

The workload-dependent MAP/PH/1 queue with infinite/finite workload capacity[☆]



Mehmet Akif Yazici, Nail Akar^{*}

Electrical and Electronics Engineering Department, Bilkent University, Ankara, 06800, Turkey

ARTICLE INFO

Article history:

Received 22 May 2012

Received in revised form 5 May 2013

Accepted 19 September 2013

Available online 1 October 2013

Keywords:

Markovian arrival process

Phase-type distribution

Markov fluid queues

Workload-dependent queues

Block-LU factorization

ABSTRACT

We propose a numerical algorithm for finding the steady-state queue occupancy distribution for a workload-dependent MAP/PH/1 queue in which the arrival process and the service rate depend continuously on the instantaneous workload in the system. Both infinite and finite queue capacity scenarios are considered, including partial rejection and complete rejection policies for the latter. Using discretization, this system is approximately described by a multi-regime Markov fluid queue for which numerical algorithms are available. The computational complexity of the proposed method is linear in the number of regimes used for discretization. We provide numerical examples to validate the proposed approach.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

We study a single-server queuing system in which the job arrivals are modeled by a workload-dependent Markovian arrival process (MAP). A workload-dependent MAP differs from an ordinary MAP described in [1,2] by its matrix parameters not being fixed but allowed to vary with the instantaneous queue occupancy. The workload brought by an individual job, namely the job size, has a phase-type (PH-type) distribution. The queue service discipline is FIFO (first-in–first-out). The queue is drained at a rate $c(x)$ when the queue occupancy takes the value $x > 0$. In the infinite queue capacity case, a new job arrival is always admitted, and it increases the queue occupancy (or workload) by the job size. In this paper, we will also study the case of finite queue capacity. Although most finite queue capacity models pose a limit on the maximum number of jobs allowed in the system, our interest in this study will be in models in which there is an upper limit on the overall workload that the buffer can hold, say B . Such buffers are called workload bounded; in such buffers, different policies can take effect depending on what is rejected when the workload limit is to be exceeded. With a partial rejection policy, if the current workload plus the job size of an arriving job exceeds the workload capacity B , then the workload is increased up to B , which amounts to rejecting part of the arriving job. In a complete rejection policy, the job is completely rejected in the same situation. The goal of this article is the numerical calculation of the steady-state distribution of the system workload in the infinite and finite queue capacity scenarios, the latter for both rejection policies. Other performance measures of interest, including the job loss probability and the workload loss probability, can then be derived from this distribution. For a discussion of various rejection policies for finite buffer systems, we refer the reader to [3].

[☆] This work is supported in part by the Science and Research Council of Turkey (Tübitak) under project grant EEEAG-111E106.

^{*} Corresponding author. Tel.: +90 312 2902337; fax: +90 312 2664192.

E-mail addresses: yazici@ee.bilkent.edu.tr (M.A. Yazici), akar@ee.bilkent.edu.tr (N. Akar).

URL: <http://www.ee.bilkent.edu.tr/~akar> (N. Akar).

In related work, the authors of [4] study an M/G/1 queue with workload-dependent arrivals and service rates for the infinite queue capacity case. The workload-bounded M/G/1 buffer under a complete rejection policy was studied in [5] with closed-form expressions for the M/M/1 case. Ref. [6] studies M/G/1 queues with finite buffers with workload-dependent arrival rate, service speed, and both partial and complete rejection policies. Level crossings and Volterra integral equations play a key role in [6], in which closed-form expressions are also given. The goal of this article is to extend the model in [6] so that we allow a more general arrival process, namely a MAP, and develop a numerically stable and computationally efficient algorithm to solve the steady-state workload distribution. On the other hand, Ref. [7] investigates a workload-bounded buffer using a complete rejection policy with an MMPP (Markov modulated Poisson process) arrivals, which is a subcase of a MAP. However, neither the MMPP nor the service speed is allowed to depend on the workload in this work. The model of [7] has also been extended to systems with multiple priority classes in [8]. Multi-class MAP arrivals with workload-dependent acceptance policies have recently been studied in [9] in the context of modeling customer impatience.

The approach we take in the current paper is to use the well-established theory of Markov fluid queues (MFQs) and in particular multi-regime MFQs (MRMFQs) [10] to study workload-dependent MAP/PH/1 queues. This approach saves us from the burden of rederiving the integral equations that would arise, and we can readily use the already existing numerically stable and efficient methods that have been proposed for the steady-state solution of MRMFQs. The connection between workload-dependent MAP/PH/1 queues and MRMFQs is obtained using sample path arguments as in [9,11,12], but the derivation of this connection in the case of complete rejection policy is one of the main contributions of this paper. Fluid queues in which a fluid acts as the input to and output of the buffer have been used extensively in various stochastic modeling contexts since their introduction in the early studies of [13]. MFQs are described by a joint Markovian process $\{(X(t), Z(t)) : t \geq 0\}$, where $X(t)$ represents the buffer level and $Z(t)$ is an underlying finite state-space continuous-time Markov chain (CTMC) that determines the drift, i.e., the rate at which the buffer content $X(t)$ changes. Ref. [14] studies MFQs with infinite queue sizes using a spectral expansion approach, whereas [15] extends this analysis to finite queue sizes. In MRMFQs, which are also called “level-dependent” [16], “multi-layer” [17,18] or “multi-threshold” [19] fluid queues, the buffer space is partitioned into a finite number of non-overlapping intervals which are called the regimes of the MRMFQ. In MRMFQs, the infinitesimal generator of the background CTMC as well as the drift of the buffer process are allowed to depend on the regime in which the buffer level resides; see [16,20,10] for more detailed description of MRMFQs as well as the boundary conditions that arise in the solution of the steady-state distribution of the buffer level. In continuous-feedback Markov fluid queues (CFMFQs), both the infinitesimal generator of the background CTMC and the drift of the buffer process are allowed to continuously depend on the instantaneous buffer level [21]. Explicit solutions for the steady-state distribution of CFMFQs are available only for specific subcases [21], and numerical solutions based on discretization are generally the basic tool for their numerical analysis [22].

The main method proposed in this study to find the steady-state distribution of the workload-dependent MAP/PH/1 queue comprises the following three main steps.

- (i) The workload-dependent MAP/PH/1 queue for the infinite and finite queue capacity cases, under partial and complete rejection policies for the latter, is described by a CFMFQ using sample path arguments.
- (ii) The resulting CFMFQ is approximated by an MRMFQ using discretization.
- (iii) The boundary conditions for this MRMFQ are solved using block-tridiagonal LU factorization [23] so as to obtain the steady-state distribution of the queue occupancy.

The main contributions of the current work are as follows.

- We extend the problem addressed in [4,6] to a more general setting with workload-dependent MAP arrivals as opposed to Poisson arrivals.
- We provide a numerical solution to the finite-capacity workload-dependent MAP/PH/1 queue with complete rejection, which, to the best of our knowledge, has not been addressed before in the literature. A complete rejection policy is essential in computer and communication systems and networks in which jobs need to be processed in their entirety.
- One of the main computational engines used for solving MRMFQs is the block-tridiagonal LU factorization stemming from the block-banded structure of the linear system of equations that arise. The entries of the block-banded matrix are obtained by using Schur decomposition and Sylvester equations, as in [10]. We also show by numerical examples that the computational complexity of the proposed algorithm depends linearly on the number of regimes used for discretizing the CFMFQ. This approach is similar in spirit to the approach taken in [9], in which algebraic Riccati equations are solved for each regime to solve again a block-banded linear system of equations.

