Sojourn Time Distributions in Modulated G-Queues with Batch Processing

Peter G Harrison and Harf Zatschler
Department of Computing, Imperial College London
{pgh,hz3}@doc.ic.ac.uk

Abstract

Quantiles on response times, given by probability distribution functions, are a critical metric for quality of service in computer networks as well as many other logistical systems. We derive explicit expressions in the time domain for the sojourn (or response) time probability distribution in a modulated, batched G-queue. More precisely, this queue is Markovian with arrival streams of both positive (normal) and negative customers. Arrivals occur in batches of geometric size and service completions also release batches of geometric size, truncated at the current queue length. All the queue's parameters are modulated by an independent, stationary, continuous time Markov chain. This highly complex queue is able to model many characteristics observed in modern distributed computer systems and telecommunications traffic, such as burstiness, autocorrelation and failures. However, previously, sojourn time distributions have not been obtained even for an MMPP/M/1 queue. We simplify a previous result for their Laplace transform which we then show takes a rational form and can be inverted to give a mixture of exponential and Erlang distributions, possibly modified with sine-factors. An algorithm is described which generates these functions from any given model parameterization and is applied to a range of problems to illustrate graphically the potentially diverse density functions that ensue.

1 Introduction

The probability distribution of many response, or sojourn, times constitutes a vital quality of service (QoS) metric in many operational systems such as computer networks, where end-to-end delay is the issue. For example, emergency services in several countries must meet precise QoS requirements, like at least 90% of ambulances called arriving at the scene within 10 minutes. For on-line transaction processing (OLTP) and other real-time systems, quantiles are often specified as quality of service metrics in Service Level Agreement contracts and industry standard benchmarks such as TPC-C, which uses the 90^{th} quantile of response time [9]. Sojourn time distributions in analytical models are therefore acquiring increasing importance.

We derive expressions for the sojourn time probability distribution in a modulated, batched queue with negative customers – the MM CPP/GE/c G-queue which we will refer to as the MBG-queue [2, 4, 1, 8]. In the past, such results have usually been obtained as transforms in the Laplace domain, which usually require numerical inversion, a computationally expensive and imprecise operation, especially in the tail where inversion algorithms have a tendency to become unstable. The MBG-queue is Markovian with Poisson arrival streams of both positive (normal) and negative customers. Arrivals occur in batches of geometric size and the departures following a service period (of exponential duration) also occur in batches of geometric size, truncated by the current queue length. All the queue's parameters are modulated by an independent, stationary, continuous time Markov chain.

This complex queue has been investigated before for its steady state properties and can be used to model many characteristics observed in present-day distributed computer systems and telecommunications networks, such as burstiness and autocorrelation in traffic, transmission errors and node failures, [1]. However, to the authors' knowledge, sojourn time distributions have not been obtained in the time domain, even for an MMPP/M/1 queue with no batches or negative customers. In this paper, we simplify a previous result obtained for the Laplace transform derived in [3], which we then invert analytically. In the next section we give the relevant background theory and results on the MBG-queue, including the solution for its equilibrium state probabilities by the method of spectral analysis (SA) [6, 7]. We also derive the aforesaid simplification of the formula for the required Laplace transform of sojourn time distribution, which depends intimately on the SA formulation. In section 3, we show that the rational form of the simplified formula can be expressed as partial fractions, using properties of the matrices involved, so giving a sojourn time distribution which is a mixture (with positive or negative coefficients) of exponential and Erlang distributions, possibly modified with sine-factors. In section 4 we present an algorithm that generates these functions from any given parameterization of an MBG-queue and this is applied to a range of problems in section 5. These numerical results illustrate the potentially diverse density functions that ensue, in particular some that do not decrease monotonically like the exponential density. The paper concludes in section 6 where we summarize our contribution and point to areas that merit further investigation.

2 A modulated multi-server with batches

We consider an infinite capacity MBG-queue of the type considered in [1, 8] with one positive customer stream, one service stream with c homogeneous servers, one negative customer stream and the following parameters:

- Q : the generator matrix of the N-phase modulating Markov process ;
- Λ : the diagonal positive arrival rate matrix, where $\lambda_i \equiv \Lambda_{ii}$ is the positive batch arrival rate in phase i, $1 \le i \le N$;
- Θ : the diagonal matrix of positive batch size geometric distribution parameters, *i.e.* the probability that an arriving positive batch in phase i has $n \ge 1$ customers is $(1-\theta_i)\theta_i^{n-1}$, where $\theta_i \equiv \Theta_{ii}$;
- M : the diagonal instantaneous service rate matrix, where $\mu_i \equiv M_{ii}$ is the batch service rate (when the queue length is non-zero) in phase i ;
- Φ : the diagonal matrix of (truncated) departure batch size geometric distribution parameters, *i.e.* the probability that a departing batch in phase i has n customers is $(1-\phi_i)\phi_i^{n-1}$ at queue lengths $l > n+c-1$, ϕ_i^{n-1} at queue length $l = n+c-1$ and 1 for $n = 1$ at $l \le c$, where $\phi_i \equiv \Phi_{ii}$;
- K : the diagonal negative arrival rate matrix, where $\kappa_i \equiv K_{ii}$ is the negative batch arrival rate in phase i ;
- R : the diagonal matrix of negative batch size geometric distribution parameters, *i.e.* the probability that an arriving negative batch in phase i has n customers is $(1-\rho_i)\rho_i^{n-1}$, where $\rho_i \equiv R_{ii}$.

The queueing discipline (for positive customers) is first-come-first-served (FCFS) and the killing strategy (for negative customers) is RCH, where customers are removed from the head of the queue, corresponding to server failures or server-detected errors in queueing models. Notice that RCH implies truncation of arriving negative batches exactly as for service batches.

2.1 Sojourn time distribution

Consider the sojourn time distribution of a particular, 'tagged', individual customer that arrives in phase k of the modulating Markov process at an MBG-queue with j customers in front, *i.e.*

$$F_{kj}(t) = P(T \le t | I(0) = k, A(0) = j)$$

where the random variable $I(x)$ is the phase of the modulation process at time x, $A(x)$ denotes the number of customers ahead of the tagged customer at time x and T is the remaining sojourn time of the tagged customer, without loss of generality (by the memoryless property) at time 0.

