Queueing Systems 41, 271–298, 2002

The MM CPP/GE/c G-Queue: Sojourn Time Distribution

P.G. HARRISON pgh@doc.ic.ac.uk
Department of Computing, Imperial College, London, UK

Received 15 June 2001; Revised 5 January 2002

Abstract. We obtain the sojourn time probability distribution function at equilibrium for a Markov modulated, multi-server, single queue with generalised exponential (GE) service time distribution and compound Poisson arrivals of both positive and negative customers. Such arrival processes can model both burstiness and correlated traffic and are well suited to models of ATM and other telecommunication networks. Negative customers remove (ordinary) customers in the queue and are similarly correlated and bursty. We consider both the cases where negative customers remove positive customers from the front and the end of the queue and, in the latter case, where a customer currently being served can and cannot be killed by a negative customer. These cases can model an unreliable server or load balancing respectively. The results are obtained as Laplace transforms and can be inverted numerically. The MM CPP/GE/c G-Queue therefore holds the promise of being a viable building block for the analysis of queues and queueing networks with bursty, correlated traffic, incorporating load balancing and node-failures, since the equilibrium behaviour of both queue lengths and response times can be determined in a tractable way.

Keywords: ATM, burstiness, correlated traffic, negative customers, queueing theory

1. Introduction

We consider a multi-server queue with generalised exponential (GE) service times and with both positive and negative arrival streams, each of which is a (special case of a) compound Poisson process (CPP), i.e. a Poisson point process with batch arrivals of geometrically distributed size. In other words, inter-arrival times are also GE random variables. In addition, all three GE distributions (for positive and negative inter-arrival times and for service time) are modulated by a continuous time Markov phase process (CTMP). This queue is termed the MM CPP/GE/c G-queue and its equilibrium queue length probability distribution is obtained in [2], along with a study of its departure process. Negative customers remove (positive) customers in the queue and have been used to model random neural networks, task termination in speculative parallelism and faulty components in manufacturing systems [1,3–5]. The name G-queue has been adopted for queues with negative customers and a collection of papers on the subject may be found in [6]. The MM CPP/GE/c queueing model can account for burstiness and correlation, but in addition, the negative customers can represent additional behaviours such as breakdowns of unreliable servers, killing signals and load balancing. We derive the sojourn time probability distribution for positive customers that are not killed as a Laplace transform, which can be inverted numerically. We consider both the cases where negative customers remove positive customers from the front and from the end (tail) of the queue and, in the latter case, the two sub-cases in which a customer currently being served can and cannot be killed by a negative customer – i.e. are *immune.* The former case can be used to model an unreliable server where a negative arrival causes the task in service to be lost and also some others in the front part of the queue. The second case, especially the subcase where customers in service are immune, can model load balancing: Markov phases representing high load cause the injection of negative customers together with corresponding positive customers elsewhere in a network. The MM CPP/GE/c G-Queue therefore holds the promise of being a viable building block for the analysis of queues and hence (approximately) queueing networks with bursty, correlated traffic, incorporating load balancing and node-failures, since the equilibrium behaviour of both queue lengths and response times can be determined in a tractable way.

Sojourn time distributions in G-queues were first considered by Harrison and Pitel who considered single Markovian queues with negative arrivals in [9] and tandem networks thereof in [10], with numerically tractable results in the former and in certain special cases of the latter. The results presented here are initially obtained conditional on the state of the system seen by a new positive arrival and then deconditioned using the steady state queue length probability distribution. This is found using the method of spectral expansion in [2]. In the next section, the MM CPP/GE/c G-queue is fully defined, including the alternative killing semantics discussed above. Sections 3 and 4 consider sojourn time distributions when customers are killed at the front and rear of the queue respectively. The paper concludes in section 5 with a short discussion of the implications of our results on the modelling of various types of telecommunication networks. The appendix gives some technical details arising in the analysis of inimmune killing at the front of the queue.

2. Problem definition

2.1. Modulation

The entire system is 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

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

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:

$$\mathbf{r}Q = 0; \qquad \mathbf{r}\mathbf{e}_N = 1$$

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

2.2. *The arrival processes*

We consider two CPP arrival processes of positive and negative customers, modulated by the same Markov chain, with N phases and generator matrix Q. A negative customer removes a positive customer in the queue or being served according to a specified *killing discipline.* We consider here the RCH killing discipline, i.e. removal of customers from the head (or front) of the queue, and two variants of the RCE killing discipline, *viz.* removal of customers from the end of the queue, where the most recent positive arrivals are removed. The first variant of RCE 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. The second variant 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 idle servers. Let (λ_i, θ_i) and (κ_i, ρ_i) be the parameters of the GE inter-arrival time distributions in phase i for the positive and negative arrival streams respectively, i.e. the respective inter-arrival time probability distribution functions are $1 - (1 - \theta_i)\mathrm{e}^{-\lambda_i t}$ and $1 - (1 - \rho_i)\mathrm{e}^{-\kappa_i t}$ in phase i $(t \geqslant 0)$. Thus, the arrival *point-processes* are Poisson with batch arrivals at each point having geometric size – specifically, size $k + 1$ with probability $(1 - \theta_i)\theta_i^k$ and $(1 - \rho_i)\rho_i^k$ for positive and negative customers, respectively, in phase i. The total average positive and negative arrival rates, $\overline{\lambda}$ and $\overline{\kappa}$, are therefore

$$\overline{\lambda} = \sum_{i=1}^{N} \frac{r_i \lambda_i}{1 - \theta_i} \quad \text{and} \quad \overline{\kappa} = \sum_{k=1}^{N} \frac{r_i \kappa_i}{1 - \rho_i}.$$

In the sequel, we use the notation

$$\begin{aligned}
\Lambda &= diag(\lambda_1, \lambda_2, \ldots, \lambda_N),\\
K &= diag(\kappa_1, \kappa_2, \ldots, \kappa_N),\\
\Theta &= diag(\theta_1, \theta_2, \ldots, \theta_N),\\
R &= diag(\rho_1, \rho_2, \ldots, \rho_N).
\end{aligned}$$

2.3. The server

We consider c parallel identical servers, each having GE-distributed service times with parameters (μ_i, ϕ_i) in phase i of the modulating Markov chain. We do not lose generality by associating the parameters of the server's GE distribution with the phase of the arrival processes: if they were independently modulated, we could take the Kronecker product of the three independent constituent Markov chains to modulate both arrival processes and service times. The service discipline is FCFS and each server serves at most one customer at a time. We write

$$M = diag(\mu_1, \mu_2, \ldots, \mu_N),$$
$$\Phi = diag(\phi_1, \phi_2, \ldots, \phi_N).$$

The operation of the GE server is similar to that described for the CPP arrival process except that 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 $j \geqslant c$ (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. Thus, the probability that a departing batch has size s is $(1-\phi)\phi^{s-1}$ for $1 \leqslant s \leqslant j-c$ and ϕ^{j-c} for $s = j-c+1$. When $j \leqslant c$, the departing batch has size 1 with probability one. 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. The steady state solution for queue lengths

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 process, and $J(t) \geqslant 0$ represents the number of customers in the system at time t, including any in service. The system evolves as a continuous time, discrete state Markov chain on a semi-infinite lattice strip. We denote the steady state probability for state (k, j) by

$$p_{k,j} = \lim_{t\to\infty} \mathrm{Prob}\big(I(t) = k,\ J(t) = j\big)$$

and for queue length $j \geqslant 0$, we define the vector $\mathbf{v}_j = (p_{1j}, \ldots, p_{Nj})$. The method of spectral expansion [11,12] then yields the solution

$$\mathbf{v}_j = \sum_{k+1}^{N} a_k \boldsymbol{\psi}_k \xi_k^j, \quad j = c, c+1, \ldots, \tag{1}$$

where the ξ_k $(k = 1, 2, \ldots, N)$ are the N eigenvalues strictly within the unit circle and the $\boldsymbol{\psi}_k$ are the corresponding left-eigenvectors of the matrix polynomial equation

$$\boldsymbol{\psi}\big(Q_0 + Q_1\xi + Q_2\xi^2 + Q_3\xi^3\big) = 0 \tag{2}$$

for certain matrices Q_0, Q_1, Q_2 and Q_3. The a_k are arbitrary constants so chosen that all the balance equations are satisfied [11]. An alternate method of solution is that of [13]. However, this method is based on the matrix-geometric method [14] and leads to a solution for the steady state probabilities in terms of *matrix* powers. We shall see that a solution expressed in terms of scalar powers is required for our analysis of sojourn time distributions and so the method of spectral analysis is preferred.

3. RCH killing discipline

We interpret RCH killing as representing the failure of one of the c servers with loss of any customer in service there and possible loss of customers waiting to commence service, if any. Thus, the negative arrival process in phase k is regarded as a superposition of c independent compound Poisson processes with parameters $(\kappa_k/c, \rho_k)$ at each server.[1] We consider the passage of a special "tagged" customer through the queue. Ultimately, this customer will either be served or killed by a negative customer; we require the probability distribution function of the (virtual) sojourn time of customers that are not killed, i.e. the time elapsed between the arrival instant and the completion of service. Let the random variable $A(x)$ denote the number of customers ahead of the tagged customer at time x and the random variable T denote the time remaining, without loss of generality at time 0, up to the departure of the tagged customer. For $j \geqslant 0$, we define the probability distributions $\mathbf{F}_j(t) = (F_{1j}(t), \ldots, F_{Nj}(t))$ where, for $1 \leqslant k \leqslant N$,

$$F_{kj}(t) = P\big(T \leqslant t \mid I(0) = k,\ A(0) = j\big).$$

Now, when the state is (k, j), we consider an initial small interval of length h and derive an expression for $\mathbf{F}_j(t+h)$ in terms of $\{\mathbf{F}_a(t) \mid a \geqslant 0\}$. By the Markov property and stationarity, we can write, for $j \geqslant c$:

$$\begin{aligned}\mathbf{F}_j(t+h) = {} & (I + Qh - Kh - cMh)\mathbf{F}_j(t) \\ & + h \sum_{s=1}^{j-c+1} K(I-R)R^{s-1}\mathbf{F}_{j-s}(t) + hKR^{j-c+1}\mathbf{0} \\ & + h \sum_{s=1}^{j-c+1} cM(I-\Phi)\Phi^{s-1}\mathbf{F}_{j-s}(t) + hcM\Phi^{j-c+1}\mathbf{e}_N + \mathrm{o}(h) \end{aligned} \tag{3}$$

where $\mathbf{0}$ is the zero-vector of appropriate length. Notice that if a batch of negative customers arrives with size $j-c+2$ or more (removing the one in service, $j-c$ queueing in front of the tagged customer and the tagged customer), the tagged customer is killed and so does *not* complete service in time less than t – i.e. does so with probability 0. Similarly, if there is a batch service completion of $j-c+2$ or more customers, the tagged customer completes service in time less than t with probability 1. Taking the term $I\mathbf{F}j(t)$

