Transient and First Passage Time Distributions of First- and Second-order Multi-regime Markov Fluid Queues via ME-fication

Nail Akar · **Omer Gursoy** · **Gabor Horvath** ·
Miklos Telek

Abstract We propose a numerical method to obtain the transient and first passage time distributions of first- and second-order Multi-Regime Markov Fluid Queues (MRMFQ). The method relies on the observation that these transient measures can be computed via the stationary analysis of an auxiliary MRMFQ. This auxiliary MRMFQ is constructed from the original one, using sample path arguments, and has a larger cardinality stemming from the need to keep track of time. The conventional method to approximately model the deterministic time horizon is Erlangization. As an alternative, we propose the so-called *ME-fication* technique, in which a Concentrated Matrix Exponential (CME) distribution replaces the Erlang distribution for approximating deterministic time horizons. ME-fication results in much lower state-space dimensionalities for the auxiliary MRMFQ than would be with Erlangization. Numerical results are presented to validate the effectiveness of ME-fication along with the proposed numerical method.

Keywords Multi-regime Markov fluid queues · Matrix exponential distributions · Transient distribution · First passage time distribution

N. Akar and O. Gursoy
Electrical and Electronics Engineering Department
Bilkent University, Bilkent 06800, Ankara, Turkey
Tel.: +90-312-2902337
Fax: +90-312-2664192
E-mail: {akar,gursoy}@ee.bilkent.edu.tr

G. Horváth
Budapest University of Technology and Economics
Department of Networked Systems and Services
Magyar Tudosok krt. 2, 1117 Budapest, Hungary
Tel.: +36-1-4633254
Fax: +36-1-4633263
E-mail: ghorvath@hit.bme.hu

M. Telek
MTA-BME Information Systems Research Group
Magyar Tudosok krt. 2, 1117 Budapest, Hungary
Tel.: +36-1-4632084
Fax: +36-1-4633263
E-mail: telek@hit.bme.hu

1 Introduction

In the vast majority of stochastic models, the stationary analysis is much simpler than analyzing the transient behavior. Over the past decades, several solution methodologies (matrix-analytic methods, invariant subspace methods, Schur decomposition-based methods, etc.) have been developed to obtain the stationary solution of a large class of structured Markovian systems in a numerically efficient way. There are several results available for the analysis of transient quantities including first passage times and these methods typically provide the solution in Laplace transform domain.

The focus of this paper is the transient analysis of a wide class of Markov Fluid Queues (MFQs). Transient queue lengths and waiting time distributions are approximatively obtained by Houdt and Blondia (2005) using stationary analysis for a discrete-time queue. In Velthoven et al. (2007), an algorithm is proposed to assess the transient performance measures for every possible initial configuration of a Quasi-Birth-and-Death (QBD) Markov chain by means of the stationary solution of another properly constructed Markov chain. Similarly, a numerical method has been proposed by Yazici and Akar (2017) for finding the ruin probabilities for a general continuous-time risk problem using the stationary solution of a certain MFQ. In this paper, we generalize these techniques to the general class of MFQs by which several transient quantities including first passage times are obtained from the stationary solution of an appropriately constructed auxiliary MFQ. The main benefits of this method are that it provides the transient quantity directly (no Laplace transform inversion is involved), and that the availability of the mature stationary solvers ensure the superior numerical stability of the proposed method.

MFQs have been studied for a long time (see Kulkarni (1997) for an early survey) and have been proven to be efficient modeling techniques for several telecommunication systems and insurance risk related practical problems among several further application areas Anick et al. (1982); Gerber and Shiu (1998). Based on such practical motivations, special attention has been devoted to the transient analysis of MFQs. In general, the transient behavior of MFQs can be expressed in a simple way by partial matrix-differential equations. The numerical solution methods to these differential equations can be classified as transform domain and direct (also called time domain) methods. The transform domain methods essentially describe the underlying partial differential equations (PDEs) in Laplace transform domain together with the boundary conditions and apply certain numerical inverse Laplace transformation methods for computing time domain results. In more recent analytical approaches, the transform domain description is based on matrix analytic methods Ahn et al. (2007); Ahn and Ramaswami (2005). One of the seminal results of the direct time domain analysis methods is a randomization-based numerical solution of the underlying PDEs, proposed by Sericola (1998). This solution method has several desirable properties (efficiency, numerical stability, sign change free, etc.), but is restricted to the analysis starting from an idle fluid buffer. To overcome this restriction, later on, transient analysis methods have been proposed based on the concept of replacing the (deterministic) time horizon with a phase type (PH-type) distributed random time and solving the obtained model with the underlying Markov chain modified accordingly. To get a reasonable accurate solution, the PH-type distributed random time horizon must be as close to deterministic as possible. The least varying PH-distributed random variable of order n is the order-n Erlang distribution Aldous and Shepp (1987), which is the sum of n i.i.d. exponentially distributed random times with the same parameter. Consequently, the PH-distributed random time horizon is commonly assumed to be order-n Erlang distributed and the specific properties of the Erlang distribution are applied in the numerical analysis of such methods Asmussen et al. (2002); Ramaswami et al. (2008a). This solution methodology is commonly referred to as *Erlangization*. Erlangization is a robust procedure, but, to obtain high accuracy, high order Erlang distributions need to be used that can increase the model size to a large extent. Recent results suggest that

Concentrated Matrix Exponential (CME) distributions can approximate deterministic time horizons more efficiently Horváth et al. (2016), which we propose to utilize in this work. Inspired by Erlangization-based methods, we call this particular method as *ME-fication* that uses CMEs to approximate a deterministic time horizon.

All the above-mentioned methods, both the transform domain-based and the time domain-based ones, were targeting basic MFQs. In this paper, our aim is to study a wider class of MFQs, where multiple regimes and second order fluid flows are both present. Multi-regime MFQs (also called level-dependent or multi-layer fluid queues) have been investigated for a while to model load-dependent systems and congestion control Le et al. (2007). Two different solution methods are available for their stationary analysis; the matrix-analytical approach da Silva Soares and Latouche (2009) and the Schur decomposition-based approach Kankaya and Akar (2008). Second-order fluid models (Asmussen (1995), also called as Markov-Modulated Brownian Motion, MMBM) are popular extensions of the basic MFQs. The procedure proposed by Horváth and Telek (2017) is able to provide the stationary solution of systems allowing both multiple regimes and second-order fluid flows by the matrix-analytic method. To the best of our knowledge, transient solution has never been developed for this particular system. In this work, our contribution is two-fold. First, we adapt the methods in the literature to compute the transient measures by a stationary solver to second-order multi-regime MFQs. Next, we introduce the *ME-fication* method which provides more accurate results in case of equal model sizes when compared to Erlangization.

The rest of the paper is organized as follows. In Section 2, the preliminaries are introduced: Section 2.1 gives a short overview on the phase-type and the matrix-exponential distributions, Section 2.2 summarizes the approximation of deterministic variables with phase-type and matrix-exponential distributions, and Section 2.3 introduces the existing stationary solution methods of multi-regime MFQs. The proposed auxiliary MFQ based transient analysis of multi-regime MFQs is described in Section 3 while the first passage time analysis is presented in Section 4. Section 5 presents various extensions of the proposed auxiliary MFQ based analysis framework. Section 6 provides a number of numerical examples for validating the effectiveness of the proposed method. Finally, we conclude.

2 Preliminaries

2.1 Phase-Type and Matrix Exponential Distributions

To describe a phase-type (PH) distribution, a continuous-time Markov chain is defined on the state space $\{1, \ldots, N, N+1\}$ with state $N+1$ being absorbing and all other states being transient. The initial probability vector is of the form $(\alpha, 0)$, and the infinitesimal generator of the Markov chain is

$$\begin{bmatrix} A & A^0 \\ 0 & 0 \end{bmatrix}.$$

Here, α is a row probability vector of size N, A is a $N \times N$ transient generator matrix, e denotes a column vector of ones with appropriate size, and $A^0 = -Ae$ is a column vector of size N containing the transition rates to the absorbing state. The fact that α is a probability vector implies that $\alpha \geq 0$ and $\alpha e = 1$; and the fact that S is a transient generator matrix implies that its diagonal elements are strictly negative, the off-diagonal elements are non-negative and $Ae \leq 0$ holds. Let Θ denote the time till absorption into the absorbing state $N+1$. Then, the distribution of Θ is called PH-type, i.e., $\Theta \sim PH(\alpha, A)$ with order N, where the notation $\sim$ is synonymous with "distributed according to". For a detailed study of PH-type distributions, we refer the reader to Neuts (1981). The cumulative distribution function (cdf) and probability density function (pdf) of $\Theta \sim PH(\alpha, A)$, denoted by $F_\Theta(x)$

and $f_{\Theta}(x)$, respectively, are given as:

$$F_{\Theta}(x) = 1 - \alpha e^{Ax} e, \quad f_{\Theta}(x) = -\alpha e^{Ax} Ae, \quad \text{for } x \geq 0. \tag{1}$$

A generalization of the PH distribution is the so-called Matrix Exponential (ME) distribution Asmussen and Bladt (1996); see also Bladt and Neuts (2003); Fackrell (2003); He and Zhang (2007) for a detailed description of ME distributions and their properties. We say $\Theta \sim ME(\alpha, A)$ with order N if the pdf of the random variable Θ is in the form of (1), however, for ME distributions, the parameters α and A do not have to satisfy the sign constraints that apply to PH distributions. The only constraint is that $f_{\Theta}(x)$ must be a legitimate density function. Thus, $f_{\Theta}(x) \geq 0, \forall x \geq 0$. In those general cases, ME distributions do not possess the stochastic interpretation of that of PH distributions.

2.2 Approximating Deterministic Variables

According to the method proposed in this paper, the deterministic time horizon is approximated by a PH or by an ME distribution. If Θ is deterministic with $\Theta = t$, its cdf is a unit step function located at t. PH distributions are able to approximate $\Theta = t$ arbitrarily well. In the Erlangization method (see Asmussen et al. (2002), Ramaswami et al. (2008b)), $\Theta = t$ is approximated by $\tilde{\Theta}_N \sim PH(\alpha_N, A_N)$ (also called Erlang-N with order N) where

$$\alpha_N = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, A_N = \frac{N}{t} \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & & -1 \end{bmatrix}.$$

As the order N increases, $\tilde{\Theta}_N$ converges to $\Theta = t$ in distribution but with relatively slow convergence rate since the squared coefficient of variation (SCV) of $\tilde{\Theta}_N$ is $1/N$. With ME distributions of order N, it is possible to achieve a much lower SCV than that of the Erlang-N distribution. Unfortunately, neither the structure providing the minimal SCV nor the explicit formula for the minimal SCV are known for ME distributions. In Horváth et al. (2016), a family of ME distributions (called concentrated ME distributions or CME in short) is considered, and the parameters providing the minimal SCV are obtained numerically. The main result of Horváth et al. (2016) is that the asymptotic behavior of the SCV of the proposed CME distribution of order N is $2/N^2$.

