RAIRO Operations Research
RAIRO Oper. Res. **38** (2004) 173–191
DOI: 10.1051/ro:2004018

CHARACTERIZATION OF THE DEPARTURE PROCESS FROM AN *ME/ME/1* QUEUE *

Jayesh Kumaran[1], Kenneth Mitchell[1] and Appie van de Liefvoort[1]

Abstract. In this paper we propose a family of finite approximations for the departure process of an *ME/ME/1* queue indexed by a parameter k defined as the system size of the finite approximation. The approximations capture the interdeparture times from an *ME/ME/1* queue exactly and preserve the lag correlations of inter-event times of the departures from an *ME/ME/1* queue up to lag $(k-1)$.

1. Introduction

Modern high-speed networks, combined with new and different high-speed transmission and switching technologies have attracted heterogeneous mixtures of services and applications. A number of high-quality, high-resolution measurements of different traffic in various networks such as multimedia traffic in high-speed networks, packet streams in local area networks (LAN), cell streams from variable bit rate (VBR) video streams in ATM networks, etc., have been carried out and analyzed. These reveal the presence of correlations, either strong and short term, or small, long term, and very persistent [4,31,33]. Furthermore, theoretical evidence and empirical studies have established that these correlations can have significant effect on the queueing behavior [26,32].

Queueing networks have frequently been used for the performance analysis of computer and communication systems, and there are many methods available for

* *This work has been supported in part by NSF under grant No. ANI 0106640.*

[1] Division of Computer Science and Electrical Engineering, School of Computing and Engineering, University of Missouri-Kansas City, Missouri 64110, USA;
e-mail: mitchellke@umkc.edu

either the exact or approximate solution of such queueing networks, see *e.g.* [6]. There are no exact analytic results available for open queueing networks with general arrival processes that may be correlated and with general service demands at FCFS service centers.

Approximations for (open) queueing networks have appeared in the literature where service centers are treated as *GI/G/1* queues with renewal input. Gelenbe (with his coworkers) extended the diffusion approximation technique to study the behavior of a single node in a queueing network, and uses the first and second moments of both the arrival time and the service time distribution to find an approximation of the queue length distribution at a service center. These individual approximations are then combined together into an approximation for the queueing network as a whole. Gelenbe then incorporates this approximation in performance models for computer and communication networks, with results now available for multiple classes at each service center [10, 13, 14] (see also [12]), and uses the results to successfully study ATM Call Admission Control [11].

Kühn developed a decomposition method for open queueing networks, where again the service centers are treated as *GI/G/1* queues with renewal input, and again only the first two moments of both the arrival time and the service time distributions are used to find an approximation to the performance measures of a service center, this time often by iteration [21]. Whitt incorporated and extended (merging, *e.g.*) these ideas in the Queueing Network Analyzer (QNA) [36], which in turn stimulated a variety of special situation extensions. See [6, 17, 34] for an overview of the various methods. There are only a few papers reporting on incorporating the correlation structure in approximations, see *e.g.* [29].

These techniques are based on decomposing the network into subnetworks, and approximating the solution for the original network by the aggregation of the solutions of these subnetworks. Such approximation methods have been referred to as "flow-equivalence" methods or Norton's Theorem methods, and include the Nearly Completely Decomposable (NCD) methods, where the subnetworks in the decomposition are only weakly coupled to the remainder of the network. After solving such a subnetwork, this subnetwork is then replaced by a single service center that has similar performance qualities as the subnetwork it replaces, like flow equivalency. These approximation models are not sufficient when network behavior shows that a small, but persistent correlation over several time scales is present.

In this paper, we present a finite approximation for an infinite *GI/G/1* queue such that the (marginal) interdeparture distributions of both models are equal, and such that the correlations in the departures match each other closely. These finite models can then be incorporated into larger non-product queueing networks, which are often NCD when persistent correlation is observed, and can be used in iterative methods for the solution of the model.

It is well known that the departure process of an *M/M/s* system forms a Poisson process (Burke [7]). Daley [8] considers the *M/G/1* and *G/M/1* systems and finds the stationary distribution for the departure process of those queues. Extensive analysis of departure processes from an *M/G/1//N* queue (finite system space) can be found in [9,19,20]. Whitt [35,37] uses renewal approximations for the departure process and uses these in the analysis of general queueing networks. Berstimas and Nakazato [5] establish a close connection between the departure process and the idle time. Hu [18] uses a recursive procedure to calculate the MacLaurin series from which he derives the moments and covariances of the departure process of a *GI/G/1* queue.

In recent years, studies of departure processes relaxed the assumption that the arrival process is renewal. Girish and Hu [15] propose an approximation for the departure process of a *G/G/1* queue with Markov modulated arrivals by using the MacLaurin series to derive the interdeparture moments and the lag-1 autocorrelations. Green [16] provides a family of approximations indexed by parameter k for the departure process from a *MAP/PH/1* queue by approximating the busy period with a phase-type random variable to get a MAP description for the approximation of the departure process. He proves that this family of approximations capture the interdeparture distribution of the departure process from a *MAP/PH/1* queue and capture the lag-i correlations of the interdeparture times for $i < k$. Sadre and Haverkort [34] provide a MAP approximation for the departure process from a *MAP/MAP/1* queue.

In this paper, we present an approximation to the departure process of a *GI/G/1* queue that preserves the marginal interdeparture time distribution and also matches the lag-i correlations for $i < k$. Our approximation uses the waiting time distribution that can be obtained by solving for the sum space rather than the product space that has been used in the approximation defined by Green [16]. We compare our results with that of Green and with a truncated system *GI/G/1//N*.

The rest of the paper is organized as follows. In Section 2 we present the basic description of the model. In Section 3 we derive the analytic description of the departure process and in Section 4 we provide numerical results and comparisons to other approximations. Section 5 concludes the paper.

2. Model description

2.1. Matrix exponential process

We use Linear Algebraic Queueing Theory (LAQT) to study the second order statistics of the departing stream [28, 30]. Here, we briefly review the needed material. A matrix exponential (ME) distribution [27] is defined as a probability distribution whose density can be written as

$$f(t) = \boldsymbol{p} \exp\left(-\boldsymbol{B}t\right) \boldsymbol{B}\boldsymbol{e}', \qquad t \geq 0, \tag{1}$$

where $\boldsymbol{p}$ is the starting operator for the process, $\boldsymbol{B}$ is the process rate operator, and $\boldsymbol{e}'$ is a summing operator a vector usually consisting of all 1's. The nth moment of the matrix exponential distribution is given by $E[X^n] = n!\boldsymbol{p}\boldsymbol{V}^n\boldsymbol{e}'$, where $\boldsymbol{V}$ is the inverse of $\boldsymbol{B}$. The class of matrix exponential distributions is identical to the class of distributions that possess a rational Laplace-Stieltjes transform. As such, it is more general than continuous phase type distributions which have a similar appearance.