[1] Recall that negative arrivals at an idle server have no effect. The rate at each server is constant, κ_k/c in phase k for all queue lengths, consistent with a breakdown rate.

to the left-hand side in equation (3) and dividing by h now yields, for $j \geqslant c$, the vector differential-difference equation:

$$\frac{\mathrm{d}\mathbf{F}_j(t)}{\mathrm{d}t} = (Q - K - cM)\mathbf{F}_j(t) + K(I - R)\sum_{s=1}^{j-c+1} R^{s-1}\mathbf{F}_{j-s}(t) + cM(I - \Phi)\sum_{s=1}^{j-c+1} \Phi^{s-1}\mathbf{F}_{j-s}(t) + cM\Phi^{j-c+1}\mathbf{e}_N. \tag{4}$$

For $j < c$, the tagged customer is in service and so his progress is influenced solely by the negative arrivals and the service completion time at his particular server. Thus, $\mathbf{F}_j(t) = \mathbf{F}_0(t)$ for all $0 \leqslant j < c$ and we have

$$\mathbf{F}_0(t + h) = \big(I + Qh - Mh - (K/c)h\big)\mathbf{F}_0(t) + hM\mathbf{e}_N + h(K/c)\mathbf{0} + \mathrm{o}(h).$$

This yields the vector differential equation

$$\frac{\mathrm{d}\mathbf{F}_0(t)}{\mathrm{d}t} = (Q - M - K/c)\mathbf{F}_0(t) + M\mathbf{e}_N, \tag{5}$$

i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathrm{e}^{-(Q-M-K/c)t}\mathbf{F}_0(t)\big) = \mathrm{e}^{-(Q-M-K/c)t}M\mathbf{e}_N$$

which has solution

$$\mathbf{F}_0(t) = \big(1 - \mathrm{e}^{-(M+K/c-Q)t}\big)(M + K/c - Q)^{-1}M\mathbf{e}_N$$

since $\mathbf{F}_0(0) = \mathbf{0}$.[2]

We define the Laplace transform vector of the distribution functions $\mathbf{F}_j(t)$ by

$$\mathbf{L}_j(s) = \left(\int_0^\infty \mathrm{e}^{-st}F_{1j}(t)\,\mathrm{d}t, \ldots, \int_0^\infty \mathrm{e}^{-st}F_{Nj}(t)\,\mathrm{d}t\right).$$

Then, the Laplace transform of the derivative with respect to t, i.e. of the vector of probability density functions $\mathbf{F}'_j(t) = (F'_{1j}(t), \ldots, F'_{Nj}(t))$, is $s\mathbf{L}_j(s)$ by a simple integration by parts, since $\mathbf{F}_j(0) = 0$ for all $j \geqslant 0$.

3.1. Recurrence formula and its solution

We can now derive recurrence formulas for the $\mathbf{L}_j(s)$ which we solve using the generating function method. The Laplace transform of the equilibrium sojourn time distribution

[2] When $A(0) = 0$, it is understood that the tagged customer is actually undergoing a nonzero service period, i.e. that the state is not instantaneous.

then follows directly in terms of these generating functions. To this end, we define the generating function vector $\mathbf{D}$ (one component per phase) by

$$\mathbf{D}(z, s) = \sum_{j=c}^{\infty} \mathbf{L}_j(s) z^j.$$

The following proposition determines this generating function.

Proposition 1. The generating function $\mathbf{D}(z, s)$ is given by the equation

$$C(z, s)\mathbf{D}(z, s) = \mathbf{B}(z, s)$$

where

$$\begin{aligned}
C(z, s) &= S - Q + K + cM - K(I - R)(I - Rz)^{-1}z - cM(I - \Phi)(I - \Phi z)^{-1}z,\\
\mathbf{B}(z, s) &= \frac{cz^c}{s} M\Phi(I - \Phi z)^{-1}\mathbf{e}_N + \frac{z^c}{s}\big[K(I - R)(I - Rz)^{-1}\\
&\quad + cM(I - \Phi)(I - \Phi z)^{-1}\big](S - Q + M + K)^{-1} M\mathbf{e}_N,\\
S &= sI.
\end{aligned}$$

Proof. Taking the Laplace transform of equation (4), we get, for $j \geqslant c$:

$$\begin{aligned}
s\mathbf{L}_j(s) &= (Q - K - cM)\mathbf{L}_j(s) + K(I - R)\sum_{s=1}^{j-c+1} R^{s-1}\mathbf{L}_{j-s}(s)\\
&\quad + cM(I - \Phi)\sum_{s=1}^{j-c+1} \Phi^{s-1}\mathbf{L}_{j-s}(s) + \frac{c}{s} M\Phi^{j-c+1}\mathbf{e}_N.
\end{aligned} \tag{6}$$

Multiplying by z^j and summing from $j = c$ to ∞ now gives (omitting function arguments for brevity, changing the order of summation, replacing the summation variable j by $j + s$ and summing the last term)

$$\begin{aligned}
&(S - Q + K + cM)\mathbf{D}\\
&\quad = K(I - R)\sum_{s=1}^{\infty}\sum_{j=c-1}^{\infty} R^{s-1}z^s\mathbf{L}_j z^j + cM(I - \Phi)\sum_{s=1}^{\infty}\sum_{j=c-1}^{\infty} \Phi^{s-1}z^s\mathbf{L}_j z^j\\
&\qquad + \frac{cz^c}{s} M\Phi(I - \Phi z)^{-1}\mathbf{e}_N\\
&\quad = K(I - R)(I - Rz)^{-1}z\big[\mathbf{L}_{c-1}z^{c-1} + \mathbf{D}\big]\\
&\qquad + cM(I - \Phi)(I - \Phi z)^{-1}z\big[\mathbf{L}_{c-1}z^{c-1} + \mathbf{D}\big] + \frac{cz^c}{s} M\Phi(I - \Phi z)^{-1}\mathbf{e}_N.
\end{aligned} \tag{7}$$

Taking the Laplace transform of equation (5) gives

$$s\mathbf{L}_0(s) = (Q - M - K)\mathbf{L}_0(s) + \frac{M\mathbf{e}_N}{s}$$

so that

$$\mathbf{L}_0(s) \equiv \mathbf{L}_{c-1}(s) = (S - Q + M + K)^{-1} \frac{M\mathbf{e}_N}{s}.$$

The result then follows by collecting the $\mathbf{D}$ terms in equation (7). □

3.2. Existence and uniqueness of the solution

A unique solution for $\mathbf{D}(z, s)$ exists if and only if the matrix $C(z, s)$ of proposition 1 has an inverse for all z and s, i.e. if its determinant is always nonzero. It is easy to verify that, given any z and s and omitting these arguments for brevity, $C = -(Q + \Delta)$ where $\Delta = diag(d_0, \dots, d_{N-1})$ is a diagonal matrix with every diagonal element $d_i < 0$ $(0 \leqslant i \leqslant N - 1)$.

Proposition 2. The determinant of C is nonzero.

Proof.[3] Let $\mathbf{x}$ be a right-eigenvalue of $Q + \Delta$ and suppose that $\mathbf{x}$ has eigenvalue equal to zero, so that $(Q + \Delta)\mathbf{x} = \mathbf{0}$. Let integer m $(0 \leqslant m \leqslant N - 1)$ be such that $|x_m| \geqslant |x_i|$ for i, $0 \leqslant i \leqslant N - 1$. Then,

$$|q_{mm} + d_m|\,|x_m| = \left|\sum_{j \neq m} q_{mj} x_j\right| \leqslant \sum_{j \neq m} |q_{mj}|\,|x_j| \leqslant |x_m| \sum_{j \neq m} |q_{mj}|.$$

Hence, since $x_m \neq 0$ and Q is a stochastic matrix,

$$-q_{mm} - d_m \leqslant \sum_{j \neq m} q_{mj}$$

and so $-d_m \leqslant 0$ which is a contradiction. Therefore the eigenvalues of $Q + \Delta$, and hence of C, are all nonzero. □

A unique solution for $\mathbf{D}(z, s)$ therefore always exists.

3.3. Equilibrium response time distribution

Let the Laplace transform of the unconditional sojourn time density function of a randomly selected arrival be denoted $L(s)$. This requires the probability mass function vector $\boldsymbol{\alpha}_j$ for the number of customers in front of the tagged arrival, j, at its arrival instant in each phase (corresponding to the components of this vector). Thus we require

$$L(s) = \sum_{j=0}^{\infty} \boldsymbol{\alpha}_j \cdot \mathbf{L}_j(s).$$

[3] My thanks to Ash Argent-Katwala, my Ph.D. student, for this proof.

Theorem 1.

$$L(s) = (I-\Theta)\mathbf{L}_0(s)\cdot\sum_{j=0}^{c-1}\Theta^j\sum_{q=0}^{j}\Theta^{-q}\mathbf{v}'_q + (I-\Theta)\mathbf{D}(\Theta,s)\cdot\sum_{q=0}^{c-1}\Theta^{-q}\mathbf{v}'_q$$
$$+ (I-\Theta)\Theta^{-1}\sum_{k=1}^{N}a_k\xi_k\big(I-\xi_k\Theta^{-1}\big)^{-1}\big[\big(\xi_k\Theta^{-1}\big)^{c-1}\mathbf{D}(\Theta,s)-\mathbf{D}(\xi_k,s)\big]\cdot\psi'_k$$

where $\mathbf{D}(\Theta,s) = (D_1(\theta_1,s),\ldots,D_N(\Theta_N,s))$, $\mathbf{v}'_q = \Lambda\mathbf{v}_q/\lambda^*$ and λ^* is that mean batch arrival time.