2.3 Multi-Regime First- and Second-Order Markov Fluid Queues

Conventional Markov Fluid Queues (MFQs) are described by a joint Markovian process $\mathbf{X}(t) = (X_f(t), X_d(t)), t \geq 0$, where $0 \leq X_f(t) \leq B$ represents the fluid level in the buffer, B denotes the buffer capacity, and the discrete modulating phase process $X_d(t) \in \{1, 2, \ldots, n\}$ is a Continuous Time Markov Chain (CTMC) with state space cardinality n and generator matrix Q. Throughout the paper, we assume finite capacity MFQs, i.e., $B < \infty$. In Section 5, we discuss how to handle the infinite buffer capacity case. In MFQs, the net rate of fluid change (or drift) is r_i when the phase of the modulating process $X_d(t)$ is i. The drift matrix R is the diagonal matrix of drifts: $R = \mathbf{diag}\{r_1, r_2, \ldots, r_n\}$ and the process $\mathbf{X}(t)$ is fully characterized with the pair (Q, R) and the initial state $\mathbf{X}(0) = (X_f(0), X_d(0))$. Above, the **diag** operator stands for the diagonal concatenation of its scalar input arguments. When the

arguments are diagonal matrices, **diag** operator stands for the block diagonal concatenation of its matrix input arguments:

$$\mathbf{diag}\{A_1, A_2, \ldots, A_l\} = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & A_l \end{bmatrix},$$

for diagonal $A_j, 1 \leq j \leq l$.

In Level Dependent MFQs (LDMFQ), the drift matrix does not only depend on $X_d(t)$ but it also depends on the instantaneous fluid level $X_f(t)$. Moreover, in LDMFQs, the generator is also allowed to depend on $X_f(t)$. Therefore, LDMFQs are characterized with a pair of level dependent generator and drift matrices $(Q(x), R(x))$ for $0 \leq x \leq B$; see Scheinhardt et al. (2005), da Silva Soares and Latouche (2009). A sub-case of LDMFQs is MRMFQs in which the buffer is partitioned into a finite number of non-overlapping intervals (referred to as regimes) and the drift matrix and the generator matrix are allowed to depend on the regime only and are fixed in each regime; see Kankaya and Akar (2008), Mandjes et al. (2003). Specifically, in MRMFQs, the buffer is partitioned into $K > 1$ regimes with the boundaries $0 = T^{(0)} < T^{(1)} < \cdots < T^{(K-1)} < T^{(K)} = B < \infty$. When $T^{(k-1)} < X_f(t) < T^{(k)}$, the fluid process is said to be in regime k at time t. The MRMFQ is characterized with the level-dependent pair of matrices $(Q(x), R(x))$ which turn out to have the following specific form:

$$Q(x) = \begin{cases} Q^{(k)} & \text{if } T^{(k-1)} < x < T^{(k)}, \quad k = 1, 2, \ldots, K, \\ \tilde{Q}^{(k)} & \text{if } x = T^{(k)}, \quad k = 0, 1, \ldots, K, \end{cases} \tag{2}$$

$$R(x) = \begin{cases} R^{(k)} & \text{if } T^{(k-1)} < x < T^{(k)}, \quad k = 1, 2, \ldots, K, \\ \tilde{R}^{(k)} & \text{if } x = T^{(k)}, \quad k = 0, 1, \ldots, K, \end{cases} \tag{3}$$

where the regime-k generator and drift matrices are denoted by $Q^{(k)}$ and $R^{(k)}$, respectively, and the boundary-k generator and drift matrices are denoted by $\tilde{Q}^{(k)}$ and $\tilde{R}^{(k)}$, respectively.

Another generalization of MFQs allows the so-called second order fluid accumulation which follows a Brownian motion with drift. In this case, apart from the generator matrix Q and the fluid drift matrix R, the system has a third parameter matrix, namely a diagonal matrix $S = \mathbf{diag}\{s_1, s_2, \ldots, s_n\}$ describing the variance of the Brownian motion in various states of the background process. Those states where $s_i = 0$ holds are called first-order states and behave as described before. For the second-order states with $s_i > 0$, the fluid increment in an infinitesimally small time interval $(t, t+\Delta)$ is normally distributed with mean $r_i\Delta$ and variance $s_i\Delta$. For the second-order states, two types of boundary behavior are commonly assumed in the literature: absorbing and reflecting boundaries. In case of the absorbing behavior, when the Brownian motion representing the fluid level reaches a boundary, it sticks to the boundary till the next state transition. In case of the reflecting behavior, the fluid process is reflected by the boundary. Both have been studied extensively in the literature; see Asmussen (1995) and Karandikar and Kulkarni (1995).

Second-order fluid models can be studied in the multi-regime setting as well. The definition of these systems is similar to the above described first-order case: matrices Q, R, and S are regime-dependent. At the internal boundaries of second-order MRMFQs $(T^{(1)}, \ldots, T^{(K-1)})$ we might have a number of different behaviors (transition, reflection, absorption; potentially different when reaching the boundary from below or from above) as it is detailed in Horváth and Telek (2017). Here we only note that those different boundary behaviors can be described by $\tilde{Q}^{(k)}$ and $\tilde{R}^{(k)}$ (as in the first order case) and the variance parameters right at the boundary ($\tilde{S}^{(k)}$) are not needed for the model behavior.

The regime-k steady-state joint probability density function (pdf) vector $f^{(k)}(x)$ of a first- or second-order MRMFQ is defined as

$$f^{(k)}(x) = \left[f_1^{(k)}(x) \; f_2^{(k)}(x) \; \cdots \; f_n^{(k)}(x) \right], \tag{4}$$

where

$$f_i^{(k)}(x) = \lim_{t \to \infty} \frac{d}{dx} \Pr\{X_f(t) \le x, X_d(t) = i\}, \quad T^{(k-1)} < x < T^{(k)}, 1 \le k \le K. \tag{5}$$

Similarly, the steady-state boundary-k probability mass (pma) vector $c^{(k)}$ is defined as

$$c^{(k)} = \left[c_1^{(k)} \; c_2^{(k)} \; \cdots \; c_n^{(k)} \right], \quad c_i^{(k)} = \lim_{t \to \infty} \Pr\{X_f(t) = T^{(k)}, X_d(t) = i\}, \quad 0 \le k \le K. \tag{6}$$

For the first-order MRMFQs, a matrix-analytical algorithm has been proposed in Kankaya and Akar (2008) to obtain the joint pdf vector given in (4) in matrix exponential form and the joint pma vector in (6). This numerical algorithm requires the solution of a linear matrix equation of at most size $n(2K+1)$ for an MRMFQ with n states and K regimes. The computational complexity of the proposed algorithm can be reduced to $O(n^3 K)$ on the basis of the observation that the linear matrix equation is in block tri-diagonal form Yazici and Akar (2013). Moreover, Yazici and Akar (2013) show that the more general LDMFQs can effectively be approximated by their MRMFQ counterparts by properly discretizing the level-dependent generator and drift matrices thanks to the linear dependence of the computational complexity on the number of regimes.

For the stationary solution of second-order MRMFQs, the parameters of the matrix-exponential solution have been derived in Horváth and Telek (2017). According to that procedure, matrix-quadratic equations are solved to obtain the matrix coefficients, while the vector parameters of the solution are given by a set of linear equations of size $n(K+1) + \sum_{k=1}^{K} n_+^{(k)} + \sum_{k=1}^{K} n_-^{(k)} + 2\sum_{k=1}^{K} n_\sigma^{(k)}$, where $n_-^{(k)}$ denotes the number of first-order states with negative rate, $n_+^{(k)}$ the number of first-order states with positive rate, and $n_\sigma^{(k)}$ the number of second-order states in regime k, respectively. As in the first-order case, the set of linear equations can be re-ordered to a tri-diagonal form to enable faster numerical solution. We note that in the first-order case, $n_+^{(k)} + n_-^{(k)} = n, n_\sigma^{(k)} = 0$, and $n(K+1) + \sum_{k=1}^{K} n_+^{(k)} + \sum_{k=1}^{K} n_-^{(k)} + 2\sum_{k=1}^{K} n_\sigma^{(k)}$ simplifies to $n(2K+1)$.

3 Transient Solution of Second-Order Multi-regime Markov Fluid Queues

Throughout the paper, the focus will be on second-order MRMFQs which will shortly be referred to as MFQs for convenience since a first-order MFQ is a special sub-case. We are given the MFQ process $\mathbf{X}(t) = (X_f(t), X_d(t))$ with the fluid level process $0 \le X_f(t) \le B$ and the modulating process $X_d(t) \in \{1, 2, \ldots, n\}$. We assume that the MFQ is characterized with three piece-wise constant matrices $(Q_X(x), R_X(x), S_X(x))$. Let us be given a time horizon Θ which is PH distributed of order N, characterized with the pair (α, A) with $A^0 = -Ae$ and e being a $N \times 1$ column vector of ones. The level process is assumed to start operation at $X_f(0) = a$ and the phase process at $X_d(0) = i$.

We are interested in finding the following joint pdf and joint pma when the MFQ evolves until the random time horizon Θ expires:

$$f_\Theta^{a,i}(j, x) = \frac{d}{dx} \Pr\{X_f(\Theta) \le x, X_m(\Theta) = j | X_f(0) = a, X_d(0) = i\}, \tag{7}$$

$$c_\Theta^{a,i}(j, x) = \Pr\{X_f(\Theta) = x, X_m(\Theta) = j | X_f(0) = a, X_d(0) = i\}. \tag{8}$$
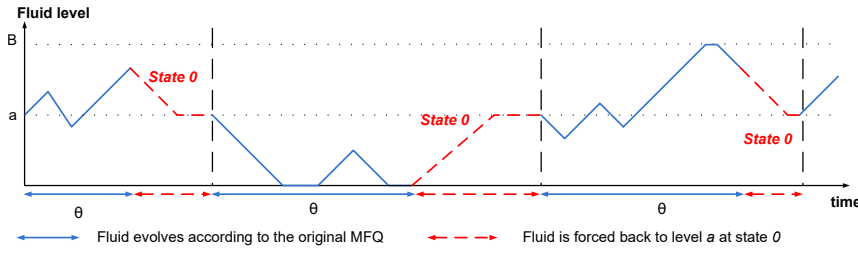
Fig. 1 A sample path of the auxiliary MFQ $\mathbf{Y}(t)$ for the transient distribution of the original MFQ $\mathbf{X}(t)$.