We define the vector of conditional sojourn time distribution functions by

$$\mathbf{F}_j(t) = (F_{1j}(t), \ldots, F_{Nj}(t))$$

with corresponding Laplace-Stieltjes transform vector

$$\mathbf{L}_j(s) = (L_{1j}(s), \ldots, L_{Nj}(s))$$

where $L_{kj}(s) = \int_0^\infty e^{-st} \mathrm{d}F_{kj}(t)$, $1 \le k \le N$, also the Laplace transform of the conditional sojourn time probability density function, $f_{kj}(t) = \frac{dF_{kj}(t)}{dt}$. Hence we seek the unconditional sojourn time distribution, $\mathbf{F}(t) = \sum_{j=0}^\infty \boldsymbol{\alpha}_j \mathbf{F}_j(t)$, with Laplace transform $\mathbf{L}(s) = \sum_{j=0}^\infty \boldsymbol{\alpha}_j \mathbf{L}_j(s)$, where the k^{th} component of the vector $\boldsymbol{\alpha}_j$, $\alpha_{j,k}$, is the equilibrium probability that a tagged customer arrives in phase k of the modulating process to find $j \ge 0$ customers in front, including any ahead within the same arriving batch. In [3], the N-vectors $\boldsymbol{\alpha}_j$ are expressed in terms of the equilibrium probability distribution vector $\mathbf{v}'_q$ that the queue length faced in each phase by an arriving *batch* is $q \ge 0$, $\mathbf{v}'_q = \mathbf{v}_q \Lambda / \lambda^*$, a vector of ratios of probability fluxes. Here, $\mathbf{v}_q$ is the equilibrium probability row-vector that the queue length is q in each phase at a random instant and λ^* is the average batch arrival rate with respect to the modulating Markov process, *i.e.* $\lambda^* = \boldsymbol{\pi}\Lambda\mathbf{e}$, where $\boldsymbol{\pi}$ is the equilibrium probability row-vector of the modulating Markov process and $\mathbf{e}$ is the N-component column-vector with unit components $(1, \ldots, 1)$.

All individual arriving customers that join a queue of length less than c, including any other customers in front within the same arriving batch, can enter service immediately. They therefore experience the same sojourn time stochastically (have identical sojourn time distribution), since negative arrivals affect each of the busy servers uniformly and positive arrivals do not pre-empt. The Laplace transform of this distribution is determined in [3] as:

(1) $$\mathbf{L}_{0\ldots c-1}(s) = (sI - Q + M + K/c)^{-1} M\mathbf{e}/s$$

For the so-called repeating region of queue lengths $j \geq c$, the generating function $\mathbf{D}(z,s) = \sum_{j=c}^{\infty} \mathbf{L}_j z^j$ is used. It is determined by the equation

$$\mathbf{D}(z,s) = C^{-1}(z,s)\mathbf{b}(z,s) \tag{2}$$

where

$$\begin{aligned} C(z,s) = \quad & sI - Q + K - K(I-R)(I-Rz)^{-1}z \\ & +cM - cM(I-\Phi)(I-\Phi z)^{-1}z \qquad (3) \\ \mathbf{b}(z,s) = \quad & z^c \left[K(I-R)(I-Rz)^{-1} + cM(I-\Phi)\right. \\ & \left.(I-\Phi z)^{-1}\right]\mathbf{L}_0(s) + \tfrac{cz^c}{s} M\Phi(I-\Phi z)^{-1}\mathbf{e} \qquad (4) \end{aligned}$$

The Laplace transform of the unconditional sojourn time density function can then be shown to be (after simple manipulation of the expression given in [3]):

$$\begin{aligned} L(s) = & \left(\sum_{j=0}^{c-1}\left(\sum_{q=0}^{j} \mathbf{v'}_q \Theta^{-q}\right)\Theta^j\right)(I-\Theta)\Lambda\mathbf{L}_0(s) \\ & + \left(\sum_{q=0}^{c-1} \mathbf{v'}_q \Theta^{-q}\right)(I-\Theta)\Lambda\mathbf{D}(\Theta,s) \qquad (5) \\ & - \left(\sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi'}_k (\xi_k I - \Theta)^{-1}\right)\Theta^{1-c}(I-\Theta)\Lambda\mathbf{D}(\Theta,s) \\ & + \sum_{k=1}^{N} a_k \xi_k \boldsymbol{\psi'}_k (\xi_k I - \Theta)^{-1}(I-\Theta)\Lambda\mathbf{D}(\xi_k,s) \end{aligned}$$

Here, ξ_k and $\boldsymbol{\psi}_k$ are the k^{th} eigenvalue-eigenvector pair used in the spectral expansion method to find the equilibrium probabilities $\mathbf{v}_q$ for $q \geq c$. Notice that $(\mathbf{v}_j\Lambda)_k = 0$ whenever $\Lambda_{kk} = 0$ for some $k, 1 \leq k \leq N$, for all $j \geq 0$. Hence, once having already computed the $\mathbf{v}_q$ and $\boldsymbol{\psi}_k$ for use in equation 5, with these zero k^{th}-components, we may assume, without loss of generality, that $\Lambda_{kk} \neq 0$ for every $k, 1 \leq k \leq N$.

Specifically, as reported in [1], in the repeating region, the equilibrium probability vectors satisfy the transformed balance equations :

$$\mathbf{v}_q Q_0 + \mathbf{v}_{q+1} Q_1 + \mathbf{v}_{q+2} Q_2 + \mathbf{v}_{q+3} Q_3 = \mathbf{0} \tag{6}$$

for given $N \times N$ matrices $Q_n, n = 0,1,2,3$, defined by the model parameters. The characteristic equation

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

has N eigenvalue-eigenvector solutions (eigenvalues within the unit disc) $(\xi_k, \boldsymbol{\psi}_k)$, $1 \leq k \leq N$ and then, for $q \geq c$,

$$\mathbf{v}_q = \sum_{k=1}^{N} a_k \xi_k^q \boldsymbol{\psi}_k$$

2.2 A simplification of the Laplace transform

It is now shown that the expression for the Laplace transform of the sojourn time density function in equation (5) simplifies greatly in that the $\mathbf{D}(\Theta,s)$ terms cancel. We use some auxiliary results relating to the eigensystem used in the SA solution for the equilibrium state probabilities, which we prove in Appendix A.

Consider the transformed balance equations (6) in the repeating region. These are obtained in [1] by an elimination process that repeatedly subtracts multiples of one (untransformed) equation from the equation at an adjacent *level* (corresponding to a queue length that differs by one). In particular, let us consider the elimination with respect to arrival batches and define the (left-associative) *backwards level-difference operator* B which, when applied to any symbolic expression containing variables denoting levels, reduces these variables by one. Let the transformed balance equation for level $j \geq 0$ be $E_j = \mathbf{0}$, so that, for example, $E_q \equiv \mathbf{v}_q Q_0 + \mathbf{v}_{q+1} Q_1 + \mathbf{v}_{q+2} Q_2 + \mathbf{v}_{q+3} Q_3$ for $q \geq c$. Then we have, by the elimination procedure of [1]:

$$E_q = E'_q - E'_{q-1}\Theta = E'_q(1 - B\Theta)$$

where $E'_q = \mathbf{0}$ is the untransformed balance equation at level q. By inverting this transformation, we obtain (lemma 4) the untransformed balance equation E'_{c-1} as

$$E'_{c-1}\Theta = \sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} X + \sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi}_k Y_k \Theta$$

where $Y_k = \sum_{j=0}^{n-1} \xi_k^j \sum_{i=j}^{n-1} Q_{i+1}\Theta^{i-j}$ and $X = \sum_{i=0}^{n} Q_i \Theta^i$. (Here, $n = 3$, but the result in Appendix A is more general.)

