \sum_{k=1}^{2N} b_k \boldsymbol{\gamma}_k \xi_k^{L-1-j} \qquad (c \le j \le L-1) \tag{18}$$

where λ_k $(k = 1, 2, \ldots, N)$ are the N eigenvalues of least absolute value of the matrix polynomial $Z(\lambda)$ and ξ_k $(k = 1, 2, \ldots, 2N)$ are the $2N$ eigenvalues of least absolute value of the matrix polynomial $\overline{Z}(\xi)$. $\boldsymbol{\psi}_k$ and $\boldsymbol{\gamma}_k$ are the left-eigenvectors of $Z(\lambda)$ and $\overline{Z}(\xi)$ respectively, corresponding to the eigenvalues λ_k and ξ_k respectively. a_k, b_k are arbitrary constants to be determined later.

Observe that the matrix $Q_0 + Q_1 + Q_2 + Q_3$ is singular, so that $\lambda = 1$ is an eigenvalue on the unit-circle for both $Z(\lambda)$ and $\overline{Z}(\xi)$. If condition (2) of section 2.6 is satisfied, the number of eigenvalues of $Z(\lambda)$ that are strictly within the unit circle is N. If (2) is not satisfied, that number is $N-1$. The properties of these eigenvalues, eigenvectors and the relevant spectral analysis are dealt with in [4, 24]. Some of them are summarised below. Let the rank of Q_0 be $N - n_0$ and that of Q_3 be $N - n_3$. Then,

(a) $Z(\lambda)$ would have $d = 3N - n_3$ eigenvalues of which n_0 are zero eigenvalues, whereas $\overline{Z}(\xi)$ would have n_3 zero eigenvalues and $3N - n_0 - n_3$ non-zero eigenvalues.

(b) If $(\lambda \neq 0, \boldsymbol{\psi})$ is a non-zero eigenvalue-eigenvector pair of $Z(\lambda)$, then there exists a corresponding non-zero eigenvalue-eigenvector pair, $(\xi = \frac{1}{\lambda}, \boldsymbol{\gamma} = \boldsymbol{\phi})$ for $\overline{Z}(\xi)$. Thus, the non-zero eigenvalues of these two matrix polynomials are mutually reciprocal.

(c) The N eigenvalues of least absolute value of $Z(\lambda)$ and the $2N$ eigenvalues of least absolute value of $\overline{Z}(\xi)$ lie either strictly inside, or on, their respective unit-circles, but not outside.

(d) There is no problem posed by multiple eigenvalues, i.e. independent eigenvectors having coincident eigenvalues, since each pair $(\lambda, \boldsymbol{\psi})$ is *distinct.*

If the unknowns a_k, b_k could be determined in such a way that all the balance equations were satisfied, then the vectors $\mathbf{v}_j$ would be the unique steady state solution. Using the spectral expansion solution for (18), the vectors $\mathbf{v}_j$ $(j = L-3, L-2, L-1)$ are already expressed as linear sums of known vectors with the unknown scalar coefficients a_k, b_k. Using (44), $\mathbf{v}_L$ can also be expressed in the same way as:

$$\mathbf{v}_L = \mathbf{y}_L \left[R(E-\Delta)(E-\Theta\Delta)(\Delta-\Phi)\right]^{-1} \tag{19}$$

where,

$$\begin{aligned}
\mathbf{y}_L \;=\; & -\mathbf{v}_{L-3}\left[\Sigma-(Q-C-R)\Theta\right] \\
& -\mathbf{v}_{L-2}\left[Q(E+\Theta\Phi)-\Sigma(E+\Phi)-C(E+\Theta)-R(E+\Theta+\Theta\Phi-\Theta\Delta)\right] \\
& -\mathbf{v}_{L-1}\left[-Q\Phi+\Sigma\Phi+C+R(E+(\Phi-\Delta)(E+\Theta-\Theta\Delta))\right]
\end{aligned}$$

Now that each $\mathbf{v}_j$ $(c \le j \le L)$ is expressed as a linear sum of known vectors, with a_k, b_k as unknown coefficients, using equations (45,18), we can also express $\mathbf{v}_{c-1}$ as such a linear sum:

$$\mathbf{v}_{c-1} = \mathbf{y}_{c-1}\left[\Sigma-(Q-C_{c-1}-R)\Theta\right]^{-1} \tag{20}$$

where

$$\begin{aligned}
\mathbf{y}_{c-1} \;=\; & -\mathbf{v}_c\left[Q-\Sigma-C(E+\Theta\Phi)-R(E+\Theta-\Theta\Delta)\right] \\
& -\sum_{s=1}^{L-c-1}\left[\sum_{k=1}^{N} a_k\boldsymbol{\psi}_k\lambda_k^s+\sum_{k=1}^{2N} b_k\boldsymbol{\gamma}_k\xi_k^{L-1-c-s}\right] \\
& \qquad \left[C\Phi^{s-1}(E-\Phi-\Theta\Phi)+R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right] \\
& -\mathbf{v}_L\left[C\Phi^{L-c-1}(E-\Phi-\Theta\Phi)+R(E-\Delta)\Delta^{L-c-1}(E-\Theta\Delta)\right]
\end{aligned} \tag{21}$$

The middle term in the above sum on the right hand side can be simplified as,

$$\begin{aligned} &-\sum_{k=1}^{N} a_k \boldsymbol{\psi}_k \left[C\Omega_{1,k,c-1}(E-\Phi-\Theta\Phi) + R\Omega_{2,k,c-1}(E-\Delta)(E-\Theta\Delta) \right] \\ &-\sum_{k=1}^{2N} b_k \boldsymbol{\gamma}_k \left[C\Omega_{3,k,c-1}(E-\Phi-\Theta\Phi) + R\Omega_{4,k,c-1}(E-\Delta)(E-\Theta\Delta) \right] \end{aligned} \quad (22)$$

where,

$$\begin{aligned} \Omega_{1,k,c-1} &= \text{Diag}\left[\frac{\lambda_k - \lambda_k^{L-c}\phi_1^{L-c-1}}{1-\lambda_k\phi_1}, \ldots, \frac{\lambda_k - \lambda_k^{L-c}\phi_N^{L-c-1}}{1-\lambda_k\phi_N} \right] \\ \Omega_{2,k,c-1} &= \text{Diag}\left[\frac{\lambda_k - \lambda_k^{L-c}\delta_1^{L-c-1}}{1-\lambda_k\delta_1}, \ldots, \frac{\lambda_k - \lambda_k^{L-c}\delta_N^{L-c-1}}{1-\lambda_k\delta_N} \right] \\ \Omega_{3,k,c-1} &= \text{Diag}\left[\frac{\xi_k^{L-c-2} - \xi_k^{-1}\phi_1^{L-c-1}}{1-\frac{\phi_1}{\xi_k}}, \ldots, \frac{\xi_k^{L-c-2} - \xi_k^{-1}\phi_N^{L-c-1}}{1-\frac{\phi_N}{\xi_k}} \right] \\ \Omega_{4,k,c-1} &= \text{Diag}\left[\frac{\xi_k^{L-c-2} - \xi_k^{-1}\delta_1^{L-c-1}}{1-\frac{\delta_1}{\xi_k}}, \ldots, \frac{\xi_k^{L-c-2} - \xi_k^{-1}\delta_N^{L-c-1}}{1-\frac{\delta_N}{\xi_k}} \right] \end{aligned}$$

The above can be computed quite easily even for large or infinite L.

Following a similar procedure as above, $\mathbf{v}_{c-2}, \mathbf{v}_{c-3}$ can also be expressed as linear sums of known vectors with unknown coefficients a_k, b_k, using equations (46,47) respectively, together with (18).