Next, we describe the queuing model in more detail. The arrivals to the queuing system are modeled by a MAP with ℓ phases [1,2,24,25], which is characterized by a matrix pair (D_0, D_1) . The matrices D_0 and D_1 are $\ell \times \ell$ matrices, D_0 has negative diagonal elements and non-negative off-diagonal elements, D_1 is non-negative, and $D = D_0 + D_1$ is an irreducible infinitesimal generator. The matrix D_0 (D_1) governs transitions among phases of the MAP without (with) arrivals. Let π denote the stationary probability vector of the phase process with generator D , i.e., $\pi D = \mathbf{0}^T$, $\pi \mathbf{1} = 1$, where $\mathbf{0}$ and $\mathbf{1}$ denote a column vector of zeros and ones, respectively, of appropriate size. We will also denote an $m \times n$ matrix of zeros and the identity matrix of size n by $0_{m \times n}$ and I_n , respectively. The mean job arrival rate λ for the MAP is written as $\lambda = \pi D_1 \mathbf{1}$ [24]. A MAP model can be obtained to fit observation data as in [26]. If the underlying MAP behavior depends on the instantaneous

workload x in the queue, then the MAP is characterized with the matrix pair $(D_0(x), D_1(x))$ for $x \geq 0$, which is then referred to as a workload-dependent MAP.

The job size is represented by the random variable U , which is modeled by a phase-type (PH-type) distribution [27]. Consider a Markov process on the states $\{1, 2, \dots, h + 1\}$ with initial probability vector $[\alpha \ 0]$, $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_h]$ and infinitesimal generator $\begin{bmatrix} T & T^0 \\ \mathbf{0}^T & 0 \end{bmatrix}$, where T is an $h \times h$ nonsingular matrix, T^0 is an $h \times 1$ matrix, and $T\mathbf{1} + T^0 = \mathbf{0}$. The random variable U is defined as the time till absorption into the absorbing state $h + 1$. In this case, U is said to possess a PH-type distribution, which is characterized by the matrix pair (α, T) . As in [4], we assume that the service speed is represented by a function $c(x) > 0$ when the current workload takes the value $x > 0$, and $c(0) = 0$. When $c(x) = c$, then the service time has a phase-type distribution with characterizing pair (α, cT) , but otherwise there is no direct relationship between the job size and the service time.

The rest of the paper is organized as follows. In Section 2, Markov fluid queues and their variations are briefly described. The workload-bounded and workload-dependent buffer problems as well as their connection to Markov fluid queues and methods for finding their steady-state solutions are provided in Section 3. We provide numerical examples to validate the proposed approach in Section 4. Finally, we conclude in Section 5.

2. Markov fluid queues

2.1. Single-regime Markov fluid queues

A single-regime Markov fluid queue is defined through a finite state-space continuous-time Markov chain $\{Z(t) : t \geq 0\}$ that modulates the buffer through a drift function $r(Z(t))$ [14,20]. Let $X(t)$ be the buffer level at time t . Let $Z(t)$ have the state space $S = \{1, \dots, M\}$, and let Q denote its infinitesimal generator. We also define the diagonal rate matrix $R = \text{diag}\{r(1), \dots, r(M)\}$. Let $F(x) = [F_1(x) \ \dots \ F_M(x)]$ denote the steady-state cumulative distribution function (cdf) vector for $X(t)$, where $F_m(x)$ is the joint cdf at state m . Actually, $F_m(x) = \lim_{t \rightarrow \infty} P\{Z(t) = m, X(t) \leq x\}$. Spectral expansion based methods to find $F(x)$ for the infinite and finite buffer cases, i.e., $X(t)$ cannot exceed a certain threshold B , are available in [28,15], respectively, for the single-regime fluid queue.

2.2. Multi-regime Markov fluid queues (MRMFQs)

The following generalization is based on [20,10]. A multi-regime Markov fluid queue has more boundaries than the usual two terminal boundary points, 0 and B , that a single-regime Markov fluid queue has. Consider a buffer with capacity B with $K - 1$ intermediate boundaries. Together with the upper and lower terminal boundaries 0 and B , there exists a total of $K + 1$ boundaries, and K regimes. Let $T^{(k)}$, $0 \leq k \leq K$ denote these boundaries. We have $0 = T^{(0)} < T^{(1)} < \dots < T^{(K-1)} < T^{(K)} = B$, and we say that the system is operating in regime k when the buffer level is between $T^{(k-1)}$ and $T^{(k)}$. In regime k , we denote the infinitesimal generator of the background process by $Q^{(k)}$, and the drift matrix by $R^{(k)}$. We denote the infinitesimal generator and the drift matrix at boundary $T^{(k)}$ by $\tilde{Q}^{(k)}$ and $\tilde{R}^{(k)}$, respectively. We denote the steady-state probability mass accumulation vector at $T^{(k)}$ by $c^{(k)} = [c_1^{(k)} \ \dots \ c_M^{(k)}]$. Note that $c_m^{(k)} = \lim_{t \rightarrow \infty} P\{Z(t) = m, X(t) = T^{(k)}\}$. We also denote the steady-state probability density function (pdf) vector in regime k by $f^{(k)}(x) = [f_1^{(k)}(x) \ \dots \ f_M^{(k)}(x)]$, $T^{(k-1)} < x < T^{(k)}$. The corresponding steady-state cdf vector in regime k is denoted by $F^{(k)}(x) = [F_1^{(k)}(x) \ \dots \ F_M^{(k)}(x)]$, where $F_m^{(k)}(x) = \lim_{t \rightarrow \infty} P\{Z(t) = m, X(t) \leq x, T^{(k-1)} \leq x < T^{(k)}\}$. Let $F(x) = F^{(k)}(x)$ when $T^{(k-1)} \leq x < T^{(k)}$. Similarly, we define $f(x) = f^{(k)}(x)$ when $T^{(k-1)} < x < T^{(k)}$. Note that $F(T^{(k)-}) + c^{(k)} = F(T^{(k)})$, $1 \leq k \leq K$. Let $S_0^{(k)}$, $S_-^{(k)}$, and $S_+^{(k)}$ denote the set of states in which the drift is 0, negative, and positive in regime k , respectively. Similarly, $\tilde{S}_0^{(k)}$, $\tilde{S}_-^{(k)}$, and $\tilde{S}_+^{(k)}$ denote the set of states in which the drift is 0, negative, and positive, at the boundary point $T^{(k)}$, respectively. Obviously, $S_0^{(k)} \cup S_-^{(k)} \cup S_+^{(k)} = S$ for $1 \leq k \leq K$, and we have $\tilde{S}_0^{(k)} \cup \tilde{S}_-^{(k)} \cup \tilde{S}_+^{(k)} = S$ for $0 \leq k \leq K$.