With the lemmas of Appendix A, we can give a much simpler form of the Laplace transform of the sojourn time density function, where the terms involving the generating function vector $\mathbf{D}(\Theta,s)$ (*i.e.* evaluated at Θ) vanish.

Proposition 1. *The unconditional sojourn time density in the MBG-queue considered has Laplace transform:*

$$\begin{aligned} L(s) = & \left(\sum_{j=0}^{c-1}\left(\sum_{q=0}^{j} \mathbf{v}_q \Theta^{-q}\right)\Theta^j\right)(I-\Theta)(\Lambda/\lambda^*)\mathbf{L}_0(s) \\ & + \sum_{k=1}^{N} a_k \xi_k \boldsymbol{\psi}_k (\xi_k I - \Theta)^{-1}(I-\Theta)(\Lambda/\lambda^*)\mathbf{D}(\xi_k,s) \qquad (7) \end{aligned}$$

Proof From equation (5), remembering that the lowest power of z in $\mathbf{D}(z,s)$ is z^c, it suffices to show that (assuming the arrival rates Λ_{kk} are non-zero in all phases k)

$$\sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} = \sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi}_k H_k^{-1}\Theta$$

where $H_k = \xi_k I - \Theta$. This is equivalent to proving

$$\sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} X = \sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi}_k H_k^{-1} X \Theta$$

again assuming non-zero arrival rates and also almost surely finite batch sizes, so that X has an inverse. By lemma 2, this is equivalent to

$$\sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} X = -\sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi}_k Y_k \Theta$$

The proposition therefore follows from lemma 4 and the untransformed balance equation $E'_{c-1} = 0$. ■

In the absence of arrival batches, *i.e.* when Θ is the zero-matrix, the proposition simplifies further to:

(8)
$$L(s) = \sum_{j=0}^{c-1} \mathbf{v}_j (\Lambda/\lambda^*) \mathbf{L}_0(s) + \sum_{k=1}^{N} a_k \psi_k (\Lambda/\lambda^*) \mathbf{D}(\xi_k, s)$$

3 Algebraic Laplace inversion

In this section, it is shown that the Laplace transforms of the sojourn time densities are always algebraically invertible and we develop a more general, automated formulation incorporating multiple streams of all types. For practical purposes, the expression (7) needs to be inverted to yield the function $F(t)$, the sojourn time distribution, in the time domain. Traditionally, this was either deemed too difficult and not done at all, or was performed by way of expensive numerical inversion techniques. However, because the components of $L(s)$ take a rational form – as shown in Appendix B – exact sojourn time densities can be found without need for numerical inversion.

Proposition 2. *The Laplace transform $L(s)$ is analytically invertible and gives an unconditional sojourn time density function which is a mixture of exponential and Erlang densities as well as those of the types $e^{at}\cos(bt)$ and $e^{at}\sin(bt)$, for real numbers $a \le 0$ and $b > 0$.*

Proof The expression (7) for $L(s)$ is a sum of terms, each of which depends on s through either the function $\mathbf{L}_0(s)$ or $\mathbf{D}(z, s)$, for some particular value of z which is an eigenvalue in the spectral analysis solution. These eigenvalues are either real or occur in complex conjugate pairs. Appendix B shows that both of these functions are sums of rational functions of the form $u/(sv)$, where u and v are polynomials in s with real coefficients. Consequently, v has real roots, $\{x_1, \ldots, x_k\}$ say, and roots that form complex conjugate pairs, $\{y_1 \pm iz_1, \ldots, y_l \pm iz_l\}$, where all $x_i, y_j < 0$ for $1 \le i \le k$ and $1 \le j \le l$. The polynomial v is of higher order than u and it is therefore possible to re-write each term using partial fractions as follows:

(9)
$$\begin{aligned} \frac{u}{sv} &= \frac{c_0}{s} + \frac{c_1}{(s-x_1)^{m_{1,1}}} + \ldots + \frac{c_k}{(s-x_k)^{m_{1,k}}} \\ &+ \frac{d_1}{(s-(y_1+z_1))^{m_{2,1}}} + \ldots + \frac{d_l}{(s-(y_l+z_l))^{m_{2,l}}} \\ &+ \frac{e_1}{(s-(y_1-z_1))^{m_{2,1}}} + \ldots + \frac{e_l}{(s-(y_l-z_l))^{m_{2,l}}} \end{aligned}$$

Each of these terms can be algebraically inverted using the rule that the unique inverse of a Laplace transform $l(s) = \frac{1}{(s-a)^n}$ is $f(t) = \frac{1}{(n-1)!} t^{n-1} e^{at}$. Hence, every term of $L(s)$ can be algebraically inverted. The real roots x_k are negative, so that their contribution to the resulting unconditional sojourn time density $F(t)$ consists of Erlang-n densities $\frac{1}{(n-1)!} t^{n-1} e^{at}$, where a is non-positive. Only when a very specific relationship exists between the eigenvalues of the spectral analysis do we find $n > 1$.

The roots that occur in complex conjugate pairs can be inverted in the same way. The two Laplace terms $\frac{d_k}{s-(a+ib)}$ and $\frac{e_k}{s-(a-ib)}$ have as their unique inverses, respectively

$$\begin{aligned} d_k e^{(a+ib)t} &= d_k e^{at} e^{ibt} = d_k e^{at}(\cos(bt) + i\sin(bt)) \\ e_k e^{(a-ib)t} &= e_k e^{at} e^{-ibt} = e_k e^{at}(\cos(bt) - i\sin(bt)) \end{aligned}$$

The sojourn time distribution is a purely real quantity and so the imaginary parts must cancel for all values of t. This cancellation can only occur when the coefficients d_k and e_k also form a complex conjugate pair, *i.e.* $e_k = \overline{d}_k$. It then follows that $d_k e^{(a+ib)t} + e_k e^{(a-ib)t} = 2\left[\Re(d_k) e^{at}\cos(bt) - \Im(d_k) e^{at}\sin(bt)\right]$. ■

The occurrence of complex conjugate coefficients and roots of the denominators revealed in this proof is illustrated in one of the examples of the next section. The composition of the mixture of the 3 different types of density functions is given by the following result.