We note, also for the rest of the paper, that the derivative in (7) is computed only for x different from any regime boundary $T^{(k)}$, $0 \leq k \leq K$, and $c_\Theta^{a,i}(j,x)$ can be different from zero only at regime boundaries. To avoid pathological cases, we assume that a is not a regime boundary and $a \neq x$. For $x = a$ and a is not a regime boundary, $f_\Theta^{a,i}(j,a)$ can be approximated as $f_\Theta^{a,i}(j,a) \simeq f_\Theta^{a,i}(j,a+\delta)/2 + f_\Theta^{a,i}(j,a-\delta)/2$ with sufficiently small δ. When Θ is deterministic with $\Theta = t$, then (7) and (8) provide expressions for the joint cdf and pma of the transient behavior of the MFQ at time t.

We now construct an auxiliary MFQ with a larger state space whose steady-state solution provides an approximate solution for the distribution given in (7) and (8) of the original MFQ. Consider the sample path of the auxiliary fluid process in Fig. 1 (which is a purely first-order example, for simplicity). The fluid process given in Fig. 1 is actually an MFQ process denoted by $\mathbf{Y}(t) = (Y_f(t), Y_d(t))$ where the discrete modulating process $Y_d(t)$ has $1 + Nn$ states where Nn of these states correspond to the pairs $(k,\ell), 1 \leq k \leq N, 1 \leq \ell \leq n$, k keeping track of the phase of the time horizon Θ whereas ℓ being the phase of the modulating process $X_d(t)$ of the original MFQ. We order the states of $Y_d(t)$ as $(0, (1,1), (1,2), \ldots, (1,n), (2,1), \ldots, (N,n))$, where state 0 is an auxiliary state used to reset the $\mathbf{Y}(t)$ process to the initial state of the $\mathbf{X}(t)$ process and the PH distributed time horizon Θ.

A sample path of $\mathbf{Y}(t)$ is depicted in Fig. 1. This fluid level process starts at level a and evolves according to the original MFQ and to its level-dependent generator, rate, and variance matrices $Q_X(x), R_X(x)$, and $S_X(x)$, respectively, until the time horizon $\Theta \sim PH(\alpha, A)$ is reached (solid/blue line). When the timer expires, the fluid level process is forced back to the initial value a via an auxiliary state designated as state 0 (dashed/red line). This is achieved by a negative drift (positive drift) and zero variance at state 0 if the fluid level was above a (below a) at the epoch of timer expiration. Once the fluid level reaches level a at state 0, the level is forced to stay at the boundary a with zero drift for an exponentially distributed duration with unit mean.[1] The fluid process then escapes from state 0 to state $(k,i), 1 \leq k \leq N$, with probability α_k and subsequently this pattern repeats forever in Fig. 1. This approach is indeed the adaptation of the methods introduced in Houdt and Blondia (2005) and Yazici and Akar (2017) to second-order MRMFQ systems.

We decompose the state space of $Y_d(t)$ into two subsets: $\{0\}$ of size 1 and $\{(1,1), (1,2), \ldots, (1,n), (2,1), \ldots, (N,n)\}$ of size Nn. The subset-based matrix blocks of the

---

[1] With respect to the transient analysis of the process $\mathbf{X}(t)$, it is not necessary that the auxiliary process stays at level a for an exponentially distributed time. It would also be sufficient if the auxiliary process jumps to the next Θ-long phase immediately after reaching level a. The reason why we do not apply that approach is that forced, immediate state transition of the background process at a fluid level is a very special feature in fluid queues which are not supported by general fluid model solvers, and requires the application of a special fluid model solver introduced by Horvath and Van Houdt (2012). Allowing an exponentially distributed sojourn with rate 1 at level a does not give rise to any inaccuracy and moreover makes the auxiliary process to be a standard MRMFQ (without immediate transitions) for which solution methods and implemented codes are more commonly available.

characterizing matrices of the auxiliary MFQ process $\mathbf{Y}(t)$ are given as:

$$Q_Y(x) = \left[\begin{array}{c|c} 0 & 0 \\ \hline A^0 \otimes e & I_N \otimes Q_X(x) + A \otimes I_n \end{array}\right], \text{ if } x \neq a, \tag{9}$$

$$R_Y(x) = \begin{cases} \mathbf{diag}\{1, I_N \otimes R_X(x)\} & \text{if } x < a, \\ \mathbf{diag}\{-1, I_N \otimes R_X(x)\} & \text{if } x > a, \end{cases} \tag{10}$$

$$S_Y(x) = \mathbf{diag}\{0, I_N \otimes S_X(x)\}, \text{ if } x \neq a. \tag{11}$$

The process $\mathbf{Y}(t)$ has an extra boundary at $x = a$ with parameters

$$Q_Y(a) = \left[\begin{array}{c|c} -1 & \alpha \otimes e_i \\ \hline A^0 \otimes e & I_N \otimes Q_X(a) + A \otimes I_n \end{array}\right], \tag{12}$$

$$R_Y(a) = \mathbf{diag}\{0, I_N \otimes R_X(a)\}, \tag{13}$$

$$S_Y(a) = \mathbf{diag}\{0, 0_{Nn \times Nn}\}, \tag{14}$$

where I_l is the identity matrix of size l, $0_{l \times k}$ is an $l \times k$ matrix of zeros, e_i is the size-n row vector of zeros with the only non-zero element in position i being equal to 1. We note that the subscripts indicative of the sizes are dropped throghout the paper when the sizes are clear from the context.

The first row of $Q_Y(x)$ in (9) can be interpreted as follows. As long as $Y_d(t)$ stays in the first subset, that is in state 0, and the fluid level is different from a, the process remains in state 0. The second row of $Q_Y(x)$ in (9) ensures that the process stays in the second subset for a θ long interval and during this period, it follows the behaviour of $X_d(t)$. At the completion of the Θ-long interval, indicated by a state transition with rate A^0, $Y_d(t)$ moves to state 0. The only case when $Y_d(t)$ can leave state 0 is at fluid level a according to the first row of $Q_Y(a)$ in (12) and the exit rate is 1. Leaving state 0, $Y_d(t)$ starts according to the initial distribution of the PH distribution in a state representing $X_d(t) = i$. The interpretations of $R_Y(x)$ and $S_Y(x)$ follow the same pattern.

We need the following definitions for the steady-state distributions of the auxiliary MFQ $\mathbf{Y}(t)$. Let $f_Y(s,x)$ and $c_Y(s,x)$ denote the steady-state joint pdf and pma of the MFQ $\mathbf{Y}(t)$:

$$f_Y(s,x) = \frac{d}{dx} \lim_{t \to \infty} \Pr\{Y_f(t) \leq x, Y_d(t) = s\}, \tag{15}$$

$$c_Y(s,x) = \lim_{t \to \infty} \Pr\{Y_f(t) = x, Y_d(t) = s\}, \tag{16}$$

for $s = 0$ or $s = (k, \ell), 1 \leq k \leq N, 1 \leq \ell \leq n$.

Theorem 1 *The transient density, $f_{\Theta}^{a,i}(j,x)$, and the transient probability masses at the boundaries, $c_{\Theta}^{a,i}(j,T^{(k)})$, are obtained from the related stationary density and probability masses of $\mathbf{Y}(t)$ as*

$$f_{\Theta}^{a,i}(j,x) = \frac{\sum_u f_Y((u,j),x) A_u^0}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x) dx + \sum_{v=0}^K c_Y((u,\ell),T^{(v)}) \right) A_u^0}, \tag{17}$$

$$c_{\Theta}^{a,i}(j,T^{(k)}) = \frac{\sum_u c_Y((u,j),T^{(k)}) A_u^0}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x) dx + \sum_{v=0}^K c_Y((u,\ell),T^{(v)}) \right) A_u^0}, \tag{18}$$

where A_u^0 denotes the uth entry of A^0.

Proof As it is depicted in Fig. 1, $\mathbf{Y}(t)$ follows a cyclic behaviour of stochastically identical cycles (separated by vertical dashed lines in Fig. 1). In each cycle, $\mathbf{Y}(t)$ starts from level a and MRMFQ background state i and spends a $PH(\alpha,A)$ distributed time in subset $\{(1,1),(1,2),\ldots,(1,n),(2,1),\ldots,(N,n)\}$. We are interested in the system state at the end of the $PH(\alpha,A)$ distributed phase of a cycle (at the end of the blue/solid line in Fig. 1), based on the stationary behaviour of $\mathbf{Y}(t)$.

Due to the independent and stochastically identical cycles of $\mathbf{Y}(t)$, we have

$$f_{\Theta}^{a,i}(j,x)\Delta + \sigma(\Delta) = \tag{19}$$

$$= \Pr\{\text{at a transition to 0 the fluid level is in } (x,x+\Delta) \text{ and the MRMFQ state is } j\}$$

$$= \lim_{\delta \to 0}\lim_{t \to \infty} \frac{\Pr\{\text{transition to 0 in } (t,t+\delta),}{\text{at time } t \text{ the fluid level is in } (x,x+\Delta) \text{ and the MRMFQ state is } j\}}{\Pr\{\text{transition to 0 in } (t,t+\delta)\}}$$

$$= \lim_{\delta \to 0} \frac{\sum_u \left(f_Y((u,j),x)\Delta + \sigma(\Delta)\right)\left(A_u^0\delta + \sigma(\delta)\right)}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x)dx + c_Y((u,\ell),0) + \sum_{v=0}^K c_Y((u,\ell),T^{(v)})\right)\left(A_u^0\delta + \sigma(\delta)\right)}$$

$$= \frac{\sum_u \left(f_Y((u,j),x)\Delta + \sigma(\Delta)\right)A_u^0}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x)dx + c_Y((u,\ell),0) + \sum_{v=0}^K c_Y((u,\ell),T^{(v)})\right)A_u^0},$$

where $\sigma(\cdot)$ is an error term for which $\lim_{\delta \to 0}\sigma(\delta)/\delta = 0$. Dividing the left and right hand sides of (19) by Δ and taking the limit as $\Delta \to 0$, we obtain the identity (17). The proof of (18) follows the same pattern. □

4 First Passage Times in Multi-regime Markov Fluid Queues

Consider the same MFQ process $\mathbf{X}(t) = (X_f(t),X_d(t))$ as described in the previous section characterized with matrices $(Q_X(x),R_X(x),S_X(x))$. Similar to the previous section, the MFQ process is assumed to start operation at $X_f(0) = a$ and $X_d(0) = i$. Let $\tau^{a,i,b}$ denote the first passage time from level a to level b, defined as

$$\tau^{a,i,b} = \inf_t\{X_f(t) = b | X_f(0) = a, X_d(0) = i\}. \tag{20}$$

We are interested in the probability defined by

$$F_\tau^{a,i,b}(\Theta) = \Pr\{\tau^{a,i,b} < \Theta\}, \tag{21}$$

for PH-distributed Θ of order N characterized with the pair (α,A). When Θ is deterministic with $\Theta = t$, then Eqn. (21) provides an expression for the cdf of the first passage time from level a to level b at time t. Without loss of generality we assume that $b > a$. If $b < a$, one can transform the original problem by mirroring the fluid level.