It is shown in [10] that the steady-state pdf vector of the K -regime fluid queue satisfies

$$\frac{d}{dx} f^{(k)}(x) R^{(k)} = f^{(k)}(x) Q^{(k)}, \quad T^{(k-1)} < x < T^{(k)}, \quad 1 \leq k \leq K, \tag{1}$$

along with a set of boundary conditions. In this paper, we assume that $R^{(k)}$ is invertible for each regime, i.e., $S_0^{(k)} = \emptyset, \forall k$, which does not lead to any loss of generality from the viewpoint of workload-dependent buffers with non-zero service speeds. Moreover, in each state of the modulating process, the sign of the service speed remains the same for all regimes, which allows us to simplify the set of boundary conditions listed in [10] to obtain

$$c_m^{(0)} = 0, \quad \forall m \in S_+^{(1)} \tag{2}$$

$$c_m^{(k)} = 0, \quad \forall m \in \left(S_+^{(k)} \cap S_+^{(k+1)} \right) \cup \left(S_-^{(k)} \cap S_-^{(k+1)} \right), \quad 1 \leq k < K \tag{3}$$

$$c_m^{(K)} = 0, \quad \forall m \in S_-^{(K)} \tag{4}$$

$$f^{(1)}(0+)R^{(1)} = c^{(0)}\tilde{Q}^{(0)} \tag{5}$$

$$f^{(k+1)}(T^{(k)}+)R^{(k+1)} - f^{(k)}(T^{(k)}-)R^{(k)} = c^{(k)}\tilde{Q}^{(k)}, \quad 1 \leq k < K \tag{6}$$

$$f^{(K)}(B-)R^{(K)} = -c^{(K)}\tilde{Q}^{(K)} \tag{7}$$

$$\left(\sum_{k=1}^K \int_{T^{(k-1)}+}^{T^{(k)}-} f^{(k)}(x)dx + \sum_{k=0}^K c^{(k)} \right) \mathbf{1} = 1. \tag{8}$$

The solution to the steady-state pdf vector is then given in [10] in mixed matrix-exponential form for $1 \leq k \leq K$:

$$f^{(k)}(x) = a^{(k)}V^{(k)}(x) = a^{(k)} \begin{bmatrix} L_0^{(k)} \\ e^{A_-^{(k)}(x-T^{(k-1)})}L_-^{(k)} \\ e^{-A_+^{(k)}(T^{(k)}-x)}L_+^{(k)} \end{bmatrix}, \quad T^{(k-1)} < x < T^{(k)}, \tag{9}$$

where $a^{(k)} = [a_0^{(k)} \quad a_-^{(k)} \quad a_+^{(k)}]$ is a $1 \times M$ coefficient vector that needs to be solved for, using the boundary conditions (2)–(8).

Above, the matrix defined by $(Y^{(k)})^{-1} = [L_0^{(k)T} \quad L_-^{(k)T} \quad L_+^{(k)T}]^T$ is a similarity transformation that puts $A^{(k)} = Q^{(k)}(R^{(k)})^{-1}$ into block-diagonal form:

$$(Y^{(k)})^{-1}A^{(k)}Y^{(k)} = \begin{bmatrix} 0 & & \\ & A_-^{(k)} & \\ & & A_+^{(k)} \end{bmatrix},$$

where $A_-^{(k)}$ has all its eigenvalues in the open left half plane and $A_+^{(k)}$ has all its eigenvalues in the open right half plane. A numerically efficient and stable algorithm to calculate $Y^{(k)}$ based on ordered Schur decomposition along with a Sylvester equation is given in detail in [10]. In this paper, we propose to use the same algorithm.

The only step that remains for the complete solution of the finite-capacity MRMFQ is finding the coefficient vectors $a^{(k)}$ and $c^{(k)}$. Using (2)–(8), one can form a system of linear equations in $a^{(k)}$, $1 \leq k \leq K$, and $c^{(k)}$, $0 \leq k \leq K$. Since each $a^{(k)}$ appears in equations involving only $a^{(k-1)}$, $a^{(k+1)}$, $c^{(k-1)}$, and $c^{(k)}$, and since each $c^{(k)}$ appears in equations involving only $a^{(k)}$ and $a^{(k+1)}$, the system of linear equations can be made block banded by ordering the unknown coefficient vectors as

$$[c^{(0)} \quad a^{(1)} \quad c^{(1)} \quad a^{(2)} \quad c^{(2)} \quad \dots \quad a^{(K-1)} \quad c^{(K-1)} \quad a^{(K)} \quad c^{(K)}].$$

The block-banded structure can be exploited to cut down the time required for the complete solution dramatically, as will be detailed later.

Solving infinite buffers where $T^{(K)} = \infty$ requires a similar procedure. Instead of the boundary conditions (4) and (7) that describe the behavior of the system at the upper boundary $T^{(K)} = B$ when $B < \infty$, we need to have conditions to ensure stability. Therefore, we should have $a_0^{(K)} = 0$, $a_+^{(K)} = \mathbf{0}^T$, so that the solution given in (9) remains bounded for all values of x . Moreover, there cannot be any probability mass accumulation at infinity which gives $c^{(K)} = \mathbf{0}^T$. The condition for the existence of a steady-state distribution for the infinite buffer case is that the mean drift in regime K should be strictly negative, i.e., $\pi^{(K)}R^{(K)}\mathbf{1} < 0$, where $\pi^{(k)}$ denotes the stationary distribution of the CTMC with the infinitesimal generator $Q^{(k)}$ [10]. We also assume that $\pi^{(k)}R^{(k)}\mathbf{1} \neq 0$ for all regimes k .

2.3. Continuous-feedback Markov fluid queues (CFMFQs)

In CFMFQs, the transition rates within the states of the background process and/or the drifts depend on the buffer level in a continuous fashion. Therefore, instead of the regime-dependent $Q^{(k)}$ and $R^{(k)}$ matrices of MRMFQs, we have $Q(x)$ and $R(x)$ matrices that depend continuously on the buffer level x . In this case, the CFMFQ is characterized with the matrix pair $(Q(x), R(x))$. The analytical treatment of such systems requires the solution to the differential equation

$$\frac{d}{dx} (f(x)R(x)) = f(x)Q(x), \quad 0 < x < B. \tag{10}$$

The work in [29] employs three different numerical discretization methods to solve (10), whereas [22] extends this study to accommodate sign changes in the drift matrix $R(x)$, which leads to potential probability masses at points of sign change. In this paper, we propose approximating CFMFQs with MRMFQs. For this purpose, the buffer space is divided into a number of regimes in any of which the parameters of the modulating process are held constant. We then use the numerically stable and efficient algorithms to solve MRMFQs which are already available in the literature, for example in [10]. Obviously, as one uses more and more regimes, the accuracy of the MRMFQ approximation increases. We propose a method to solve the system in linear time with respect to the number of regimes. On the other hand, the question of how our method compares to those of [29,22,9] in terms of storage and computational run time is left out of the scope of this study.

Lastly, we present the condition for the existence of a steady-state distribution for the infinite buffer case. We consider the most general scenario in which $D_0(x), D_1(x)$, and $c(x)$ all depend on the buffer level x . Assuming convergence, we define $\lim_{x \rightarrow \infty} D_0(x) = \bar{D}_0$, $\lim_{x \rightarrow \infty} D_1(x) = \bar{D}_1$, and $\lim_{x \rightarrow \infty} c(x) = \bar{c}$. Let $\bar{\pi}$ satisfy $\bar{\pi}(\bar{D}_0 + \bar{D}_1) = \mathbf{0}^T$, $\bar{\pi} \mathbf{1} = 1$. Then, the distribution exists if $\beta \bar{\pi} \bar{D}_1 \mathbf{1} / \bar{c} < 1$, where β is the mean job size. For further exploration of this topic, see [30].

3. Workload-dependent MAP/PH/1 queue

We study three different versions of the workload-dependent MAP/PH/1 queue; the infinite buffer (IB) case, then the case of a finite buffer with partial rejection (FB-PR) policy, and finally the case of a finite buffer with complete rejection (FB-CR) policy. In all three versions, the arrival process is a MAP characterized with the pair $(D_0(x), D_1(x))$ with ℓ phases, the job size has a PH-type distribution characterized with the pair (α, T) with h phases, and the workload depletion rate $c(x)$ is a function of the instantaneous workload x in the buffer. Let t_{ij} , $1 \leq i, j \leq h$, t_i^0 , $1 \leq i \leq h$, and α_i , $1 \leq i \leq h$, denote the entries of T , T^0 , and α , respectively.