Proposition 3. *In the absence of Erlang-2 and higher densities, the unconditional sojourn time density function of a queue with N modulation states is a mixture of N_1 exponential densities (e^{at}), one of which is a constant ($a = 0$), and N_2 densities of types $e^{at}\cos(bt)$ and $e^{at}\sin(bt)$, where $N_1 + 2N_2 = N^2 + N + 1$.*

Proof When there are no Erlang densities, the roots x_i of $v(s)$ are distinct and $m_{i,j} = 1$ for all i, j in expression (9). Therefore, $F(t)$ is a mixture of exponential densities as well as $e^{at}\cos(bt)$- and $e^{at}\sin(bt)$-type densities. The generating function $\mathbf{D}(z, s)$ is evaluated for particular values of z, which are the eigenvalues ξ_i found during spectral expansion. In an N-phase modulated infinite queue, there are N distinct eigenvalues within the unit circle. Now, by

Lemma 9, N of the roots of the denominator of $\mathbf{D}(z,s)$ depend on z, while the other $N+1$ do not. Thus, taken together, $\mathbf{D}(\xi_1,s),\ldots,\mathbf{D}(\xi_N,s)$ contribute N^2+N+1 distinct denominator-roots, and hence contributions to the distribution. The $e^{at}\cos(bt)$ and $e^{at}\sin(bt)$ densities arise from pairs of complex conjugate roots, so that they too must occur in pairs.

The addition of $\mathbf{L}_0(s)$ does not give any additional exponentials, because its denominator shares its roots with those found in $\mathbf{b}(z,s)$ (see Lemma 8), which are already included by $\mathbf{D}(z,s)$. ■

Notice that when there are Erlang-n densities for $n>1$, they arise from a contribution of the form $1/(s-a)^n$. In partial fractions, this yields an exponential, Erlang-2, ..., and Erlang-n contributions. The (non-exponential) Erlang distributions therefore occur in addition to the exponential distributions and their number depends on the multiplicities of the associated roots.

3.1 Multiple streams

To be able to estimate the response time distribution on a path through a network of queues, using the sojourn time distributions of individual queues, we need to extend the formulae to allow for multiple streams. Most important is the inclusion of multiple positive and negative customer arrival streams, which occur when a number of queue outputs are fed into a given queue. Multiple service streams, whilst it is not immediately obvious what role they would have in a real network, are also included in our derivation, due to the simplicity of the changes necessary to accommodate them.

Multiple positive arrival streams

The effect of positive arrivals on the sojourn time is very limited, because, using the RCH killing paradigm, we only need to keep track of customers *in front* of the tagged customer. Positive arrivals can only happen *behind* the tagged customer, so that the inclusion of multiple positive arrival streams only has an impact on the equilibrium state probabilities, $\mathbf{v}_q$, computed by the SA algorithm.

When the positive arrival rate matrices for each stream are given by Λ_k, we have $\lambda^* = \sum_{k=1}^{n^{arr}} \boldsymbol{\pi}\Lambda_k\mathbf{e}$ where n^{arr} is the number of positive arrival streams. Essentially, the expression (7) stands, the only change being to replace the two occurrences of Λ by $\sum_{k=1}^{n^{arr}} \Lambda_k$

Multiple negative arrival streams

The inclusion of multiple killing streams is done by augmenting the killing matrix K and its associated batch distribution matrix R using a sum over the number of negative arrival streams, n^{kill}. Within the processing region, the Laplace transform of the distribution becomes

$$\mathbf{L}_{0\ldots c-1}(s) = (sI - Q + M + \tfrac{1}{c}\textstyle\sum_{k=0}^{n^{kill}} K_k)^{-1}M\mathbf{e}/s$$

As before, we use a generating function $\mathbf{D}(z,s) = \sum_{j=c}^{\infty}\mathbf{L}_j z^j$ for representing the repeating region. Its value is determined by the equation $\mathbf{D}(z,s) = C^{-1}(z,s)\mathbf{B}(z,s)$ where

$$\begin{aligned}
C(z,s) = \;& sI - Q + cM - cM(I-\Phi)(I-\Phi z)^{-1}z \\
& - \textstyle\sum_{k=1}^{n^{kill}} K_k(I + (I-R_k)(I-R_kz)^{-1}z) \\
\mathbf{b}(z,s) = \;& (cz^c/s)M\Phi(I-\Phi z)^{-1}\mathbf{e} \\
& + \Big[\textstyle\sum_{k=1}^{n^{kill}} K_k(I-R_k)(I-R_kz)^{-1} + cM \\
& (I-\Phi)(I-\Phi z)^{-1}\Big] \times\; cz^c\mathbf{L}_0(s)
\end{aligned}$$

Multiple service streams

Once multiple negative customer streams are accommodated, the addition of multiple service streams is straightforward. When the killing paradigm is RCH, the arrival of a negative customer batch is indistinguishable, to all but the tagged customer, from a batch service event with the same parameters. It is therefore possible to re-label all but one of the n^{serv} service streams to negative customer streams. The resulting system has one service stream (with rate matrix M and batch matrix Φ) as well as $n^{kill}+n^{serv}-1$ negative customer streams, consisting of the original as well as the relabelled ones.

The expressions for multiple negative customer streams derived in the previous section can now be applied to derive the sojourn time distribution conditional on the tagged customer completing service in the one remaining service stream. By re-labelling the specified service streams so that each is the single service stream in turn, the unconditional sojourn time distribution can be determined simply.

4 Automation of the sojourn time calculation

When used in conjunction with the automated generation of steady-state probabilities using spectral expansion as in [8], the sojourn time calculations lend themselves to a straightforward automated implementation. The use of spectral expansion as opposed to other matrix geometric methods is central, as the eigenvalues found during this process are subsequently used for the values of z in the generating function $\mathbf{D}(z,s)$, and would not necessarily be immediately available were other matrix analytic methods being used. The algorithm to calculate the sojourn time of a geometrically batched infinite queue can then be specified as follows

- Use the spectral expansion method to derive the steady-state probability distribution of the queue. The probabilities of the infinite repeating region are given in terms of eigenvalues and eigenvectors.
- Compute the vector generating function $\mathbf{D}(z,s)$ using equation (2).

- Calculate the unconditional Laplace transform $L(s)$ from expression (7).
- Symbolically transform $L(s)$ into partial fractions in s.
- Pattern match each fraction to get the unique density function.

The computational cost is dominated by the derivation of the steady-state probability distribution, which requires the finding of eigenvalues of a matrix. Finding $L(s)$ involves multiplication by diagonal and inversion of diagonally dominant $N \times N$ matrices, which is both efficient and numerically stable. Finally, the partial fraction step is done by solving a linear system for the coefficients c_i.

Due to the availability of our *BlueSolver* package [8] written in Mathematica® that automates the spectral expansion portion of the work, we have implemented the remaining steps using the same environment, which has proven itself both efficient and stable.

5 Numerical examples

Exponential distributions Our first example demonstrates the solution process on a $c = 3$ server infinite queue with $N = 2$ modulation states and modulation matrix

$$Q = \begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix}$$

The arrival process is given by $\Lambda = diag(2,4)$, $\Theta = (1/10)I$, negative customers arrive according to $K = (1/2)I$, $R = (2/10)I$ and service occurs with parameters $M = I$ and $\Phi = (1/10)I$.