Proof. We may write, omitting the argument s for brevity,

$$L = \sum_{j=0}^{c-1}\boldsymbol{\alpha}_j\cdot\mathbf{L}_j + \sum_{j=c}^{\infty}\boldsymbol{\alpha}_j\cdot\mathbf{L}_j.$$

Now, at equilibrium, the probability that the phase is k and the queue length is q, just before an arrival instant (of a batch), is equal to the probability flux out of the state (k,q) divided by the total flux due to arrivals (in all states). This probability is therefore $\lambda_k p_{kq}/\lambda^*$ where

$$\lambda^* = \sum_{h=1}^{N}\sum_{q=0}^{\infty}\lambda_h p_{hq} = \sum_{h=1}^{N}r_h\lambda_h$$

is the average batch arrival rate. Recall that p_{kq} is the steady state probability distribution for the phase (k) and queue length (q), as per section 2.4, where we wrote $\mathbf{v}_q = (p_{1q},\ldots,p_{Nq})$. We now define

$$\mathbf{v}'_q = \left(\frac{\lambda_1 p_{1q}}{\lambda^*},\ldots,\frac{\lambda_N p_{Nq}}{\lambda^*}\right) = \frac{\Lambda\mathbf{v}_q}{\lambda^*}.$$

The total number of customers in front of the tagged arrival (j) is the sum of the queue length already present and the number in front *within* the arriving batch. Because batch sizes are geometric, the latter number of customers is distributed as the whole batch size. Hence,

$$\boldsymbol{\alpha}_j = (I-\Theta)\Theta^j\sum_{q=0}^{j}\Theta^{-q}\mathbf{v}'_q.$$

For $j \geqslant c$, this may be written

$$\boldsymbol{\alpha}_j = (I-\Theta)\Theta^j\left[\sum_{q=0}^{c-1}\Theta^{-q}\mathbf{v}'_q + \sum_{q=c}^{j}\sum_{k=1}^{N}a_k\boldsymbol{\psi}'_k\Theta^{-q}\xi_k^q\right]$$

where $\boldsymbol{\psi}'_k = \Lambda\boldsymbol{\psi}_k/\lambda^*$. Now, $L_j = L_0 = (S - Q + M + K)^{-1}M\mathbf{e}_N/s$ for all $j \leqslant c - 1$, and so

$$\begin{aligned} L &= (I-\Theta)\mathbf{L}_0 \cdot \sum_{j=0}^{c-1}\Theta^j \sum_{q=0}^{j}\Theta^{-q}\mathbf{v}'_q + (I-\Theta)\sum_{j=c}^{\infty}\Theta^j\mathbf{L}_j \cdot \sum_{q=0}^{c-1}\Theta^{-q}\mathbf{v}'_q \\ &\quad + (I-\Theta)\sum_{j=c}^{\infty}\Theta^j\mathbf{L}_j \cdot \sum_{q=c}^{j}\sum_{k=1}^{N} a_k\boldsymbol{\psi}'_k\big(\xi_k\Theta^{-1}\big)^q \\ &= (I-\Theta)\mathbf{L}_0 \cdot \sum_{j=0}^{c-1}\Theta^j \sum_{q=0}^{j}\Theta^{-q}\mathbf{v}'_q + (I-\Theta)\mathbf{D}(\Theta, s) \cdot \sum_{q=0}^{c-1}\Theta^{-q}\mathbf{v}'_q \\ &\quad + (I-\Theta)\sum_{k=1}^{N} a_k\boldsymbol{\psi}'_k\big(\xi_k\Theta^{-1}\big) \cdot \sum_{j=c}^{\infty}\Theta^j\mathbf{L}_j\big(I - \xi_k\Theta^{-1}\big)^{-1}\Big[\big(\xi_k\Theta^{-1}\big)^{c-1} - \big(\xi_k\Theta^{-1}\big)^j\Big] \end{aligned}$$

and the result then follows by definition of $\mathbf{D}(\cdot, s)$ evaluated at Θ and ξ_k. □

3.4. Special case

To compare with a known result, suppose that there are no negative customers ($K = 0$), no modulation ($Q = 0$, $N = 1$) and no burstiness ($\Theta = \Phi = 0$). The result should be that of the classical M/M/c queue, see, for example, [7]. The vector of queue length probabilities has one component, say $\mathbf{v}_j = (p_j)$ and $\mathbf{v}'_j = \mathbf{v}_j$. Then, since $\Theta = 0$, $\alpha_j = p_j$ (as expected by the Random Observer Property in the classical queue) and $p_j = a\xi^j$ for $j \geqslant c$ where $\xi = \lambda/(c\mu)$, $\Lambda = (\lambda)$ and $M = (\mu)$. The constant a is such that the queue length probabilities sum to one, i.e.

$$\sum_{j=0}^{c-1} p_j + \frac{a\xi^c}{1-\xi} = 1$$

yielding $a = \xi^{-c}(1-\xi)(1 - \sum_{j=0}^{c-1} p_j)$. The one-component generating function vector $\mathbf{D}(z, s) = (D(z, s))$ and similarly $\mathbf{L}_j(s) = (L_j(s))$ for $j \geqslant 0$. Thus, since $\mathbf{D}(z, s)/z^k \to \mathbf{0}$ as $z \to 0$ for $k < c$,

$$L(s) = L_0(s)\sum_{j=0}^{c-1} p_j + aD(\xi, s)$$

since the unit eigenvector $\psi_1 = (1)$. Proposition 1 now gives $C(z, s)D(z, s) = B(z, s)$ where $C(z, s) = s + c\mu(1 - z)$ and $B(z, s) = (z^c/s)c\mu^2(s + \mu)^{-1}$. Hence

$$D(\xi, s) = \frac{\xi^c c\mu^2}{s(s+\mu)(s + c\mu - \lambda)}$$

giving the result

$$L(s) = \frac{\mu}{s(s+\mu)}\left[q + (1-q)\frac{c\mu - \lambda}{s + c\mu - \lambda}\right]$$

where $q = \sum_{j=0}^{c-1} p_j$ is the probability that an arrival does not have to queue. This is the standard result required.

4. RCE killing discipline

When customers are removed from the end of the queue, any that may be behind the tagged customer influence its progress, its probability of being killed and hence its sojourn time; these customers may be seen as offering the tagged customer "protection" from negative arrivals, which cannot occur under RCH. To account for this, we define the random variables $A(x)$ and $B(x)$ to be the numbers of customers ahead of and behind the tagged customer respectively at time x. Thus, $J(x) = A(x) + B(x) + 1$ in the notation of section 2.4. Let the random variable T denote the time remaining, without loss of generality at time 0, up to the departure of the tagged customer. For $i, j \geqslant 0$, we define the probability distributions $\mathbf{F}_{ij}(t) = (F_{1ij}(t), \ldots, F_{Nij}(t))$ where, for $1 \leqslant k \leqslant N$,

$$F_{kij}(t) = P\big(T \leqslant t \mid I(0) = k, B(0) = i, A(0) = j\big).$$

Similarly to the RCH case, when the state is (k, i, j), we consider an initial small interval of length h and derive an expression for $\mathbf{F}_{ij}(t+h)$ in terms of $\{\mathbf{F}_{ab}(t) \mid a, b \geqslant 0\}$. We can then write, for $j \geqslant c$:

$$\begin{aligned}\mathbf{F}_{ij}(t+h) = {} & (I + Qh - \Lambda h - Kh - cMh)\mathbf{F}_{ij}(t) + h\sum_{s=1}^{\infty} \Lambda(I-\Theta)\Theta^{s-1}\mathbf{F}_{i+s,j}(t) \\ & + h\sum_{s=1}^{i} K(I-R)R^{s-1}\mathbf{F}_{i-s,j}(t) + hKR^{i}\mathbf{0} \\ & + h\sum_{s=1}^{j-c+1} cM(I-\Phi)\Phi^{s-1}\mathbf{F}_{i,j-s}(t) + hcM\Phi^{j-c+1}\mathbf{e}_N + \mathrm{o}(h). \end{aligned} \tag{8}$$

Notice in this case that if a batch of negative customers arrives with size $i+1$ or more, the tagged customer is killed and so does *not* complete service in time less than t – i.e. does so with probability 0. As with RCH killing, if there is a batch service completion of $j-c+2$ or more customers (the one in service, $j-c$ queueing in front of the tagged customer and the tagged customer), the tagged customer completes service in time less than t with probability 1. These equations hold for $j \geqslant c$ whether or not customers in service can be killed and yield:

$$\begin{aligned}\frac{\mathrm{d}\mathbf{F}_{ij}(t)}{\mathrm{d}t} &= (Q-\Lambda-K-cM)\mathbf{F}_{ij}(t)+\Lambda(I-\Theta)\sum_{s=1}^{\infty}\Theta^{s-1}\mathbf{F}_{i+s,j}(t)\\ &\quad + K(I-R)\sum_{s=1}^{i}R^{s-1}\mathbf{F}_{i-s,j}(t)\\ &\quad + cM(I-\Phi)\sum_{s=1}^{j-c+1}\Phi^{s-1}\mathbf{F}_{i,j-s}(t)+cM\Phi^{j-c+1}\mathbf{e}_N.\end{aligned} \tag{9}$$

For $j \leqslant c-1$, the tagged customer is in service and we now consider the cases where customers in service cannot or may be killed by a negative arrival. In the immune case – customers in service cannot be killed – the remaining sojourn time of any customer in service is independent of both arrival processes and of the service completions at other servers. Thus, $\mathbf{F}_{ij}(t)=\mathbf{F}_{00}(t)$ for all $0\leqslant j\leqslant c-1$, $i\geqslant 0$, and we have

$$\mathbf{F}_{00}(t+h)=(I+Qh-Mh)\mathbf{F}_{00}(t)+hM\mathbf{e}_N+\mathrm{o}(h).$$

