NUMERICAL METHODS FOR THE TRANSIENT ANALYSIS OF MULTI-REGIME MARKOV FLUID QUEUES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Ömer Gürsoy

January 2019

Numerical Methods for the Transient Analysis of Multi-Regime Markov Fluid Queues
By Ömer Gürsoy
January 2019

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Nail Akar(Advisor)

Tuğrul Dayar

Mehmet Akif Yazıcı

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

NUMERICAL METHODS FOR THE TRANSIENT ANALYSIS OF MULTI-REGIME MARKOV FLUID QUEUES

Ömer Gürsoy
M.S. in Electrical and Electronics Engineering
Advisor: Nail Akar
January 2019

Markov fluid queue models have served as one of the main tools for the performance analysis of computer and communication systems and networks. These models have also been used in other disciplines such as insurance risk, finance, inventory control, etc. This thesis focuses on the time dependent (transient) analysis of Markov fluid queue models. In particular, a numerical method is proposed to obtain both the transient and first passage time distributions of a Multi-Regime Markov Fluid Queue (MRMFQ). The proposed method is based on obtaining the steady-state solution of an auxiliary MRMFQ that is to be constructed from the original MRMFQ which then leads to the related transient measures of interest. First, in order to model the deterministic time horizon, the Erlangization method is used. Then, as an alternative to Erlangization, ME-fication technique which efficiently replaces the Erlang distribution with a Concentrated Matrix Exponential (CME), is used. As an application of the proposed method, an M/M/S+G queue with generally distributed impatience times is modeled by using MRMFQs and our transient analysis method is subsequently applied to obtain the time dependent distributions. Numerical examples are given to show the effectiveness of the proposed transient analysis method while employing ME-fication.

Keywords: queuing analysis, Markov fluid queues, transient analysis, first passage time distribution, Matrix exponential distribution, M/M/S+G queue.

ÖZET

ÇOK REJİMLİ MARKOV AKIŞKAN KUYRUKLARININ ZAMANA BAĞLI ANALİZİ İÇİN NUMERİK YÖNTEMLER

Ömer Gürsoy
Elektrik ve Elektronik Mühendisliği, Yüksek Lisans
Tez Danışmanı: Nail Akar
January 2019

Markov akışkan kuyrukları iletişim ağlarının performans analizinde kullanılan kayda değer modelleme tekniklerinden biridir. Bu modelleme tekniği sigorta risk dağılımı, finans modellemeleri ve envanter kontrolü gibi bir çok alanda da kullanılabilir. Bu tezde Markov akışkan kuyruklarının zamana bağlı analizleri ele alınmıştır. Çok rejimli Markov akışkan kuyruklarının zamana bağlı ve ilk geçiş dağılımlarını bulmak için sayısal bir yöntem teklif edilmiştir. Bu yöntem orjinal Markov akışkan kuyruğundan zaman kuşağı eklenerek elde edilen yardımcı bir Markov akışkan kuyruğunun kararlı hal çözümünün elde edilmesine dayanır. Bu çözüm gerekli zamana bağlı dağılımların elde edilmesine öncülük etmektedir. Probleme zaman kuşağının eklenmesinde ilk önce Erlangizasyon tekniği kullanılmıştır. Daha sonra, Erlang dağılımını Konsantre Matris üstel dağılımı ile değiştiren matris üstel tekniği kullanılmıştır. Buna ek olarak Markov akışkan kuyrukları kullanılarak genel dağılımlı sabırsız müşterileri olan M/M/S+G kuyrukları modellenmiş ve yukarıda bahsedilen teknik bu modelin zamana bağlı dağılımlarını bulmak için uygulanmıştır. Matris üstel tekniği ve teklif edilen zamana bağlı sayısal yöntemin geçerliliği sayısal örneklerle gösterilmiştir.

Anahtar sözcükler: kuyruk analizi, Markov akışkan kuyrukları, zamana bağlı analiz, ilk geçiş dağılımı, matris üstel dağılımı, M/M/S+G kuyruğu.

Acknowledgement

I would like to express my sincere and special gratefulness to my advisor Prof. Nail Akar for his extensive guidance and insight throughout my study. His door was always open whenever I ran into trouble in my research. Without his assistance, this work would never have come to existence.

I would like to thank Prof. Tuğrul Dayar and Prof. Mehmet Akif Yazıcı for agreeing to be on my thesis committee.

Finally, I would like to thank my family members, my father Cüneyt, my mother Bülvin and my sister Gökçe for their continuous support and encouragement throughout my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Queueing systems have successfully been used for performance modeling in communication networks such as satellite communications, optical networks, Internet. These systems also have numerous applications in other disciplines such as finance, insurance, storage, inventory control. The focus of this thesis is a specific class of queueing systems called Markov Fluid Queues (MFQ) in which the rate of change (called drift), of the fluid at any moment depends on the state of a finite state Continuous Time Markov Chain (CTMC). MFQs are joint Markovian processes that consist of the following two main components [1]:

- A finite state CTMC process that is considered as the background process. Each state of this process dictates the rate of change (drift) of the fluid.
- Continuous-valued buffer level that changes with respect to a drift which is dictated by the state of the background process and drift rates. The buffer capacity can either be finite or infinite.