We propose to construct an auxiliary MFQ denoted by $\mathbf{Z}(t)$, whose steady-state solution provides a solution for the probability given in (21) for the original MFQ $\mathbf{X}(t)$. The main idea is to construct an MFQ, with similar cyclic behavior as before, which stops the fluid process in every cycle once it reaches level b. In case of a first-order system, consider the sample path of the auxiliary fluid process $\mathbf{Z}(t)$ in Fig. 2. This process starts at level a and evolves according to the original MFQ and to its level-dependent generator and rate matrices $Q_X(x)$ and $R_X(x)$, until either the time horizon $\Theta \sim PH(\alpha,A)$ is reached or the process hits level b before Θ (solid/blue line). In the former situation, when the timer expires, the fluid process is forced back to the desired initial level a through the auxiliary state 0 spending an exponentially distributed time duration with unit mean at the particular level a (dashed/red
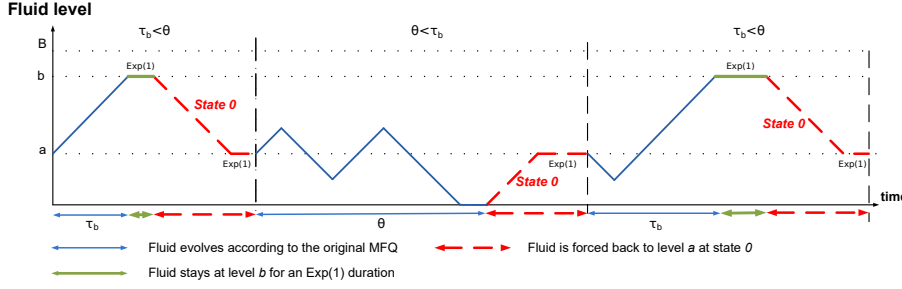
Fig. 2 A sample path of the auxiliary MFQ $\mathbf{Z}(t)$ constructed for the first passage times of the original MFQ $\mathbf{X}(t)$.

line). The second cycle in Fig. 2 is an example for this type of situation. However, the fluid process may also reach level b before the timer expires as in the first and third cycles of Fig. 2. When this happens, the fluid process is forced to stay at fluid level b for an exponentially distributed time duration with unit mean before transitioning to state 0 (thick/green line)[2]. This pattern of cycles repeats as shown in Fig. 2.

The same idea can be generalized to second-order systems as well. The only additional feature to introduce to the auxiliary MFQ $\mathbf{Z}(\mathbf{t})$ is that the boundary at level b has to be absorbing for the second order states, meaning that the fluid process stays there once it reaches level b.

The fluid process defined accordingly is actually an MFQ process denoted by $\mathbf{Z}(t) = (Z_f(t), Z_d(t))$ with the same state-space as that of the MFQ process $\mathbf{Y}(t)$ of the previous section. Actually, the $(Q_Z(x), R_Z(x), S_Z(x))$ parameters are the same as the $(Q_Y(x), R_Y(x), S_Y(x))$ parameters, for $0 \le x < b$. When the fluid level is b, the fluid process needs to stay at this level for an exponentially distributed duration with unity mean before eventually escaping to state 0. Therefore, the generator of the background process at fluid level b is

$$Q_Z(b) = \left[\begin{array}{c|c} 0 & 0 \\ \hline e & -I_{Nn} \end{array} \right], \tag{22}$$

and the matrix of the fluid rates is

$$R_Z(b) = \mathbf{diag}\{-1, 0_{Nn \times Nn}\}. \tag{23}$$

As observed in Fig. 2, the overall trajectory is cyclic. All cycles terminate with the fluid level staying at level a for an exponentially distributed duration of time with unit mean (dashed/red line). In cycles where the fluid level reaches the level b before the timer expires, the fluid process visits the level b for an exponentially distributed duration of time with unit mean (thick/green line). We define the probability masses for MFQ $\mathbf{Z}(t)$ similar to (16). For this purpose, let $c_Z((k, \ell), b)$ denote the steady-state probability mass at level b for the state $(k, \ell), 1 \le k \le m, 1 \le \ell \le n$ for the auxiliary MFQ $\mathbf{Z}(t)$ and similarly $c_Z(0, a)$ denotes the steady-state probability mass at level a for state 0.

Theorem 2 *The first passage time probability, $F_\tau^{a,i,b}(\Theta)$, is obtained from the stationary probability masses of the auxiliary MFQ $\mathbf{Z}(t)$ as*

$$F_\tau^{a,i,b}(\Theta) = \frac{\sum_k \sum_\ell c_Z((k, \ell), b)}{c_Z(0, a)}. \tag{24}$$

---

[2] Also in this case the exponentially distributed delay helps to avoid immediate state transitions of the fluid model upon reaching level b.

Proof Similar to $\mathbf{Y}(t)$, $\mathbf{Z}(t)$ also follows a cyclic behaviour of stochastically identical cycles. In a cycle where level b is not reached (e.g., the second cycle of Fig. 2) $\mathbf{Z}(t)$ spends zero time at level b and spends one unit of time on the average at level a in state 0 [3]. In a cycle where level b is reached (e.g., the first cycle of Fig. 2) $\mathbf{Z}(t)$ spends one unit of time at level b and one unit of time at level a on the average. As a result, the ratio of the mean time spent at level b to that spent at level a gives $F_\tau^{a,i,b}(\Theta)$. Since $\mathbf{Z}(t)$ is ergodic, the time average can also be obtained from the stationary probabilities at level b and level a according to (24). □

5 Extensions of the Basic Models

5.1 Transient Analysis with Random Initial State

In the previous sections, we assumed that the MFQ process $\mathbf{X}(t)$ starts from a deterministic state $X_f(0) = a$ and $X_d(0) = i$. In this section, we provide a similar auxiliary MFQ based analysis of the case when $X_f(0)$ is Finite PH (FPH) distributed on $(0, B)$ and $X_d(0)$ is discrete distributed on $\{1, 2, \ldots, n\}$ such that

$$\Pr\{X_f(0) < x\} = \frac{1 - \beta e^{Mx} e}{1 - \beta e^{MB} e}, \ 0 \leq x \leq B,$$

(see Ramaswami and Viswanath (2014) and He et al. (2019)) and $\Pr\{X_d(0) = i\} = \pi_i$. In this case, we say that the initial fluid level is a random variable denoted by Ψ which is $FPH(\beta, M, B)$-distributed where β is a row vector of size m and M is of size $m \times m$, and the size n row vector composed of the probabilities π_i is denoted by π. We are interested in finding the following joint pdf and joint pma:

$$f_\Theta^{\Psi,\pi}(j,x) = \frac{d}{dx} \Pr\{X_f(\Theta) \leq x, X_m(\Theta) = j | X_f(0) \sim \Psi, X_d(0) \sim \pi\}, \qquad (25)$$

$$c_\Theta^{\Psi,\pi}(j,x) = \Pr\{X_f(\Theta) = x, X_m(\Theta) = j | X_f(0) \sim \Psi, X_d(0) \sim \pi\}, \qquad (26)$$

where $\Theta \sim PH(\alpha, A)$.

To evaluate the transient behavior of this system, we introduce the auxiliary MFQ process $\mathbf{V}(t)$ with state space $\{0\} \cup \{1, 2, \ldots, m\} \cup \{(1,1), (1,2), \ldots, (1,n), (2,1), \ldots, (N,n)\}$. State 0 has the same role as before. To reset the fluid level to an initial value, which is level zero in case of $\mathbf{V}(t)$, states $1, 2, \ldots, m$ set the finite PH distributed initial fluid level. These states are visited as long as the initial fluid level is reached. The fluid rate in these states is one which ensures that the initial fluid level is identical with the time spent in the set of states $\{1, 2, \ldots, m\}$. States $(1,1), (1,2), \ldots, (1,n), (2,1), \ldots, (N,n)$ represent the evolution of the original MFQ, similar to $\mathbf{Y}(t)$.

A sample path of the auxiliary fluid process $\mathbf{V}(t)$ is depicted in Fig. 3. $\mathbf{V}(t)$ also follows a cyclic behaviour with stochastically identical cycles. In each cycle, the process $\mathbf{V}(t)$ starts at level 0 and in the first phase (dotted/yellow line), it sets the $FPH(\beta, M, B)$ distributed initial fluid level. To set the initial fluid level to be finite PH distributed, we need to consider that the transient process characterized by generator M can last longer than B and in this case the fluid level needs to be reset to zero with the use of state 0 as it is exemplified in the second cycle in Fig. 3. If the transient process characterized by generator M concludes within B, then we have an exit transition according to M^0 before time B after which the process evolves according to the original MFQ indicated by the solid/blue line in Fig. 3.