The next step is to express $\mathbf{v}_{c-4}, \ldots, \mathbf{v}_1$ as similar linear sums of known vectors with the same unknown coefficients, using (50), as:

$$\mathbf{v}_{j-1} = \mathbf{y}_{j-1} \left[\Sigma - (Q - C_{j-1} - R)\Theta \right]^{-1} \quad (2 \le j \le c-3) \; ;$$

where

$$\begin{aligned} \mathbf{y}_{j-1} \;=\; & -\mathbf{v}_j \left[Q(E+\Theta\Delta) - \Sigma(E+\Delta) - C_j(E+\Theta+\Theta\Delta) - R(E+\Theta) \right] \\ & -\mathbf{v}_{j+1} \left[-Q\Delta + \Sigma\Delta + C_{j+1}(E+\Delta+\Theta\Delta) + R \right] + \mathbf{v}_{j+2} C_{j+2} \Delta \; . \end{aligned}$$

Using the equations (51,18) and the same procedure as in (20) for $\mathbf{v}_{c-1}$, $\mathbf{v}_0$ too can be expressed as a similar linear sum.

We still have 3 vector equations, (43,6,52) and a scalar equation (15). By substituting the expressions we already have for all the $\mathbf{v}_j$'s as linear sums of known vectors and the unknown scalars a_k, b_k in these equations, we obtain $3N + 1$ linear simultaneous scalar equations in $3N$ unknowns, the a_k, b_k's. Out of these $3N + 1$ linear equations, only $3N$ (including the normalising equation) are independent. Hence, they can be solved for the a_k, b_k's.

Since the eigenvalues and eigenvectors are either real or complex conjugate pairs, so are the a_k, b_k's. An efficient procedure for computing the eigenvalue-eigenvectors,the a_k, b_k's and the required steady state probabilities and performance measures is given in [4].

4.1 System with infinite queueing capacity

So far the analysis has been for the case of finite L. A corresponding analysis for the case of infinite queueing capacity yields the spectral expansion solution as:

$$\mathbf{v}_j = \sum_{k=1}^{N} a_k \boldsymbol{\psi}_k \lambda_k^{j-c} \quad (j = c, c+1, \ldots) \ . \tag{23}$$

corresponding to equation (14). Here, we need only the N relevant eigenvalue-eigenvector pairs of $Z(\lambda)$, and the N a_k's (which are reals or complex conjugate pairs) are to be computed. Notice that equation (23) is the same as (18) when the limit $L \to \infty$ is taken. Notice too that the computation time for this case would be much less than that for finite L.

4.2 Examples

We now return to the examples used in section 3.2 and consider the solutions for their steady state probabilities derived by the SEM above. We work with the case of infinite L and so only need to consider the polynomial $Z(\lambda)$.

4.2.1 M/M/1 G-queue

In this example, the simplest of queues with negative customers, proposition 1 gave $Q_0 = \sigma$, $Q_1 = -(\sigma+\mu+\rho)$, $Q_2 = \mu+\rho$, $Q_3 = 0$. The matrix polynomial $Z(\lambda)$ has dimension one and equating its determinant to zero reduces to the scalar equation

$$\sigma - (\sigma + \mu + \rho)\lambda + (\mu + \rho)\lambda^2 = 0$$

As we observed, $\lambda = 1$ is always a solution and we factorise the equation as

$$(\lambda - 1)((\mu + \rho)\lambda - \sigma) = 0$$

Hence the eigenvalue we require (inside the unit disk) is $\lambda = \sigma/(\mu+\rho)$ and we obtain the solution

$$p_j = a_1 \left(\frac{\sigma}{\mu+\rho}\right)^{j-1}$$

for $j \geq 1$. This is precisely the standard result for the M/M/1 G-queue with no batches.

4.2.2 MMCPP/M/1 queue

For this queue, we obtained $Q_0 = \Sigma$, $Q_1 = Q - \Sigma - C - R$, $Q_2 = C + R$, $Q_3 = 0$, giving matrix polynomial

$$Z(\lambda) = \Sigma + (Q - \Sigma - C - R)\lambda + (C+R)\lambda^2$$

The equation $|Z(\lambda)| = 0$ has $2N$ solutions, one of which is unity, when there are N phases. For $N = 2$ we could derive a cubic equation for the other three roots but the procedure would add little insight. There are no known closed form solutions for this queue.

4.2.3 M/M/1 G-queue with batch arrivals

This example is perhaps the most interesting since the equations arising from proposition 1 are not the raw balance equations. As in example 1, the matrices are 1×1 and there is one eigenvalue λ, the least solution of the cubic equation $Q_0 + Q_1\lambda + Q_2\lambda^2 + Q_3\lambda^3 = 0$. Since $\mu > 0$, $Q_3 = 0$ if and only if $\delta = 0$, i.e. the negative arrivals are all single or absent; we consider this case below. Otherwise, we know that $\lambda = 1$ is a solution, i.e. $Q_0 + Q_1 + Q_2 + Q_3 = 0$, and so the cubic factorises into

$$(1-\lambda)(Q_0 - (Q_2 + Q_3)\lambda - Q_3\lambda^2) = 0$$

Hence the root we seek is the smaller of those of the quadratic

$$\mu\delta\lambda^2 - (\sigma\delta + \mu(1+\theta\delta) + \rho)\lambda + \sigma + (\mu+\rho)\theta = 0 \tag{24}$$

There is no standard solution published for the M/M/1 G-queue with batches of both positive and negative customers, as far as we know. Therefore we compare our solution with those for the M/M/1 queue with geometric batch arrivals, but no negative customers, and for the M/M/1 G-queue with batch negative arrivals and single positive arrivals.

No negative customers

Here we have $\rho = \delta = 0$ and so $Q_0 = \sigma + \mu\theta,\ Q_1 = -\sigma - \mu(1+\theta),\ Q_2 = \mu,\ Q_3 = 0$. Notice that if we also had negative *single* arrivals, i.e. $\rho > 0,\ \delta = 0$, the only change to Q_0, Q_1, Q_2, Q_3 would be to replace μ by $\mu + \rho$. This is exactly as expected since the effect of the negative arrivals would only be to 'help' the server, effectively increasing its rate as in simple M/M/1 G-queues.

Hence we require the smaller root of the quadratic equation

$$\mu\lambda^2 - (\sigma + \mu + \mu\theta)\lambda + \sigma + \mu\theta = 0$$

which is $\lambda = \theta + \sigma/\mu$. It then follows that the equilibrium probability $p_j = a_1\lambda^{j-1}$ for $j \geq 1$.

Now, the standard solution for an M/M/1 queue with bulk arrivals of geometric size is given in terms of the equilibrium queue length probability generating function

$$\Pi(z) = \frac{(1-\nu)(1-\theta z)}{1-\lambda' z}$$

where $\nu = \frac{\sigma}{(1-\theta)\mu}$ and $\lambda' = 1 - (1-\nu)(1-\theta) = \lambda$; see, for example, [14]. Rearranging, we have

$$\frac{\Pi(z)}{1-\nu} = \frac{1}{1-\lambda z} + \frac{\theta}{\lambda}\left(1 - \frac{1}{1-\lambda z}\right) = 1 + (1-\theta/\lambda)\sum_{i=1}^{\infty}\lambda^i z^i$$

if $\lambda z < 1$. The two solutions are therefore compatible and if we considered the equation at level zero in proposition 1, they would be seen to be identical.

Single positive arrivals

In this case we have $\theta = 0$ in equation (24) giving

$$\mu\delta\lambda^2 - (\sigma\delta + \mu + \rho)\lambda + \sigma = 0$$

Now, it is shown in [12] that the M/M/1 G-queue with bulk negative arrivals having batch size probability generating function $H(z)$ has equilibrium probability mass function $p_j = (1-\lambda)\lambda^j$ for $j \geq 0$ where λ is the least solution of the fixpoint equation in z

$$z = \frac{\sigma}{\mu + \rho h(z)}$$

where $h(z) = \frac{1-H(z)}{1-z}$. For our geometric batch sizes, we have

$$H(z) = \frac{(1-\delta)z}{1-\delta z}$$

and so our fixpoint equation is

$$z = \frac{\sigma(1-\delta z)}{\mu(1-\delta z)+\rho}$$

This simplifies to $\mu\delta z^2 - (\sigma\delta + \mu + \rho)z + \sigma = 0$ giving the same λ as above.