3.1. Infinite buffer (IB)

Consider the operation of the infinite buffer. When the buffer is depleted, it stays empty until the next job arrival. When an arrival occurs, the buffer level increases abruptly by an amount equal to the arriving job size. Replacing these jumps by durations of linear increase with unity slope, one can obtain from the workload (or buffer level) process of the MAP/PH/1 queue, denoted by $Y(t)$, a transformed process $X(t)$, which can be modeled as a Markov fluid queue [31]. Since a sample path of $Y(t)$ can be obtained from the sample path of $X(t)$ by deleting the time segments in which $X(t)$ is increasing, solving for the steady-state distribution of $X(t)$ will suffice as far as the steady-state distribution of $Y(t)$ is concerned. Therefore, workload-dependent buffers can be analyzed using the paradigm of Markov fluid queues as in [32,9,11,12,31].

Next, we define the CFMFQ associated with the workload-dependent MAP/PH/1 queue for which $Y(t)$ denotes the workload at time t . Let $X(t)$ be the process obtained from $Y(t)$ via the transformation described, and let $Z(t)$ be its modulator. The states of $Z(t)$ will be made up of the phases of the arrival process and those of the job size. When a job arrives, $Z(t)$ will transit into one of the PH states, and $X(t)$ starts increasing. When absorption occurs, $Z(t)$ will return to the MAP state that the arrival takes the MAP into. The system needs to remember this state; therefore we should have ℓ replicas of the PH-type distribution. Consequently, the process $X(t)$ can be described by a CFMFQ with $h\ell + \ell$ states with the infinitesimal generator $Q(x)$ and the rate matrix $R(x)$ given as follows:

$$Q(x) = \begin{bmatrix} I_\ell \otimes T & I_\ell \otimes T^0 \\ D_1(x) \otimes \alpha & D_0(x) \end{bmatrix}, \quad R(x) = \begin{bmatrix} I_{h\ell} & \\ & -c(x)I_\ell \end{bmatrix}, \quad 0 \leq x < \infty. \tag{11}$$

The first $h\ell$ states represent the PH-type replicas, in which the drift is $+1$, and the states $h\ell + 1$ through $h\ell + \ell$ represent the MAP phases with drifts $-c(x)$. We also stick with the notation introduced in Section 2, with the size of the state space being $M = h\ell + \ell$.

Let $f(x) = [f_1(x) \cdots f_{h\ell+\ell}(x)]$ and $F(x) = [F_1(x) \cdots F_{h\ell+\ell}(x)]$ denote the pdf and cdf vectors for this CFMFQ, assuming that the steady-state distribution exists. In order to find the steady-state distribution of the workload, we propose the following MRMFQ approximation to this system. Let K denote the number of regimes to be employed in the approximation, and let $T^{(k)}$, $0 \leq k \leq K$, denote the regime boundaries. Since $T^{(K)} = \infty$, we have to choose the $T^{(K-1)}$ value beyond which we will assume that the $Q(x)$ and $R(x)$ matrices are approximately held constant. To this end, we propose selecting $T^{(K-1)}$ the minimum value satisfying

$$\max \left\{ \left| Q(T^{(K-1)}) - \lim_{x \rightarrow \infty} Q(x) \right|_\infty, \left| R(T^{(K-1)}) - \lim_{x \rightarrow \infty} R(x) \right|_\infty \right\} \leq \epsilon \tag{12}$$

for a given $\epsilon > 0$. After selecting a suitable $T^{(K-1)}$, we let $\delta = \frac{T^{(K-1)}}{K-1}$. We then uniformly choose the boundary points as $T^{(k)} = \delta k$, $0 \leq k \leq K - 1$. The parameter matrices of the approximative MRMFQ are then chosen as

$$Q^{(k)} = \tilde{Q}^{(k-1)} = Q(T^{(k-1)}), \quad 1 \leq k \leq K, \tag{13}$$

$$R^{(k)} = R(T^{(k-1)}), \quad 1 \leq k \leq K, \tag{14}$$

$$\tilde{R}^{(k)} = R(T^{(k)}), \quad 0 \leq k \leq K - 1. \tag{15}$$

Next, we examine the boundary conditions for the infinite buffer. We have to solve $c^{(k)}$, $0 \leq k \leq K$, and $a^{(k)}$, $1 \leq k \leq K$. We immediately obtain $c^{(k)} = \mathbf{0}^T$, $1 \leq k \leq K$, and $a_0^{(K)} = 0$, $a_+^{(K)} = \mathbf{0}^T$. Defining

$$c_-^{(0)} = [c_{h\ell+1}^{(0)} \cdots c_{h\ell+\ell}^{(0)}], \tag{16}$$

$$H = \begin{bmatrix} W^{(0)} & & & & & \\ -V^{(1)}(0) & V^{(1)}(T^{(1)}) & & & & \\ & -V^{(2)}(T^{(1)}) & V^{(2)}(T^{(2)}) & & & \\ & & & \ddots & & \\ & & & & V^{(K)}(T^{(K)}) & \\ & & & & & -W^{(K)} \end{bmatrix}, \tag{26}$$

$$c_+^{(K)} = [c_1^{(K)} \dots c_{h\ell}^{(K)}],$$

$$W^{(K)} = -[I_{h\ell} \quad 0_{h\ell \times \ell}] \tilde{Q}^{(K)} (R^{(K)})^{-1},$$

and $c_-^{(0)}$, $W^{(0)}$, and $V^{(k)}(x)$ are given by (16), (19), and (9), respectively. Again, the equation $zH = \mathbf{0}^T$ can be solved using block-tridiagonal LU factorization.

The steady-state joint pdf vector of the buffer level, $g(y) = [g_1(y) \dots g_\ell(y)]$, is defined in the same way as in the IB case, and, similar to (21), it is given by

$$g(y) = \frac{[f_{h\ell+1}(y) \dots f_{h\ell+\ell}(y)]}{F(B) [0_{1 \times h\ell} \quad 1_{1 \times \ell}]^T}. \tag{27}$$

For the FB-PR case, one can define the workload loss probability, denoted by $P_{w,FB-PR}$, as the ratio of the workload lost to the overall amount of workload that has arrived at the buffer. In terms of $g(y)$, $P_{w,FB-PR}$ is expressed as

$$P_{w,FB-PR} = \frac{1}{-\alpha T^{-1} \mathbf{1}} \int_0^B \int_{B-y}^\infty (w - B + y) g(y) D_1(y) \mathbf{1} \alpha e^{T w} T^0 dw dy. \tag{28}$$