The joint density function of the first k-successive intervals between events describes a matrix exponential process (MEP):

$$f_k(x_1, \ldots, x_k) = \boldsymbol{p} \exp(-x_1\boldsymbol{B})\boldsymbol{L} \ldots \exp(-x_k\boldsymbol{B})\boldsymbol{L}\boldsymbol{e'}. \tag{2}$$

The matrix $\boldsymbol{L}$ is the event generator matrix. Examples for such processes are a Poisson process ($\boldsymbol{B} = [\lambda]$, $\boldsymbol{L} = [\lambda]$), a renewal process ($\boldsymbol{L} = \boldsymbol{B}\boldsymbol{e'}\boldsymbol{p}$), and a Markov Arrival Processes (MAP) ($\boldsymbol{B} = -\boldsymbol{D}_0$, $\boldsymbol{L} = \boldsymbol{D}_1$). Note that $\boldsymbol{B}$ and $\boldsymbol{L}$ are not limited to being Markovian rate matrices, so every MAP is an MEP, but not *vice versa* (see also [24]). We assume the process to be covariance stationary, so that $\boldsymbol{p}$ is the stationary vector for the process at embedded event points (*i.e.* $\boldsymbol{p}\boldsymbol{V}\boldsymbol{L} = \boldsymbol{p}$). The expression for the lag-k covariance, the covariance between the first interval and the kth is

$$\text{cov}[X_0, X_k] = \boldsymbol{p}\boldsymbol{V}(\boldsymbol{V}\boldsymbol{L})^k\boldsymbol{V}\boldsymbol{e'} - (\boldsymbol{p}\boldsymbol{V}\boldsymbol{e'})^2, \qquad k \geq 1. \tag{3}$$

The auto-correlation at lag-k, r[k], can be found by dividing cov[X_0, X_k] by the variance

$$\text{var[X]} = 2\boldsymbol{p}\boldsymbol{V}^2\boldsymbol{e'} - (\boldsymbol{p}\boldsymbol{V}\boldsymbol{e'})^2. \tag{4}$$

Finally, the marginal process is matrix exponential with density given in equation (1).

2.2. Basic introduction to the model

Let the arrival and the service process be renewal processes with matrix exponential representations $\langle \boldsymbol{p}_a, \boldsymbol{B}_a, \boldsymbol{e}'_a \rangle$ and $\langle \boldsymbol{p}_s, \boldsymbol{B}_s, \boldsymbol{e}'_s \rangle$ with dimensions m_a and m_s respectively, so that $\boldsymbol{L}_a = \boldsymbol{B}_a\boldsymbol{e}'_a\boldsymbol{p}_a$ and $\boldsymbol{L}_s = \boldsymbol{B}_s\boldsymbol{e}'_s\boldsymbol{p}_s$. For the analytic development of the model we introduce the following notations:

$$\widehat{\boldsymbol{B}_a} = \boldsymbol{B}_a \otimes \boldsymbol{I}_s, \tag{5}$$
$$\widehat{\boldsymbol{L}_a} = \boldsymbol{L}_a \otimes \boldsymbol{I}_s, \tag{6}$$
$$\widehat{\boldsymbol{p}_s} = \boldsymbol{I}_a \otimes \boldsymbol{p}_s, \tag{7}$$
$$\widehat{\boldsymbol{B}_s} = \boldsymbol{I}_a \otimes \boldsymbol{B}_s, \tag{8}$$
$$\widehat{\boldsymbol{L}_s} = \boldsymbol{I}_a \otimes \boldsymbol{L}_s, \tag{9}$$

where $\otimes$ denotes the Kronecker product, and $\boldsymbol{I}_x$ is the identity matrix with dimension m_x. The *ME/ME/1* queue is a quasi-birth death process, whose state diagram is shown in Figure 1. The associated infinitesimal generator matrix $\boldsymbol{Q}$ is

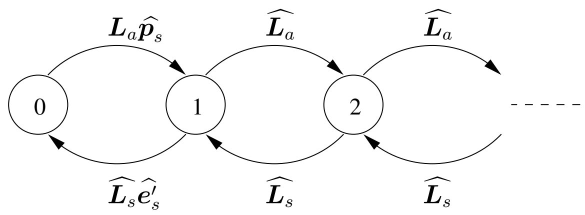


FIGURE 1. Markov state diagram for a *G/G/1* queue.

as follows:

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{B}_a & \boldsymbol{L}_a\widehat{\boldsymbol{p}_s} & \boldsymbol{0} & \boldsymbol{0} & \cdots \\ \widehat{\boldsymbol{L}_s}\widehat{\boldsymbol{e}'_s} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}) & \widehat{\boldsymbol{L}_a} & \boldsymbol{0} & \cdots \\ \boldsymbol{0} & \widehat{\boldsymbol{L}_s} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}) & \widehat{\boldsymbol{L}_a} & \cdots \\ \vdots & \ddots & \ddots & \ddots & \cdots \end{bmatrix}.$$

The transitions that define departures from the queue are those that occur at a completion of service. Thus the departure process itself is an MEP. The queue departure transitions are included in what we will later define as the $\boldsymbol{L}_d$ matrix and all other queue transitions are included in what we will later define as the $\boldsymbol{B}_d$ matrix. Since the buffer size is infinite, there is no exact finite description for the departure process. The departure process can be approximated by truncating the queue to some finite size and thus truncating the infinite matrices defined above. The characteristics of the arrival process and the utilization of the system determine how well the truncated approximation characterizes the departure process. The family of approximations presented in this paper limits the buffer to size k where the service distribution in the global state k is replaced by a matrix exponential distribution with representation $\langle \boldsymbol{p}_b, \boldsymbol{B}_b, \boldsymbol{e}_b\rangle$, where $\boldsymbol{B}_b$ is defined as the rate matrix of the waiting time distribution. The vector $\boldsymbol{p}_b = \frac{\boldsymbol{\pi}(k-1)\widehat{\boldsymbol{L}_a}\widehat{\boldsymbol{e}'_a}}{\boldsymbol{\pi}(k-1)\widehat{\boldsymbol{L}_a}\widehat{\boldsymbol{e}'}}$, where $\boldsymbol{\pi}(k-1)$ is the vector describing the steady state of the internal phases of the arrival and service processes in the state $(k-1)$.

The waiting time distribution itself is a zero modified matrix exponential distribution with dimension equal to the dimension of the service time distribution, see [3, 25]. Van de Liefvoort constructs $\boldsymbol{B}_b$ from the spectral decomposition of a coupling matrix $\boldsymbol{C}$ which is defined as a matrix in the sum space as [25]

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{B}_a & -\boldsymbol{B}_a\boldsymbol{e}'_a\boldsymbol{p}_s \\ \boldsymbol{B}_s\boldsymbol{e}'_s\boldsymbol{p}_a & -\boldsymbol{B}_s \end{bmatrix}, \tag{10}$$

while Asmussen constructs $\boldsymbol{B}_b$ by first solving for $\boldsymbol{T}$ the fixed point problem of the form

$$\boldsymbol{T} = \boldsymbol{D} + \int_0^\infty \exp(\boldsymbol{T}u)\boldsymbol{A}(\mathrm{d}u) \tag{11}$$