5 The matrix-geometric approach

An alternate solution for (14) is to extend Naoumov's method [25] for QBD processes, based on the matrix-geometric solution method [26]. That solution is:

$$\mathbf{v}_j = \mathbf{w}T^{j-c} + \mathbf{u}_1 S_1^{L-1-j} + \mathbf{u}_2 S_2^{L-1-j} \qquad (c \leq j \leq L-1)\ , \tag{25}$$

where T is a minimal non-negative matrix solution to the matrix equation,

$$Q_0 + TQ_1 + T^2Q_2 + T^3Q_3 = 0\ , \tag{26}$$

S_1, S_2 are two independent minimal non-negative matrix solutions to the matrix equation,

$$Q_3 + SQ_2 + S^2Q_1 + S^3Q_0 = 0 \tag{27}$$

and $\mathbf{w}, \mathbf{u}_1, \mathbf{u}_2$ are arbitrary vector constants which become the unknowns to be found in order to satisfy all the other balance equations.

The two solution methods are mathematically very similar. One way of obtaining T, S_1, S_2 is from the eigenvalues and eigenvectors obtained in Section 4. Let Λ be the $N \times N$ diagonal matrix whose diagonal elements are the N eigenvalues (λ's) of least absolute value of $Z(\lambda)$, arranged in some order. Let Ψ be the $N \times N$ matrix whose i^{th} row is the eigenvector of $Z(\lambda)$ corresponding to the eigenvalue that is the i^{th} diagonal element of Λ. Similarly, let Ξ_1, Ξ_2 be $N \times N$ diagonal matrices whose diagonal elements comprise all the $2N$ eigenvalues of least absolute value of $\overline{Z}(\xi)$ and Υ_1, Υ_2, corresponding to Ξ_1, Ξ_2 respectively, be $N \times N$ matrices whose rows are the eigenvectors of $\overline{Z}(\xi)$ corresponding to the eigenvalues in Ξ_1, Ξ_2, in the same order. Then, by elementary linear algebra,

$$T = \Psi^{-1}\Lambda\Psi\ ;\ S_1 = \Upsilon_1^{-1}\Xi_1\Upsilon_1\ ;\ S_2 = \Upsilon_2^{-1}\Xi_2\Upsilon_2\ . \tag{28}$$

Using the matrix-geometric solution (25), it would be possible to develop a similar procedure as in Section 4 to solve for the unknowns and hence for the steady state probabilities. Since there are many ways of splitting the above mentioned $2N$ eigenvalues into two distinct groups of N each, there may be many distinct solutions for Ξ_1 and Ξ_2 and hence for S_1 and S_2. We believe it is possible to compute T and one pair of minimal non-negative S_1, S_2 directly from equations (26,27) by extending Naoumov's procedure.

Substantial work has already been done in this direction by Bini and Meini [2, 3] who use an efficient cyclic reduction based LU factorisation and certain FFT based methods to accomplish the task of finding T, S_1, S_2. Indeed, most eigenvalue-eigenvector routines, such as are used in spectral expansion, also use LU factorisation. Hence there is another obvious intersection between the two methods. Indeed, a great deal of sustained effort has already gone into developing efficient eigenvalue-eigenvector routines by control theoreticians over decades, and advantage should be taken of these results.

It may also be noted that, if the matrix-geometric expansion for $\mathbf{v}_j$ (25) is used, the computation of $\mathbf{y}_{c-1}$ (and similarly of $\mathbf{y}_{c-2}, \mathbf{y}_{c-3}$) (21) can be computationally expensive for large or infinite L. Also, some preliminary, limited studies suggest that the spectral expansion method is computationally more efficient than the matrix geometric method with folding procedures [24, 18]. However, the MGM has been carefully tuned for efficiency and numerical stability – Bini and Meini's work is a recent example – and the authors do not advocate either one. A comprehensive comparison of the two methods, both mathematically and in terms of numerical stability and efficiency, for a range of problems (not just the queue considered here), would be of considerable interest but this is not the subject of the present paper.

6 RCE-immune servicing

The preceding analysis assumes that the last positive customer in the system will be killed by a negative arrival whether or not it is actually receiving service at the instant of the negative arrival. However, if a negative arrival is to represent the transfer of work out of a queue due to some load balancing strategy, it makes sense for only customers waiting to commence service to be susceptible to killing—i.e. only customers not in the first c positions in the queue. In this case we say that customers in service are *immune* from killing by negative customers.

In the case of RCE-immune servicing, the downward transition rates become:

$$\begin{aligned}
C_{j+s,j} &= C(E-\Phi)\Phi^{s-1} + R(E-\Delta)\Delta^{s-1} \\
&\qquad (c+1 \le j \le L-1\ ;\ s = 1,2,\ldots,L-j) \\
&= C(E-\Phi)\Phi^{s-1} + R\Delta^{s-1} \quad (j = c\ ;\ s = 1,2,\ldots,L-c) \\
&= C\Phi^{s-1} \qquad (j = c-1\ ;\ s = 1,2,\ldots,L-c+1) \\
&= 0 \qquad (c \ge 2\ ;\ 0 \le j \le c-2\ ;\ s \ge 2) \\
&= C_{j+1} \qquad (c \ge 2\ ;\ 0 \le j \le c-2\ ;\ s = 1)
\end{aligned}$$

The balance equations[3] can be expressed, following the procedure in Section 3, as (for $L > c+4$):

(i) For $c+2 \le j \le L-3$,

$$\mathbf{v}_{j-1}Q_0 + \mathbf{v}_jQ_1 + \mathbf{v}_{j+1}Q_2 + \mathbf{v}_{j+2}Q_3 = 0$$

(ii) For $j = L, L-1, L-2$,

$$\sum_{s=1}^{L} \mathbf{v}_{L-s}\Theta^{s-1}\Sigma + \mathbf{v}_L(Q-C-R) = 0$$

$$\begin{aligned}
&\mathbf{v}_{L-2}\left[\Sigma - (Q-C-R)\Theta\right] \\
&+\mathbf{v}_{L-1}\left[Q - \Sigma - C(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Delta)\right] \\
&+\mathbf{v}_L\left[C(E-\Phi)(E-\Theta\Phi) + R(E-\Delta)(E-\Theta\Delta)\right] = 0
\end{aligned}$$

$$\begin{aligned}
&\mathbf{v}_{L-3}\left[\Sigma - (Q-C-R)\Theta\right] \\
&+\mathbf{v}_{L-2}\left[Q - \Sigma - C(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Delta)\right] \\
&+\mathbf{v}_{L-1}\left[C(E-\Phi)(E-\Theta\Phi) + R(E-\Delta)(E-\Theta\Delta)\right] \\
&+\mathbf{v}_L\left[C(E-\Phi)\Phi(E-\Theta\Phi) + R(E-\Delta)\Delta(E-\Theta\Delta)\right] = 0
\end{aligned}$$

[3] These balance equations are also valid for the more general case in which customers that are not in service can be killed by negative customers in an arbitrary order.