This yields the vector differential equation

$$\frac{\mathrm{d}\mathbf{F}_{00}(t)}{\mathrm{d}t}=-(M-Q)\mathbf{F}_{00}(t)+M\mathbf{e}_N, \tag{10}$$

i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathrm{e}^{(M-Q)t}\mathbf{F}_{00}(t)\big)=\mathrm{e}^{(M-Q)t}M\mathbf{e}_N$$

which has solution

$$\mathbf{F}_{00}(t)=\big(1-\mathrm{e}^{-(M-Q)t}\big)(M-Q)^{-1}M\mathbf{e}_N$$

since $\mathbf{F}_{00}(t)=\mathbf{0}$.[4]

In the inimmune case, for $j\leqslant c-1$, a service completion may occur in front of or behind the tagged customer, but neither can cause the tagged customer itself to depart since it occupies its own server. Since the customers in service are susceptible to being killed, the state of the queue behind the tagged customer influences the probability of the tagged customer being killed and so failing to complete service. The behaviour of this part of the queue must therefore be accounted for in the analysis and we obtain, for $j\leqslant c-1$:

(a) $i\geqslant c-j$

$$\begin{aligned}\mathbf{F}_{ij}(t+h) &= (I+Qh-\Lambda h-Kh-cMh)\mathbf{F}_{ij}(t)\\ &\quad + h\sum_{s=1}^{\infty}\Lambda(I-\Theta)\Theta^{s-1}\mathbf{F}_{i+s,j}(t)+h\sum_{s=1}^{i}K(I-R)R^{s-1}\mathbf{F}_{i-s,j}(t)\end{aligned}$$

[4] When $A(0)=B(0)=0$, it is understood that the tagged customer is actually undergoing a nonzero service period, i.e. that the state is not instantaneous.

$$
\begin{aligned}
&+ hjM \sum_{s=1}^{i+j+1-c} (I-\Phi)\Phi^{s-1}\mathbf{F}_{i-s+1,j-1}(t) \\
&+ hjM\Phi^{i+j+1-c}\mathbf{F}_{c-j-1,j-1}(t) \\
&+ h(c-j-1)M \sum_{s=1}^{i+j+1-c} (I-\Phi)\Phi^{s-1}\mathbf{F}_{i-s,j}(t) \\
&+ h(c-j-1)M\Phi^{i+j+1-c}\mathbf{F}_{c-j-2,j}(t) + hM\mathbf{e}_N + \mathrm{o}(h),
\end{aligned}
\tag{11}
$$

(b) $i \leqslant c - j - 1$

$$
\begin{aligned}
\mathbf{F}_{ij}(t+h) = {} & \big(I + Qh - \Lambda h - Kh - (i+j+1)Mh\big)\mathbf{F}_{ij}(t) \\
&+ h\sum_{s=1}^{\infty} \Lambda(I-\Theta)\Theta^{s-1}\mathbf{F}_{i+s,j}(t) + h\sum_{s=1}^{i} K(I-R)R^{s-1}\mathbf{F}_{i-s,j}(t) \\
&+ hjM\mathbf{F}_{i,j-1}(t) + hiM\mathbf{F}_{i-1,j}(t) + hM\mathbf{e}_N + \mathrm{o}(h).
\end{aligned}
\tag{12}
$$

Equations (11) and (12) lead to similar differential equations to (9) which will be used in section 4.1.2. Let the Laplace transform vector of the distribution functions $\mathbf{F}_{ij}(t)$ be:

$$
\mathbf{L}_{ij}(s) = \left(\int_0^\infty \mathrm{e}^{-st} F_{1ij}(t)\,\mathrm{d}t, \ldots, \int_0^\infty \mathrm{e}^{-st} F_{Nij}(t)\,\mathrm{d}t\right).
$$

Then, the Laplace transform of the vector of probability density functions, $\mathbf{F}'_{ij}(t) = (F'_{1ij}(t), \ldots, F'_{Nij}(t))$, is $s\mathbf{L}_{ij}(s)$, analogous to the RCH case.

4.1. Generating functions

As with the RCH case, we now derive recurrence formulas for the $\mathbf{L}_{ij}(s)$, which we solve using the generating function method. The Laplace transform of the equilibrium sojourn time distribution then follows in terms of these generating functions, the analysis being a little more complicated than in the RCH case due to the extra dimension introduced by the customers behind the tagged customer. We define the generating function vector $\mathbf{G}(y, z, s) = (G_1(y, z, s), G_2(y, z, s), \ldots, G_N(y, z, s))$ by

$$
\mathbf{G}(y, z, s) = \sum_{i=0}^{\infty}\sum_{j=c}^{\infty} \mathbf{L}_{ij}(s) y^i z^j.
$$

The following proposition determines this generating function, up to a certain function of y and s, viz. $\mathbf{Z}(y, s)$, whether or not customers in service can be killed. This function is considered below in sections 4.1.1 and 4.1.2.

Proposition 3. The generating function $\mathbf{G}(y, z, s)$ is given by the equation

$$V(y, z, s)\mathbf{G}(y, z, s) = cz^c M(I - z\Phi)^{-1}\left((I - \Phi)\mathbf{Z}(y, s) + \frac{\Phi \mathbf{e}_N}{s(1 - y)}\right) \\ - \Lambda(I - \Theta)(yI - \Theta)^{-1}\mathbf{G}(\Theta, z, s)$$

for all $y \neq \theta_i$ $(1 \leqslant i \leqslant N)$, where

$$\begin{aligned} V(y, z, s) &= S - Q + \Lambda + K + cM - \Lambda(I - \Theta)(yI - \Theta)^{-1} \\ &\quad - yK(I - R)(I - yR)^{-1} - zcM(I - \Phi)(I - z\Phi)^{-1}, \\ \mathbf{Z}(y, s) &= \sum_{i=0}^{\infty} \mathbf{L}_{i,c-1}(s) y^i, \\ \mathbf{G}(\Theta, z, s) &= \big(G_1(\theta_1, z, s), G_2(\theta_2, z, s), \ldots, G_N(\theta_N, z, s)\big), \\ S &= sI. \end{aligned}$$

Proof. Taking the Laplace transform of equation (9), we get

$$\begin{aligned} &(S - Q + \Lambda + K + cM)\mathbf{L}_{ij}(s) \\ &\quad = \Lambda(I - \Theta)\sum_{s=1}^{\infty}\Theta^{s-1}\mathbf{L}_{i+s,j}(s) + K(I - R)\sum_{s=1}^{i} R^{s-1}\mathbf{L}_{i-s,j}(s) \\ &\quad\quad + cM(I - \Phi)\sum_{s=1}^{j-c+1}\Phi^{s-1}\mathbf{L}_{i,j-s}(s) + \frac{cM}{s}\Phi^{j-c+1}\mathbf{e}_N. \end{aligned} \tag{13}$$

Multiplying throughout by $y^i z^j$ and summing over the domain $i \geqslant 0$ and $j \geqslant c$, we obtain:

$$\begin{aligned} &(S - Q + \Lambda + K + cM)\mathbf{G}(y, z, s) \\ &\quad = \Lambda(I - \Theta)\sum_{j=c}^{\infty}\sum_{i=0}^{\infty}\sum_{s=1}^{\infty}\Theta^{s-1}\mathbf{L}_{i+s,j}(s) y^i z^j \\ &\quad\quad + K(I - R)\sum_{j=c}^{\infty}\sum_{i=0}^{\infty}\sum_{s=1}^{i} R^{s-1}\mathbf{L}_{i-s,j}(s) y^i z^j \\ &\quad\quad + cM(I - \Phi)\sum_{j=c}^{\infty}\sum_{i=0}^{\infty}\sum_{s=1}^{j-c+1}\Phi^{s-1}\mathbf{L}_{i,j-s}(s) y^i z^j + \frac{cM}{s}\left(\sum_{j=c}^{\infty}\sum_{i=0}^{\infty}\Phi^{j-c+1} y^i z^j\right)\mathbf{e}_N. \end{aligned} \tag{14}$$

For the first term on the right-hand side, we change the summation variable i to $k = i + s$ and sum over the domain $1 \leqslant s \leqslant k$ and $k \geqslant 1$, leaving the domain of j unchanged, to get the term (omitting the arguments s for brevity):

$$\Lambda(I - \Theta)y^{-1} \sum_{j=c}^{\infty} \sum_{k=1}^{\infty} \sum_{s=1}^{k} (\Theta/y)^{s-1} \mathbf{L}_{k,j} y^k z^j = \Lambda(I - \Theta) \sum_{j=c}^{\infty} \sum_{k=1}^{\infty} \frac{y^k I - \Theta^k}{yI - \Theta} \mathbf{L}_{k,j} z^j$$

where the matrix $yI - \Theta$ in the denominator denotes multiplication by its inverse.[5] This simplifies to $\Lambda(I - \Theta)(yI - \Theta)^{-1}(\mathbf{G}(y, z, s) - \mathbf{G}(\Theta, z, s))$, as required for the Θ-terms. The second term is handled similarly by changing the summation domain to $\sum_{j=c}^{\infty} \sum_{s=1}^{\infty} \sum_{i=s}^{\infty}$ and then changing the last summation variable from i to $i + s$ so that the sum over s can be separated out. For the third term, the summation domain is written $\sum_{i=0}^{\infty} \sum_{s=1}^{\infty} \sum_{j=s+c-1}^{\infty}$ and the last summation variable is changed from j to $j + s$ giving

$$\begin{aligned} cM(I - \Phi) \sum_{i=0}^{\infty} \sum_{s=1}^{\infty} \sum_{j=c-1}^{\infty} z(z\Phi)^{s-1} \mathbf{L}_{i,j}(s) y^i z^i \\ = zcM(I - \Phi)(I - z\Phi)^{-1} \big(\mathbf{G}(y, z, s) - z^{c-1} \mathbf{Z}(y, s)\big). \end{aligned}$$

The last term is straightforward. □

Questions of existence and uniqueness of a solution for $\mathbf{G}$ are considered in section 4.2. The difference in the solution between the models in which customers in service can or cannot be killed lies in the computation of the function $\mathbf{Z}(y, s)$. To compute this we consider the recurrence formulas for queue lengths $j \leqslant c - 1$.