3.3. Finite buffer with complete rejection (FB-CR)

In this model, the buffer capacity B is finite, and arrivals with job sizes exceeding the available buffer space are rejected entirely. We will use the same approach as in the IB and FB-PR cases; however, due to the complete rejection policy, the underlying CFMFQ will be different. We will now show that the characterizing matrices of the associated CFMFQ for the FB-CR case are given as

$$Q(x) = \begin{bmatrix} I_\ell \otimes \tilde{T}(x) & I_\ell \otimes \tilde{T}^0(x) \\ D_1(x) \otimes \tilde{\alpha}(x) & \tilde{D}_0(x) \end{bmatrix}, \quad 0 \leq x \leq B, \tag{29}$$

$$R(x) = \begin{bmatrix} I_{h\ell} & \\ & -c(x)I_\ell \end{bmatrix}, \quad 0 \leq x < B, \quad R(B) = \begin{bmatrix} 0_{h\ell \times h\ell} & \\ & -c(B)I_\ell \end{bmatrix}, \tag{30}$$

where

$$\begin{aligned} \tilde{T}^0(x) &= [\tilde{t}_i^0(x)], & \tilde{t}_i^0(x) &= \frac{t_i^0}{1 - v^{(i)} e^{T(B-x)} \mathbf{1}}, & 1 \leq i \leq h, \\ \tilde{T}(x) &= [\tilde{t}_{ij}(x)], & \tilde{t}_{ij}(x) &= t_{ij} \frac{1 - v^{(j)} e^{T(B-x)} \mathbf{1}}{1 - v^{(i)} e^{T(B-x)} \mathbf{1}}, & 1 \leq i, j \leq h, i \neq j, \\ \tilde{t}_{ii}(x) &= - \sum_{j=1, j \neq i}^h \tilde{t}_{ij}(x) - \tilde{t}_i^0(x), & & & 1 \leq i \leq h, \\ \tilde{\alpha}(x) &= [\tilde{\alpha}_i(x)], & \tilde{\alpha}_i(x) &= \alpha_i (1 - v^{(i)} e^{T(B-x)} \mathbf{1}), & 1 \leq i \leq h, \\ \tilde{D}_0(x) &= D_0(x) + D_1(x) \alpha e^{T(B-x)} \mathbf{1}, & & & \end{aligned}$$

and $v^{(i)}$ denotes a row vector of zeros except for a value of 1 at position i . To see this, let us first assume that the associated CFMFQ is in a state at time τ during which the buffer level is increasing. Denoting the phase for the job size at time τ by $s(\tau) = i$, and the buffer level with $X(\tau)$, we already know that the remaining job size, denoted by $u(\tau)$, satisfies $u(\tau) < B - X(\tau)$. Hence, as $\Delta\tau \rightarrow 0$, the quantity $P\{s(\tau + \Delta\tau) = h + 1 \mid s(\tau) = i, u(\tau) < B - X(\tau)\}$

$$\begin{aligned} &= \frac{P\{s(\tau + \Delta\tau) = h + 1, u(\tau) < B - X(\tau) \mid s(\tau) = i\}}{P\{u(\tau) < B - X(\tau) \mid s(\tau) = i\}}, \\ &= \frac{t_i^0 \Delta\tau}{1 - v^{(i)} e^{T(B-X(\tau))} \mathbf{1}}. \end{aligned} \tag{31}$$

Similarly, the probability $P\{s(\tau + \Delta\tau) = j \mid s(\tau) = i, u(\tau) < B - X(\tau)\}$, where $1 \leq j \leq h$, reduces as $\Delta\tau \rightarrow 0$ to

$$\begin{aligned} &= \frac{P\{u(\tau) < B - X(\tau), s(\tau + \Delta\tau) = j \mid s(\tau) = i\}}{P\{u(\tau) < B - X(\tau) \mid s(\tau) = i\}}, \\ &= \frac{P\{u(\tau + \Delta\tau) < B - X(\tau + \Delta\tau) \mid s(\tau + \Delta\tau) = j\} t_{ij} \Delta\tau}{P\{u(\tau) < B - X(\tau) \mid s(\tau) = i\}}, \\ &= \frac{1 - v^{(j)} e^{T(B-X(\tau))} \mathbf{1}}{1 - v^{(i)} e^{T(B-X(\tau))} \mathbf{1}} t_{ij} \Delta\tau + O(\Delta\tau^2). \end{aligned} \tag{32}$$

The north-east and north-west blocks of the characterizing matrix $Q(x)$ in (29) follow from (31) and (32), respectively. Let us now look into epochs of new job arrivals when the buffer level takes the value x . A new job is rejected if its size exceeds the available buffer space $B - x$, which occurs with probability $\alpha e^{T(B-x)} \mathbf{1}$. Therefore, transitions associated with rejected arrivals should contribute to transitions without arrivals. This observation is reflected in the south-east corner of $Q(x)$, given in (29). On the other hand, using Bayes' rule, an accepted job's service will start at phase i with probability $\tilde{\alpha}_i(x) = \alpha_i (1 - v^{(i)} e^{T(B-x)} \mathbf{1})$, which is the probability of being initially at phase i given that the PH-type random variable is less than $B - x$, from which the south-west corner of $Q(x)$ in (29) follows.

After the CFMFQ is defined, its MRMFQ approximation with discretization is carried out exactly as in the FB-PR case, except for a slight modification. Notice that \tilde{T} and \tilde{T}^0 approach infinity as x approaches B . Therefore, we cannot use (22) towards $\tilde{Q}^{(K)}$. In order to be able to carry out numerical calculations, however, we need to use a $\tilde{Q}^{(K)}$ matrix with finite entries. Moreover, arbitrarily large choices as the entries of $\tilde{Q}^{(K)}$ have the potential to lead to an ill-conditioned system of equations. Therefore, we opted for using $\tilde{Q}^{(K)} = Q(T^{(K-1)} + \delta/2)$. We obtained reasonably good results with this midpoint strategy, so we stuck to this method throughout the study.

At this point, we have everything we need for solving the buffer level distribution $g(y)$, which is defined in (20). The solution procedure is exactly the same as in the FB-PR case, and (27) still holds for $g(y)$. Now that we have the buffer level distribution, the job loss probability $P_{\text{FB-CR}}$ can be expressed as

$$P_{\text{FB-CR}} = \frac{\int_0^B g(y) D_1(y) \mathbf{1} \alpha e^{T(B-y)} \mathbf{1} dy}{\int_0^B g(y) D_1(y) \mathbf{1} dy}. \tag{33}$$

On the other hand, the workload loss probability $P_{w, \text{FB-CR}}$ can be written as

$$P_{w, \text{FB-CR}} = \frac{1}{-\alpha T^{-1} \mathbf{1}} \int_0^B \int_{B-y}^\infty w g(y) D_1(y) \mathbf{1} \alpha e^{T w} T^0 dw dy. \tag{34}$$

Note that (34) differs from (28) only by the term w inside the integral instead of the term $w - (B - y)$, which reflects the different rejection policies of these two schemes.

4. Numerical examples

In this section, we present numerical examples to validate the proposed approach. The first example addresses the FB-CR case for which we fix $B = 10$ and $c(x) = 1 - \frac{1}{2} \sin(\frac{2\pi}{B}x)$, which takes values in $[0.5, 1.5]$ in a single period within $[0, B]$. This function is selected since it drives the buffer level towards the middle of the buffer space $[0, B]$. The job size distribution is of phase type, characterized with the pair $(\mathbf{1} \ 0], \begin{bmatrix} -2 & 2 \\ 0 & -3 \end{bmatrix})$. We now define the following function, $\mathcal{D}_{\gamma, B}(x) = \frac{1 - e^{\gamma x/B}}{1 - e^\gamma}$, which takes values in $[0, 1]$ for $x \in [0, B]$. The function $\mathcal{D}_{\gamma, B}(x)$ is an increasing function of x in $[0, B]$, and it is convex (concave) in x when $\gamma > 0$ ($\gamma < 0$). We use $\mathcal{D}_{\gamma, B}(x)$ to experiment with general nonlinear continuous functions which can be increasing or decreasing, and convex or concave. For this example, we set

$$D_0(x) = \begin{bmatrix} -d_{11}(x) & \mathcal{D}_{2, 10}(x) \\ 1 + \mathcal{D}_{-1, 10}(x) & -d_{22}(x) \end{bmatrix}, \quad D_1(x) = \begin{bmatrix} 2 - \mathcal{D}_{-4, 10}(x) & 1 - \mathcal{D}_{-3, 10}(x) \\ 1 - \mathcal{D}_{1, 10}(x) & \frac{1}{2}(1 - \mathcal{D}_{1, 10}(x)) \end{bmatrix}, \tag{35}$$

where $d_{11}(x)$ and $d_{22}(x)$ are selected such that $D_0(x) + D_1(x)$ is stochastic. Note that there are all possible combinations of increasing/decreasing and convex/concave functions in (35). The steady-state density of the buffer level seen at an arbitrary time, which we denote by $\mathbf{g}(y) = g(y)\mathbf{1}$, is plotted for varying number of regimes K in Fig. 1, along with the simulation results. It is clear that the analysis results converge to simulation results as K is increased, with the difference between the two vanishing for $K \geq 1024$. Throughout the rest of the numerical examples, we fix $K = 1024$.