(iii) For $j = c+1, c, c-1$,

$$
\begin{aligned}
\mathbf{v}_c \left[\Sigma - (Q - C)\Theta\right] & \\
+\mathbf{v}_{c+1} \left[Q - \Sigma - C(E + \Theta - \Theta\Phi) - R(E + \Theta)\right] & \\
+ \sum_{s=1}^{L-c-1} \mathbf{v}_{c+1+s} \left[C(E - \Phi)\Phi^{s-1}(E - \Theta\Phi) + R\Delta^{s-1}(E - \Delta - \Theta\Delta)\right] &= 0 \\
(for\ c \geq 1) &
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}_{c-1} \left[\Sigma - (Q - C_{c-1})\Theta\right] & \\
+\mathbf{v}_c \left[Q - \Sigma - C(E + \Theta)\right] & \\
+ \sum_{s=1}^{L-c} \mathbf{v}_{c+s} \left[C\Phi^{s-1}(E - \Phi - \Theta\Phi) + R\Delta^{s-1}\right] &= 0 \\
(for\ c \geq 2) &
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{v}_{c-2} \left[\Sigma - (Q - C_{c-2})\Theta\right] & \\
+\mathbf{v}_{c-1} \left[Q - \Sigma - C_{c-1}(E + \Theta)\right] & \\
+ \sum_{s=1}^{L-c+1} \mathbf{v}_{c-1+s} \left[C\Phi^{s-1}\right] &= 0 \\
(for\ c \geq 3) &
\end{aligned}
$$

(iv) For $c - 2 \geq j \geq 2$,

$$\mathbf{v}_{j-1} \left[\Sigma - (Q - C_{j-1})\Theta\right] + \mathbf{v}_j \left[Q - \Sigma - C_j(E + \Theta)\right] + \mathbf{v}_{j+1}C_{j+1} = 0$$

(v) For $j = 1$,

$$\mathbf{v}_0 \left[\Sigma - Q\Theta\right] + \mathbf{v}_1 \left[Q - \Sigma - C_1(E + \Theta)\right] + \mathbf{v}_2 C_2 = 0 \quad (for\ c \geq 3)$$

$$\mathbf{v}_0 \left[\Sigma - Q\Theta\right] + \mathbf{v}_1 \left[Q - \Sigma - C_1(E + \Theta)\right] + \sum_{s=1}^{L-1} \mathbf{v}_{1+s} C\Phi^{s-1} = 0 \quad (for\ c = 2)$$

$$
\begin{aligned}
\mathbf{v}_0 \left[\Sigma - Q\Theta\right] + \mathbf{v}_1 \left[Q - \Sigma - C_1(E + \Theta)\right] & \\
+ \sum_{s=1}^{L-1} \mathbf{v}_{1+s} \left[C\Phi^{s-1}(E - \Phi - \Theta\Phi) + R\Delta^{s-1}\right] = 0 \qquad & (for\ c = 1)
\end{aligned}
$$

(vi) For $j = 0$,

$$\begin{aligned} \mathbf{v}_0 [Q - \Sigma] + \mathbf{v}_1 [C_1] &= 0 \qquad (for\ c \geq 2) \\ \mathbf{v}_0 [Q - \Sigma] + \sum_{s=1}^{L} \mathbf{v}_s C \Phi^{s-1} &= 0 \qquad (for\ c = 1) \end{aligned}$$

(vii) Summation

$$\sum_{j=0}^{L} \mathbf{v}_j \mathbf{e}_N = 1$$

where Q_0, Q_1, Q_2, Q_3 are as given in Section 3. The above set of equations can be solved in the same way as in Section 4. Notice that in the case of the M/M/1 G-queue with no batches, the only difference from the equations derived in section 3.2.1 is that we now have

$$-p_0 \sigma + p_1 \mu = 0$$

rather than

$$-p_0 \sigma + p_1 (\mu + \rho) = 0$$

This reflects the fact that customers in service cannot be killed.

7 The RCH killing discipline

In this section we consider the RCH (Removal of Customers from the head of the queue) discipline. This is appropriate for modelling server breakdowns, where a customer in service will be lost for sure and maybe also a portion of the queue of waiting customers, cf. [15]. We consider the case of a single server ($c = 1$), that is, the MM CPP/GE/1/L G-Queue, although the analysis can be extended straightforwardly to the case of a multiserver ($c > 1$). The system is ergodic when L is finite, but the stability condition when $L = \infty$ is different from the one in Section 2.6.

The transition rates relating to positive arrivals, $B_{i,j,j+s}$, are unchanged from the RCE case, section 3. To determine the rates $C_{i,j+s,j}$, consider the system in the state $(i, j + s)$, where $j + s \leq L$. The rate at which a batch of l negative customers arrives is $(1 - \delta_i)\delta_i^{l-1}\rho_i$ ($l = 1, 2, \ldots$). If $l \geq j + s$ all the jobs will be removed by the negative customers. If $l < j + s$, then the job in service plus $l - 1$ jobs *waiting* for service will be removed, leaving only $j+s-l$ jobs. However, due the nature of the generalised exponential service times of the server, a certain number of customers may be 'leaked' out, i.e. serviced *instantaneously* just after the killing takes place. Alternatively, we may redefine the operation of the system such that this leakage does not occur, i.e. immediately after a negative arrival, the next customer in the queue (if any) *cannot* skip service, in contrast to a normal service completion. In this 'no leakage' case, the equilibrium state probabilities are identical to those of the case of RCE inimmune servicing, i.e. variant 1.

Assuming leakage may occur, the number leaked is k ($k = 0, 1, \ldots, j+s-l-1$) with probability $(1-\phi_i)\phi_i^k$ and $j+s-l$ with probability $\sum_{k=j+s-l}^{\infty}(1-\phi_i)\phi_i^k = \phi_i^{j+s-l}$. Also, the rate at which a batch of l ($l \geq j+s$) negative customers arrives is $\sum_{l=j+s}^{\infty}(1-\delta_i)\delta_i^{l-1}\rho_i = \delta_i^{j+s-1}\rho_i$. The rates $C_{i,j+s,j}$ can now be derived as,

$$\begin{aligned}
C_{i,j+s,j} &= (1-\phi_i)\phi_i^{s-1}\mu_i + (1-\delta_i)\rho_i(1-\phi_i)\{\phi_i\delta_i\}_{s-1} \\
&\qquad (\forall i \; ; \; 1 \leq j \leq L-1 \; ; \; s \geq 1) \; ; \\
&= \phi_i^{s-1}\mu_i + (1-\delta_i)\rho_i\{\phi_i\delta_i\}_{s-1} + \rho_i\delta_i^s \\
&\qquad (\forall i \; ; \; j = 0 \; ; \; s \geq 2) \; ; \\
&= \mu_i + \rho_i \qquad (\forall i \; ; \; j = 0 \; ; \; s = 1) \; ;
\end{aligned}$$

where

$$\begin{aligned}
\{\phi_i\delta_i\}_{s-1} &= \sum_{k=0}^{s-1}\phi_i^{s-1-k}\delta_i^k \qquad (s \geq 2) \; ; \\
&= 1 \qquad (s = 1) \; .
\end{aligned}$$

The above can be written in matrix form as,

$$\begin{aligned}
C_{j+s,j} &= (E-\Phi)\Phi^{s-1}M + (E-\Delta)R(E-\Phi)\{\Phi\Delta\}_{s-1} \\
&\qquad (1 \leq j \leq L-1 \; ; \; s \geq 1) \; ; \\
&= \Phi^{s-1}M + (E-\Delta)R\{\Phi\Delta\}_{s-1} + R\Delta^s \quad (j = 0 \; ; \; s \geq 2) \; ; \\
&= M + R \qquad (j = 0 \; ; \; s = 1) \; ;
\end{aligned}$$

where,

$$\begin{aligned}
\{\Phi\Delta\}_{s-1} &= \sum_{k=0}^{s-1}\Phi^{s-1-k}\Delta^k \qquad (s \geq 2) \; ; \\
&= E \qquad (s = 1) \; .
\end{aligned}$$

Notice that

$$\{\Phi\Delta\}_{s-1} = \Phi\{\Phi\Delta\}_{s-2} + \Delta^{s-1} \quad (s \geq 2) \ .$$

Following the same procedure as in Section 3, the balance equations in this case can be derived as,

(i) For $2 \leq j \leq L-3$,

$$\mathbf{v}_{j-1}Q_0 + \mathbf{v}_j Q_1 + \mathbf{v}_{j+1}Q_2 + \mathbf{v}_{j+2}Q_3 = 0 \ ;$$