4.1.1. Immune servicing

When customers in service are immune from killing, taking the Laplace transform of equation (10) yields

$$s\mathbf{L}_{00}(s) = (Q - M)\mathbf{L}_{00}(s) + \frac{M\mathbf{e}_N}{s}$$

so that

$$\mathbf{L}_{00}(s) = (S - Q + M)^{-1} \frac{M\mathbf{e}_N}{s}.$$

[5] Notice that by the hypothesis that $y \neq \theta_i$, the inverse exists. If $y = \theta_i$, say, the first component of the expression becomes

$$\Lambda(I - \Theta) \sum_{j=c}^{\infty} \sum_{k=1}^{\infty} k L_{k,j,1} y^{k-1} z^j = \Lambda(I - \Theta) \frac{\mathrm{d}G_1}{\mathrm{d}y}.$$

Since $\mathbf{L}_{ij}(s) = \mathbf{L}_{00}(s)$ for all $j < c$ and $i \geqslant 0$, we have

$$\mathbf{Z}(y, s) = \sum_{i=0}^{\infty} \mathbf{L}_{i,c-1}(s) y^i = \frac{(S - Q + M)^{-1} M \mathbf{e}_N}{s(1 - y)}. \tag{15}$$

Substituting into proposition 3 then yields

$$\begin{aligned} &V(y, z, s)\mathbf{G}(y, z, s) \\ &\quad = cz^c M(I - z\Phi)^{-1} \left(\frac{(I - \Phi)(S - Q + M)^{-1} M}{s(1 - y)} + \frac{\Phi}{s(1 - y)} \right) \mathbf{e}_N \\ &\quad - \Lambda(I - \Theta)(yI - \Theta)^{-1} \mathbf{G}(\Theta, z, s) \end{aligned} \tag{16}$$

for all $y \neq \theta_i$ $(1 \leqslant i \leqslant N)$.

4.1.2. Inimmune servicing

When customers in service can be killed, we transform the differential equations arising from equations (11) and (12) to get, for $j \leqslant c - 1$ (dropping the argument s for brevity):

(a) If $i \geqslant c - j$

$$\begin{aligned} &(S - Q + \Lambda + K + cM)\mathbf{L}_{ij} \\ &\quad = \Lambda(I - \Theta) \sum_{s=1}^{\infty} \Theta^{s-1} \mathbf{L}_{i+s,j} + K(I - R) \sum_{s=1}^{i} R^{s-1} \mathbf{L}_{i-s,j} \\ &\quad + jM(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1} \mathbf{L}_{i-s+1,j-1} + jM\Phi^{i+j+1-c} \mathbf{L}_{c-j-1,j-1} \\ &\quad + (c - j - 1)M(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1} \mathbf{L}_{i-s,j} \\ &\quad + (c - j - 1)M\Phi^{i+j+1-c} \mathbf{L}_{c-j-2,j} + \frac{M\mathbf{e}_N}{s}. \end{aligned} \tag{17}$$

(b) If $i \leqslant c - j - 1$

$$\begin{aligned} &\big(S - Q + \Lambda + K + (i + j + 1)M\big) \mathbf{L}_{ij} \\ &\quad = \Lambda(I - \Theta) \sum_{s=1}^{\infty} \Theta^{s-1} \mathbf{L}_{i+s,j} + K(I - R) \sum_{s=1}^{i} R^{s-1} \mathbf{L}_{i-s,j} \\ &\quad + jM\mathbf{L}_{i,j-1} + iM\mathbf{L}_{i-1,j} + \frac{M\mathbf{e}_N}{s}. \end{aligned} \tag{18}$$

We then define the generating function

$$\mathbf{H}(y, z, s) = \sum_{j=0}^{c-1} \sum_{i=0}^{\infty} \mathbf{L}_{ij}(s) y^i z^j.$$

The dynamic behaviour of the system starting from states with queue length less than c can then be summarised in terms of the function $\mathbf{H}$. Since the applications of this variant are obscure, we do not provide a full solution here. However, proposition 4 in the appendix expresses the function $\mathbf{H}$ in terms of five finite polynomials in y and z, with coefficients given by the conditional Laplace transforms $\mathbf{L}_{ij}$. For the special case of a single server, i.e. $c = 1$, the result is simpler since the five polynomials all vanish to give[6]

$$W(y,s)\mathbf{H}(y,z,s) = \frac{M\mathbf{e}_N}{(1-y)s} - \Lambda(I-\Theta)(yI-\Theta)^{-1}\mathbf{H}(\Theta,z,s)$$

where

$$\begin{aligned} W(y,s) = {} & S - Q + M + \Lambda\big[I - (I-\Theta)(yI-\Theta)^{-1}\big] \\ & + K\big[I - y(I-R)(I-yR)^{-1}\big]. \end{aligned}$$

When $c \geqslant 2$, let

$$\mathbf{H}_j(y,s) = \sum_{i=0}^{\infty} \mathbf{L}_{ij}(s) y^i$$

so that

$$\mathbf{H}(y,z,s) = \sum_{j=0}^{c-1} \mathbf{H}_j(y,s) z^j.$$

Comparing coefficients of z^j in proposition 4 then gives, for $0 \leqslant j \leqslant c-1$,[7]

$$\begin{aligned} & \big\{S - Q + \Lambda\big[I - (I-\Theta)(yI-\Theta)^{-1}\big] + K\big[I - y(I-R)(I-yR)^{-1}\big] \\ & \quad + M\big[cI - (c-j-1)y(I-\Phi)(I-y\Phi)^{-1}\big]\big\}\mathbf{H}_j(y,s) \\ & \quad - j(I-\Phi)(I-y\Phi)^{-1}M\mathbf{H}_{j-1}(y,s) \\ & = \frac{M\mathbf{e}_N}{(1-y)s} - \Lambda(I-\Theta)(yI-\Theta)^{-1}\mathbf{H}_j(\Theta,s) \\ & \quad + \big[(c-j)I - (c-j-1)y(I-\Phi)(I-y\Phi)^{-1}\big]M \sum_{i=0}^{c-j-2} \mathbf{L}_{ij} y^i \\ & \quad + j\big[I - (I-\Phi)(I-y\Phi)^{-1}\big]M \sum_{i=0}^{c-j-1} \mathbf{L}_{i,j-1} y^i - (1-y)M \sum_{i=0}^{c-j-2} (i+1)\mathbf{L}_{ij} y^i \\ & \quad + (c-j-1)y^2\Phi(I-y\Phi)^{-1}M\mathbf{L}_{c-j-2,j} y^{c-j-2} \\ & \quad + jy\Phi(I-y\Phi)^{-1}M\mathbf{L}_{c-j-1,j-1} y^{c-j-1}. \end{aligned} \tag{19}$$

These equations must be solved for the $\mathbf{L}_{ij}(s)$ and hence $\mathbf{Z}(y,s)$; see section 4.4.2.

[6] In this case the result follows more easily from lemma 1, which precedes proposition 4.
[7] Comparing the coefficients of z^c yields the identity $\mathbf{H}_{c-1}(y,s) = \mathbf{Z}(y,s)$.

4.2. Equilibrium response time distribution

Again, let the unconditional sojourn time density function of a randomly selected arrival have Laplace transform $L(s)$. This requires the joint probability mass function vector $\boldsymbol{\alpha}_{ij}$ for the number of customers behind, i, and ahead, j, seen by the tagged arrival in each phase (corresponding to the components of the vector). Thus we require

$$L(s) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \boldsymbol{\alpha}_{ij} \cdot \mathbf{L}_{ij}(s).$$

Theorem 2. In the case of immune customer servicing,

$$\begin{aligned} L(s) = (I-\Theta)\mathbf{L}_{00}(s) \cdot \sum_{j=0}^{c-1} \Theta^{j} \sum_{q=0}^{j} \Theta^{-q}\mathbf{v}'_q + (I-\Theta)^2\mathbf{G}(\Theta,\Theta,s) \cdot \sum_{q=0}^{c-1} \Theta^{-q}\mathbf{v}'_q \\ + (I-\Theta)^2\Theta^{-1} \sum_{k=1}^{N} a_k\xi_k \big(I-\xi_k\Theta^{-1}\big)^{-1}\big[\big(\xi_k\Theta^{-1}\big)^{c-1}\mathbf{G}(\Theta,\Theta,s) \\ - \mathbf{G}(\Theta,\xi_k,s)\big] \cdot \boldsymbol{\psi}'_k \end{aligned}$$

where $\mathbf{G}(\Theta,\Theta,s) = (G_1(\theta_1,\theta_1,s), G_2(\theta_2,\theta_2,s), \ldots, G_N(\theta_N,\theta_N,s))$.