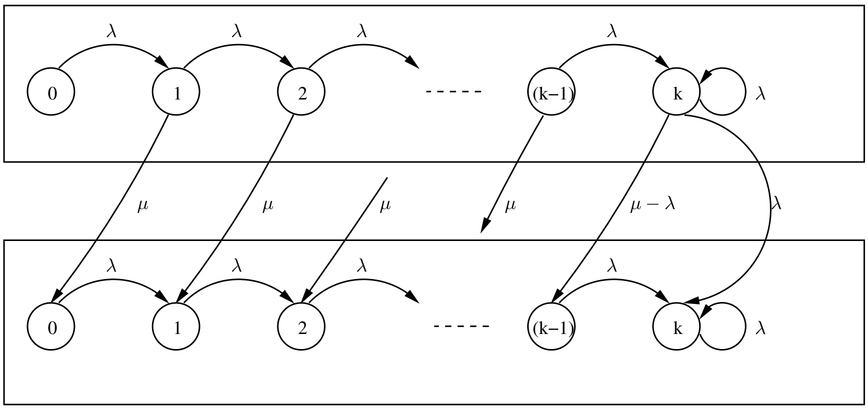


FIGURE 2. Markov state diagram for adjusted *M/M/1//k* queue.

for a suitable matrix kernel $\boldsymbol{A}(\mathrm{d}u)$ and a matrix $\boldsymbol{D}$ that determine the distribution of the process away from the boundary, and then letting $\boldsymbol{T} = -\boldsymbol{B}_b$. Asmussen and Moller obtain the rate matrix of the waiting time distribution in a multi-server queue by transforming the solution $\boldsymbol{T}$, see [2].

The departure transitions from the state k are adjusted in our approximation such that with the rate equal to $\boldsymbol{B}_b\boldsymbol{e}'_b$ a departure transition is made to state $(k-1)$ where a normal service process is started. With rate equal to $(\boldsymbol{B}_s - \boldsymbol{B}_b)\boldsymbol{e}'_b$ a departure transition is made to state k. An arrival event when in state $(k-1)$ has the effect of keeping intact the internal phase at which the service process was active at the instance of the arrival, but the rates of the service departure transitions in state k are adjusted.

2.3. Exponential example

This approximation can best be explained by looking at an example using an *M/M/1* queue as depicted in Figure 2. Let the arrival and the service process be exponentially distributed with mean rates λ and μ respectively. The waiting time distribution in the system is exponentially distributed with the mean $(\mu - \lambda)$. Let the utilization of the system be defined by $\rho = \lambda/\mu$. Thus the matrix exponential representations are given as follows.

$$\boldsymbol{p}_a = [1], \quad \boldsymbol{B}_a = [\lambda], \quad \boldsymbol{e}'_a = [1] \tag{12}$$

$$\boldsymbol{p}_s = [1], \quad \boldsymbol{B}_b = [\mu], \quad \boldsymbol{e}'_s = [1] \tag{13}$$

$$\boldsymbol{p}_b = [1], \, \boldsymbol{B}_b = [\mu - \lambda], \, \boldsymbol{e}'_b = [1]. \tag{14}$$

The steady-state solutions are given by

$$\begin{aligned}\boldsymbol{\pi}(i) &= (1-\rho)\rho^i \ , 0 \le i < (k-1)\\ \boldsymbol{\pi}(k) &= \rho^k.\end{aligned} \tag{15}$$

Observe that the adjustment at level k has the impact that the steady state probabilities are matched exactly to the steady state probabilities of the infinite queue up to level $(k-1)$ and the probability at level k sums up the probabilities of the infinite tail from level k and above. The $\langle \boldsymbol{B}_d, \boldsymbol{L}_d \rangle$ for the finite MEP approximation for the departure process of the *M/M/1* queue are given by

$$\boldsymbol{B}_d = \begin{bmatrix} \lambda & -\lambda & 0 & \cdots & 0\\ 0 & (\lambda+\mu) & -\lambda & 0 & 0\\ 0 & 0 & (\lambda+\mu) & -\lambda & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & (\lambda+\mu) & -\lambda\\ 0 & \cdots & 0 & 0 & \mu \end{bmatrix}, \ \boldsymbol{L}_d = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0\\ \mu & 0 & 0 & \cdots & 0\\ 0 & \mu & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \cdots & 0\\ 0 & \cdots & \mu & 0 & 0\\ 0 & 0 & \cdots & (\mu-\lambda) & \lambda \end{bmatrix}.$$

In this case, the stationary vector $\boldsymbol{p}$ at embedded event points is given by

$$\boldsymbol{p} = \left[(1-\rho), (1-\rho)\rho, \cdots, (1-\rho)\rho^{(k-1)}, \rho^k\right]. \tag{16}$$

Define $\boldsymbol{V}_d$ as the inverse of the matrix $\boldsymbol{B}_d$. Observe that the vector $\boldsymbol{p}\boldsymbol{V}_d$ is the eigenvector for the matrix $\boldsymbol{V}_d\boldsymbol{L}_d$ resulting in the $\mathrm{cov}[X_0, X_i] = 0$ for all $i \ge 1$.

2.4. Approximation for the departure process of an *ME/ME/1* queue

Let the arrival and the service process be renewal processes with matrix exponential representations $\langle \boldsymbol{p}_a, \boldsymbol{B}_a, \boldsymbol{e}'_a \rangle$ and $\langle \boldsymbol{p}_s, \boldsymbol{B}_s, \boldsymbol{e}'_s \rangle$ and let m_a and m_s be the dimensions respectively. The waiting time distribution in this case is matrix exponential with the number of phases equal to the dimension of the service distribution. The appendix describes how the matrix $\boldsymbol{B}_b$ is obtained from the coupling matrix defined in equation (10). It should be observed that the solution to the waiting time distribution from [25] is a sum space solution. It can also be computed as the solution to (see [3])

$$\boldsymbol{B}_b = -\boldsymbol{B}_s + \int_0^\infty \exp(-\boldsymbol{B}_b u)\boldsymbol{L}_s \ (\boldsymbol{p}_a \exp(-\boldsymbol{B}_a u)\boldsymbol{L}_a \boldsymbol{e}'_a)\mathrm{d}u. \tag{17}$$

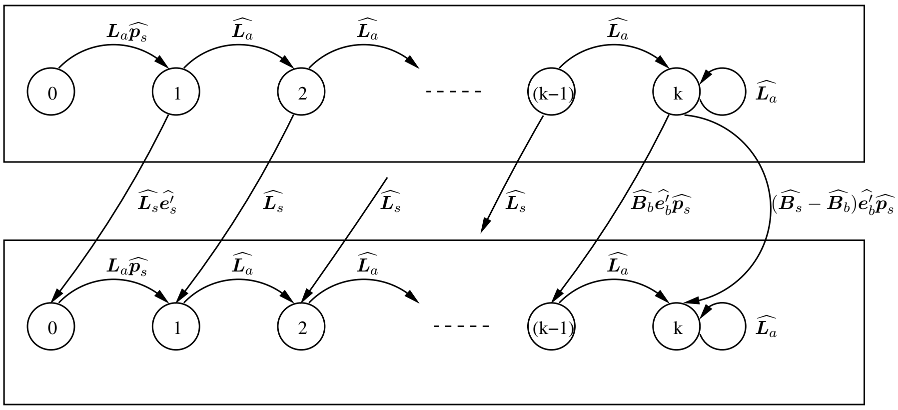


FIGURE 3. Markov state diagram for adjusted *ME/ME/1//k* queue.