(ii) For $j = L, L-1, L-2$,

$$\sum_{s=1}^{L} \mathbf{v}_{L-s}\Theta^{s-1}\Sigma + \mathbf{v}_L(Q - M - R) = 0 \ ;$$

$$\begin{aligned} \mathbf{v}_{L-2}\,[\Sigma - (Q-M-R)\Theta] \\ +\mathbf{v}_{L-1}\,[Q - \Sigma - M(E+\Theta-\Theta\Phi) - R(E+\Theta-\Theta\Phi-\Theta\Delta+\Theta\Phi\Delta)] \\ +\mathbf{v}_L\,[M(E-\Phi)(E-\Theta\Phi) + R(E-\Delta)(E-\Phi)(E-\Phi\Theta-\Delta\Theta)] = 0 \ ; \end{aligned}$$

$$\begin{aligned} \mathbf{v}_{L-3}\,[\Sigma - (Q-M-R)\Theta] \\ +\mathbf{v}_{L-2}\,[Q(E+\Theta\Phi) - \Sigma(E+\Phi) - M(E+\Theta) \\ - R(E+\Theta-\Theta\Delta+\Theta\Phi\Delta)] \\ +\mathbf{v}_{L-1}\,[-Q\Phi + \Sigma\Phi + M \\ + R(E-\Delta-\Theta\Delta+\Phi\Delta+\Theta\Phi\Delta+\Delta^2\Theta-\Delta^2\Theta\Phi)\big] \\ +\mathbf{v}_L\,[R(E-\Delta)(E-\Phi)\Delta(E-\Theta\Delta)] &= 0 \ ; \end{aligned}$$

(iii) For $j = 1, 0$

$$\begin{aligned} \mathbf{v}_0\,[\Sigma - Q\Theta] + \mathbf{v}_1\,[Q - \Sigma - M(E+\Theta) - R(E+\Theta)] \\ + \sum_{s=1}^{L-1} \mathbf{v}_{1+s}\,[C_{1+s,1} - C_{1+s,0}\Theta] &= 0 \ ; \end{aligned}$$

$$\mathbf{v}_0\,[Q - \Sigma] + \sum_{s=1}^{L} \mathbf{v}_s C_{s,0} = 0 \ .$$

(iv) Normalisation

$$\sum_{j=0}^{L} \mathbf{v}_j \mathbf{e}_N = 1 \; ;$$

where,

$$\begin{aligned}
Q_0 &= \Sigma - (Q - M - R)\Theta \; ; \\
Q_1 &= Q(E + \Theta\Phi + \Theta\Delta) - \Sigma(E + \Delta + \Phi) - M(E + \Theta + \Theta\Delta) \\
&\quad -R(E + \Theta + \Theta\Phi\Delta) \; ; \\
Q_2 &= -Q(\Phi + \Delta + \Theta\Phi\Delta) + \Sigma(\Phi + \Delta + \Delta\Phi) + M(E + \Delta + \Theta\Delta) \\
&\quad +R(E + \Phi\Delta + \Theta\Phi\Delta) \; ; \\
Q_3 &= Q\Phi\Delta - \Sigma\Phi\Delta - M\Delta - R\Phi\Delta \; .
\end{aligned}$$

These balance equations can be solved following the procedure of Section 4.

7.1 Condition for stability

Here, due to the presence of leakage, the condition for stability (when $L = \infty$) is different from the previous variants. That can be derived as follows. The average leakage rate increases with the queue length and its maximum (that is, when $J(t) \to \infty$) is given by,

$$\overline{\varphi} = \sum_{i=1}^{N} \frac{r_i \rho_i \phi_i}{1 - \phi_i} \tag{29}$$

Hence, when $L = \infty$, the system is stable if,

$$\overline{\sigma} < \overline{\rho} + \overline{\mu} + \overline{\varphi} \tag{30}$$

Obviously, if there is no leakage, the stability condition remains that of the RCE variants, (2).

8 Departure process

8.1 Departure burst size distribution

From Section 4, we have the solution for the steady state probabilities, $p_{i,j}$. The marginal probabilities, $p_{i.}$ and $p_{.j}$ are then defined as:

$$p_{i.} = \sum_{j=0}^{L} p_{i,j} \qquad ; \qquad p_{.j} = \sum_{i=1}^{N} p_{i,j} \; . \tag{31}$$

Now consider the system in the state (i,j), where $j > c$. Here, all the c servers are busy, with $j-c$ unattended positive customers in the queue. In this state, the departure rate associated with a batch size of s is $(1-\phi_i)\phi_i^{s-1}c\mu_i$ for $1 \le s \le j-c$ and $\phi_i^{j-c}c\mu_i$ for $s = j-c+1$. Hence, the average rate at which batches of size n, for $2 \le n \le L-c+1$, depart from the queue is,

$$\nu_n = \sum_{i=1}^{N}\sum_{j=c+n}^{L} p_{i,j}(1-\phi_i)\phi_i^{n-1}c\mu_i + \sum_{i=1}^{N} p_{i,c+n-1}\phi_i^{n-1}c\mu_i \qquad (n = 2,3,\ldots,L-c+1) \tag{32}$$

The average rate of single departures, ν_1, is

$$\nu_1 = \sum_{i=1}^{N}\sum_{j=1}^{c} p_{i,j}j\mu_i + \sum_{i=1}^{N}\sum_{j=c+1}^{L} p_{i,j}(1-\phi_i)c\mu_i \tag{33}$$

Thus, by the Law of Large Numbers for Markov chains, $\frac{\nu_n}{\sum_{s=1}^{L-c+1}\nu_s}$ is the equilibrium probability that the burst size is n $(n = 1,2,\ldots,L-c+1)$. The number of batch departures per unit time ν is

$$\nu = \sum_{s=1}^{L-c+1} \nu_s \tag{34}$$

and the average departure rate of positive customers $\overline{\nu}$ is

$$\overline{\nu} = \sum_{s=1}^{L-c+1} s\nu_s \tag{35}$$

Hence the loss rate of positive customers due to either overflow or being killed by negative customers is $\overline{\sigma} - \overline{\nu}$.

8.2 System at departure epochs

Let $\beta_{i,j}$ be the probability that the state of the system is (i,j) immediately after a (batch) departure epoch. Then, $\beta_{i,j}$ is proportional to the probability flux into state (i,j) *due to a departure*, i.e. $\beta_{i,j} \propto f_{i,j}$ where, for $1 \le i \le N$,

$$\begin{aligned}
f_{i,j} &= \sum_{n=1}^{L-j} p_{i,j+n}c\mu_i(1-\phi_i)\phi_i^{n-1} \qquad (j \ge c) \\
f_{i,c-1} &= \sum_{n=1}^{L-c+1} p_{i,c-1+n}c\mu_i\phi_i^{n-1} \\
f_{i,j} &= p_{i,j+1}(j+1)\mu_i \qquad (0 \le j \le c-2)
\end{aligned}$$

The normalisation constant is the reciprocal of the sum of all the $f_{i,j}$, i.e. the reciprocal of

$$\sum_{i=1}^{N}\left[\sum_{j=c}^{\infty}\sum_{n=1}^{L-j} p_{i,j+n} c\mu_i (1-\phi_i)\phi_i^{n-1} + \sum_{n=1}^{L-c+1} p_{i,c-1+n} c\mu_i \phi_i^{n-1} + \sum_{j=0}^{c-2} p_{i,j+1}(j+1)\mu_i\right]$$

Define $\beta_{i.} = \sum_{j=0}^{L} \beta_{i,j}$ $(i = 1, 2, \ldots, N)$ and $\beta_{.j} = \sum_{i=1}^{N} \beta_{i,j}$ $(j = 0, 1, \ldots, L)$.

8.3 Inter-batch departure intervals

Consider the system in steady state and let t_1 be the time instant of a batch departure. Let the next departure occur at time instant t_2 $(> t_1)$ and define the random variable $\tau = t_2 - t_1$. Let $G(s)$ be the Laplace transform of the equilibrium probability density function of the inter-departure interval τ. In order to derive an expression for $G(s)$, we define $\Gamma_{i,j}(s)$ to be the Laplace transform of the density of the random variable τ, *given that* the state of the system at t_1 is (i, j), i.e. $I(t_1) = i$ and $J(t_1) = j$.

We now consider the case of RCE-*immune servicing*. If $j \geq c$, then $\Gamma_{i,j}(s)$ is independent of j; denoted by $\Gamma_i(s)$, say. Then,

$$\Gamma_i(s) = \frac{c\mu_i}{c\mu_i + q_i + s} + \sum_{l=1}^{N} \frac{q_{i,l}\Gamma_l(s)}{c\mu_i + q_i + s} \quad (i = 1, 2, \ldots, N) \tag{36}$$

(Note that $q_{i,i} = 0$.)