In the second example, for the two finite buffer models FB-PR and FB-CR, we present three scenarios in which the job size distribution is exponential, Erlangian (E_{10}), and hyper-exponential (H_2) with a coefficient of variation of 10 with balanced means [33], all having mean job sizes of 1. We test the proposed approach in two different loading scenarios. For this purpose, the MAP in this example is characterized by the matrices in (35) for the low-loading case, and $D_1(x)$ is doubled for each x for the high-loading case along with the corresponding modifications in $d_{ii}(x)$, $i = 1, 2$. The service speed and the buffer capacity

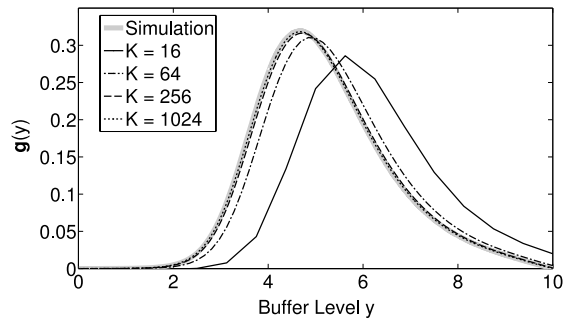


Fig. 1. Steady-state buffer level pdf for varying number of regimes.

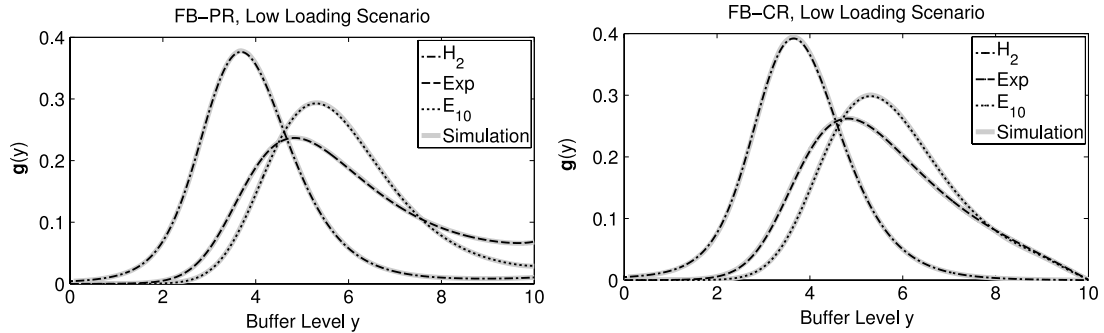


Fig. 2. Steady-state buffer level pdf for the FB-PR and FB-CR policies under low loading.

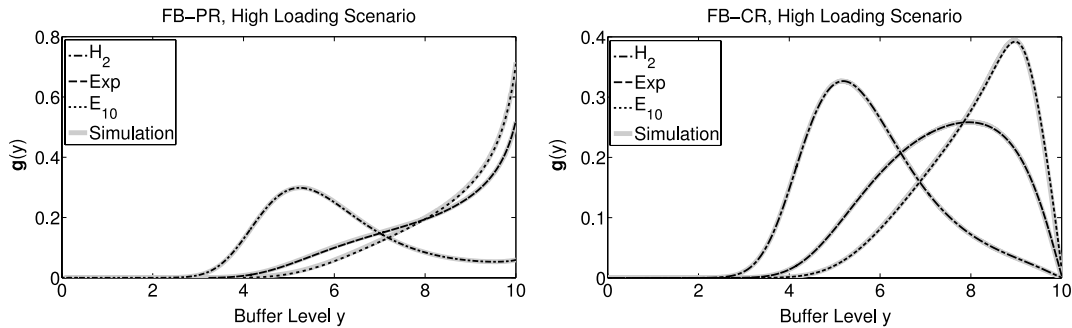


Fig. 3. Steady-state buffer level pdf for the FB-PR and FB-CR policies under high loading.

are chosen as in the previous example. The steady-state buffer level density is given in Figs. 2 and 3 for low-loading and high-loading scenarios, respectively, for the FB-CR and FB-PR rejection policies. The proposed approach perfectly captures the pdf of the buffer level for all the tested job size distributions and for both loading scenarios without any numerical stability problems. Moreover, in contrast to the FB-PR case, the pdf of the buffer level for the FB-CR case vanishes as the buffer level approaches B due to the complete rejection policy. The workload loss probabilities produced by the FB-PR (FB-CR) policy and the corresponding simulation results are given in Table 1 (Table 2), the latter table also presenting the job loss probability. Note that, with the FB-CR policy, larger jobs are more likely to be rejected. This situation is magnified with increased service time variability. If a rejection policy favors smaller jobs as opposed to larger jobs, this would have adverse effect on the workload loss probability but may lead to improved overall job loss probability. For example, in Table 2, in the low-loading case, the job loss rate is not even monotonically increasing with the service time variability for the FB-CR policy.

For validating the approach for the IB scenario, we use the previous example, but with a MAP arrival process characterized by

$$D_0(x) = \begin{bmatrix} -d_{11}(x) & 2 - e^{-x} \\ 3 - e^{-x/2} & -d_{22}(x) \end{bmatrix}, \quad D_1(x) = \begin{bmatrix} e^{-x} & 2e^{-x/2} \\ \frac{1}{2}e^{-x/10} & e^{-x/5} \end{bmatrix},$$

where $d_{11}(x)$ and $d_{22}(x)$ are again selected such that $D_0(x) + D_1(x)$ is stochastic. For the high-loading case, we again doubled $D_1(x)$. The service speed is $c(x) = 3 - \frac{1}{2}e^{-x} \cos(2\pi x)$, which has no particular significance other than having a limit as

Table 1

Workload loss probabilities for the FB-PR policy: simulation results represent 10 runs with 10^6 units of simulated time and 99% confidence intervals.

		$P_{w,FB-PR}$	Simulation
High loading	H_2	4.9714×10^{-1}	$4.9124 \times 10^{-1} \pm 3.1493 \times 10^{-3}$
	Exp	2.9519×10^{-1}	$2.9185 \times 10^{-1} \pm 7.0134 \times 10^{-4}$
	E_{10}	2.6532×10^{-1}	$2.6206 \times 10^{-1} \pm 5.3100 \times 10^{-4}$
Low loading	H_2	4.7235×10^{-1}	$4.6966 \times 10^{-1} \pm 4.0205 \times 10^{-3}$
	Exp	5.9215×10^{-2}	$5.8706 \times 10^{-2} \pm 2.7172 \times 10^{-4}$
	E_{10}	1.3103×10^{-2}	$1.2856 \times 10^{-2} \pm 1.3227 \times 10^{-4}$

Table 2

Workload and job loss probabilities for the FB-CR policy: simulation results represent 10 runs with 10^6 units of simulated time and 99% confidence intervals.