---

[3] It is possible that $\mathbf{Z}(t)$ spends some time at level 0, but not in state 0, as it is exemplified by the second cycle of Fig. 4.
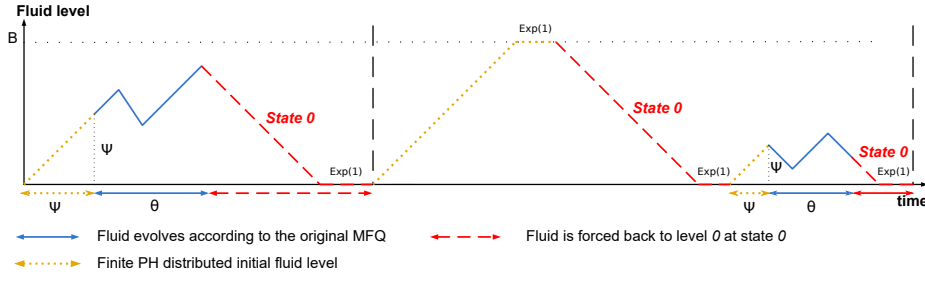
Fig. 3 A sample path of the auxiliary MFQ $\mathbf{V}(t)$ constructed for the transient analysis of the original MFQ $\mathbf{X}(t)$ with random initial state.

This phase concludes when the time horizon $\Theta \sim PH(\alpha,A)$ is reached. At this point, the fluid process is forced back to level 0 through the auxiliary state 0 (dashed/red line) in the last phase of the cycle. This pattern of phases repeats in all cycles as shown in Fig. 3.

According to these three subsets of states, the blocks of the characterizing matrices of $\mathbf{V}(t)$, $(Q_V(x), R_V(x), S_V(x))$, are given as

$$
Q_V(x) = \left[\begin{array}{c|c|c}
0 & 0 & 0 \\\hline
0 & M & M^0(\alpha \otimes \pi) \\\hline
A^0 \otimes e & 0 & I_N \otimes Q_X(x) + A \otimes I_n
\end{array}\right],
\tag{27}
$$

$$
R_V(x) = \mathbf{diag}\{-1, I_m, I_N \otimes R_X(x)\},
\tag{28}
$$

$$
S_V(x) = \mathbf{diag}\{0, 0_m, I_N \otimes S_X(x)\},
\tag{29}
$$

for $0 < x < B$, and for the boundaries we have

$$
Q_V(0) = \left[\begin{array}{c|c|c}
-1 & \beta & 0 \\\hline
0 & M & M^0(\alpha \otimes \pi) \\\hline
A^0 \otimes e & 0 & I_N \otimes Q_X(0) + A \otimes I_n
\end{array}\right],
\tag{30}
$$

$$
R_V(0) = \mathbf{diag}\{0, I_m, I_N \otimes R_X(0)\},
\tag{31}
$$

$$
Q_V(B) = \left[\begin{array}{c|c|c}
0 & 0 & 0 \\\hline
e & -I_m & 0 \\\hline
A^0 \otimes e & 0 & I_N \otimes Q_X(B) + A \otimes I_n
\end{array}\right],
\tag{32}
$$

$$
R_V(B) = \mathbf{diag}\{-1, 0_m, I_N \otimes R_X(B)\}.
\tag{33}
$$

Let $f_V((u,j),x)$ and $c_V((u,j),x)$ denote the steady-state joint pdf and pma of the MFQ $\mathbf{V}(t)$ for the particular state $(u,j), 1 \le u \le N, 1 \le j \le n$. Similar to the deterministic initial state, from sample path arguments, it follows that the transient density $f_\Theta^{\Psi,\pi}(j,x)$ can be obtained from the stationary densities $f_V((\cdot,j),x)$ embedded at the completion epochs of the PH distributed time horizon Θ. The following theorem presents our result.

Theorem 3 *With initial fluid level Ψ and initial state distribution π, the transient density, $f_\Theta^{\Psi,\pi}(j,x)$, and the transient probability masses at the boundaries, $c_\Theta^{\Psi,\pi}(j,T^{(k)})$, are obtained from the stationary behaviour of $\mathbf{V}(t)$ as follows:*

$$
f_\Theta^{\Psi,\pi}(j,x) = \frac{\sum_u f_V((u,j),x)A_u^0}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x)dx + \sum_{v=0}^K c_Y((u,\ell),T^{(v)})\right)A_u^0},
$$

$$
c_\Theta^{\Psi,\pi}(j,T^{(k)}) = \frac{\sum_u c_V((u,j),T^{(k)})A_u^0}{\sum_u \sum_\ell \left(\int_{x=0}^B f_Y((u,\ell),x)dx + \sum_{v=0}^K c_Y((u,\ell),T^{(v)})\right)A_u^0},
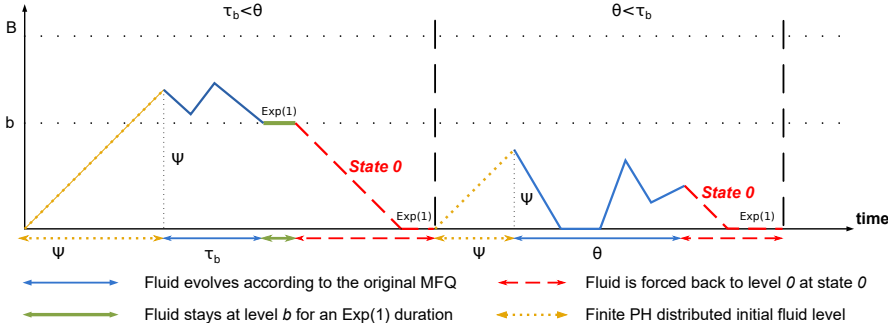$$

Fig. 4 A sample path of the auxiliary MFQ $\mathbf{W}(t)$ constructed for the first passage time analysis of the original MFQ $\mathbf{X}(t)$ with random initial state.

where A_u^0 denotes the uth entry of A^0, as before.

Proof The proof is practically identical with the proof of Theorem 1.   □

5.2 First Passage Time with Random Initial State

Let us assume $X_f(0)$ is $FPH(\beta, M, B)$ distributed and $X_d(0)$ is discrete distributed according to π. Let the first passage time be defined as:

$$\tau^{\Psi,\pi,b} = \inf_t \{X_f(t) = b | X_f(0) \sim \Psi, X_d(0) \sim \pi\}. \tag{34}$$

We define the following first passage time distribution

$$F_\tau^{\Psi,\pi,b}(\Theta) = \Pr\{\tau^{\Psi,\pi,b} < \Theta\}, \tag{35}$$

which can be obtained by an auxiliary MFQ process $\mathbf{W}(t)$ with the same state space of the MFQ process $\mathbf{V}(t)$ described in the previous subsection. A sample path of the auxiliary fluid process $\mathbf{W}(t)$ is depicted in Fig. 4. The sample path of $\mathbf{W}(t)$ is a rather straightforward combination of the sample path of $\mathbf{Z}(t)$ and the part of the sample path of $\mathbf{V}(t)$ which sets the initial fluid level to be finite PH distributed. The process $\mathbf{W}(t)$ starts at level 0 and in the first phase (dotted/yellow line) it sets the $FPH(\beta, M, B)$ distributed initial fluid level, which might happen without reaching level B or with a random number of visits to level B. Fig. 4 demonstrates only the first case. A sample path where the $FPH(\beta, M, B)$ distributed initial fluid level is set with reaching level B once is depicted in the second cycle in Fig. 3. In the next phase (solid/blue line), the process evolves according to the original MFQ up to reaching level b or the time horizon Θ. If level b is reached first (first period in Fig. 3) then an exponentially distributed time with unit mean is needed (thick/green line) to move the system to state 0 which resets the fluid level in the next phase (dashed/red line) to 0. If the $PH(\alpha, A)$ distributed time horizon concludes first (second cycle in Fig. 3) due to a transition to state 0 according to A^0, the fluid level is reset to 0 in the next (dashed/red line) phase. Fig. 4 emphasizes that the $FPH(\beta, M, B)$ distributed initial fluid level can be smaller (second cycle in Fig. 3) as well as larger (first cycle in Fig. 3) than b. If the first passage time starting from $(0, b)$ is of interest then the initial fluid level can be set to be $FPH(\beta, M, b)$-distributed using the same approach.

The characterizing matrices of $\mathbf{W}(t)$ are identical with $(Q_V(x), R_V(x), S_V(x))$ pertaining to the process $\mathbf{V}(t)$ for $0 \leq x < b$ and $b < x \leq B$. When $x = b$, the characterizing matrices are given as

$$Q_W(b) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & M & M^0(\alpha \otimes \pi) \\ e & 0 & -I_{Nn} \end{bmatrix},$$ (36)

$$R_W(b) = \mathbf{diag}\{-1, I_m, 0_{Nn}\}.$$ (37)

Based on the steady-state solution of MFQ $\mathbf{W}(t)$ and letting $c_W((k,\ell),b)$ and $c_W(0,0)$ denote the corresponding steady-state probability mass accumulation at state $(k,\ell), 1 \leq k \leq N, 1 \leq \ell \leq n$ and state 0, and at boundaries b and 0, respectively, the cdf of the first passage time to b is given by the following theorem.

Theorem 4 *With initial fluid level Ψ and initial state distribution π, the first passage time probability, $F_\tau^{\Psi,\pi,b}(\Theta)$, are obtained from the stationary behaviour of $\mathbf{W}(t)$ as*

$$F_\tau^{\Psi,\pi,b}(\Theta) = \frac{\sum_k \sum_\ell c_W((k,\ell),b)}{c_W(0,0)(1 - \beta e^{MB} e)}.$$ (38)

Proof The proof follows the pattern of the proof of Theorem 2. The only significant difference comes from the fact that $\mathbf{W}(t)$ can visit level 0 in state 0 in a period more than ones, as it is in the second cycle of Fig. 3. The mean number of visits to level 0 in state 0 in a period is $1/(1 - \beta e^{MB} e)$ and the second term of the denominator renormalizes the ratio accordingly. □

5.3 Infinite Buffer Case

The case when the fluid buffer is infinite, $B = \infty$, can be handled with the same approach due to the fact that the stationary analysis of MRMFQ with infinite buffer is available as well (see e.g. Horváth and Telek (2017)). In the infinite buffer case, the only additional requirement is that the average drift of the last regime from T^{K-1} to ∞ should be negative, i.e., if $\gamma^{(K)}$ is the stationary solution to $\gamma^{(K)} Q^{(K)} = 0, \gamma^{(K)} e = 1$, then $\gamma^{(K)} R^{(K)} e$, should be negative.

Similarly, the approach for starting the analysis from a random initial fluid level is applicable when the buffer is infinite. In this case, the initial fluid level is PH distributed with parameters (β, M), and the expressions for finite B remain valid when $B \to \infty$ as well. For example, the term $(1 - \beta e^{MB} e)$ in (38) converges to one as $B \to \infty$.