If $1 \leq j \leq c-1$, the next significant event that changes the state of the system can be either (i) a single departure making $J(t) = j - 1$, or (ii) a new batch arrival of positive customers making $j < J(t) < c$, or (iii) a new batch arrival of positive customers making $j < J(t) \geq c$, or (iv) a phase change of the modulating process. Note that the arrival of negative customers makes no difference to the analysis since we are assuming that the service is immune and the next service completion cannot, therefore, be affected. The transition rates of these events are $j\mu_i$, $\sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1}\sigma_i$, $(1 - \sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1})\sigma_i$ and q_i respectively. Hence, we can write,

$$\begin{aligned} \Gamma_{i,j}(s) \;=\; & \frac{j\mu_i}{j\mu_i + \sigma_i + q_i + s} + \frac{\sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1}\sigma_i\Gamma_{i,j+n}(s)}{j\mu_i + \sigma_i + q_i + s} + \\ & \frac{(1 - \sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1})\sigma_i\Gamma_i(s)}{j\mu_i + \sigma_i + q_i + s} + \frac{\sum_{l=1}^{N} q_{i,l}\Gamma_{l,j}(s)}{j\mu_i + \sigma_i + q_i + s} \\ & (1 \leq i \leq N\ ;\ 1 \leq j \leq c-1)\,. \end{aligned} \tag{37}$$

If $j = 0$ at the instant t_1, then the next significant event can be either (i) a batch arrival of positive customers making $0 < J(t) < c$, or (ii) a batch arrival of positive customers making $0 < J(t) \geq c$, or (iii) a phase change of the modulating process. Hence we have,

$$\Gamma_{i,0}(s) = \frac{\sum_{n=1}^{c-1}(1-\theta_i)\theta_i^{n-1}\sigma_i\Gamma_{i,n}(s) + (1-\sum_{n=1}^{c-1}(1-\theta_i)\theta_i^{n-1})\sigma_i\Gamma_i(s)}{\sigma_i+q_i+s} + \frac{\sum_{l=1}^{N} q_{i,l}\Gamma_{l,0}(s)}{\sigma_i+q_i+s} \quad (i = 1, 2, \ldots, N) \ . \tag{38}$$

The Laplace transform $G(s)$ can now be written as,

$$G(s) = \sum_{i=1}^{N}\sum_{j=0}^{c-1}\beta_{i,j}\Gamma_{i,j}(s) + \sum_{i=1}^{N}(\beta_{i.} - \sum_{j=0}^{c-1}\beta_{i,j})\Gamma_i(s) \tag{39}$$

In principle, we could now compute $G(s)$ at all points s that may be required by solving the linear simulataneous equations given by equations (36), (37) and (38). Such a computation would be required, for example, if the probability density of interdeparture time were required and the Laplace transform had to be inverted numerically. However, this would be very expensive—perhaps prohibitively—and here we concentrate instead on the moments of the interdeparture time distribution.

Let $\Gamma_{i,j}^{(h)}(s)$ be the h^{th} derivative of $\Gamma_{i,j}(s)$ with respect to s and $\Gamma_i^{(h)}(s)$ be that of $\Gamma_i(s)$. Then, by differentiating equation (39) h times successively, we get,

$$G^{(h)}(s) = \sum_{i=1}^{N}\sum_{j=0}^{c-1}\beta_{i,j}\Gamma_{i,j}^{(h)}(s) + \sum_{i=1}^{N}(\beta_{i.} - \sum_{j=0}^{c-1}\beta_{i.j})\Gamma_i^{(h)}(s) \ . \tag{40}$$

If g_h is the h^{th} moment of τ, then we have,

$$g_h = (-1)^h G^{(h)}(s)|_{s=0} \ .$$

In order to compute g_h, we need the values of $\Gamma_{i,j}^{(h)}(s)$ and $\Gamma_i^{(h)}(s)$, at $s = 0$. Differentiating equations (36, 37, 38) h times successively with respect to s and performing some algebraic simplification, we get the following equations:

$$\Gamma_i^{(h)}(s) = \frac{\sum_{l=1}^{N} q_{i,l}\Gamma_l^{(h)}(s) - h\Gamma_i^{(h-1)}(s)}{c\mu_i + q_i + s} \quad (h = 1, 2, \ldots) \tag{41}$$

$$\Gamma_{i,j}^{(h)}(s) = \frac{\sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1}\sigma_i\Gamma_{i,j+n}^{(h)}(s)}{j\mu_i+\sigma_i+q_i+s} + \frac{(1-\sum_{n=1}^{c-j-1}(1-\theta_i)\theta_i^{n-1})\sigma_i\Gamma_i^{(h)}(s)+\sum_{l=1}^{N}q_{i,l}\Gamma_{l,j}^{(h)}(s)}{j\mu_i+\sigma_i+q_i+s} - \frac{h\Gamma_{i,j}^{(h-1)}(s)}{j\mu_i+\sigma_i+q_i+s} \quad (1\le i\le N\ ;\ 1\le j\le c-1\ ;\ h=1,2,\ldots)\ . \tag{42}$$

$$\Gamma_{i,0}^{(h)}(s) = \frac{\sum_{n=1}^{c-1}(1-\theta_i)\theta_i^{n-1}\sigma_i\Gamma_{i,n}^{(h)}(s)+(1-\sum_{n=1}^{c-1}(1-\theta_i)\theta_i^{n-1})\sigma_i\Gamma_i^{(h)}(s)}{\sigma_i+q_i+s} + \frac{\sum_{l=1}^{N}q_{i,l}\Gamma_{l,0}^{(h)}(s)-h\Gamma_{i,0}^{(h-1)}(s)}{\sigma_i+q_i+s} \quad (i=1,2,\ldots,N\ ;\ h=1,2,\ldots)$$

If $\Gamma_{i,j}^{(h-1)}(0), \Gamma_i^{(h-1)}(0)$ $(\forall i,j)$ are known, then equations (41, 42, 43), on substituting $s=0$, become $(c+1)N$ linear simultaneous equations in the $(c+1)N$ unknowns, $\Gamma_{i,j}^{(h)}(0)$ $(i=1,2,\ldots,N\ ;\ j=0,1,\ldots,c-1)$ and $\Gamma_i^{(h)}(0)$ $(i=1,2,\ldots,N)$, for any $h\ge 1$. We do have $\Gamma_{i,j}^{(0)}=1\,(\forall(i,j))$ and $\Gamma_i^{(0)}=1\,(\forall i)$. Hence, to find g_k, these sets of linear simultaneous equations need to be solved successively for $h=1,2,\ldots,k$. Thus, g_h can be computed for any given value of h exactly. Using these moments of τ, an appropriate batch Markov renewal process may be identified to approximate the departure process. The work of Ryden [28] may be useful to accomplish such a task.

9 Conclusions

We have derived an exact result for the equilibrium state probabilities of a Markov modulated multi-server queue with unbounded or bounded capacity, with generalised exponential service times and with compound Poisson arrivals of both positive and negative customers—the MM CPP/GE/c/L G-queue. This is a highly representative queue that can account for many types of traffic and processing patterns, e.g. correlated, bursty traffic, environmentally sensitive service times, unreliable servers and load balancing. It generalises significantly both queues with MMPP arrivals [5, 21] and with generalised exponential service times [19, 20]. Moreover, it also generalises results on G-queues which had been restricted to exponential service times and bulk-Poisson arrivals [6], apart from the intractable extension to the M/G/1 G-queue [17].

By considering the departure process of the queue, we would have the basis of a building block for analysing *networks* of such queues in terms of the internal arrival processes at each constituent queue. An interesting approximate approach is to consider all queues in isolation with positive arrival streams determined by the busy/idle status of the source nodes and the negative arrival streams determined by the equilibrium dynamic behaviour of certain queue lengths so as to facilitate load balancing. For example, if a certain queue's length passes a given threshold, customers will be transferred out of it until the length goes below a lower threshold. Similarly, the positive arrival rate at underutilised queues will increase. Transfers of work could be represented by a combination of negative arrivals at one queue and extra positive arrivals at another when these queues are over- and under-utilised respectively. The dynamics of the utilisation levels can be described in a modulating CTMC, although synchronised redistribution of load cannot be thus modelled. In practice, this need not be a problem since the information about the loading of nodes downstream, which is used in the load balancing algorithm, cannot be current and is typically not transmitted upstream continually.