Once this matrix $\boldsymbol{B}_b$ is derived, introduce

$$\widehat{\boldsymbol{B}_b} = \boldsymbol{I}_a \otimes \boldsymbol{B}_b, \tag{18}$$

$$\widehat{\boldsymbol{p}_b} = \boldsymbol{I}_a \otimes \boldsymbol{p}_s, \text{and} \tag{19}$$

$$\widehat{\boldsymbol{e}'_b} = \boldsymbol{I}_a \otimes \boldsymbol{e}'_s. \tag{20}$$

The approximation is now constructed by truncating the infinite state space to a system of size $k+1$ where the activities while in the last state (*i.e.* k) are adjusted. Figure 3 shows the Markov state diagram for an *ME/ME/1//k* with the adjustment at level k. The steady state balance equation for this approximation depicted in Figure 3 can be written as follows

$$\boldsymbol{\pi}(0)\boldsymbol{B}_a = \boldsymbol{\pi}(1)\widehat{\boldsymbol{L}_s}\widehat{\boldsymbol{e}'_s}, \tag{21}$$

$$\boldsymbol{\pi}(1)(\widehat{\boldsymbol{B}_a} + \widehat{\boldsymbol{B}_s}) = \boldsymbol{\pi}(0)\boldsymbol{L}_a\widehat{\boldsymbol{p}_s} + \boldsymbol{\pi}(2)\widehat{\boldsymbol{L}_s}, \tag{22}$$

$$\boldsymbol{\pi}(i)(\widehat{\boldsymbol{B}_a} + \widehat{\boldsymbol{B}_s}) = \boldsymbol{\pi}(i-1)\widehat{\boldsymbol{L}_a} + \boldsymbol{\pi}(i+1)\widehat{\boldsymbol{L}_s},\ 1 < i < (k-1), \tag{23}$$

$$\boldsymbol{\pi}(k-1)(\widehat{\boldsymbol{B}_a} + \widehat{\boldsymbol{B}_s}) = \boldsymbol{\pi}(k-2)\widehat{\boldsymbol{L}_a} + \boldsymbol{\pi}(k)\widehat{\boldsymbol{B}_b}\widehat{\boldsymbol{e}'_b}\widehat{\boldsymbol{p}_b}, \tag{24}$$

$$\boldsymbol{\pi}(k)(\widehat{\boldsymbol{B}_a} + \widehat{\boldsymbol{B}_s} - \widehat{\boldsymbol{L}_a}) = \boldsymbol{\pi}(k-1)\widehat{\boldsymbol{L}_a} + \boldsymbol{\pi}(k)(\widehat{\boldsymbol{B}_s} - \widehat{\boldsymbol{B}_b})\widehat{\boldsymbol{e}'_b}\widehat{\boldsymbol{p}_b}. \tag{25}$$

The solution to the steady state balance equations are given by

$$\boldsymbol{\pi}(i) = \tilde{\boldsymbol{\pi}}(i), \quad 0 < i < (k-1), \tag{26}$$

$$\boldsymbol{\pi}(k)\boldsymbol{e}' = \sum_{n=k}^{\infty} \tilde{\boldsymbol{\pi}}(n)\boldsymbol{e}', \tag{27}$$

where $\tilde{\boldsymbol{\pi}}(i)$ is defined as the steady state probability of the *ME/ME/1* infinite queue at state i. The steady state solution for being in states $0\ldots(k-1)$ is identical to the steady state solution of the infinite system. The arrival process is active in level k even though an actual arriving customer does not cause a change in the state. This keeps track of the state of the arrival process at the instant that the chain moves from level k to $(k-1)$ and preserves the correlation of the arrival process. The MEP descriptors $\langle \boldsymbol{B}_d, \boldsymbol{L}_d\rangle$ for the departure process from the finite queue are given as

$$\boldsymbol{B}_d = \begin{bmatrix} \boldsymbol{B}_a & \boldsymbol{L}_a\widehat{\boldsymbol{p}_s} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}) & \widehat{\boldsymbol{L}_a} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}) & \widehat{\boldsymbol{L}_a} & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}) & \widehat{\boldsymbol{L}_a} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & (\widehat{\boldsymbol{B}_a}+\widehat{\boldsymbol{B}_s}-\widehat{\boldsymbol{L}_a}) \end{bmatrix}, \tag{28}$$

$$\boldsymbol{L}_d = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \widehat{\boldsymbol{L}_s}\widehat{\boldsymbol{e}'_s} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \widehat{\boldsymbol{L}_s} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \cdots & \mathbf{0} \\ \mathbf{0} & \cdots & \widehat{\boldsymbol{L}_s} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \widehat{\boldsymbol{B}_b}\widehat{\boldsymbol{e}'_b}\widehat{\boldsymbol{p}_s} & (\widehat{\boldsymbol{B}_s}-\widehat{\boldsymbol{B}_b})\widehat{\boldsymbol{e}'_b}\widehat{\boldsymbol{p}_s} \end{bmatrix}. \tag{29}$$

In Appendix B we have provided a brief review of Green's approximation, defined in [16]. This approximation works by solving for the steady state queue distribution of the infinite queue, which is in the product space.

3. Numerical analysis and discussion

The model developed in the previous section allows us to study the first- and second-order characteristics of a departing stream from a single server queue with matrix exponential arrival and service distributions. In this section, we present numerical results for different arrival processes and study the impact of the server distributions on these streams. We compare our results with those obtained by using the flat truncation and Green's approximation [16]. We will refer to these methods as flat truncation and Green's truncation respectively. We perform extensive numerical experiments to explore various combinations and present the interesting and insightful results, while discussing trends and special cases.

3.1. Experimental setup

We characterize distributions using their first-order characteristics. We specify the mean rate, λ, and squared coefficient of variation, c^2, for the construction of the arrival processes. When $c^2 = 1$, we use the Poisson process with $\boldsymbol{B} = [\lambda]$ and $\boldsymbol{L} = [\lambda]$. For non-exponential renewal processes, we use the hyper exponentials with balanced means, H_2, represented in LAQT as

$$\boldsymbol{p} = \left[\, p,\, 1-p \,\right], \qquad \boldsymbol{B} = \lambda \begin{bmatrix} 2p & 0 \\ 0 & 2(1-p) \end{bmatrix}, \qquad \boldsymbol{L} = \boldsymbol{B}\boldsymbol{e}'\boldsymbol{p}, \tag{30}$$

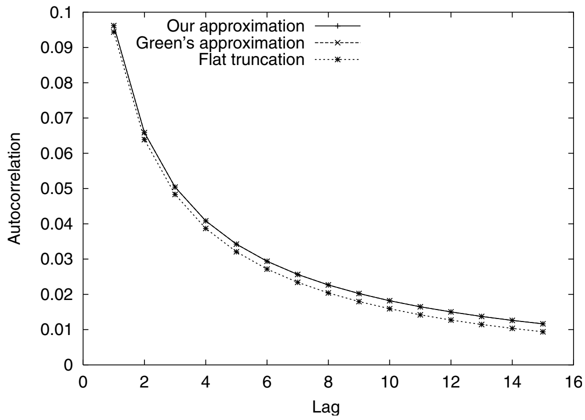
where

$$p = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{c^2-1}{c^2+1}}. \tag{31}$$