Calculating $\mathbf{L_0(s)}$

Using the formulae from section 2, we find

$$\begin{aligned} \mathbf{L}_0(s) &= A_{L0}^{-1}(s)M\mathbf{e}/s \quad \text{where} \\ A_{L0}(s) &= sI - Q + M + K/c \\ &= \begin{pmatrix} s+\frac{19}{6} & -2 \\ -1 & s+\frac{13}{6} \end{pmatrix} \end{aligned}$$

To calculate $A_{L0}^{-1}(s)$, we need its determinant and cofactors. As it is a 2×2 matrix, the cofactors are simple to determine. We also find $\det A_{L0} = (s+\frac{7}{6})(s+\frac{25}{6})$, so that

$$A_{L0}^{-1}(s) = \frac{1}{(s+\frac{7}{6})(s+\frac{25}{6})} \begin{pmatrix} s+\frac{15}{6} & 2 \\ 1 & s+\frac{19}{6} \end{pmatrix}$$

Hence, we find

$$\begin{aligned} \mathbf{L}_0(s) &= \frac{1}{s(s+\frac{7}{6})(s+\frac{25}{6})} \begin{pmatrix} s+\frac{13}{6} & 2 \\ 1 & s+\frac{19}{6} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{s(s+\frac{7}{6})(s+\frac{25}{6})} \begin{pmatrix} s+\frac{25}{6} \\ s+\frac{25}{6} \end{pmatrix} \end{aligned}$$

Our particular values of K and M allow for the $(s+\frac{25}{6})$-factor to cancel, so that, after using partial fractions,

$$\mathbf{L}_0(s) = \frac{1}{s(s+\frac{7}{6})}\mathbf{e} = \left(\frac{\frac{6}{7}}{s} - \frac{\frac{6}{7}}{s+\frac{7}{6}}\right)\mathbf{e}$$

It can therefore be seen that the Laplace transform $\mathbf{L}_0$ contributes an exponential distribution $c_1 e^{-\frac{7}{6}t}$ as well as a constant term $c_0 e^{0t} = c_0$ to the solution for $F(t)$.

Calculating $\mathbf{b(z,s)}$

The evaluation of $\mathbf{b}(z,s)$ requires $\mathbf{L}_0$ and a series of straightforward matrix multiplications by diagonal, s-independent matrices. The roots in s in the denominators of $b_i(z,s)$ are the same as those of $L_{0,i}$. Indeed we find,

$$\mathbf{b}(z,s) = \frac{z^3(345 - 6s(z-5) - 65z)}{2s(s+\frac{7}{6})(z-10)(z-5)}\mathbf{e}$$

Calculating $\mathbf{C(z,s)}$

The repeating region (levels $j \geq c$) is represented using the generating function $\mathbf{D}(z,s) = C^{-1}(z,s)\mathbf{b}(z,s)$, which is given by equation (2). By substituting the appropriate parameter values, we find

$$C = \begin{pmatrix} s+\frac{69}{2}+\frac{280z-1450}{(z-5)(z-10)} & -2 \\ -1 & s+\frac{67}{2}+\frac{280z-1450}{(z-5)(z-10)} \end{pmatrix}$$

To invert this matrix, we calculate the determinant and adjoint matrix:

$$\det(C) = \left(s - \frac{-71z^2+505z-650}{2(z-5)(z-10)}\right)\left(s + \frac{65(z-1)(z-\frac{70}{13})}{2(z-5)(z-10)}\right)$$

$$C^{adj} = \begin{bmatrix} s+\frac{67}{2}+\frac{280z-1450}{(z-5)(z-10)} & 2 \\ 1 & s+\frac{69}{2}+\frac{280z-1450}{(z-5)(z-10)} \end{bmatrix}$$

As expected, both roots of the determinant depend on z and we find

$$\mathbf{D}(z,s) = \frac{z^3(345 - 6s(z-5) - 65z)\mathbf{e}}{2s(s+\frac{7}{6})(s+\frac{65(z-1)(z-\frac{70}{13})}{2(z-5)(z-10)})(z-5)(z-10)}$$

The denominator of each component of $\mathbf{D}(z,s)$ is a polynomial in s, that shares its roots with those found for $C^{-1}(z,s)$ and $\mathbf{b}(z,s)$. The total number of roots is one less than the sum of the constituent parts because, once again, we were able to cancel one[1] of them due to the simple nature of M and K. Had no cancellation been possible throughout the calculations – as is to be expected in the general case – we would be left with polynomials in s of degree $2N+1$ in the denominator of $\mathbf{D}(z,s)$.

In our example, the eigenvalues had to be found numerically during spectral expansion and they are (displaying 5 significant digits only) $\xi_1 = 0.37135$ and $\xi_2 = 0.951724$.

[1] $\det(C(z,s))$ had a root $\left(s - \frac{-71z^2+505z-650}{2(z-5)(z-10)}\right)$

The eigenvectors are $\psi_1 = (0.94818, -0.31773)$ and $\psi_2 = (0.43885, 0.90096)$. Finally, the two corresponding coefficients are $a_1 = 0.019201$ and $a_2 = 0.037792$.

Substituting these values into (7) and using partial fractions, the Laplace transform of the unconditional sojourn time density is found as

$$L(s) = \frac{0.84304}{s} - \frac{0.87724}{0.12031+s} + \frac{0.034159}{\frac{7}{6}+s} + \frac{0.000043590}{2.0601+s}$$

After inversion, we find that

$$F(t) = P(T < t) = 0.84304 - 0.87724e^{-0.12031t} + 0.034159e^{-\frac{7}{6}t} + 0.000043590e^{-2.0601t}$$

From this expression, it is apparent that $\lim_{t\to\infty} P(T < t) = 0.84304$. This is the probability that a customer is not killed. To obtain the probability density function, $f(t)$, of the sojourn time of a customer *given that it is not killed*, we divide $F(t)$ by this probability, and differentiate with respect to t. This probability density is shown in Figure 1. It is not monotonic, so that the most likely sojourn time is not zero, but approximatively 1.245.

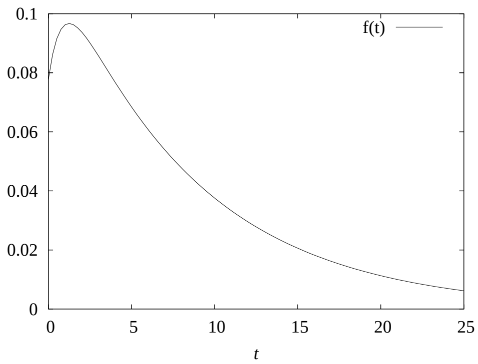

Figure 1. Probability density of the customer sojourn time, if it is not killed.