Appendix

Proofs of non-repeating cases in Proposition 1

There are four non-repeating cases, corresponding to the balance equations for levels near to the upper bound L, near to the threshold c, below the threshold and near to 0 (empty queue).

Case 2: Levels $L-1, L-2$

Substitute $j = L-1$ in (10) to obtain,

$$\begin{aligned} &\mathbf{v}_{L-2}\left[\Sigma - (Q - C - R)\Theta\right] \\ +&\mathbf{v}_{L-1}\left[Q - \Sigma - C(E + \Theta - \Theta\Phi) - R(E + \Theta - \Theta\Delta)\right] \\ +&\mathbf{v}_{L}\left[C(E - \Phi)(E - \Theta\Phi) + R(E - \Delta)(E - \Theta\Delta)\right] = 0 \end{aligned} \quad (43)$$

Substitute $j = L - 2$ in (12) to obtain,

$$\begin{aligned}
& \mathbf{v}_{L-3}\left[\Sigma - (Q - C - R)\Theta\right] \quad (44)\\
+ & \mathbf{v}_{L-2}\left[Q(E + \Theta\Phi) - \Sigma(E + \Phi) - C(E + \Theta) - R(E + \Theta\Phi + \Theta - \Theta\Delta)\right]\\
+ & \mathbf{v}_{L-1}\left[-Q\Phi + \Sigma\Phi + C + R(E + (\Phi - \Delta)(E + \Theta - \Theta\Delta))\right]\\
+ & \mathbf{v}_{L}\left[R(E - \Delta)(E - \Theta\Delta)(\Delta - \Phi)\right] = 0
\end{aligned}$$

Case 3: Levels $c, c-1, c-2$

For the level $j = c$:

If $c = 1$, then *goto* **Case 5**.
If $c > 1$, then we have

$$C_{c-1} = (c-1)M \ ; \ C_c = C \ ; \ I_{c>0} = 1 \ ;$$
$$C_{c,c-1} = C + R(E - \Delta) \ ;$$
$$C_{c+s,c} = C(E - \Phi)\Phi^{s-1} + R(E - \Delta)\Delta^{s-1} \ ;$$
$$C_{c+s,c-1} = C\Phi^{s} + R(E - \Delta)\Delta^{s}$$

Substituting the above in (9) for $j = c$, we get,

$$\begin{aligned}
& \mathbf{v}_{c-1}\left[\Sigma - (Q - C_{c-1} - R)\Theta\right] \quad (45)\\
+ & \mathbf{v}_{c}\left[Q - \Sigma - C(E + \Theta) - R(E + \Theta - \Theta\Delta)\right]\\
+ & \sum_{s=1}^{L-c} \mathbf{v}_{c+s}\left[C\Phi^{s-1}(E - \Phi - \Theta\Phi) + R(E - \Delta)\Delta^{s-1}(E - \Theta\Delta)\right] = 0
\end{aligned}$$

For the level $j = c - 1$:

If $c \leq 2$, then *goto* **Case 5**. If $c > 2$, then we have,

$$\begin{aligned}
C_{c-1,c-2} &= C_{c-1} + R(E - \Delta) \ ;\\
C_{c-1+s,c-1} &= C\Phi^{s-1} + R(E - \Delta)\Delta^{s-1} \ ; \ C_{c-1+s,c-2} = R(E - \Delta)\Delta^{s}
\end{aligned}$$

Substituting the above in (9) for $j = c - 1$, we get,

$$\begin{aligned}
& \mathbf{v}_{c-2}\left[\Sigma - (Q - C_{c-2} - R)\Theta\right] \quad (46)\\
+ & \mathbf{v}_{c-1}\left[Q - \Sigma - C_{c-1}(E + \Theta) - R(E + \Theta - \Theta\Delta)\right]\\
+ & \sum_{s=1}^{L-c+1} \mathbf{v}_{c-1+s}\left[C\Phi^{s-1} + R(E - \Delta)\Delta^{s-1}(E - \Theta\Delta)\right] = 0
\end{aligned}$$

For the level $j = c - 2$:

If $c \leq 3$, then *goto* **Case 5**. If $c > 3$, then we have,

$$\begin{aligned}
C_{c-2,c-3} &= C_{c-2} + R(E - \Delta) \; ; \; C_{c-2+s,c-3} = R(E-\Delta)\Delta^s \; ; \\
C_{c-2+s,c-2} &= C_{c-1} + R(E - \Delta) \qquad \text{(for } s = 1) \; ; \\
&= R(E - \Delta)\Delta^{s-1} \qquad \text{(for } s \geq 2)
\end{aligned}$$

Substituting the above in (9), we get,

$$\begin{aligned}
\mathbf{v}_{c-3}\left[\Sigma - (Q - C_{c-3} - R)\Theta\right] \\
+\mathbf{v}_{c-2}\left[Q - \Sigma - C_{c-2}(E + \Theta) - R(E + \Theta - \Theta\Delta)\right] \\
+\mathbf{v}_{c-1}\left[C_{c-1} + R(E - \Delta)(E - \Theta\Delta)\right] \\
+\sum_{s=2}^{L-c+2} \mathbf{v}_{c-2+s}\left[R(E - \Delta)\Delta^{s-1}(E - \Theta\Delta)\right] &= 0
\end{aligned} \tag{47}$$

Case 4: Levels $c - 3, \ldots, 2$

If $c \leq 4$, then *goto* **Case 5**. If $c > 4$, consider the j-range $2 \leq j \leq c - 2$. For this range, we have,

$$\begin{aligned}
C_{j,j-1} &= C_j + R(E - \Delta) \; ; \; C_{j+s,j-1} = R(E - \Delta)\Delta^s \; ; \\
C_{j+s,j} &= C_{j+1} + R(E - \Delta) \qquad \text{(for } s = 1) \\
&= R(E - \Delta)\Delta^{s-1} \qquad \text{(for } s \geq 2)
\end{aligned}$$

Substituting the above in (9), we get, for the level j,

$$\begin{aligned}
\mathbf{v}_{j-1}\left[\Sigma - (Q - C_{j-1} - R)\Theta\right] \\
+\mathbf{v}_{j}\left[Q - \Sigma - C_{j}(E + \Theta) - R(E + \Theta - \Theta\Delta)\right] \\
+\mathbf{v}_{j+1}\left[C_{j+1} + R(E - \Delta)(E - \Theta\Delta)\right] \\
+\sum_{s=2}^{L-j} \mathbf{v}_{j+s}\left[R(E - \Delta)\Delta^{s-1}(E - \Theta\Delta)\right] &= 0 \\
\text{(for } 2 \leq j \leq c - 2)
\end{aligned} \tag{48}$$

The balance equation for the $(j+1)^{th}$ level is obtained by replacing j by $j+1$ in the above, as

$$\mathbf{v}_{j}\left[\Sigma - (Q - C_{j} - R)\Theta\right] \tag{49}$$

$$
\begin{aligned}
+\mathbf{v}_{j+1}\left[Q-\Sigma-C_{j+1}(E+\Theta)-R(E+\Theta-\Theta\Delta)\right]& \\
+\mathbf{v}_{j+2}\left[C_{j+2}+R(E-\Delta)(E-\Theta\Delta)\right]& \\
+\sum_{s=2}^{L-j-1}\mathbf{v}_{j+1+s}\left[R(E-\Delta)\Delta^{s-1}(E-\Theta\Delta)\right]&=0 \\
(\text{for } 2\le j+1\le c-2 \text{ or } 1\le j\le c-3)&
\end{aligned}
$$