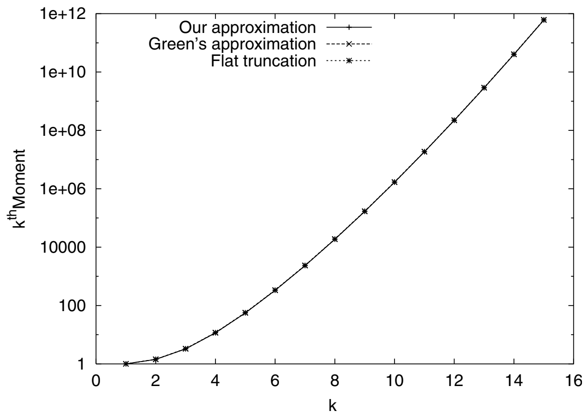
We use arrival and service distributions constructed using equation (30) and study the impact of various server distributions on these streams by looking at the squared coefficient of variation (c_d^2) and the correlation structure of the departing streams. We consider the case where the arrivals are Poisson and the service is assumed to follow an Erlang-10 or H_2 distribution. The mean arrival rate of the arrival process was fixed at $\lambda = 1$ and the utilization at the queue was fixed at $\rho = 0.8$. Figure 4a shows the lag correlations and Figure 4b shows the moments of the departing stream for the three approximations for buffer size of $k = 10$. Observe that the first $(k-1)$ lag correlations of the approximations defined in this paper and in [16] match exactly. There is a jump in the autocorrelation coefficient at lag k. It should be noted that Green proves that these lag-i, $i < k$, correlations are equal to the lag-i, $i < k$, correlations of the infinite *GI/G/1* queue. A proof that our approximations are equal can be found in [22]. Figure 5 shows the marginal distribution of the departure process.

We next consider the case where the service distribution is a hyper exponential distribution constructed using equation (30). The mean arrival rate was fixed at $\lambda = 1.0$, the squared coefficient of variation of the arrival process was fixed at $c^2 = 4.0$, and the buffer size was fixed at $k = 10$. The utilization of the queue was set at $\rho = 0.8$. We observe the lag correlations and the moments of the departing stream when the squared coefficient of variation of the service process was set at $c_s^2 = 1.0$ and 4.0, Figures 6a and 6b. An important observation is the effect of the c_s^2 on the lag correlations and its effectiveness in reducing the correlations of the departing stream. From Figures 4b, 7a, and 7b, we observe that the moments match exactly with Green's truncation, and both methods capture the marginals exactly (in particular they are independent of the buffer size k). Note that the flat truncated model fails to capture the lag correlations of the departing stream.

In the next case, we study the impact of the c_a^2 of the arrival process on the correlations of the departing stream by varying c_a^2. The mean arrival rate was fixed at $\lambda = 1.0$ and the utilization of the queue was fixed at $\rho = 0.8$. The service distribution was assumed to be an Erlang-10 distribution. Figures 8a and 8b show the lag-1 correlations and lag-2 correlations of the departing streams for varying c^2



(a) Lag correlations of the departing stream.



(b) Moments of the departing stream.

FIGURE 4. *M/E10/1* queue.

of the arrival process in which the buffer size was fixed at $k = 10$. It is interesting to observe the decrease in correlation with an increase in the burstiness of the arrival stream.

In the following case we look at the c^2 and lag-1 correlation of the departing stream for various utilizations ranging from 0.05 to 0.95. The service process was assumed to follow an Erlang-10 distribution. The mean arrival rate was fixed at $\lambda = 1.0$ and the squared coefficient of variation was fixed at $c^2 = 4.0$. Figure 9a corresponds to the lag-1 autocorrelation and Figure 9b the squared coefficient of variation of the departing stream for various utilizations. Note the lag-1 correlation increasing and the squared coefficient of variation decreasing with an increase in the utilization of the queue. We can also observe from the above experiments that the flat truncation does not adequately approximate the departure process.

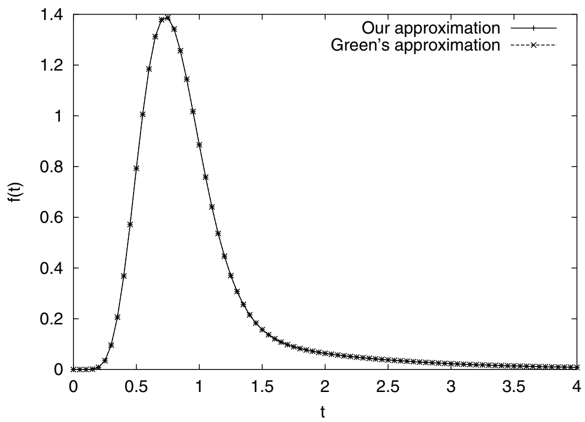
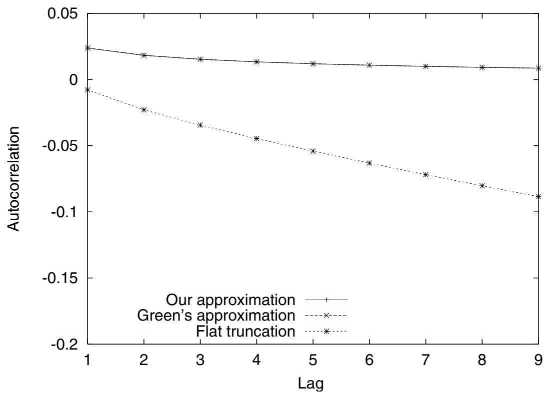
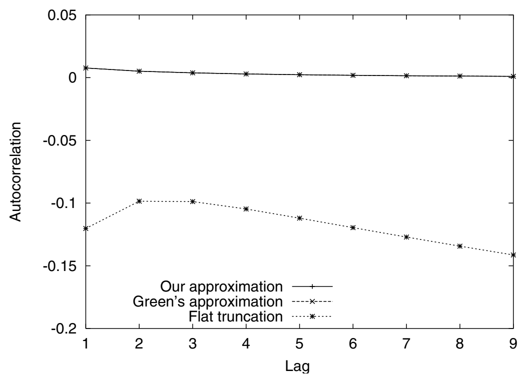


FIGURE 5. *M/E10/1* queue: marginal distribution of the departing stream.

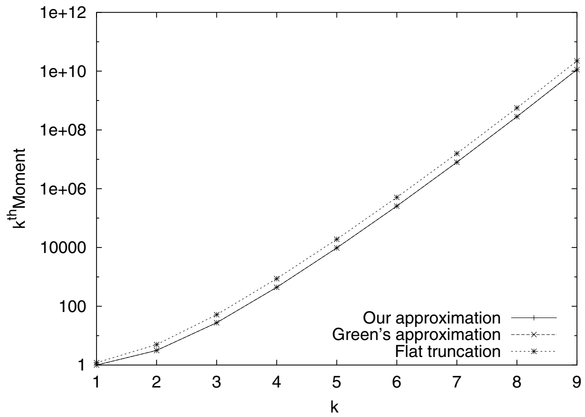


(a) Squared coefficient of variation of the service $c^2 = 1.0$.