5.4 ME-distributed Time Horizon

One of the intended contributions of this paper is the replacement of the PH distribution (or its least varying member, namely the Erlang distribution) with the concentrated ME (CME) distribution in the description of the time horizon Θ. It is beneficial, because the SCV of the Erlang distribution linearly depends on the order while the SCV of the CME distribution quadratically depends on the order Horváth et al. (2016).

The findings of Bean and Nielsen (2010); Buchholz and Telek (2010, 2012) ensure the applicability of ME distributions in place of PH distributions in the introduced auxiliary MRMFQs. This way, the analysis procedure of Horváth and Telek (2017) can be also applied when α and A do not obey the sign constraints described in Section 2.1, but $-\alpha e^{Ax} A e$ is a valid (non-negative) density function for $x > 0$; and the construction of the CME distributions in Horváth et al. (2016) ensures the non-negativity of $-\alpha e^{Ax} A e$.

6 Numerical Examples

In this section, we present numerical examples for both first- and second-order MFQs. For the first-order scenario, variations of a "benchmark" problem appearing in several research studies is studied. For the case of the second-order scenario, however, there are no alternative methods in the literature for comparison, and since simulating second-order systems is rather challenging, the validation of the model is not trivial. In order to make the numerical experiments reproducible, we made our Matlab implementation of the procedures and the examples of the second-order scenario publicly available at http://www.hit.bme.hu/~ghorvath/software.

6.1 First-order MFQ Examples

In the first numerical example, we study the case of 10 statistically identical traffic sources that are multiplexed into a single buffer of size 100, a case which is studied in Akar and Sohraby (2004) and Ahn et al. (2007). Each individual source in this example is modeled by a three-state Markov fluid source with one OFF state and two ON states. Particularly, the ON time is assumed to have a hyper-exponential distribution with mean 2, and coefficient of variation of 4 with balanced means as described in Tijms (1994). The probability of the source being in the ON state is 0.4 and each ON source generates traffic at a unit rate. The initial vector of a source, π_{source}, is chosen so that the stationary distribution of the modulating Markov chain is restricted to reside in one of the positive drift states. Consequently, a single source is characterized by

$$Q_{source} = \begin{bmatrix} -0.9697 & 0 & 0.9697 \\ 0 & -0.0303 & 0.0303 \\ 0.3232 & 0.0101 & -0.3333 \end{bmatrix}, \quad R_{source} = \mathbf{diag}\{1,1,0\}. \qquad (39)$$

The drain rate of the fluid queue is then set to a value so as to meet a desired overall utilization of 0.95. This example leads to an original MFQ $\mathbf{X}_1(t)$ characterized with the matrix pair (Q_{X_1}, R_{X_1}) with $n = 66$ states (representing the distribution of the 10 sources in one of the three states). As in Ahn et al. (2007), we assume that the buffer is empty at the beginning, i.e., $a = 0$, and the initial vector π is chosen so that the modulating Markov chain is initially restricted to reside in one of the positive drift states according to the steady-state vector of Q_{X_1}.

Erlang-N and CME-N distributions are used for approximating the deterministic time horizon in all the numerical examples. We note here that the order of CME distributions is taken to be always odd; see Horváth et al. (2016). The complementary cdf (ccdf) of the buffer content is

$$G_t^{0,\pi}(x) = \Pr\{X_f(t) > x | X_f(0) = 0, X_d(0) \sim \pi\}, \qquad (40)$$

which is approximately computed for various values of x and for two values of $t = 100, 1000$, and the results are tabulated in Tables 1 and 2, respectively, along with the numerical results reported in Ahn et al. (2007) and also with simulation results with 95% confidence intervals. We observe the following:

– The convergence with CME-N is quite rapid when compared to that obtained by Erlang-N.
– The CME-21 results appear to outperform those of Erlang-100 based on the numerical results obtained by Ahn et al. (2007) which clearly shows that ME-fication is an effective alternative to Erlangization with much lesser computational complexity.
– The results obtained with the particular CME-101 are in line with the results of Ahn et al. (2007) up to four digits.

We also modify the original MFQ in such a way that each ON source generates traffic at a rate of 1.5 (1.25 respectively) when the queue occupancy is less than a threshold $B_T = 50$ and it is kept unchanged at the rate of 1 above this threshold which gives rise to an original MFQ called $\mathbf{X}_2(t)$ ($\mathbf{X}_2^*(t)$ respectively) with two regimes. The ccdf $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_2(t)$ is presented in Tables 3 and 4 with respect to varying values of x for the cases $t = 100$ and $t = 1000$, respectively, along with the simulation results we obtained with 95% confidence intervals. In this way, the proposed method for transient analysis is also validated for the two-regime example and the accuracy of the CME-N based ME-fication method is very good with rapid convergence with respect to the order parameter N. Moreover, the results obtained with the particular CME-101 approximation reside within the simulation confidence intervals for all the cases.

Table 1 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_1(t)$ when $t = 100$ with respect to varying x.

| | | | Erlang-N | | CME-N | | | | |
x	Simulation	Ahn et al. (2007)	$N = 50$	$N = 100$	$N = 21$	$N = 41$	$N = 61$	$N = 81$	$N = 101$
0	0.8005798±0.0000798	0.8005439	0.8006440	0.8005932	0.8005748	0.8005507	0.8005466	0.8005453	0.8005447
10	0.5146968±0.0000754	0.5146490	0.5144087	0.5145341	0.5145469	0.5146283	0.5146410	0.5146448	0.5146465
20	0.4087070±0.0000867	0.4086868	0.4080080	0.4083574	0.4084831	0.4086421	0.4086687	0.4086772	0.4086809
30	0.3286545±0.0000946	0.3285634	0.3274791	0.3280335	0.3282759	0.3284980	0.3285364	0.3285490	0.3285546
40	0.2634564±0.0000882	0.2634434	0.2620412	0.2627537	0.2630950	0.2633622	0.2634095	0.2634252	0.2634322
50	0.2095425±0.0000815	0.2095419	0.2079325	0.2087444	0.2091570	0.2094505	0.2095034	0.2095212	0.2095291
60	0.1648205±0.0000813	0.1647741	0.1630696	0.1639226	0.1643756	0.1646780	0.1647334	0.1647521	0.1647605
70	0.1275776±0.0000697	0.1275597	0.1258657	0.1267065	0.1271688	0.1274644	0.1275191	0.1275378	0.1275461
80	0.0963567±0.0000584	0.0963395	0.0947629	0.0955402	0.0959784	0.0962507	0.0963015	0.0963189	0.0963267
90	0.0689573±0.0000544	0.0689034	0.0675842	0.0682333	0.0686029	0.0688292	0.0688717	0.0688862	0.0688927
100-	0.0261213±0.0000324	0.0261030	0.0256553	0.0258796	0.0260011	0.0260782	0.0260925	0.0260973	0.0260995

Table 2 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_1(t)$ when $t = 1000$ with respect to varying x.

| | | | Erlang-N | | CME-N | | | | |
x	Simulation	Ahn et al. (2007)	$N = 50$	$N = 100$	$N = 21$	$N = 41$	$N = 61$	$N = 81$	$N = 101$
0	0.8291758±0.0000878	0.8292083	0.8292038	0.8292064	0.8292036	0.8292075	0.8292080	0.8292082	0.8292082
10	0.5859997±0.0000878	0.5859580	0.5859471	0.5859535	0.5859286	0.5859530	0.5859563	0.5859571	0.5859575
20	0.4930738±0.0000811	0.4930416	0.4930291	0.4930364	0.4930091	0.4930359	0.4930396	0.4930406	0.4930410
30	0.4195304±0.0000794	0.4195613	0.4195483	0.4195559	0.4195282	0.4195556	0.4195594	0.4195604	0.4195608
40	0.3562679±0.0000973	0.3562558	0.3562429	0.3562504	0.3562239	0.3562504	0.3562539	0.3562549	0.3562553
50	0.2999885±0.0000809	0.2999119	0.2998997	0.2999068	0.2998826	0.2999069	0.2999102	0.2999111	0.2999114
60	0.2489487±0.0000995	0.2488533	0.2488422	0.2488487	0.2488274	0.2488489	0.2488518	0.2488525	0.2488529
70	0.2019001±0.0000815	0.2018935	0.2018838	0.2018895	0.2018714	0.2018897	0.2018922	0.2018928	0.2018931
80	0.1577762±0.0000690	0.1578151	0.1578072	0.1578118	0.1577973	0.1578120	0.1578140	0.1578146	0.1578148
90	0.1142023±0.0000524	0.1142349	0.1142291	0.1142325	0.1142219	0.1142327	0.1142342	0.1142346	0.1142347
100-	0.0397383±0.0000261	0.0397938	0.0397920	0.0397931	0.0397895	0.0397931	0.0397936	0.0397937	0.0397938

Table 3 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_2(t)$ when $t = 100$ with respect to varying values of x.

		CME-N				
x	Simulation	$N = 21$	$N = 41$	$N = 61$	$N = 81$	$N = 101$
0	0.9989738±0.0000847	0.9989467	0.9989647	0.9989679	0.9989689	0.9989694
10	0.9915537±0.0002184	0.9913360	0.9915069	0.9915350	0.9915439	0.9915478
20	0.9808783±0.0002018	0.9804597	0.9807409	0.9807878	0.9808028	0.9808094
30	0.9607752±0.0001046	0.9602078	0.9605846	0.9606477	0.9606680	0.9606770
40	0.9187809±0.0002092	0.9181303	0.9185703	0.9186443	0.9186681	0.9186787
50	0.7086112±0.0009859	0.7081774	0.7086226	0.7086997	0.7087250	0.7087363
60	0.4354735±0.0007387	0.4350771	0.4356319	0.4357318	0.4357654	0.4357804
70	0.3219587±0.0006860	0.3213813	0.3219654	0.3220720	0.3221080	0.3221241
80	0.2373893±0.0004176	0.2368594	0.2374089	0.2375099	0.2375442	0.2375596
90	0.1661971±0.0005445	0.1658028	0.1662493	0.1663316	0.1663596	0.1663721
100-	0.0591348±0.0002722	0.0589744	0.0591116	0.0591365	0.0591450	0.0591487

We also tabulate the ccdf of the first passage time

$$G_\tau^{0,\pi,b}(t) = \Pr\{\tau^{0,\pi,b} > t\}, \tag{41}$$

where

$$\tau^{0,\pi,b} = \inf_t\{X_f(t) = b | X_f(0) = 0, X_d(0) \sim \pi\}$$

for the MFQ $\mathbf{X}_1(t)$ for $t = 100$ and $t = 1000$, respectively, in Tables 5 and 6, with respect to varying values of the parameter b. The same ccdf of the first passage time for the two-regime MFQ $\mathbf{X}_2^*(t)$ is presented in Tables 7 and 8 for $t = 100$ and $t = 1000$, respectively. In all these tables, simulation results with 95% confidence results are also reported. The results clearly demonstrate the CME-N approximations are very effective for obtaining the first passage time distributions and all the approximate analytical results lie inside the confidence intervals except for one single instance which is the last row of Table 7.