Modifing (48) by subtracting from it equation (49) post-multiplied by Δ, we get

$$
\begin{aligned}
\mathbf{v}_{j-1}\left[\Sigma-(Q-C_{j-1}-R)\Theta\right]& \qquad (50)\\
+\mathbf{v}_{j}\left[Q(E+\Theta\Delta)-\Sigma(E+\Delta)-C_j(E+\Theta+\Theta\Delta)-R(E+\Theta)\right]& \\
+\mathbf{v}_{j+1}\left[-Q\Delta+\Sigma\Delta+C_{j+1}(E+\Delta+\Theta\Delta)+R\right]& \\
+\mathbf{v}_{j+2}\left[-C_{j+2}\Delta\right]=0 \qquad (2\le j\le c-3)&
\end{aligned}
$$

Case 5: Levels $1,0$

For the level $j=1$, we have,

$$
\begin{aligned}
C_0 &= 0\ ;\ C_1=M\ ;\ C_{1,0}=C_1+R\ ; \\
C_{1+s,0} &= C\Phi^s+R\Delta^s \qquad (c=1)\ ; \\
&= R\Delta^s \qquad (c>1)\ ; \\
C_{1+s,1} &= C(E-\Phi)\Phi^{s-1}+R(E-\Delta)\Delta^{s-1} \qquad (c=1) \\
&= C\Phi^{s-1}+R(E-\Delta)\Delta^{s-1} \qquad (c=2) \\
&= C_2+R(E-\Delta) \qquad (c\ge 3\ ;\ s=1) \\
&= R(E-\Delta)\Delta^{s-1} \qquad (c\ge 3\ ;\ s\ge 2)
\end{aligned}
$$

Substituting the above in (9) for $j=1$, we get,

$$
\begin{aligned}
\mathbf{v}_0\left[\Sigma-Q\Theta\right]+\mathbf{v}_1\left[Q-\Sigma-(C_1+R)(E+\Theta)\right]& \qquad (51)\\
+\sum_{s=1}^{L-1}\mathbf{v}_{1+s}\left[C_{1+s,1}-C_{1+s,0}\Theta\right]&=0
\end{aligned}
$$

For the level $j=0$, we have,

$$
\begin{aligned}
C_{1,0} &= C_1+R \\
C_{s,0} &= R\Delta^{s-1} \qquad (c>1) \\
&= C\Phi^{s-1}+R\Delta^{s-1} \qquad (c=1)
\end{aligned}
$$

Substituting the above in (7) for $j = 0$, we get

$$\mathbf{v}_0 [Q - \Sigma] + \mathbf{v}_1 [C_1 + R] + \sum_{s=2}^{L} \mathbf{v}_s \left[C\Phi^{s-1}(1 - I_{c-1>0}) + R\Delta^{s-1} \right] = 0 \quad (52)$$

References

[1] M. Bhabuta and P.G. Harrison. Analysis of ATM traffic on the London MAN, Proceedings of the 4th International Conference on Performance Modelling and Evaluation of ATM Networks, Ilkely, Chapman and Hall, 1997.

[2] D. Bini and B. Meini. On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems, Proceedings of the 2nd International Workshop on the Numerical Solution of Markov Chains, Raleigh, North Carolina, pp.21-38, 1995.

[3] D. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems, *SIAM Journal on Matrix Analysis and Applications*, Vol. 17, No. 4, pp. 906-926, 1996.

[4] R. Chakka. *Performance and Reliability Modelling of Computing Systems Using Spectral Expansion*, Ph.D. Thesis, University of Newcastle upon Tyne, 1995. Available from the PDCS project, Department of Computing Science, University of Newcastle upon Tyne, as Technical Report No. 179.

[5] R. Chakka and P.G. Harrison. Analysis of MMPP/M/c/L queues, Proceedings of the Twelfth UK Computer and Telecommunications Performance Engineering Workshop, Edinburgh, pp. 117-128, 1996.

[6] J.M. Fourneau, E. Gelenbe and R. Suros. G-networks with multiple classes of positive and negative customers, *Theoretical Computer Science*, Vol. 155, pp. 141–156, 1996.

[7] J.M. Fourneau, M. Hernandez. Modelling defective parts in a flow system using G-networks. *Second International Workshop on Performability Modelling of Computer and Communication Systems*, Le Mont Saint-Michel, June 1993.

[8] Fretwell, R.J. and Kouvatsos, D.D. Correlated Traffic Modelling and Batch Renewal Markov Modulated Processes, Proc. 4th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks, Chapman & Hall, pp. 20-44, 1997.

[9] R. Fretwell and D. Kouvatsos. ATM traffic burst lengths are geometrically bounded, *Proceedings of the 7 th IFIP Workshop on Performance Modelling and Evaluation of ATM & IP Networks*, Antwerp, Belgium, 1999.

[10] E. Gelenbe. Random neural networks with positive and negative signals and product form solution, *Neural Computation*, Vol. 1, No. 4, pp. 502-510, 1989.

[11] E. Gelenbe. Queueing networks with negative and positive customers, *Journal of Applied Probability* **28**, pp. 656-663, 1991.

[12] E. Gelenbe. G-Networks with signals and batch removal, *Probability in the Engineering and Informational Sciences* **7**, pp. 335-342, 1993.

[13] E. Gelenbe, P. Glynn and K. Sigman. Queues with negative arrivals, *Journal of Applied Probability* **28**, pp. 656-663, 1991.

[14] P.G. Harrison and N.M. Patel. *Performance modelling of communication networks and computer architectures*, Addison-Wesley, 1993.

[15] P.G. Harrison, N.M. Patel and E. Pitel. Reliability modelling using G-queues, *European Journal of Operations Research*, **126**, pp.273-287, 2000.

[16] P.G. Harrison and E. Pitel. Sojourn times in single server queues with negative customers, *Journal of Applied Probability* **30**, pp. 943-963, 1993.

[17] P.G. Harrison and E. Pitel. The $M/G/1$ queue with negative customers, *Advances in Applied Probability* **28**, pp. 540-566, 1996.

[18] B. Haverkort and A.Ost. Steady state analyses of infinite stochastic petri nets: A comparison between the Spectral Expansion and the Matrix Geometric methods, in Proceedings of the 7th International Workshop on Petri Nets and Performance Models, Saint Malo, France, pp. 335-346, 1997.

[19] D. Kouvatsos. Entropy maximisation and queueing network models, *Annals of Operations Research*, **48**, pp. 63-126, 1994.

[20] D. Kouvatsos, J. Wilkinson, P.G. Harrison and M.K. Bhabuta. Performance Analysis of Buffered Banyan ATM Switch Architectures, in *Performance Modelling and Evaluation of ATM Networks, Vol. II*, D. Kouvatsos, ed., Chapman-Hall, 1996.

[21] K.S. Meier-Hellstern. The analysis of a queue arising in overflow models, *IEEE Transactions on Communications* **37**, pp. 367–372, 1989.

[22] B.B. Mandelbrot and J.W.V. Ness. Fractional Brownian Motions, fractional noises and applications, *SIAM Review* **10,4** pp. 422-437, 1968.

[23] I. Mitrani. *Probabilistic Modelling*, Cambridge University Press, 1998.

[24] I. Mitrani and R. Chakka. Spectral expansion solution for a class of Markov models: Application and comparison with the matrix-geometric method, *Performance Evaluation* **23** pp. 241-260, 1995.

[25] V. Naoumov, U. Krieger and D. Wagner. Analysis of a multi-server delay-loss system with a general Markovian arrival process, in *Matrix-analytical methods in Stochastic models*, S.R. Chakravarthy and A.S. Alfa eds., Marcel Dekker, pp. 43-66, 1997.

[26] M. Neuts. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Dover Publications, 1995.

[27] I. Norros. A storage model with self-similar input, *Queueing Systems and their Applications* **16** pp. 387-396, 1994.

[28] T. Ryden. An EM algorithm for estimation in Markov-modulated Poisson processes, *Computational Statistics & Data Analysis* 21 pp. 431-447, 1996.

[29] W.J. Stewart. *Introduction to Numerical Solution of Markov Chains*, Princeton University Press, 1994.