Proof. Again omitting the argument s for brevity, we have

$$L = \sum_{i=0}^{\infty}\sum_{j=0}^{c-1} \boldsymbol{\alpha}_{ij} \cdot \mathbf{L}_{ij} + \sum_{i=0}^{\infty}\sum_{j=c}^{\infty} \boldsymbol{\alpha}_{ij} \cdot \mathbf{L}_{ij}.$$

As in section 3.3, the equilibrium probability that the state is (k,q) just before an arrival instant is $\lambda_k p_{kq}/\lambda^*$ where $\lambda^* = \sum_{h=1}^{N} r_h\lambda_h$ is the average batch arrival rate. Again, we define

$$\mathbf{v}'_q = \left(\frac{\lambda_1 p_{1q}}{\lambda^*}, \ldots, \frac{\lambda_N p_{Nq}}{\lambda^*}\right) = \frac{\Lambda \mathbf{v}_q}{\lambda^*}.$$

The only way the tagged customer can see arrivals behind is if they are in the same batch and similarly the total number seen ahead is the sum of the queue length already present and the number in front *within* the arriving batch. Moreover, because batch sizes are geometric, the numbers of customers in the same batch ahead of and behind the randomly selected tagged arrival are independent and distributed as the whole batch size. We therefore have

$$\boldsymbol{\alpha}_{ij} = (I-\Theta)^2\Theta^{i+j} \sum_{q=0}^{j} \Theta^{-q}\mathbf{v}'_q.$$

For $j \geqslant c$ this may be written

$$\boldsymbol{\alpha}_{ij} = (I - \Theta)^2 \Theta^{i+j} \left[\sum_{q=0}^{c-1} \Theta^{-q} \mathbf{v}'_q + \sum_{q=c}^{j} \sum_{k=1}^{N} a_k \boldsymbol{\psi}'_k \Theta^{-q} \xi_k^q \right]$$

where $\boldsymbol{\psi}'_k = \Lambda \boldsymbol{\psi}_k / \lambda^*$. For immune customer servicing, we know that $L_{ij} = L_{00}$ for all i and $j \leqslant c - 1$, where L_{00} is known. Similarly, the function $\mathbf{Z}(y, s)$ of proposition 3 was easily determined in this case. We then obtain

$$L = (I - \Theta)\mathbf{L}_{00} \cdot \sum_{j=0}^{c-1} \Theta^j \sum_{q=0}^{j} \Theta^{-q} \mathbf{v}'_q + (I - \Theta)^2 \sum_{i=0}^{\infty} \sum_{j=c}^{\infty} \Theta^{i+j} \mathbf{L}_{ij} \cdot \sum_{q=0}^{c-1} \Theta^{-q} \mathbf{v}'_q$$
$$+ (I - \Theta)^2 \sum_{i=0}^{\infty} \sum_{j=c}^{\infty} \Theta^{i+j} \mathbf{L}_{ij} \cdot \sum_{q=c}^{j} \sum_{k=1}^{N} a_k \boldsymbol{\psi}'_k \big(\xi_k \Theta^{-1}\big)^q.$$

The result follows after some algebraic manipulation. □

4.3. Numerical solution

It now remains to find the vector function $\mathbf{G}(y, z, s)$ at $y = \Theta$. This can be done from proposition 3 using the analyticity of $\mathbf{G}$ inside the unit y-disk. The proposition can be written $V\mathbf{G} = \mathbf{U}$, or $\mathbf{G} = V^{-1}\mathbf{U}$, where

$$\mathbf{U}(y, z, s) = cz^c M(I - z\Phi)^{-1} \left((I - \Phi)\mathbf{Z}(y, s) + \frac{\Phi \mathbf{e}_N}{s(1 - y)} \right)$$
$$- \Lambda(I - \Theta)(yI - \Theta)^{-1} \mathbf{G}(\Theta, z, s).$$

This is a set of N equations for any particular choice of y, z, s. Thus, if $V(y, z, s)$ is singular for any choice of y, z with $|y| < 1$, $|z| < 1$, there must be a linear dependence amongst the equations on both sides, i.e. there must exist a vector $\boldsymbol{\eta}(y, z, s)$ such that $\boldsymbol{\eta} V = \mathbf{0}$ and $\boldsymbol{\eta} \cdot \mathbf{U} = 0$, for the equations to be consistent. The vector $\boldsymbol{\eta}(y, z, s)$ depends solely on $V(y, z, s)$ – essentially it is its left eigenvector for eigenvalue 0. This approach is a generalisation of that used in the scalar situation where we would have $G = U/V$ so that we must have $U = 0$ whenever $V = 0$ for any values of y, z inside their unit disks. In this way $G(\theta, z, s)$ would be determined for scalar θ.

In our case, we have N unknowns, $\{G_i(\theta_i, z, s) \mid 1 \leqslant i \leqslant N\}$ and one equation for each (y, z) pair in the unit disks that renders V singular. We therefore solve the equation $|V| = 0$ for y as a function of z. Suppose we obtain solutions inside the unit disks $y_i(z)$ for $i = 1, \ldots, N$. This yields corresponding vectors $\boldsymbol{\eta}_i(z)$ and substituting y_i for y in $\mathbf{U}$ we obtain N linear equations in the required vector $\mathbf{G}(\Theta, z, s)$, *viz.* $\boldsymbol{\eta}_i(z) \cdot \mathbf{U} = 0$, which can be solved.

The equation $|V(y, z, s)| = 0$ in y is of degree $2N$ and we postulate that there are exactly N solutions inside the unit y-disk in order that there exists a unique solution for $\mathbf{G}$. If there were less than N solutions in the unit disk, $\mathbf{G}(\Theta, z, s)$ would not be fully defined. If there were more than N, at most N of them could yield independent equations $\boldsymbol{\eta}_i(z) \cdot \mathbf{U} = 0$, otherwise there would be an inconsistency.

Results on the location of the eigenvalues would be interesting theoretically, but appear intractable. The more important issue is to find a numerical algorithm to compute the required N eigenvalues and eigenvectors efficiently and accurately. If it turned out that sufficient (independent) eigenvalues could not be found in any particular numerical application, the method would fail. If sufficient eigenvalues can be found, a solution is straightforward, at least in principle. This is exactly the same argument as is used in the Spectral Analysis method [11] and many other eigenvalue problems in engineering. In our case, many sets of these N eigenvalues are needed, one for each s value utilised by a Laplace transform inverter. This work is in progress. Fortunately, in our case, the dimension N of Q, i.e. the number of phases, is small, typically less than 8, and results have proved easy to obtain for a 2-phase model.

4.4. *Special cases*

4.4.1. *Immune servicing*

When customers in service cannot be killed, we consider the same special case as we did in section 3.4 for RCH killing, *viz.* the M/M/c queue. Since there are no negative arrivals, the result derived for L here must be the same as in section 3.4, i.e. the classical result. We therefore have the parameterisation $N = 1$, $Q = K = \Theta = \Phi = 0$. The vector of queue length probabilities has one component, say $\mathbf{v}_j = (p_j) = \mathbf{v}'_j$, and $p_j = a\xi^j$ for $\xi \geqslant c$ where $\xi = \lambda/(c\mu)$, $\Lambda = (\lambda)$ and $M = (\mu)$; the eigenvector $\boldsymbol{\psi}_1 = (1)$. As in section 3.4, $a = \xi^{-c}(1-\xi)(1-\sum_{j=0}^{c-1} p_j)$. The one-component generating function vector $\mathbf{G}(y,z,s) = (G(y,z,s))$ and similarly $\mathbf{L}_{ij}(s) = (L_{ij}(s))$ for $i, j \geqslant 0$. Thus, since $\mathbf{G}(y,z,s)/z^k \to \mathbf{0}$ as $z \to 0$ for $k < c$,

$$L = L_{00} \sum_{j=0}^{c-1} p_j + aG(0, \xi).$$

There is no modulation and so the analyticity conditions are conventional scalar ones rather than involving an analysis of eigenvectors as described in section 4.2. Proposition 3 yields $VG = U$ where $V = s + \lambda + c\mu(1-z) - \lambda/y$ and $U = c\mu z^c Z(y,s) - \lambda G(0,z,s)/y$ with $Z(y,s) = \mu/(s(s+\mu)(1-y))$. The analyticity condition is that $U = 0$ when $V = 0$, i.e. when

$$y = \frac{\lambda}{s + \lambda + c\mu(1-z)}.$$

This yields

$$G(0,z,s) = \frac{c\mu^2 z^c}{s(s+\mu)(s+c\mu(1-z))}$$

so that

$$G(0, \xi, s) = \frac{c\mu^2 \xi^c}{s(s+\mu)(s+c\mu-\lambda)}.$$

Since L_{00} is the same as the L_0 of section 3.4 and $G(0, \xi, s) = D(\xi, s)$, the same result for L follows as required.

4.4.2. Inimmune servicing

Inimmune servicing is much harder, but fortunately less important for practical applications! For the special case of a single server, i.e. $c = 1$, we had in section 4.1.2

$$W(y, s)\mathbf{H}(y, z, s) = \frac{M\mathbf{e}_N}{(1-y)s} - \Lambda(I-\Theta)(yI-\Theta)^{-1}\mathbf{H}(\Theta, z, s)$$

where

$$W(y, s) = S - Q + M + \Lambda\big[I - (I-\Theta)(yI-\Theta)^{-1}\big] + K\big[I - y(I-R)(I-yR)^{-1}\big].$$

From this, $\mathbf{H}(\Theta, z, s)$ can be determined from the singularities of the matrix $W(y, s)$, i.e. points $y = y_n(s)$ at which W is singular, as in the method described in section 4.2. For the simpler case, when there is only one phase so that $Q = 0$ and all matrices become scalars, the equation $W = 0$ is quadratic in y. Thus, if y_1 is a root inside the unit disk,

$$H(\theta, z, s) = \frac{(y_1-\theta)(1-\theta)^{-1}\lambda^{-1}\mu}{(1-y_1)s}.$$

More specifically still, comparing with our previous example of the classical M/M/c queue, when $K = 0$ (no negative customers) and $\Theta = \Phi = 0$ (no batches), $W = 0$ at $y = y_1 = \lambda/(s+\lambda+\mu)$. This yields

$$H(0, z, s) = \frac{\mu}{(s+\mu)s}$$

from which we find

$$s+\mu+\lambda\big(1-y^{-1}\big)H(y, z, s) = \frac{\mu}{s(1-y)} - \frac{\lambda\mu}{ys(s+\mu)} = \frac{\mu(s+\mu-\lambda(1-y)/y)}{s(s+\mu)(1-y)}$$

so that

$$H(y, z, s) = \frac{\mu}{s(s+\mu)(1-y)}.$$

This implies that $L_{i0} = \mu/(s(s+\mu))$ for all $i \geqslant 0$ as required. Since here (for $c = 1$), $H(y, z, s) \equiv Z(y, s)$, the situation is as in the previous subsection and the classical result follows. Now suppose that $c \geqslant 2$. The recurrences (19) then become, for $0 \leqslant j \leqslant c-1$:

$$\begin{aligned}&\{s+\lambda[1-y^{-1}]+\mu[c-(c-j-1)y]\}H_j(y,s)-j\mu H_{j-1}(y,s)\\&\quad=\frac{\mu}{(1-y)s}-\lambda y^{-1}H_j(0,s)+[c-j-(c-j-1)y]\mu\sum_{i=0}^{c-j-2}L_{ij}y^i\\&\qquad-(1-y)\mu\sum_{i=0}^{c-j-2}(i+1)L_{ij}y^i\end{aligned}\tag{20}$$

where we note that $H_j(0,s)=L_{0j}$ by definition. In particular, when $j=c-1$, we get

$$[s+\lambda(1-y^{-1})+c\mu]H_{c-1}-(c-1)\mu H_{c-2}=\frac{\mu}{(1-y)s}-\lambda y^{-1}L_{0,c-1}.\tag{21}$$