In the above examples, the time parameter t was set to either 100 or 1000. For the purpose of generality, we present the ccdf $G_t^{0,\pi}(x)$ of the buffer content for the MFQ $\mathbf{X}_1(t)$ for five different values of $t = 4, 16, 64, 256, 1024$ and four values of $x = 25, 50, 75, 100$-, in Table 9. To produce the analytical results of this example which match very well to those obtained with simulation results, ME-fication is used with the parameter N set to 61 and the initial vector π is the same as in the previous examples.

Table 4 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_2(t)$ when $t = 1000$ with respect to varying values of x.

		CME-N				
x	Simulation	$N = 21$	$N = 41$	$N = 61$	$N = 81$	$N = 101$
0	0.9995199±0.0000589	0.9995088	0.9995101	0.9995102	0.9995103	0.9995103
10	0.9956387±0.0001337	0.9956265	0.9956464	0.9956489	0.9956496	0.9956498
20	0.9885377±0.0004564	0.9885163	0.9885412	0.9885444	0.9885453	0.9885456
30	0.9727059±0.0007178	0.9726228	0.9726508	0.9726547	0.9726558	0.9726562
40	0.9356245±0.0008165	0.9356211	0.9356514	0.9356558	0.9356570	0.9356574
50	0.7431783±0.0012780	0.7431248	0.7431526	0.7431565	0.7431576	0.7431579
60	0.4956653±0.0006315	0.4957093	0.4957349	0.4957383	0.4957392	0.4957396
70	0.3845687±0.0005931	0.3844426	0.3844653	0.3844683	0.3844691	0.3844694
80	0.2936199±0.0010083	0.2935768	0.2935953	0.2935978	0.2935984	0.2935988
90	0.2091269±0.0013484	0.2092693	0.2092830	0.2092848	0.2092853	0.2092855
100-	0.0711825±0.0006067	0.0713854	0.0713899	0.0713905	0.0713907	0.0713908

Table 5 Ccdf of the first passage time $G_\tau^{0,\pi,b}(100)$ for varying values of b for the MFQ $\mathbf{X}_1(t)$.

		CME-N				
b	Simulation	$N=21$	$N=41$	$N=61$	$N=81$	$N=101$
10	0.8042713±0.0002090	0.8034712	0.8039866	0.8040790	0.8041099	0.8041238
20	0.6011305±0.0002617	0.6006640	0.6010442	0.6011120	0.6011346	0.6011447
30	0.4589881±0.0002535	0.4585449	0.4588418	0.4588944	0.4589120	0.4589198
40	0.3531216±0.0002484	0.3528003	0.3530634	0.3531105	0.3531262	0.3531332
50	0.2718268±0.0002431	0.2714250	0.2716599	0.2717025	0.2717168	0.2717232
60	0.2082608±0.0002387	0.2078471	0.2080455	0.2080819	0.2080943	0.2080998
70	0.1582910±0.0002030	0.1580051	0.1581603	0.1581892	0.1581990	0.1582034
80	0.1192100±0.0001869	0.1190517	0.1191610	0.1191817	0.1191888	0.1191920
90	0.0888781±0.0001724	0.0888135	0.0888782	0.0888909	0.0888953	0.0888972
100-	0.0655609±0.0001445	0.0655512	0.0655759	0.0655811	0.0655830	0.0655839

Table 6 Ccdf of the first passage time $G_\tau^{0,\pi,b}(1000)$ for varying values of b for the MFQ $\mathbf{X}_1(t)$.

		CME-N				
b	Simulation	$N=21$	$N=41$	$N=61$	$N=81$	$N=101$
10	0.9999943±0.0000014	0.9999295	0.9999834	0.9999905	0.9999925	0.9999933
20	0.9991528±0.0000185	0.9989481	0.9991059	0.9991301	0.9991375	0.9991407
30	0.9930792±0.0000525	0.9925446	0.9929321	0.9929976	0.9930188	0.9930282
40	0.9763902±0.0000979	0.9754984	0.9761530	0.9762680	0.9763061	0.9763230
50	0.9467890±0.0001524	0.9456124	0.9464762	0.9466308	0.9466825	0.9467055
60	0.9050731±0.0001834	0.9038069	0.9047857	0.9049624	0.9050219	0.9050484
70	0.8540751±0.0002311	0.8527577	0.8537636	0.8539463	0.8540079	0.8540354
80	0.7970408±0.0002466	0.7956567	0.7966256	0.7968022	0.7968619	0.7968886
90	0.7368456±0.0002838	0.7354763	0.7363688	0.7365318	0.7365870	0.7366116
100-	0.6756845±0.0002971	0.6746477	0.6754444	0.6755902	0.6756396	0.6756616

Table 7 Ccdf of the first passage time $G_\tau^{0,\pi,b}(100)$ for varying values of b for the MFQ $\mathbf{X}_2^*(t)$.

	Simulation	CME-N				
b		$N=21$	$N=41$	$N=61$	$N=81$	$N=101$
10	0.9883925±0.0000747	0.9879995	0.9883297	0.9883862	0.9884046	0.9884128
20	0.9441918±0.0001949	0.9435728	0.9441895	0.9442987	0.9443350	0.9443512
30	0.8896289±0.0002176	0.8888313	0.8895886	0.8897237	0.8897688	0.8897889
40	0.8312572±0.0002919	0.8303829	0.8312334	0.8313858	0.8314368	0.8314595
50	0.7710200±0.0003876	0.7700925	0.7710183	0.7711851	0.7712411	0.7712662
60	0.5241633±0.0004029	0.5236870	0.5243078	0.5244211	0.5244594	0.5244766
70	0.3704714±0.0003568	0.3701940	0.3705790	0.3706496	0.3706736	0.3706843
80	0.2688547±0.0003248	0.2685341	0.2687950	0.2688434	0.2688599	0.2688673
90	0.1961632±0.0002870	0.1959625	0.1961300	0.1961617	0.1961726	0.1961775
100-	0.1421728±0.0002650	0.1426794	0.1427660	0.1427830	0.1427890	0.1427916

6.2 Second-order MFQ Examples

We consider the same MFQ $\mathbf{X}_1(t)$ as in Section 6.1, but now we assume that traffic sources generate Markov modulated Brownian motion. As before, the traffic rate in the ON state (the drift of the Brownian motion) is 1, and the effect of the variance parameter in the ON state, denoted by σ^2, is investigated in the following case study. The drift and the variance in the OFF state are both zero.

Tables 10, 11, and 12 present the ccdf of the fluid level for $t = 100$, for three different values of σ^2. Since the solution of second-order fluid models involves larger matrices

Table 8 Ccdf of the first passage time $G_\tau^{0,\pi,b}(1000)$ for varying values of b for the MFQ $\mathbf{X}_2^*(t)$.

	Simulation		CME-N			
b		$N=21$	$N=41$	$N=61$	$N=81$	$N=101$
10	1.0000000 ± 0.0000000	0.9999702	0.9999952	0.9999984	0.9999992	0.9999996
20	1.0000000 ± 0.0000000	0.9999576	0.9999927	0.9999975	0.9999988	0.9999993
30	1.0000000 ± 0.0000000	0.9999461	0.9999907	0.9999968	0.9999985	0.9999992
40	1.0000000 ± 0.0000000	0.9999348	0.9999888	0.9999961	0.9999981	0.9999989
50	0.9999995 ± 0.0000006	0.9999233	0.9999865	0.9999950	0.9999972	0.9999982
60	0.9997140 ± 0.0000125	0.9995451	0.9996794	0.9996987	0.9997042	0.9997065
70	0.9957856 ± 0.0000612	0.9953724	0.9957204	0.9957775	0.9957957	0.9958037
80	0.9823249 ± 0.0001130	0.9815373	0.9821804	0.9822922	0.9823290	0.9823453
90	0.9559581 ± 0.0001735	0.9546932	0.9555904	0.9557501	0.9558034	0.9558272
100-	0.9165630 ± 0.0002225	0.9150828	0.9161328	0.9163221	0.9163856	0.9164140

Table 9 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_1(t)$ when $t=4,16,64,256,1024$ for four different values of $x=25,50,75,100$-. (A) represents the analytical results obtained by ME-fication with $N=61$ and (S) represents simulation results with 95% confidence intervals.

	t	$x=25$	$x=50$	$x=75$	$x=100$-
	4	0.00000	0.00000	0.00000	0.00000
	16	0.08758	0.00166	0.00000	0.00000
(A)	64	0.32642	0.15603	0.06453	0.01488
	256	0.43685	0.28298	0.16744	0.03723
	1024	0.45470	0.29991	0.17959	0.03979
	4	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
	16	0.08732 ± 0.00020	0.00166 ± 0.00003	0.00000 ± 0.00000	0.00000 ± 0.00000
(S)	64	0.32635 ± 0.00054	0.15576 ± 0.00032	0.06431 ± 0.00022	0.01491 ± 0.00010
	256	0.43653 ± 0.00046	0.28301 ± 0.00035	0.16751 ± 0.00027	0.03718 ± 0.00013
	1024	0.45511 ± 0.00048	0.30003 ± 0.00041	0.17969 ± 0.00025	0.03967 ± 0.00017

than first-order ones, we did not go beyond 61 with the order of the ME approximating the finite time horizon. The tables contain simulation results, too. However, compared to the easy to simulate first-order MFQs, simulating second-order fluid queues is not trivial at all. We have followed a discretization-based approach. In an elementary step of the simulation, when moving forward the time by a small amount, we generate the normally distributed fluid increment and add it to the fluid level. To increase the accuracy of the simulation, we made the time increment inversely proportional with the variance parameter of the current state. For this numerical example, the time increment in state i was computed by $\Delta = 10^{-4}/\max\{1,s_i\}$. The results in the tables are obtained by averaging 10^5 simulation results. We have omitted the confidence intervals, since these results have two sources of uncertainty: the discretization and the finite number of repeated executions both decrease the confidence, and while the latter one can be quantified, we can not quantify the uncertainty introduced by discretization. By such small time increments, the computational effort of the simulation is huge, underlining the importance of the analytical solution presented in the paper.