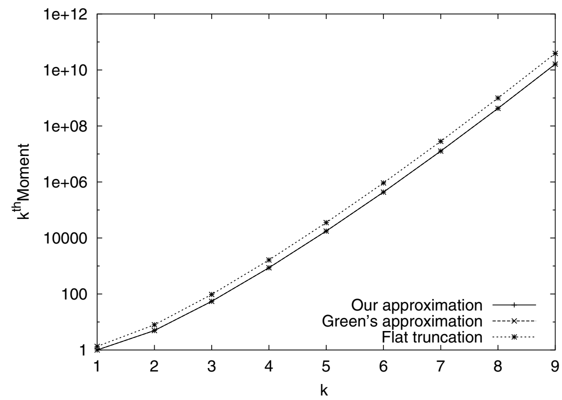


(b) Squared coefficient of variation of the service $c^2 = 4.0$.

FIGURE 6. *H2/H2/1* queue: lag correlations of the departure process.



(a) Squared coefficient of variation of the service $c^2 = 1.0$.

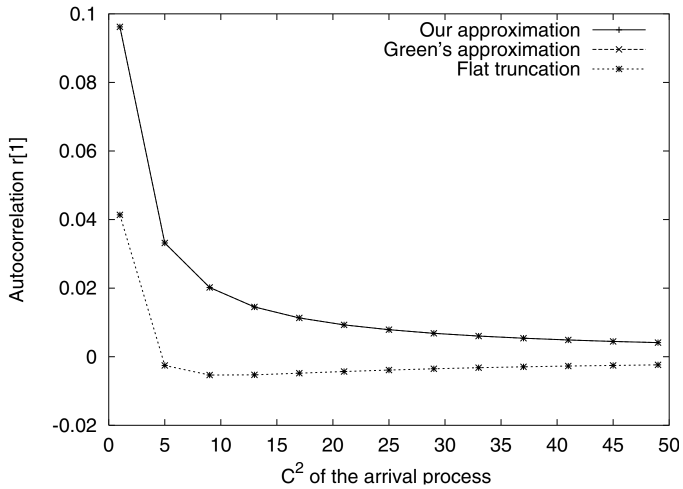


(b) Squared coefficient of variation of the service $c^2 = 4.0$.

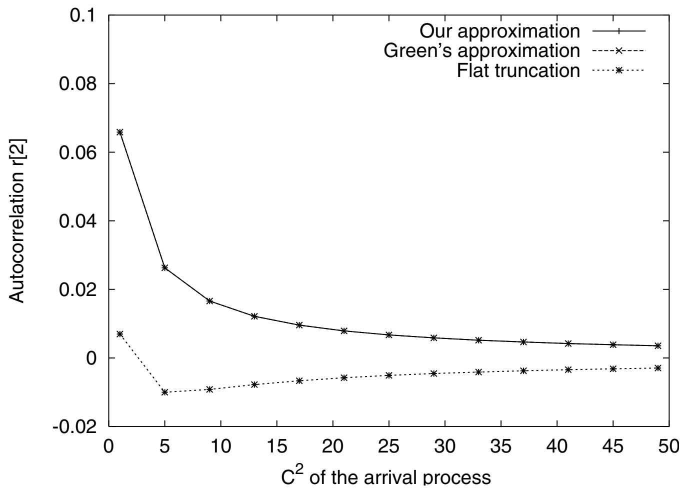
FIGURE 7. *H2/H2/1* queue: moments of the departure process.

4. Conclusions and future work

In this paper we proposed a finite QBD approximation for an infinite *GI/G/1* queue, that captures the marginal distribution exactly of a departure process that is independent of the buffer size of the approximation. Our approximation also captures the first $(k-1)$ lag correlations of the departure process for an approximation of buffer size k. The computational effort is limited to finding m_s eigenvalues in the negative half-plane of a coupling matrix $\boldsymbol{C}$ which is defined in the sum-space only. We have also compared our results with the approximation defined in [16] and have established the accuracy of our results. Further research is being conducted on using the approximation for the departure process for time dependent arrivals. Time dependent arrivals can be modeled using the matrix exponential



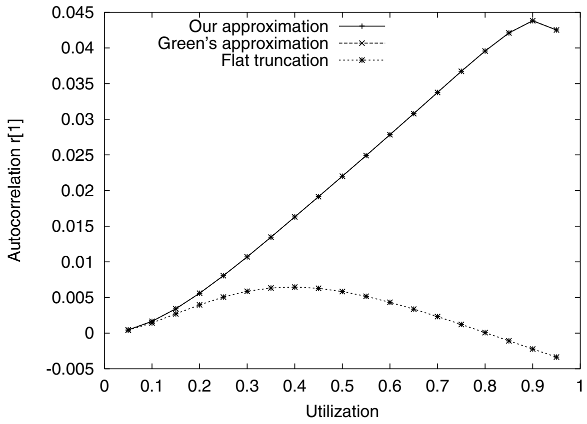
(a) Lag-1 Autocorrelation.



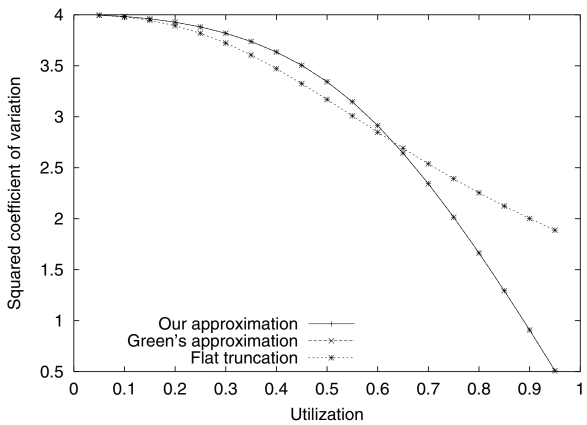
(b) Lag-2 Autocorrelation.

FIGURE 8. *H2/E10/1* queue: varying the squared coefficient of variation of the arrival stream.

process. In this case it is no longer true that the dimension of the service and the waiting time distributions match. Current work also involves using our approximation in studying the departure process of a tagged stream from a multi-class queue. Research along these lines is ongoing, where we hope to use this or similar approximations to solve large queueing networks with correlated streams between service centers.



(a) Lag-1 Autocorrelation.



(b) Squared coefficient of variation of the departing stream.

FIGURE 9. *H2/E10/1* queue: varying the utilization.