Sin- and Cos-modified exponential distributions

For this example we consider a queue whose solution process includes complex conjugate roots, which lead to the presence of $e^{at}\cos(bt)$ - type densities in the sojourn time distribution. To keep calculations simple we will not go into as much detail as before and use a queue with no batches or negative customers. The parameters are $c = 1$, $L = \infty$, $N = 3$ and the modulation process has generator matrix

$$Q = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

The arrival process is given by $\Lambda = I$ and services occur with rates 2, $3/2$ and 0, *i.e.* $M = diag(2, 3/2, 0)$. It is found that

$$\mathbf{L}_0 = \frac{1}{p(s)} \begin{pmatrix} (1+s)(13+4s) \\ 13+3s(4+s) \\ 13+4s \end{pmatrix}$$

where $p(s) = s(2s^3 + 13s^2 + 26s + 13) \approx 2s(s + 0.74684)(s + 2.8765 + 0.65466i)(s + 2.8765 - 0.65466i)$

Further,

$$\mathbf{D}(z,s) = -\mathbf{L}_0 + \frac{1}{q} \begin{pmatrix} 13+17s+4s^2-6z-6sz \\ 13+12s+3s^2-6z-6sz \\ 13+4s-6z \end{pmatrix}$$

where

$q = s(13+26s+13s^2+2s^3-19z-26sz-7s^2+6z^2+6sz^2)$

To 5 digits precision, the three eigenvalues were $\xi_1 = 0.27172 + 0.065316i$, $\xi_2 = \overline{\xi_1}$ and $\xi_3 = 0.88758$. The corresponding eigenvectors are

$$\mathbf{\Psi}_1 = \begin{pmatrix} 0.58342 \\ -0.32479 + 0.62603i \\ -0.013201 - 0.41171i \end{pmatrix} \quad \mathbf{\Psi}_2 = \overline{\mathbf{\Psi}_1}$$

$$\mathbf{\Psi}_3 = \begin{pmatrix} 0.56562 \\ 0.54298 \\ 0.62067 \end{pmatrix}$$

and the coefficients are $a_1 = 0.0085074 - 0.017098i$, $a_2 = \overline{a_1}$ and $a_3 = 0.063376$. With these quantities, we use formula (8) to find the desired $L(s)$, which, after performing partial fractions, is given by

$$L(s) = \frac{1}{s} - \frac{0.98520}{s+0.12516} - \frac{0.0073977+0.0039890i}{s+2.4791-0.83631i} - \frac{0.0073977-0.0039890i}{s+2.4791+0.83631i}$$

As expected, the coefficients of the complex roots in the partial fractions expansion are complex conjugate. The inversion to generate the unconditional sojourn time is now straightforward, *i.e.*

$$F(t) = 1 - 0.98520e^{-0.12516} - 2e^{-2.4791t}[0.0073977 \cos(0.83631t) + 0.0039890\sin(0.83631t)]$$

$$f(t) = \frac{\partial F(t)}{\partial t} = 0.12331e^{-0.12516t} + e^{-2.4791t} [0.043352\cos(0.83631t) - 0.0074052\sin(0.83631t)]$$

As there was no killing in this queue, $F(t) = P(T < t)$ reaches 1 as $t \to \infty$, and we have a cumulative distribution function. The presence of oscillating $\cos$ and $\sin$ terms in such a distribution function is unusual, because it might appear conceivable that $F(t)$ could be non-monotonic. However, the coefficients that are encountered are small and hence $F(t)$ is dominated by the exponential distribution, ensuring monotonicity. Figure 2 shows the contribution of the $\cos$ and $\sin$ terms to $f(t)$. The oscillations are heavily damped by the $e^{-2.47918t}$ term present in both functions.

Erlang-2 distributions

In general it is quite unlikely to find double or even higher order roots in the denominators of the elements of $\mathbf{D}(z,s)$. The inversion step would then produce Erlang-n distributions in the unconditional sojourn time distribution $F(t)$. Through careful choice of parameters, we can derive a case where this happens. Consider the expression for $\mathbf{D}(z,s)$ found for the first example.

$$\mathbf{D}(z,s) = \frac{z^3(345 - 6s(z-5) - 65z)\mathbf{e}}{2s(s+\frac{7}{6})(s+\frac{65(z-1)(z-\frac{70}{13})}{2(z-5)(z-10)})(z-5)(z-10)}$$

For an appropriate choice of z, we can make the z-dependent root in s of the denominator coincide with the z-independent root $s = -\frac{7}{6}$. Solving the equation $\frac{65(z-1)(z-\frac{70}{13})}{2(z-5)(z-10)} = \frac{7}{6}$ gives the two solutions

$$\begin{aligned} z_1 &= (5/94)(57 - \sqrt{1933}) \approx 0.693303 \\ z_2 &= (5/94)(57 + \sqrt{1933}) \approx 5.370526 \end{aligned}$$

If one of the eigenvalues we find during the queue solution process is equal to either of these, we have a double root at $s = \frac{7}{6}$. As we are dealing with infinite queues only, the eigenvalue 5.370526 is of course not attainable. The challenge now lies in finding appropriate parameters for our queue that lead to an eigenvalue $\xi = 0.693303$, without changing $\mathbf{D}(z,s)$. This problem is not directly or symbolically solvable, but it can be approached by varying the arrival rate at the queue through multiplying the rate matrix Λ by a factor f. An increase in $f\Lambda$ causes – in general – an increase in the eigenvalues. At the same time the expression $\mathbf{D}(z,s)$ is independent of Λ, so that we can use the intermediate value theorem to prove existence of a suitable value for $f\Lambda$ and use a simple bracketing algorithm to find an arbitrarily close approximation to a particular value of f that gives the desired eigenvalue. It has been found that a value of $f \approx 0.6587182746986538$ yields a queue solution with eigenvalues $\xi_1 = 0.291041$ and $\xi_2 = 0.693303$. As stated before, the value of $\mathbf{D}(z,s)$ is independent of this change in the arrival rate and therefore identical to that given in example 1. Using the expressions for $L(s)$ and partial fractions, we get

$$L(s) = \frac{0.850444}{s} + \frac{0.000633749}{s+2.56701} - \frac{0.85107}{s+\frac{7}{6}} - \frac{0.420691}{(s+\frac{7}{6})^2}$$

Using the inversion formulae, the sojourn time distribution is

$$\begin{aligned} F(t) = \quad & 0.850444 + 0.000633749e^{-2.56701t} \\ & -0.85107e^{-\frac{7}{6}t} - 0.420691te^{-\frac{7}{6}t} \end{aligned}$$

with the associated pdf being

$$\begin{aligned} f(t) = \quad & \frac{\partial \hat{F}(t)}{\partial t} = -0.00191293e^{-2.56701t} \\ & +0.672865e^{-\frac{7}{6}t} + 0.577117te^{-\frac{7}{6}t} \end{aligned}$$

where $\hat{F}(t) = \frac{F(t)}{0.850444}$ is the re-normalized cumulative distribution function that is not conditioned on a customer being killed. One of the components of the probability density function $f(t)$ is an Erlang-2 distribution – this component is illustrated in Figure 3.