Analyticity of H_{c-1} inside the unit disk then yields

$$(c-1)\mu H_{c-2}(y_1,s)=\lambda y_1^{-1}L_{0,c-1}-\frac{\mu}{s(1-y_1)}$$

where $y_1=\lambda/(s+\lambda+c\mu)$. When $j=0$, we obtain

$$\begin{aligned}&[s+\lambda(1-y^{-1})+\mu(c-(c-1)y)]H_0\\&\quad=\frac{\mu}{(1-y)s}-\lambda y^{-1}L_{00}+[c-(c-1)y]\mu\sum_{i=0}^{c-2}L_{i0}y^i-(1-y)\mu\sum_{i=0}^{c-2}(i+1)L_{i0}y^i.\end{aligned}\tag{22}$$

The analyticity condition is now a quadratic equation in y and yields a relationship between the L_{i0} for $0 \leqslant i \leqslant c-2$. Suppose now that $c=2$. Then the analyticity equations become

$$\mu H_0(y_1,s)=\lambda y_1^{-1}L_{01}-\frac{\mu}{s(1-y_1)}\tag{23}$$

and

$$\left[\frac{\lambda}{y_2}-\mu\right]L_{00}=\frac{\mu}{(1-y_2)s}$$

where y_2 is the smaller root of the equation

$$\mu y^2-(s+\lambda+2\mu)y+\lambda=0.$$

(Notice that the left-hand side is positive at $y=0,\infty$ and negative at $y=1$ so that the smaller root is the only one inside the unit disk.) Thus,

$$[\lambda-(\lambda+\mu)y_2+\mu y_2^2]L_{00}=\frac{\mu y_2}{s}$$

and so

$$(s+\mu)y_2L_{00} = \frac{\mu y_2}{s}$$

giving $L_{00} = \mu/(s(s+\mu))$ as in the case of immune servicing. Moreover, substituting into equation (22), with $c = 2$, we get

$$\begin{aligned}\left[s+\lambda\left(1-y^{-1}\right)+\mu(2-y)\right]H_0 &= \frac{\mu}{(1-y)s} + \left(\mu - \lambda y^{-1}\right)L_{00} \\ &= \frac{\mu(s+\mu+\mu-\lambda y^{-1}-\mu y+\lambda)}{s(s+\mu)(1-y)}.\end{aligned}$$

Cancelling in the numerators, we have $H_0 = \mu/(s(s+\mu)(1-y))$. Now, equation (23) gives

$$L_{01} = \frac{\mu y_1}{s\lambda(1-y_1)} + \frac{\mu^2 y_1}{s(s+\mu)\lambda(1-y_1)} = \frac{(s+2\mu)\mu y_1}{s(s+\mu)\lambda(1-y_1)}$$

where $y_1 = \lambda/(s+\lambda+2\mu)$. Hence, $L_{01} = L_{00} = \mu/(s(s+\mu))$.

Finally, equation (21), with $c = 2$, gives

$$\begin{aligned}\left[s+\lambda\left(1-y^{-1}\right)+2\mu\right]H_1 &= \frac{\mu}{(1-y)s} + \mu H_0 - \lambda y^{-1}L_{01} \\ &= \frac{\mu}{s(s+\mu)(1-y)}\left(s+\mu+\mu-\lambda y^{-1}(1-y)\right)\end{aligned}$$

whereupon cancellation in the numerator yields

$$H_1 = H_0 = \frac{\mu}{s(s+\mu)(1-y)}.$$

In fact, direct substitution verifies that $H_j = L/(1-y)$ $(0 \leqslant j \leqslant c-1)$ is a solution of the recurrence equations (20), provided $L = \mu/(s(s+\mu))$ which is the familiar result obtained for the classical M/M/c queue.

5. Conclusion

The MM CPP/GE/c G-queue is able to model many features of contention systems not representable in conventional queueing network models; for example, bursty, correlated traffic, unreliable servers and load balancing. Despite its great generality, all of the performance characteristics can be determined, including the often complex response time distribution function. The new queue therefore has great potential as a building block for analysing *networks* of such queues in terms of the intern arrival processes at each constituent queue. This approach allows queues to be considered in isolation so that more complex queueing disciplines and routing strategies can be analysed. In particular, networks with blocking could be investigated by generalising the approach of [8]. This was computationally rather expensive because of its need to solve directly the Markov chain associated with each switch in a Banyan network, a problem which would be obviated by the proposed methodology. Similarly, load balancing could be modelled by associating with each balanced queue a rate at which customers are removed (by negative customers) in a 'heavy traffic' phase, due to the crossing of an upward threshold in the queue length, and a corresponding rate at which (positive) customers are added in a 'low traffic' phase. This approximates the dynamics of the system faithfully, whilst not being synchronised with the instantaneous queue length. Although this inability to represent synchronisation is a disadvantage, the inevitable delays that occur in real scheduling systems precludes precise synchronisation anyway, and so our model could provide a good approximation.

Appendix. The generating function H for inimmune servicing

When customers in service may be killed by negative arrivals, the generating function of the Laplace transform of the conditional sojourn time density function, defined in section 4.1, is given by the following lemma and proposition.

Lemma 1. The generating function $\mathbf{H}(y,z,s) = \sum_{j=0}^{c-1}\sum_{i=0}^{\infty}\mathbf{L}_{ij}(s)y^i z^j$ satisfies the equation

$$\begin{aligned}
W(y,s)\mathbf{H}(y,z,s) = {}& \frac{(1-z^c)M\mathbf{e}_N}{(1-y)(1-z)s} - \Lambda(I-\Theta)(yI-\Theta)^{-1}\mathbf{H}(\Theta,z,s)\\
&+ cy(I-\Phi)(I-y\Phi)^{-1}M\mathbf{A}_1(y,z,s)\\
&+ (z-y)(I-\Phi)(I-y\Phi)^{-1}M\mathbf{A}_2(y,z,s)\\
&+ (c+1)M\mathbf{A}_3(y,z,s) - (1-z)M\mathbf{A}_4(y,z,s)\\
&- (1-y)M\mathbf{A}_5(y,z,s) + cy^2\Phi(I-y\Phi)^{-1}M\mathbf{A}_6(y,z,s)\\
&+ y(z-y)\Phi(I-y\Phi)^{-1}M\mathbf{A}_7(y,z,s)
\end{aligned}$$

where

$$\begin{aligned}
W(y,s) &= S - Q + \Lambda + cM + K\\
&\quad - \Lambda(I-\Theta)(yI-\Theta)^{-1} - yK(I-R)(I-yR)^{-1},\\
\mathbf{H}(\Theta,z,s) &= \big(H_1(\theta_1,z,s), H_2(\theta_2,z,s), \ldots, H_N(\theta_N,z,s)\big),\\
\mathbf{A}_1(y,z,s) &= \sum_{j=0}^{c-2}\sum_{i=c-j-1}^{\infty}\mathbf{L}_{ij}(s)y^i z^j,\\
\mathbf{A}_2(y,z,s) &= \sum_{j=0}^{c-2}\sum_{i=c-j-1}^{\infty}(j+1)\mathbf{L}_{ij}(s)y^i z^j,\\
\mathbf{A}_3(y,z,s) &= \sum_{j=0}^{c-2}\sum_{i=0}^{c-j-2}\mathbf{L}_{ij}(s)y^i z^j,
\end{aligned}$$

$$\mathbf{A}_4(y,z,s) = \sum_{j=0}^{c-2} \sum_{i=0}^{c-j-2} (j+1)\mathbf{L}_{ij}(s) y^i z^j,$$

$$\mathbf{A}_5(y,z,s) = \sum_{j=0}^{c-2} \sum_{i=0}^{c-j-2} (i+1)\mathbf{L}_{ij}(s) y^i z^j,$$

$$\mathbf{A}_6(y,z,s) = \sum_{j=0}^{c-2} \mathbf{L}_{c-j-2,j}(s) y^{c-j-2} z^j,$$

$$\mathbf{A}_7(y,z,s) = \sum_{j=0}^{c-2} (j+1)\mathbf{L}_{c-j-2,j}(s) y^{c-j-2} z^j.$$

Proof. Multiplying equations (17) and (18) by $y^i z^j$, summing the former over $0 \leqslant j \leqslant c-1$, $i \geqslant c-j$ and the latter over $0 \leqslant j \leqslant c-1$, $i \leqslant c-j-1$, and adding, yields:

$$\begin{aligned}
&(S - Q + \Lambda + K)\mathbf{H} + cM \sum_{j=0}^{c-1} \sum_{i=c-j}^{\infty} \mathbf{L}_{ij} y^i z^j + M \sum_{j=0}^{c-1} \sum_{i=0}^{c-j-1} (i+j+1)\mathbf{L}_{ij} y^i z^j \\
&= \Lambda(I - \Theta) \sum_{j=0}^{c-1} \sum_{i=0}^{\infty} \sum_{s=1}^{\infty} \Theta^{s-1} \mathbf{L}_{i+s,j} y^i z^j + K(I - R) \sum_{j=0}^{c-1} \sum_{i=0}^{\infty} \sum_{s=1}^{i} R^{s-1} \mathbf{L}_{i-s,j} y^i z^j \\
&\quad + \sum_{j=0}^{c-1} \sum_{i=c-j}^{\infty} jM(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1} \mathbf{L}_{i-s+1,j-1} y^i z^j \\
&\quad + \sum_{j=0}^{c-1} \sum_{i=c-j}^{\infty} jM\Phi^{i+j+1-c} \mathbf{L}_{c-j-1,j-1} y^i z^j \\
&\quad + \sum_{j=0}^{c-1} \sum_{i=c-j}^{\infty} (c-j-1)M(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1} \mathbf{L}_{i-s,j} y^i z^j \\
&\quad + \sum_{j=0}^{c-1} \sum_{i=c-j}^{\infty} (c-j-1)M\Phi^{i+j+1-c} \mathbf{L}_{c-j-2,j} y^i z^j \\
&\quad + \sum_{j=0}^{c-1} \sum_{i=0}^{c-j-1} jM\mathbf{L}_{i,j-1} y^i z^j + \sum_{j=0}^{c-1} \sum_{i=0}^{c-j-1} iM\mathbf{L}_{i-1,j} y^i z^j + \frac{M\mathbf{e}_N}{s} \sum_{j=0}^{c-1} \sum_{i=0}^{\infty} y^i z^j.
\end{aligned}$$