		$P_{w,FB-CR}$	Simulation
High loading	H_2	5.1012×10^{-1}	$5.0963 \times 10^{-1} \pm 4.6680 \times 10^{-3}$
	Exp	3.2352×10^{-1}	$3.2191 \times 10^{-1} \pm 5.1730 \times 10^{-4}$
	E_{10}	2.6643×10^{-1}	$2.6353 \times 10^{-1} \pm 4.4079 \times 10^{-4}$
Low loading	H_2	4.9944×10^{-1}	$4.9837 \times 10^{-1} \pm 4.5928 \times 10^{-3}$
	Exp	1.0427×10^{-1}	$1.0374 \times 10^{-1} \pm 4.7205 \times 10^{-4}$
	E_{10}	1.9091×10^{-2}	$1.8369 \times 10^{-2} \pm 9.7872 \times 10^{-5}$

		P_{FB-CR}	Simulation
High loading	H_2	1.1453×10^{-2}	$1.1333 \times 10^{-2} \pm 6.4506 \times 10^{-5}$
	Exp	1.3568×10^{-1}	$1.3457 \times 10^{-1} \pm 3.6109 \times 10^{-4}$
	E_{10}	2.3274×10^{-1}	$2.2964 \times 10^{-1} \pm 4.4478 \times 10^{-4}$
Low loading	H_2	4.8591×10^{-3}	$4.8388 \times 10^{-3} \pm 4.8460 \times 10^{-5}$
	Exp	3.2901×10^{-2}	$3.2606 \times 10^{-2} \pm 1.2858 \times 10^{-4}$
	E_{10}	1.5654×10^{-2}	$1.5245 \times 10^{-2} \pm 8.3769 \times 10^{-5}$

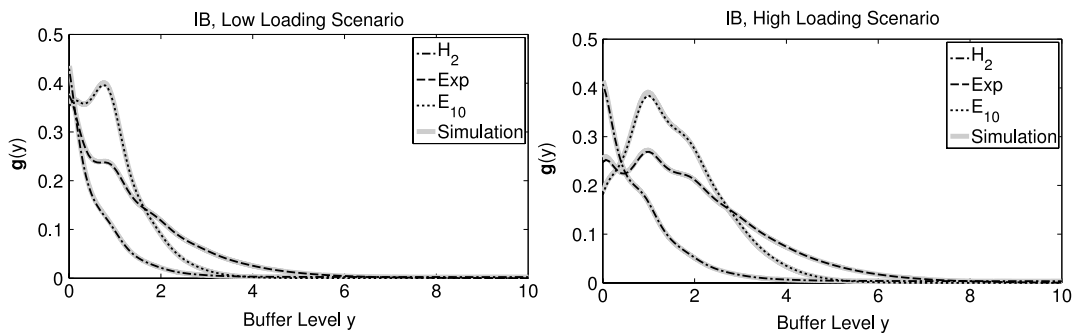


Fig. 4. Steady-state buffer level pdf for the IB case under low and high loading. The $T^{(K-1)}$ values found via (12) for low-loading and high-loading scenarios are 62.2859 and 69.1978, respectively, for $\epsilon = 10^{-3}$. The plots are truncated as the pdf vanishes.

$x \rightarrow \infty$ and being complex enough to produce non-trivial buffer level distributions. The resulting steady-state buffer level pdf obtained by analysis and simulation is depicted in Fig. 4, which clearly demonstrates the agreement between the two. The behaviors of the systems having different job size distributions are quite different, which is reflected in the pdf plots.

As for the final example, we analyze a finite buffer system with complete rejection whose arrivals occur according to an MMPP with N states. The job size has H_2 distribution with mean 5 and coefficient of variation 10 with balanced means. Enumerating the states of the MMPP from 1 to N , a state i transits into all other states with rate i , and leads to an arrival with rate $(\frac{1}{B} - \frac{2}{B} \frac{i-1}{N-1})x + \frac{i-1}{N-1}$ within $0 \leq x \leq B$, where B is the buffer capacity, selected as 10 for this example. In other words, the arrival rate in state 1 linearly goes from 0 up to 1 within $[0, B]$, the arrival rate in state N goes linearly from 1 down to 0 within $[0, B]$, and the rates of the rest of the states have equally spaced slopes that fall within $1/B$ and $-1/B$, while all rate functions intersect at the point $(B/2, 1/2)$. The service speed is taken $c(x) = 1 - \frac{1}{2} \sin(\frac{2\pi}{B}x)$, as before. The resulting pdfs for two values of N are given in Fig. 5.

Besides demonstrating that our method is able to accurately solve workload-dependent buffers with a large number of states, we also provide a computational run time analysis with this example. Run times for various values of N and K are presented in Table 3, which demonstrates the linear dependence of the computational complexity of the proposed approach on the number of regimes. The run time for each case is provided in terms of overall run time, and the three major steps that constitute the proposed method: constructing the block-tridiagonal matrix from the boundary conditions (Matrix Fill),

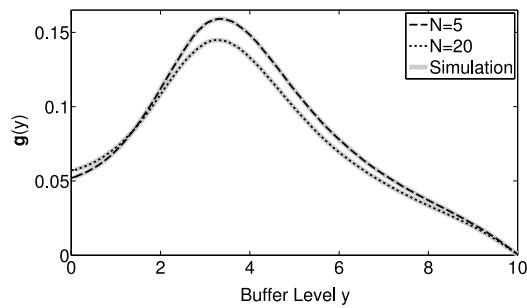


Fig. 5. Steady-state buffer level pdf for the system with MMPP arrivals for $N = 5$ and 20.

Table 3

Run times in units of seconds for various values of N and K obtained on a PC with Intel Core i7, 2.20 GHz processor, and 8 GB RAM.

N	K	Matrix Fill	Block LU	Normalization	Overall
5	256	0.6245	0.0318	0.2284	0.8918
	512	1.2618	0.0668	0.4593	1.7948
	768	1.8902	0.1063	0.6799	2.6842
	1024	2.5127	0.1472	0.8951	3.5635
10	256	0.8897	0.0622	0.2722	1.2327
	512	1.7840	0.1279	0.5401	2.4631
	768	2.6942	0.1979	0.8059	3.7114
	1024	3.6196	0.2721	1.0772	4.9853
15	256	1.2297	0.1100	0.3612	1.7133
	512	2.4828	0.2278	0.7021	3.4304
	768	3.7789	0.3559	1.0391	5.1969
	1024	5.0898	0.4861	1.3930	6.9971
20	256	1.6526	0.1654	0.4426	2.2772
	512	3.3454	0.3406	0.8695	4.5815
	768	5.0736	0.5262	1.2887	6.9237
	1024	6.8673	0.7234	1.7191	9.3542

the block-LU factorization, and the normalization of the solution stemming from (8). It is clear that the most computation-intensive step is the Matrix Fill, which consists of a Schur decomposition and a Sylvester equation for each regime in addition to the calculation of $V^{(k)}(x)$ matrices, a procedure involving matrix exponentiation. We refer to [23] for the computational complexity and stability of the associated numerical algorithms.