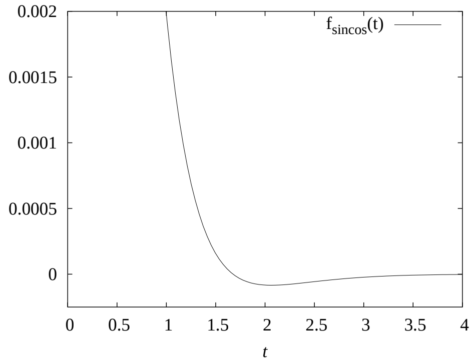

Figure 2. Oscillating components of the sojourn time probability density for the second example. Note that $f_{sincos}(0) \approx 0.043$

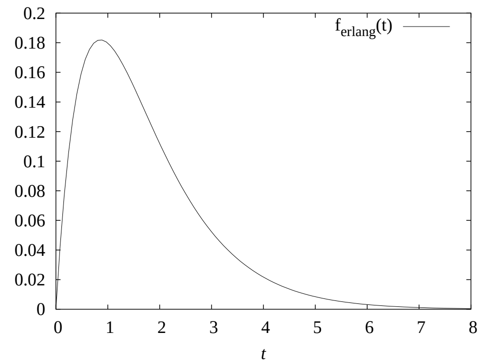

Figure 3. Erlang component of the sojourn time probability density for example 3.

6 Conclusion

Quantiles for response times have become a crucial QoS metric in diverse operational systems, from business processes and health-care to communication systems and the Internet. Analytical models have struggled to provide precise estimates in the time domain and Laplace transforms of density functions have often been the best that can be determined. Although useful for calculating moments of distributions and often numerically invertible, quantiles can usually only be found reliably in the mid-range; in the crucial tail of a density function, numerical inversion quickly becomes unstable.

We have presented and implemented a method that produces exact sojourn time densities, in the time domain, for a queue that can describe burstiness, correlation in traffic and mixed, switching modes of operation, as well as handle multiple streams of arrivals (both positive and negative) and service. This queue can model a wide variety of systems and has now become even more effective by its endowment with sojourn time quantiles. Not only have we provided an algorithm to achieve this, we have also proved that the algebraic inversion is always possible and identified the possible components of sojourn time density that can occur. Interestingly we found three such components in a Markovian system for which, hitherto, contemporary intuition could not tell whether or not Erlang densities would be present in a mixture along with negative exponentials. We have shown that this can indeed occur, although probably rarely in practice. Moreover, we also found a third fundamental component, *viz.* exponential terms modified by trigonometric functions (specifically, sine and cosine).

The extension to accommodate multiple arrival streams facilitates the modelling of nodes in networks with arrivals from many sources (e.g. other nodes) and so offers the possibility of approximating sojourn times in paths through networks. This is the subject of current research.

Appendices

A Laplace Transform simplification Lemmas

Lemma 1. *Let $Z(\xi) = \sum_{i=0}^{n} Q_i \xi^i$ for matrices Q_i $(0 \leq i \leq n)$, integer $n \geq 0$ and complex number ξ, and let the diagonal matrix $H = \xi I - \Theta$ where $\Theta_{jj} \neq \xi$ for any $j = 1, \ldots, N$. Then we may write $Z(\xi) = X + YH$ where $Y = \sum_{j=0}^{n-1} \xi^j \sum_{i=j}^{n-1} Q_{i+1}\Theta^{i-j}$ and $X = \sum_{i=0}^{n} Q_i \Theta^i$*

Proof

$$\begin{aligned} Z(\xi) &= \sum_{i=0}^{n} Q_i (H+\Theta)^i \\ &= X + \sum_{i=1}^{n} Q_i (\xi^i I - \Theta^i)(\xi I - \Theta)^{-1} H \\ &= X + \sum_{i=0}^{n-1} Q_{i+1} \sum_{j=0}^{i} \xi^j \Theta^{i-j} H \end{aligned}$$

The result follows by changing the order of summation. ■
Note that in our case, where $n = 3$, we have

$$\begin{aligned} Z(\xi) &= Q_0 + Q_1\Theta + Q_2\Theta^2 + Q_3\Theta^3 \\ &\quad + (Q_1 + Q_2\Theta + Q_3\Theta^2)H \\ &\quad + \xi(Q_2 + Q_3\Theta)H + \xi^2 Q_3 H \end{aligned}$$

Lemma 2. *When ξ is an eigenvalue of the characteristic equation $\psi Z(\xi) = 0$ with left eigenvector $\boldsymbol{\psi}$,*

$$\boldsymbol{\psi} Y = -\boldsymbol{\psi} H^{-1} X$$

Proof $\boldsymbol{\psi} Z(\xi) = \mathbf{0}$ and so $\boldsymbol{\psi} X H^{-1} = -\boldsymbol{\psi} Y$. It can be verified that the matrix X is diagonal – in our case we have $X = Q_0 + Q_1\Theta + Q_2\Theta^2 + Q_3\Theta^3 = (I-\Theta)(I-R\Theta)(I-\Phi\Theta)\Lambda$. Hence, since H is also diagonal, the matrices H^{-1} and X commute. ■

Lemma 3. *Suppose $E_q \equiv \mathbf{v}_q Q_0 + \ldots + \mathbf{v}_{q+n} Q_n$ for $q \geq c$. Then*

$$E'_q = \sum_{j=0}^{q-1} \mathbf{v}_j \Theta^{q-j} \sum_{i=0}^{n} Q_i \Theta^i + \sum_{i=0}^{n} \sum_{j=0}^{i} \mathbf{v}_{q+j} Q_i \Theta^{i-j}$$

Proof

$$\begin{aligned} E'_q &= E_q(1 - B\Theta)^{-1} = \sum_{j=0}^{\infty} E_q B^j \Theta^j \\ &= \sum_{j=0}^{\infty} E_{q-j} \Theta^j = \sum_{j=0}^{\infty} \sum_{i=0}^{n} \mathbf{v}_{q+i-j} Q_i \Theta^j \end{aligned}$$

where we define $\mathbf{v}_{-k} = \mathbf{0}$ for $k > 0$. Thus, changing the order of summation,

$$E'_q = \sum_{i=0}^{n} \sum_{j=0}^{q+i} \mathbf{v}_{q+i-j} Q_i \Theta^j = \sum_{i=0}^{n} \sum_{l=0}^{q+i} \mathbf{v}_l Q_i \Theta^i \Theta^{q-l}$$

by change of summation variable j to $l = q + i - j$. The result then follows since $X = \sum_{i=0}^{n} Q_i \Theta^i$ is a diagonal matrix. ■
The key to the simplification is now:

Lemma 4. *In the notation of the previous lemmas,*

$$E'_{c-1}\Theta = \sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} X + \sum_{k=1}^{N} a_k \xi_k^c \boldsymbol{\psi}_k Y_k \Theta$$

where $Y_k = \sum_{j=0}^{n-1} \xi_k^j \sum_{i=j}^{n-1} Q_{i+1}\Theta^{i-j}$

Proof By another change of order of summation and simple splitting of the second sum, the previous lemma may be restated as

$$\begin{aligned} E'_q &= \sum_{j=0}^{q-1} \mathbf{v}_j \Theta^{q-j} \sum_{i=0}^{n} Q_i \Theta^i + \sum_{j=0}^{n} \mathbf{v}_{q+j} Q_j \\ &\quad + \sum_{j=0}^{n} \sum_{i=j+1}^{n} \mathbf{v}_{q+j} Q_i \Theta^{i-j} \\ &= \sum_{j=0}^{q-1} \mathbf{v}_j \Theta^{q-j} X + E_q \\ &\quad + \sum_{j=0}^{n-1} \sum_{i=j}^{n-1} \mathbf{v}_{q+j} Q_{i+1} \Theta^{i-j} \Theta \end{aligned}$$

Setting $q = c$ and noting that $E_c = E'_c - E'_{c-1}\Theta$,

$$\begin{aligned} E'_{c-1}\Theta = & \sum_{k=1}^{N} \sum_{j=0}^{n-1} \sum_{i=j}^{n-1} a_k \xi_k^{c+j} \boldsymbol{\psi}_k Q_{i+1} \Theta^{i-j} \Theta \\ & + \sum_{j=0}^{c-1} \mathbf{v}_j \Theta^{c-j} X \quad ■ \end{aligned}$$

B Laplace Inversion Lemmas

For all z, the matrix $C(z,s)$ given by equation (3) is of the form $C(z,s) = \Delta(z) - Q + sI$, where $\Delta(z) = diag(d_1(z), \ldots, d_N(z))$ is a diagonal matrix with every diagonal element $d_i(z) > 0$.

Lemma 5. *The real parts of the eigenvalues of the matrix $P(z) = \Delta(z) - Q$ are positive.*

Proof The off-diagonal elements of Q are rates and therefore positive. The diagonal elements of Q are the negative row sums of the off-diagonals, *i.e.* $(Q)_{i,i} = -\sum_{j=1,j\neq i}^{N} q_{i,j}$ Consequently, $P(z)$ is strictly diagonally dominant, with a positive diagonal. Using Gershgorin's circle theorem [5] on $P(z)$ shows that the real parts of all eigenvalues are positive. ■

Let $\mathcal{P}^n(x)$ be the set of polynomials in the variable x of order n.

Lemma 6. *For all z, the elements of $C^{-1}(z,s)$ are rational functions in s of the form q/p, where $p \in \mathcal{P}^N(s)$ and $q \in \mathcal{P}^{N-1}(s)$. The roots of p depend (in general) on z and have a negative real part.*

Proof By definition of the inverse,

$$C^{-1}(z,s) = \frac{adj(C(z,s))}{\det(C(z,s))} = \frac{adj(P(z)+sI)}{\det(P(z)+sI)}$$

where the adjoint matrix, $adj(P(z)+sI)$, is the transpose of the matrix of cofactors of $P(z)+sI$. The determinant of $P(z)+sI$ is a polynomial in s of order N. All cofactors are polynomials in s of order $N-1$. Hence the elements of $C^{-1}(z,s)$ are of the form q/p, where $p \in \mathcal{P}^N(s)$ and $q \in \mathcal{P}^{N-1}(s)$.

For the second part of the lemma, we note that the roots of $\det(P(z)+sI)$ are the negatives of the eigenvalues of $P(z)$. The matrix $P(z)$ has been shown to have only eigenvalues with positive real parts. Consequently, the roots of $p(s) = \det(P(z)+sI)$ depend on z and they have negative real parts as required. ■

Lemma 7. *The elements of $\mathbf{L}_0(s)$ given in (1) are rational functions in s of the form $t/(sr)$, where $r \in \mathcal{P}^N(s)$ and $t \in \mathcal{P}^{N-1}(s)$. The roots of r have negative real parts and are independent of z.*

Proof From (1), $s\mathbf{L}_0(s) = (sI - Q + M + K/c)^{-1}M\mathbf{e}$ and so,

$$(sI - Q + M + K/c)^{-1} = \frac{adj(sI - Q + M + K/c)}{\det(sI - Q + M + K/c)}$$

The matrix $-Q + M + K/c$ is diagonally dominant with a positive diagonal and using the same argument as that in the previous lemma, it can be seen that each component of $\mathbf{L}_0$ is of the form $L_{0,i} = t/sr$ where r is a polynomial whose roots have negative real parts. $r \in \mathcal{P}^N(s)$, $t \in \mathcal{P}^{N-1}(s)$ and the roots of r are obviously independent of z. ■

Lemma 8. *For all z, the elements of $\mathbf{b}(z,s)$ given in (4) are rational functions in s of the form $t/(sr)$, $r \in \mathcal{P}^N(s)$ and $t \in \mathcal{P}^N(s)$. In addition, the roots of r are the same as those found for $\mathbf{L}_0$ in the previous Lemma.*

Proof $\mathbf{b}(z,s)$ is of the form $\mathbf{b}(z,s) = \Delta_1(z)\mathbf{e}_N/s + \Delta_2(z)\mathbf{L}_0(s)$ where the $\Delta_1(z)$ and Δ_2 are (in general) distinct diagonal matrices with non-zero values on their diagonal. The i^{th} component of $\mathbf{b}(z,s)$ is therefore of the form

$$b_i = \frac{c_{i,1}}{s} + c_{i,2}L_{0,i} = \frac{c_{i,1}}{s} + \frac{c_{i,2}t}{sr} = \frac{c_{i,1}r + c_{i,2}t}{sr}$$

where r and t are as in the previous lemma and $c_{i,1}$ and $c_{i,2}$ are independent of s. ■

Lemma 9. *For all z, the elements of $\mathbf{D}(z,s)$ are rational functions in s of the form $u/(sv)$, where $v \in \mathcal{P}^{2N}(s)$ and $u \in \mathcal{P}^{2N-1}(s)$. Of the $2N$ roots of v, exactly N are independent of z and N are dependent on z. In addition, for all z, all roots of v are either real and negative, or occur in complex conjugate pairs, with negative real parts.*

The proof follows from the previous three lemmas. ■

References

[1] 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**(11-12), pp. 881-919, 2001.

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

[3] P.G. Harrison. The MM CPP/GE/c/L G-queue: sojourn time distribution, *Queueing Systems: Theory and Applications*, 2002.

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

[5] Kreyszig, E. *Advanced Engineering Mathematics* 6th Ed. Wiley 1988 p1035

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

[7] 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.

[8] Thornley, D.J., Zatschler, H. and Harrison, P.G. An automated formulation of queues with multiple geometric batch processes in *Proceedings of HETNETS'03*, University of Bradford, July 2003

[9] Transaction Processing Performance Council. TPC benchmark C: Standard specification revision 5.2, December 2003.