APPENDIX A: COMPUTING THE $\boldsymbol{B}_b$ MATRIX OF THE WAITING TIME DISTRIBUTION

Let the interarrival time distribution and the interservice time distributions be matrix exponential distribution with representation $\langle \boldsymbol{p}_a, \boldsymbol{B}_a, \boldsymbol{e}'_a \rangle$ and $\langle \boldsymbol{p}_s, \boldsymbol{B}_s, \boldsymbol{e}'_s \rangle$ of order m_a and m_s respectively. The coupling matrix $\boldsymbol{C}$ is defined in [25] by

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{B}_a & -\boldsymbol{B}_a \boldsymbol{e}'_a \boldsymbol{p}_s \\ \boldsymbol{B}_s \boldsymbol{e}'_s \boldsymbol{p}_a & -\boldsymbol{B}_s \end{bmatrix}. \tag{A.1}$$

In [25] van de Liefvoort shows that the waiting time distribution is given by a matrix exponential distribution with representation

$$\boldsymbol{p} = [\omega_1, \omega_2, \ldots, \omega_{m_s}], \tag{A.2}$$

$$\boldsymbol{B}_{diag} = \begin{bmatrix} -z_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -z_2 & 0 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \ldots & -z_{m_s} \end{bmatrix}, \tag{A.3}$$

where z_i are the m_s eigenvalues of $\boldsymbol{C}$ in the negative half plane. The weights ω_i are given by

$$\omega_i = \Pi_j \left(1 + \frac{z_i}{s_j}\right) / \Pi_{j,j\neq i} \left(1 - \frac{z_i}{z_j}\right), \tag{A.4}$$

where s_i are the m_s eigenvalues of $\boldsymbol{B}_s$. The left eigenvector of the coupling matrix $\boldsymbol{C}$ corresponding to the eigenvalue z_i given by

$$\left[\boldsymbol{p}_a \boldsymbol{V}_a (\boldsymbol{I}_a - z_i \boldsymbol{V}_a)^{-1} \oplus A^*(-z_i) \boldsymbol{p}_s \boldsymbol{V}_s (\boldsymbol{I}_s + z_i \boldsymbol{V}_s)^{-1}\right], \tag{A.5}$$

where $A^*(z)$ is the Laplace-Stieltjes transformation of the arrival time distribution.

Define $\boldsymbol{x}_i$ as

$$\boldsymbol{x}_i = A^*(-z_i) \boldsymbol{p}_s \boldsymbol{V}_s (\boldsymbol{I}_s + z_i \boldsymbol{V}_s)^{-1}. \tag{A.6}$$

The transformed response time matrix that is used in our approximation is given by

$$\boldsymbol{B}_b = \boldsymbol{X}^{-1} (-\boldsymbol{B}_{diag}) \boldsymbol{X}, \tag{A.7}$$

where $\boldsymbol{X}$ is defined by

$$\boldsymbol{X}^T = \left[\, \boldsymbol{x}_1^t \; \boldsymbol{x}_2^t \; \cdots \; \boldsymbol{x}_{m_s}^t \,\right]. \tag{A.8}$$

The computational effort is either finding the m_s eigenvalues of the coupling matrix that are located in the negative half-plane, or by finding the $m_a - 1$ eigenvalues and eigenvectors in the positive half-plane, (followed by a deflation of the matrix).

Appendix B: Green's approximation

In this section we discuss the MAP approximation from [16] that characterizes the departure process from a *MAP/PH/1* queue. Let the interarrival time distribution and the interservice time distributions be matrix exponential distribution with representation $\langle \boldsymbol{p}_a, \boldsymbol{B}_a, \boldsymbol{e}'_a \rangle$ and $\langle \boldsymbol{p}_s, \boldsymbol{B}_s, \boldsymbol{e}'_s \rangle$ of order m_a and m_s respectively. The MAP descriptors $\langle \boldsymbol{B}_g, \boldsymbol{L}_g \rangle$ for the finite approximation for a buffer size of k

is given by

$$\boldsymbol{B}_g = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_0 & & & & & \\ \boldsymbol{0} & \boldsymbol{A}_1 & \boldsymbol{A}_0 & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \boldsymbol{0} & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \\ & & & & \boldsymbol{0} & \boldsymbol{A}_1 & \boldsymbol{E}_0 \\ & & & & & \boldsymbol{0} & \boldsymbol{E}_1 \end{bmatrix}, \ \boldsymbol{L}_g = \begin{bmatrix} \boldsymbol{0} & & & & & \\ \boldsymbol{B}_2 & \boldsymbol{0} & & & & \\ & \boldsymbol{A}_2 & \boldsymbol{0} & & & \\ & & \ddots & \ddots & & \\ & & & \boldsymbol{A}_2 & \boldsymbol{0} & \\ & & & & \boldsymbol{E}_2 & \boldsymbol{E}_3 \end{bmatrix},$$

where the matrices $\boldsymbol{B}_0$, $\boldsymbol{B}_1$, $\boldsymbol{B}_2$, $\boldsymbol{A}_0$, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{A}_2$, $\boldsymbol{E}_0$, $\boldsymbol{E}_1$, $\boldsymbol{E}_2$, $\boldsymbol{E}_3$ are defined as follows

$$\begin{aligned} \boldsymbol{B}_0 &= -\boldsymbol{L}_a, \\ \boldsymbol{B}_1 &= \widehat{\boldsymbol{B}_a}, \\ \boldsymbol{B}_2 &= -\widehat{\boldsymbol{L}_s}\widehat{\boldsymbol{e}'_s}, \\ \boldsymbol{A}_0 &= -\widehat{\boldsymbol{L}_a}, \\ \boldsymbol{A}_1 &= (\widehat{\boldsymbol{B}_a} + \widehat{\boldsymbol{B}_s}), \\ \boldsymbol{A}_2 &= -\widehat{\boldsymbol{L}_s}, \\ \boldsymbol{E}_0 &= -\widehat{\boldsymbol{L}_a}\widehat{\boldsymbol{e}'_a}, \\ \boldsymbol{E}_1 &= \boldsymbol{B}_s, \\ \boldsymbol{E}_2 &= \boldsymbol{B}_s\boldsymbol{e}'_s\boldsymbol{y}_{k-1}, \\ \boldsymbol{E}_3 &= \boldsymbol{L}_s\,(1 - \boldsymbol{y}_{k-1}\widehat{\boldsymbol{e}'}), \end{aligned}$$

where $\boldsymbol{y}_{k-1}$ is defined in [16] as the unconditional distribution of the return phase at level $(k-1)$ and is given by

$$\boldsymbol{y}_{k-1} = \frac{\boldsymbol{x}_{k-1}}{\sum\limits_{j=k-1}^{\infty} \boldsymbol{x}_j\widehat{\boldsymbol{e}'}}, \tag{B.1}$$

where x_{k-1} is defined as distribution of the QBD process (shown in Fig. 1) at level $k-1$ conditional on a departure having just occurred. The unique stationary distribution of the QBD is given by

$$\boldsymbol{\Psi} = \boldsymbol{\pi}_0 \left[\, \boldsymbol{R}_0, \boldsymbol{R}_0\boldsymbol{R}, \boldsymbol{R}_0\boldsymbol{R}^2, \cdots \right], \tag{B.2}$$

$$\boldsymbol{x}_{k-1} = \begin{cases} \boldsymbol{\pi}_0\boldsymbol{R}_0\boldsymbol{B}_2(\boldsymbol{\upsilon}\boldsymbol{L}_a\boldsymbol{e}'_a)^{-1}, & k = 1, \\ \boldsymbol{\pi}_0\boldsymbol{R}_0\boldsymbol{R}^{k-1}\boldsymbol{A}_2(\boldsymbol{\upsilon}\boldsymbol{L}_a\boldsymbol{e}'_a)^{-1}, & k > 1, \end{cases} \tag{B.3}$$

where $\boldsymbol{v}$ is defined as the steady state distribution of the arrival process, and $\boldsymbol{R}$ is defined as the minimal nonnegative solution to the matrix quadratic equation

$$\boldsymbol{R} = \sum_{k=0}^{2} \boldsymbol{R}^k \boldsymbol{A}_k. \tag{B.4}$$

A number of efficient iterative methods have been proposed in the literature for solving the equation (B.4) for the matrix $\boldsymbol{R}$ [1, 23].