On the left-hand side, we note that the terms for $i = c - j - 1$ in the second sum are equal to the extra terms in the first sum introduced if we sum from $i = c - j - 1$ instead of $c - j$. We also note that some summands are zero for certain values of the summation variables, and that the first two terms on the right-hand side reduce exactly as in proposition 3. Since the last term can be summed easily, we obtain:

$$
\begin{aligned}
&(S - Q + \Lambda + K)\mathbf{H} + cM \sum_{j=0}^{c-1} \sum_{i=c-j-1}^{\infty} \mathbf{L}_{ij} y^i z^j + M \sum_{j=0}^{c-2} \sum_{i=0}^{c-j-2} (i + j + 1)\mathbf{L}_{ij} y^i z^j \\
&\quad = \Lambda(I - \Theta)(yI - \Theta)^{-1}\big[\mathbf{H}(y, z, s) - \mathbf{H}(\Theta, z, s)\big] \\
&\qquad + yK(I - R)(I - yR)^{-1}\mathbf{H}(y, z, s) \\
&\qquad + \sum_{j=1}^{c-1} \sum_{i=c-j}^{\infty} jM(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1}\mathbf{L}_{i-s+1,j-1} y^i z^j && \text{(A.1)} \\
&\qquad + \sum_{j=1}^{c-1} \sum_{i=c-j}^{\infty} jM\Phi^{i+j+1-c}\mathbf{L}_{c-j-1,j-1} y^i z^j && \text{(A.2)} \\
&\qquad + \sum_{j=0}^{c-2} \sum_{i=c-j}^{\infty} (c - j - 1)M(I - \Phi) \sum_{s=1}^{i+j+1-c} \Phi^{s-1}\mathbf{L}_{i-s,j} y^i z^j && \text{(A.3)} \\
&\qquad + \sum_{j=0}^{c-2} \sum_{i=c-j}^{\infty} (c - j - 1)M\Phi^{i+j+1-c}\mathbf{L}_{c-j-2,j} y^i z^j && \text{(A.4)} \\
&\qquad + \sum_{j=1}^{c-1} \sum_{i=0}^{c-j-1} jM\mathbf{L}_{i,j-1} y^i z^j && \text{(A.5)} \\
&\qquad + \sum_{j=0}^{c-2} \sum_{i=1}^{c-j-1} iM\mathbf{L}_{i-1,j} y^i z^j && \text{(A.6)} \\
&\qquad + \frac{M\mathbf{e}_N}{s} \frac{1 - z^c}{(1 - y)(1 - z)}.
\end{aligned}
$$

The remaining six terms (involving the matrix M) are simplified similarly to the way in the proof of proposition 3, by changing the summation domains and variables as follows:

- Terms (A.1) and (A.3): $\sum_{s=1}^{\infty} \sum_{i=s+c-j-1}^{\infty}$ and then change the summation variable i to $i + s + c - j - 1$.
- Terms (A.2) and (A.4): Change the summation variable i to $i + c - j$.
- Term (A.5): Change the summation variable j to $j + 1$.
- Term (A.6): Change the summation variable i to $i + 1$.

The rest of the proof is routine algebraic manipulation. □

Observe that each of the functions $\mathbf{A}_i$ is definable in terms of sub-series of $\mathbf{H}$ and that they have the following identities:

$$\begin{aligned}
\mathbf{A}_2 &= \frac{\partial(z\mathbf{A}_1)}{\partial z} = z\frac{\partial \mathbf{A}_1}{\partial z} + \mathbf{A}_1,\\
\mathbf{A}_4 &= \frac{\partial(z\mathbf{A}_3)}{\partial z} = z\frac{\partial \mathbf{A}_3}{\partial z} + \mathbf{A}_3,\\
\mathbf{A}_5 &= \frac{\partial(y\mathbf{A}_3)}{\partial y} = y\frac{\partial \mathbf{A}_3}{\partial y} + \mathbf{A}_3,\\
\mathbf{A}_7 &= \frac{\partial(z\mathbf{A}_6)}{\partial z} = z\frac{\partial \mathbf{A}_6}{\partial z} + \mathbf{A}_6.
\end{aligned}$$

Moreover, the function $\mathbf{Z}$, required to compute the generating function $\mathbf{G}$ from proposition 3, is the coefficient of z^{c-1} in $\mathbf{H}$ and so is given by

$$\mathbf{H} = \mathbf{A}_1 + \mathbf{A}_3 + z^{c-1}\mathbf{Z}.$$

Lemma 1 therefore yields the following:

Proposition 4. The generating function $\mathbf{H}(y, z, s) = \sum_{j=0}^{c-1}\sum_{i=0}^{\infty} \mathbf{L}_{ij}(s) y^i z^j$ is given by the equation

$$\begin{aligned}
&W'(y, s)\mathbf{H} - (z-y)(I-\Phi)(I-y\Phi)^{-1}M\frac{\partial(z\mathbf{H})}{\partial z}\\
&= \frac{(1-z^c)M\mathbf{e}_N}{(1-y)(1-z)s} - \Lambda(I-\Theta)(yI-\Theta)^{-1}\mathbf{H}(\Theta, z, s)\\
&\quad - cz^c(I-\Phi)(I-y\Phi)^{-1}M\mathbf{Z}(y, s)\\
&\quad + \big[(c+1) - cy(I-\Phi)(I-y\Phi)^{-1}\big]M\mathbf{A}_3(y, z, s)\\
&\quad - \big[(1-z) + (z-y)(I-\Phi)(I-y\Phi)^{-1}\big]M\mathbf{A}_4(y, z, s)\\
&\quad - (1-y)M\mathbf{A}_5(y, z, s)\\
&\quad + cy^2\Phi(I-y\Phi)^{-1}M\mathbf{A}_6(y, z, s)\\
&\quad + y(z-y)\Phi(I-y\Phi)^{-1}M\mathbf{A}_7(y, z, s)
\end{aligned}$$

where

$$\begin{aligned}
W'(y, s) = S - Q &+ \Lambda\big[I - (I-\Theta)(yI-\Theta)^{-1}\big]\\
&+ K\big[I - y(I-R)(I-yR)^{-1}\big] + cM\big[I - y(I-\Phi)(I-y\Phi)^{-1}\big].
\end{aligned}$$

Proof. Substitute $\mathbf{A}_1$ by $\mathbf{H} - \mathbf{A}_3 - z^{c-1}\mathbf{Z}$ and $\mathbf{A}_2$ by $\partial(z\mathbf{H})/\partial z - \mathbf{A}_4 - cz^{c-1}\mathbf{Z}$ in lemma 1. Notice that the terms $\mathbf{A}_3$–$\mathbf{A}_7$ are finite sums. □

The function $\mathbf{H}(y, z, s)$ is therefore given by the functions $\mathbf{H}(\Theta, z, s)$ (as in proposition 3), $\mathbf{A}_3(y, z, s)$ and $\mathbf{A}_6(y, z, s)$ (after substituting for $\mathbf{A}_4(y, z, s)$, $\mathbf{A}_5(y, z, s)$ and

$\mathbf{A}_7(y, z, s))$. $\mathbf{A}_6(y, z, s)$ is a partial sum of $\mathbf{A}_3(y, z, s)$ which itself is a partial sum of $\mathbf{H}(y, z, s)$. In general, these functions can be sought by appealing to analyticity, as in section 4.3. A special case is solved in section 4.4.2.

References

[1] J. Artalejo, G-networks: A versatile approach for work removal in queueing networks, European J. Oper. Res. 126 (2000) 233–249.

[2] R. Chakka and P.G. Harrison, A Markov modulated multi-server queue with negative customers – The MM CPP/GE/c/L G-queue, Acta Informatica 37 (2001).

[3] J.M. Fourneau, E. Gelenbe and R. Suros, G-networks with multiple classes of positive and negative customers, Theoret. Comput. Sci. 155 (1996) 141–156.

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

[5] E. Gelenbe, Random neural networks with positive and negative signals and product form solution, Neural Comput. 1(4) (1989) 502–510.

[6] E. Gelenbe, ed., Feature issue on G-networks, European J. Oper. Res. 126 (2000).

[7] P.G. Harrison and N.M. Patel, *Performance Modelling of Communication Systems and Computer Architectures* (Addison-Wesley, Reading, MA, 1993).

[8] P.G. Harrison and A. de C. Pinto, An approximate analysis of asynchronous, packet-switched banyan networks with blocking, Performance Evaluation 19 (1994) 223–258.

[9] P.G. Harrison and E. Pitel, Sojourn times in single server queues with negative customers, J. Appl. Probab. 30 (1993) 943–963.

[10] P.G. Harrison and E. Pitel, Response time distributions in tandem G-networks, J. Appl. Probab. 32 (1995) 224–246.

[11] I. Mitrani, *Probabilistic Modelling* (Cambridge Univ. Press, Cambridge, 1998).

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

[13] 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,* eds. S.R. Chakravarthy and A.S. Alfa (Marcel Dekker, New York, 1997) pp. 43–66.

[14] M. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach* (Dover, New York, 1994).