5. Conclusions

In this article, using sample path arguments, we reduce the steady-state analysis of a workload-dependent buffer with MAP arrivals and PH-type distributed job sizes to that of a CFMFQ. We then appropriately approximate the CFMFQ using uniform discretization by a K -regime MRMFQ and use the existing boundary conditions to solve for the MRMFQ. Moreover, while solving the boundary equations of the MRMFQ, we take advantage of the block-tridiagonal structure of the equations so as to reduce the computational complexity to $O(K)$. With this method, large number of regimes (in the order of thousands) can relatively easily be used for discretization. We have demonstrated perfect match with simulation results in all scenarios we tested, provided that K is chosen large enough. Our future work will be directed towards applications of the proposed method for the analysis of real-world systems in which controlled queues play a major role. Other directions of future research are the exploration of different discretization techniques used for reducing CFMFQs to MRMFQs, and the comparison of our method to existing CFMFQ solvers in terms of accuracy, numerical stability, and computational complexity.

References

- [1] M.F. Neuts, Structured Stochastic Matrices of M/G/1 Type and Their Applications, Marcel Dekker, NY., 1989.
- [2] D.M. Lucantoni, K.S. Meier-Hellstern, M.F. Neuts, A single server queue with server vacations and a class of nonrenewal arrival processes, *Advances in Applied Probability* 22 (1990) 676–705.
- [3] D. Perry, S. Asmussen, Rejection rules in the M/G/1 queue, *Queueing Systems* 19 (1995) 105–130.
- [4] R. Bekker, S.C. Borst, O.J. Boxma, O. Kella, Queues with workload-dependent arrival and service rates, *Queueing Systems* 46 (2004) 537–556.
- [5] D. Perry, W. Stadje, S. Zacks, The M/G/1 queue with finite workload capacity, *Queueing Systems* 39 (2001) 7–22.
- [6] R. Bekker, Finite-buffer queues with workload-dependent service and arrival rates, *Queueing Systems* 50 (2005) 231–253.
- [7] V. Sharma, J.T. Virtamo, A finite buffer queue, in: *Global Telecommunications Conference, 1999. GLOBECOM '99*, pp. 1053–1057.

- [8] V. Sharma, J.T. Virtamo, A finite buffer queue with priorities, *Performance Evaluation* 47 (2002) 1–22.
- [9] B. van Houdt, Analysis of the adaptive MMAP[K]/PH[K]/1 queue: a multi-type queue with adaptive arrivals and general impatience, *European Journal of Operational Research* 220 (2012) 695–704.
- [10] H.E. Kankaya, N. Akar, Solving multi-regime feedback fluid queues, *Stochastic Models* 24 (2008) 425–450.
- [11] S. Asmussen, *Ruin Probabilities*, in: *Advanced Series on Statistical Science and Applied Probability*, World Scientific, 2000.
- [12] H. Kankaya, N. Akar, Exact analysis of single-wavelength optical buffers with feedback Markov fluid queues, *IEEE/OSA Journal of Optical Communications and Networking* (1) (2009) 530–542.
- [13] L. Kosten, Stochastic theory of data handling systems with groups of multiple sources, *Performance of Computer Communication Systems* (1984) 321–331.
- [14] D. Anick, D. Mitra, M.M. Sondhi, Stochastic theory of a data handling system with multiple sources, *Bell System Technical Journal* 61 (1982) 1871–1894.
- [15] R. Tucker, Accurate method for analysis of a packet speech multiplexer with limited delay, *IEEE Transactions on Communications* 36 (1988) 479–483.
- [16] A. da Silva Soares, G. Latouche, Fluid queues with level dependent evolution, *European Journal of Operational Research* 196 (2009) 1041–1048.
- [17] N.G. Bean, M.M. O'Reilly, Performance measures of a multi-layer Markovian fluid model, *Annals of Operations Research* 160 (2008) 99–120.
- [18] G. Horváth, B. van Houdt, A Multi-layer Fluid Queue with Boundary Phase Transitions and Its Application to the Analysis of Multi-type Queues with General Customer Impatience, in: *2012 Ninth International Conference on Quantitative Evaluation of Systems, QEST*, pp. 23–32.
- [19] A. Badescu, S. Drekcic, D. Landriault, On the analysis of a multi-threshold Markovian risk model, *Scandinavian Actuarial Journal* 2007 (2007) 248–260.
- [20] M. Mandjes, D. Mitra, W. Scheinhardt, Models of network access using feedback fluid queues, *Queueing Systems, Theory and Applications* 44 (2003) 2989–3002.
- [21] W. Scheinhardt, N. Foreest, M. Mandjes, Continuous feedback fluid queues, *Operations Research Letters* 33 (2005) 551–559.
- [22] M. Gribaudo, M. Telek, Stationary analysis of fluid level dependent bounded fluid models, *Performance Evaluation* 65 (2008) 241–261.
- [23] G.H. Golub, C.F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1996.
- [24] D.M. Lucantoni, New results for the single server queue with a batch Markovian arrival process, *Stochastic Models* 7 (1991) 1–46.
- [25] G. Latouche, V. Ramaswami, Introduction to Matrix Analytical Methods in Stochastic Modeling, in: *ASA-SIAM Series on Statistics and Applied Probability*, 2002.
- [26] P. Buchholz, P. Kemper, J. Kriege, Multi-class Markovian arrival processes and their parameter fitting, *Performance Evaluation* 67 (2010) 1092–1106.
- [27] M.F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore, MD, 1989.
- [28] V.G. Kulkarni, Fluid models for single buffer systems, in: J.H. Dshalalow (Ed.), *Frontiers in Queueing: Models and Applications in Science and Engineering*, CRC Press, 1997, pp. 321–338.
- [29] R. German, M. Gribaudo, G. Horváth, M. Telek, Stationary analysis of FSPNs with mutually dependent discrete and continuous parts, in: *IEEE International Workshop on Petri Nets and Performance Models*, 2003.
- [30] P.J. Brockwell, S.I. Resnick, R.L. Tweedie, Storage processes with general release rule and additive inputs, *Advances in Applied Probability* 14 (1982) 392–433.
- [31] T. Dzial, L. Breuer, A. da Silva Soares, G. Latouche, M. Remiche, Fluid queues to solve jump processes, *Performance Evaluation* 62 (2005) 132–146.
- [32] O. Kella, W. Whitt, A storage model with a two-state random environment, *Operation Research* 40 (1992) 257–262.
- [33] H.C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*, John Wiley and Sons, 1986.



Mehmet Akif Yazici earned his B.S. and M.S. degrees in Electrical and Electronics Engineering from Middle East Technical University, Turkey, in 2004 and 2006, respectively. He is currently a Ph.D. candidate in the Department of Electrical and Electronics Engineering, Bilkent University, Turkey. His research interests include computer networks, with emphasis on stochastic modeling, and analysis of telecommunication systems and networks.



Nail Akar received his B.S. degree from Middle East Technical University, Turkey, in 1987, and his M.S. and Ph.D. degrees from Bilkent University, Ankara, Turkey, in 1989 and 1994, respectively, all in Electrical and Electronics Engineering. From 1994 to 1996, he was a visiting scholar and a visiting assistant professor in the Computer Science Telecommunications program at the University of Missouri—Kansas City. He joined the Technology Planning and Integration group at Long Distance Division, Sprint, Overland Park, Kansas, in 1996, where he held a senior member of technical staff position from 1999 to 2000. Since 2000, he has been with Bilkent University, Turkey, currently as an associate professor in the Electrical and Electronics Engineering Department. His current research interests include performance analysis of computer and communication systems and networks, performance evaluation tools and methodologies, design and engineering of optical and wireless networks, queuing systems, and resource management.