References

[1] N. Akar, N. Oguz and K. Sohraby, A novel computational method for solving finite QBD processes. *Commun. Stat. Stochastic Models* **16** (2000) 273-311.

[2] S. Asmussen and J. Møller, Calculation of the steady state waiting time distribution in *GI/PH/c* and *MAP/PH/c* queues. *Queue. Syst. Theory Appl.* **37** (2001) 9-29.

[3] S. Asmussen and C. O'Cinneide, Representations for matrix-geometric and matrix exponential steady-state distributions with applications to many-server queues. *Commun. Statist. Stochastic Models* **14** (1998) 369-387.

[4] J. Beran, R. Sherman, M. Taqqu and W. Willinger, Long-range dependence in variable bit rate video traffic. *IEEE Trans. Commun.* **43** (1995) 1566-1579.

[5] D. Berstimas and D. Nakazato, *The departure process from a* GI/GI/1 *queue and its applications to the analysis of tandem queues.* Tech. Rep. 3725–91, Sloan School of Management (1990). Working paper.

[6] G. Bolch, S. Greiner, H. de Meer and K. Trivedi, *Queuing Networks and Markov Chains.* A Wiley-Interscience Publication, New York, John Wiley & Sons (1998).

[7] P. Burke, The output of queueing systems. *Oper. Res.* **699** (1956) 699-704.

[8] D. Daley, The correlation structure of the output process of some single server queueing systems. *Ann. Math. Stat.* **39** (1968) 1007-1019.

[9] R. Disney and P. deMorais, Covariance properties for the departure process of $M/E_k/1//N$ queues. *AIIE Transactions* **8** (1976) 169-175.

[10] E. Gelenbe, On approximate computer system models. *J. ACM* **22** (1975) 261-269.

[11] E. Gelenbe, X. Mang and Y. Feng, Diffusion cell loss estimate for ATM with multiclass bursty traffic. *Int. J. Comput. Syst. Sci. Engin.* **11** (1996) 325-333.

[12] E. Gelenbe and I. Mitrani, *Analysis and Synthesis of Computer Systems.* Academic press, London, New York (1980).

[13] E. Gelenbe and G. Pujolle, The behaviour of a single queue in a general queueing network. *Acta Informatica* **7** (1976/77) 123-136.

[14] E. Gelenbe and G. Pujolle, A diffusion model for multiple class queueing networks, in *Measuring, Modelling and Evaluating Computer Systems*, Proc. of the Third International Symposium, Bonn – Bad Godesberg, Germany, edited by H. Beilner and H. Gelenbe. North-Holland (1977) 189-199

[15] M. Girish and J.-Q. Hu, Approximations for the departure process of a *GI/GI/1* queue with Markov-modulated arrivals. *Eur. J. Oper. Res.* **134** (2001) 540-556.

[16] D. Green, *Departure Process from* MAP/PH/1 *Queues.* Ph.D. Thesis, The University of Adelaide, Department of Applied Mathematics (1999).

[17] A. Heindl, *Traffic-Based Decomposition of General Queueing Networks with Correlated Input Processes.* Ph.D. Thesis, Technical University, Berlin (2001).

[18] J.-Q. Hu, The departure process of a *GI/GI/1* queue and its MacLaurin series. *Oper. Res.* **44** (1996) 810-815.

[19] A. Ishikawa, On the joint distribution of the departure intervals in an $M/G/1//N$ queue. *J. Oper. Res. Soc. Japan* **34** (1991) 422-435.

[20] R. King, The covariance structure of the departure process from *M/G/1* queues with finite waiting lines. *J. R. Stat. Soc.* **33** (1982) 401-405.

[21] P. Kühn, Approximation analysis of general queueing networks by decomposition. *IEEE Trans. Commun.* **COM-27** (1979) 113-126.

[22] J. Kumaran, Ph.D. Thesis, University of Missouri-Kansas city, School of Computing Engineering. Forthcoming.

[23] G. Latouche and V. Ramaswami, A logarithmic reduction algorithm for quasi-birth-death process. *J. Appl. Probab.* **30** (1993) 650-674.

[24] Y. Lee, A. van de Liefvoort and V. Wallace, Modeling correlated traffic with a generalized IPP. *Perform. Eval.* **40** (2000) 99-114.

[25] A. van de Liefvoort, The waiting time distribution and its moments of the *PH/PH/1* queue. *Oper. Res. Lett.* **9** (1990) 261-269.

[26] M. Linvy, B. Melamed and A. Tsiolis, The impact of autocorrelation on queueing systems. *Manage. Sci.* **39** (1993) 332-339.

[27] L. Lipsky, *Queueing Theory: A Linear Algebraic Approach*. New York, MacMillan (1992).

[28] L. Lipsky, P. Fiorini, W. Hsin and A. van de Liefvoort, *Auto-correlation of lag-k for customers departing from semi-Markov processes*. Tech. Rep. TUM-19506, Technical University Munich (1995).

[29] K. Mitchell and A. van de Liefvoort, Approximation models of feed-forward queueing networks with correlated arrivals. *Perform. Eval.* **51** (2003) 137-152.

[30] K. Mitchell, A. van de Liefvoort and J. Place, Second-order statistics of an isolated departure stream from a shared buffer with correlated sources, in *Proc. of the Eighth International Conference on Telecommunication Systems Modeling and Analysis*, March (2000) 565-574.

[31] P. Pancha and M. Zarki, Variable bit rate video transmission. *IEEE Communications* **32** (1994) 54-66.

[32] B. Patuwo, R. Disney and D. McNickle, The effect of correlated arrivals on queues. *IIE Transactions* **25** (1993) 105-110.

[33] D. Reininger, B. Melamed and D. Raychaudhuri, Variable bit rate MPEG video: Characteristics, modeling and multiplexing, in *Proc. of the Fourteenth International Teletraffic Congress* **1** (1994) 295-306.

[34] R. Sadre and B. Haverkort, Characterising traffic streams in networks of *MAP/MAP/1* queues, in *Proc. of the Eleventh GI/ITG, Conference on Measuring Modelling and Evaluation of Computer Communication Systems* (2001) 195-208.

[35] W. Whitt, Approximating a point process by a renewal process I: Two basic methods. *Oper. Res.* **30** (1982) 125-147.

[36] W. Whitt, The queueing network analyzer. *Bell System Technical J.* **62** (1983) 2799-2815.

[37] W. Whitt, Approximations for departure processes and queues in series. *Naval Research Logistics Quarterly* **31** (1984) 499-521.