As seen from the results, the variance of the traffic has a significant impact on the transient distribution. For easier comparison, Fig. 5 compares the transient ccdf by four different settings of σ^2. The higher the variance is, the closer the transient distribution is to the uniform distribution. Another interesting property is visible in the figure, namely that there is no probability mass at the boundaries in the second order cases ($\sigma^2 > 0$) due to the use of reflecting boundaries.

Table 10 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 100$ with respect to varying values of x when each ON source sends traffic at rate 1 and variance $\sigma^2 = 1$.

		CME-N		
x	Simulation	$N = 21$	$N = 41$	$N = 61$
0	0.99705	0.99681	0.99681	0.99681
10	0.58586	0.58675	0.58684	0.58686
20	0.46004	0.46148	0.46165	0.46168
30	0.37120	0.37211	0.37235	0.37240
40	0.29912	0.30052	0.30082	0.30087
50	0.23930	0.24116	0.24149	0.24155
60	0.19017	0.19125	0.19160	0.19166
70	0.14817	0.14877	0.14911	0.14917
80	0.11119	0.11158	0.11188	0.11193
90	0.07641	0.07558	0.07579	0.07584
100-	0.00000	0.00000	0.00000	0.00000

Table 11 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 100$ with respect to varying values of x when each ON source sends traffic at rate 1 and variance $\sigma^2 = 4$.

		CME-N		
x	Simulation	$N = 21$	$N = 41$	$N = 61$
0	0.99837	0.99839	0.99839	0.99839
10	0.70582	0.70469	0.70483	0.70486
20	0.56478	0.56391	0.56414	0.56419
30	0.46245	0.46198	0.46229	0.46234
40	0.37973	0.37958	0.37995	0.38001
50	0.31006	0.30955	0.30994	0.31001
60	0.25001	0.24812	0.24851	0.24858
70	0.19370	0.19242	0.19276	0.19282
80	0.13839	0.13903	0.13929	0.13934
90	0.08147	0.08141	0.08155	0.08158
100-	0.00000	0.00000	0.00000	0.00000

Table 12 Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 100$ with respect to varying values of x when each ON source sends traffic at rate 1 and variance $\sigma^2 = 10$.

|		CME-N		
x	Simulation	$N = 21$	$N = 41$	$N = 61$
0	0.99927	0.99921	0.99921	0.99921
10	0.80096	0.79764	0.79778	0.79780
20	0.66831	0.66437	0.66460	0.66464
30	0.55937	0.55728	0.55756	0.55761
40	0.46660	0.46514	0.46545	0.46550
50	0.38430	0.38246	0.38276	0.38281
60	0.30710	0.30585	0.30612	0.30617
70	0.23423	0.23262	0.23284	0.23288
80	0.16071	0.16000	0.16016	0.16019
90	0.08500	0.08432	0.08440	0.08441
100-	0.00000	0.00000	0.00000	0.00000
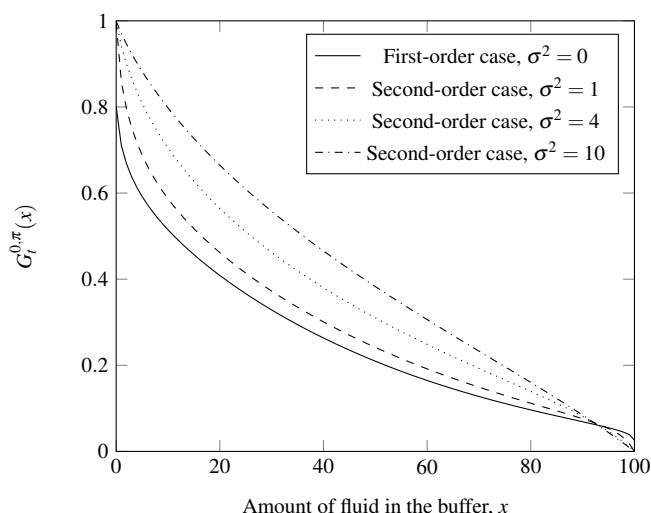
Fig. 5 Ccdf of the buffer content, $G_t^{0,\pi}(x)$, when $t = 100$ for various values of the traffic variance parameter σ^2.

7 Conclusions

We have proposed an approximate numerical solution for obtaining the transient and the first passage time distributions for both first- and second-order multi-regime MFQs which is a much more general class of problems than the ones studied in the literature. This method computes the transient measures from the stationary analysis of an auxiliary and larger-cardinality MRMFQ. Erlangization and ME-fication methods are comparatively studied for obtaining this auxiliary MRMFQ. We have shown that ME-fication is a viable alternative to Erlangization with lesser computational complexity since CME-*N* captures better the SCV of the deterministic time horizon than Erlang-*N* for a fixed order parameter *N*.

References

Ahn S, Ramaswami V (2005) Efficient algorithms for transient analysis of stochastic fluid flow models. Journal of Applied Probability 42(2):531–549

Ahn S, Badescu AL, Ramaswami V (2007) Time dependent analysis of finite buffer fluid flows and risk models with a dividend barrier. Queueing Syst Theory Appl 55(4):207–222

Akar N, Sohraby K (2004) Infinite- and finite-buffer Markov fluid queues: A unified analysis. Journal of Applied Probability 41(2):557–569

Aldous D, Shepp L (1987) The least variable phase type distribution is Erlang. Stochastic Models 3:467–473

Anick D, Mitra D, Sondhi MM (1982) Stochastic theory of a data-handling system with multiple sources. The Bell System Technical Journal 61(8):1871–1894

Asmussen S (1995) Stationary distributions for fluid flow models with or without Brownian noise. Communications in statistics Stochastic models 11(1):21–49

Asmussen S, Bladt M (1996) Renewal theory and queueing algorithms for matrix-exponential distributions. In: Alfa A, Chakravarthy S (eds) Matrix-analytic methods in stochastic models, Marcel Dekker, pp 313–341

Asmussen S, Avram F, Usabel M (2002) Erlangian approximations for finite-horizon ruin probabilities. Astin Bulletin 32(2):267–282

Bean N, Nielsen BF (2010) Quasi-birth-and-death processes with rational arrival process components. Stochastic Models 26:309–334

Bladt M, Neuts MF (2003) Matrix-exponential distributions: Calculus and interpretations via flows. Stochastic Models 19(1):113–124

Buchholz P, Telek M (2010) Stochastic Petri nets with matrix exponentially distributed firing times. Performance Evaluation 67:1373–1385

Buchholz P, Telek M (2012) Rational processes related to communicating Markov processes. Journal of Applied Probability 49:40–59

Fackrell MW (2003) Characterization of matrix-exponential distributions. PhD thesis, The University of Adelaide

Gerber H, Shiu E (1998) On the time value of ruin. North American Actuarial Journal 2(1):48–72

He Q, Horváth G, Horváth I, Telek M (2019) Moment bounds of PH distributions with infinite or finite support based on the steepest increase property. Advances in Applied Probability (AAP) 51(1)

He QM, Zhang H (2007) On matrix exponential distributions. Advances in Applied Probability 39(1):pp. 271–292

Horváth G, Telek M (2017) Matrix-analytic solution of infinite, finite and level-dependent second-order fluid models. Queueing Systems 87(3-4):325–343

Horvath G, Van Houdt B (2012) A multi-layer fluid queue with boundary phase transitions and its application to the analysis of multi-type queues with general customer impatience. In: Proceedings - 2012 9th International Conference on Quantitative Evaluation of Systems, QEST 2012, pp 23–32, DOI 10.1109/QEST.2012.12

Horváth I, Sáfár O, Telek M, Zámbó B (2016) Concentrated matrix exponential distributions. In: Fiems D, Paolieri M, Platis AN (eds) Computer Performance Engineering, Springer International Publishing, Cham, pp 18–31

Houdt BV, Blondia C (2005) Approximated transient queue length and waiting time distributions via steady state analysis. Stochastic Models 21(2-3):725–744

Kankaya HE, Akar N (2008) Solving multi-regime feedback fluid queues. Stochastic Models 24:425–450

Karandikar RL, Kulkarni VG (1995) Second-order fluid flow models: reflected Brownian motion in a random environment. Operations Research 43(1):77–88

Kulkarni VG (1997) Fluid models for single buffer systems. In: Dshalalow JH (ed) Frontiers in Queuing: Models and Applications in Science and Engineering, CRC Press, pp 321–338

Le L, Aikat J, Jeffay K, Smith FD (2007) The effects of active queue management and explicit congestion notification on web performance. IEEE/ACM Transactions On Networking 15(6):1217–1230

Mandjes M, Mitra D, Scheinhardt W (2003) Models of network access using feedback fluid queues. Queueing Syst Theory Appl 44(4):2989–3002

Neuts MF (1981) Matrix-Geometric Solutions in Stochastic Models. Johns Hopkins University Press, Baltimore, MD

Ramaswami V, Viswanath NC (2014) Phase type distributions with finite support. Stochastic Models 30(4):576–597

Ramaswami V, Woolford DG, Stanford DA (2008a) The Erlangization method for Markovian fluid flows. Annals of Operations Research 160(1):215–225

Ramaswami V, Woolford DG, Stanford DA (2008b) The Erlangization method for Markovian fluid flows. Annals of Operations Research 160(1):215–225

Scheinhardt W, van Foreest N, Mandjes M (2005) Continuous feedback fluid queues. Operations Research Letters 33(6):551 – 559

Sericola B (1998) Transient analysis of stochastic fluid models. Performance Evaluation 32(4):245 – 263

da Silva Soares A, Latouche G (2009) Fluid queues with level dependent evolution. European Journal of Operational Research 196(3):1041 – 1048

Tijms H (1994) Stochastic models: an algorithmic approach. Wiley series in probability and mathematical statistics, John Wiley & Sons

Velthoven JV, Houdt BV, Blondia C (2007) Simultaneous transient analysis of qbd markov chains for all initial configurations using a level based recursion. In: Fourth International Conference on the Quantitative Evaluation of Systems (QEST 2007), pp 79–90

Yazici MA, Akar N (2013) The workload-dependent MAP/PH/1 queue with infinite/finite workload capacity. Performance Evaluation 70(12):1047 – 1058

Yazici MA, Akar N (2017) The finite/infinite horizon ruin problem with multi-threshold premiums: a Markov fluid queue approach. Annals of Operations Research 252(1):85–99