This thesis covers the following two categories of MFQs:

- Single-Regime Markov Fluid Queues (SRMFQ): The infinitesimal generator of the background process and drift rate matrix stay the same through all buffer levels.
- Multi-Regime Markov Fluid Queues (MRMFQ): The buffer is divided into multiple regimes each of which has a different infinitesimal generator and drift rate matrix. the buffer level at any moment determines which regime the system is in.

The current thesis proposes a numerical method that can be used to obtain the transient (time dependent) and first passage time distributions of not only SRMFQs but also MRMFQs. There exist some previous research to obtain the transient solution of MFQs. Obtaining transient distributions by numerical integration of the related partial differential equations (pde) for the transient probabilities is both difficult and error-prone due to error that may raise during build up. Spectral methods have turned out to be not very successful either [2]. The most important drawback of these methods is that most of them only cover the SRMFQ case and not the MRMFQ case [3]. The proposed method is expected to be a powerful alternative to the method proposed in [3] for SRMFQs but it also provides a solution for a more general class of problems. The approach we pursue in this paper is based on constructing a new auxiliary larger MRMFQ that is to be constructed from the original MRMFQ using sample path arguments. Subsequently, obtaining the steady-state solution of this newly constructed MRMFQ will lead us to find the transient measures of interest. To achieve this, we use the PH-type distribution which is a CTMC with one absorbing state, to approximate the deterministic time horizon t. Basically, the new auxiliary MRMFQ that we construct starts like the original MRMFQ from given initial conditions and after a PH-distributed random time T with mean t, the timer expires and the MRMFQ is forced back to start from the same initial conditions and this process repeats forever. When we have the steady-state solution of the auxiliary MRMFQ, it also has an approximate transient solution at time t of the original MRMFQ. When the PH-type distribution approximates the deterministic horizon t better, then the proposed method should be expected to present a better approximation for the transient distribution at time t. In order to implement this model, we need to have the infinitesimal generator and drift matrices which characterize the original MFQ, and the PH-type distribution which is characterized by the pair $T \sim PH(\alpha, S)$. Actually, a PH-type distribution of order k is a CTMC which is defined on the state space $\{1, \ldots, k, k+1\}$ for which the absorbing state is represented by state $k+1$ and the following infinitesimal generator:

$$\begin{bmatrix} S & S^0 \\ 0 & 0 \end{bmatrix},$$

where $S^0 = -Se$, α is the initial row ptobability vector, and e is a column vector of ones of appropriate size. When the CTMC starts from a state j where $j \in \{1, 2, \ldots, k-1, k\}$ it eventually ends up at state $k+1$. We indicate this epoch as timer expired or the random time horizon T is reached. We combine the state space of the PH-type distribution with the original MFQ to construct the new auxiliary MFQ to introduce time horizon.

As the order of PH-type distribution k increases, the approximation gets better but since the size of the infinitesimal generator matrix of the PH-type distribution increases, the computation time spent for the proposed solution also increases. Erlangization is an alternative for this purpose where T is approximated by Erlang-k distribution [4]. As the order k increases, the Coefficient of Variation (CoV) decreases with $1/k$. We propose to use another alternative called Concentrated Matrix Exponential (CME) distribution introduced in [5] with parameter l denoted by CME-l with order $k = 2l + 1, l \geq 1$ whose CoV behaves approximately as $2/k^2$ and decays to zero much faster than the Erlang-k distribution with increased order of k. Our proposed model can be implemented with both Erlang and CME distributions, called as Erlangization and ME-fication methods, respectively, and thus we can study the accuracy and numerical stability of the results in both cases. There are some other applications of PH-type distributions especially in ruin and transient problems [6], [7]. Erlangization is also applied in the solution of MFQs [8].

Constructing the new auxiliary MFQ enables the option of using steady-state solution methods of MFQs since the steady-state solution of the auxiliary MFQ involves the results for transient solution of the original one. For this purpose, we use the additive decomposition method which was originally introduced in [1]. This method divides the system to three parts: constant, forward (stable) and backward (anti-stable) parts, and reduces the time complexity dramatically. Moreover, this method has a treatment for states which has zero drifts in a regime and it is able to solve problems with non-zero boundary probability masses and absorbing states as well. The computational complexity of this method can be brought down to $\mathcal{O}(N^3K)$ for an MRMFQ with N states and K regimes stemming from the block-tridiagonal structure of the arising linear equations [9]. The details of the entire proposed model will be introduced in Chapters 3 and 4.

1.2 Thesis Contribution

There are several approaches for approximating transient and first passage time distributions for MFQs and Erlangization methods are previously used in the literature. Some of the time-dependent MFQ solutions use the spectral solution method which is not very stable, some of them only found solutions for SRMFQs, and the CME distribution has not been introduced in any of these approaches available in the literature.

- The main contribution of this thesis is the introduction of a new method for approximating transient and first passage time distributions in a MRMFQ using an arbitrary PH-type distributed time horizon. The proposed method for approximation is applicable not only for SRMFQs but also the MRMFQ cases and since it uses the additive decomposition method, it can also solve cases with non-zero boundary probability masses, absorbing states and states with zero drifts. Therefore, it expands the area of application dramatically compared to most other methods.

- Concentrated Matrix Exponential (CME) distributions are proposed to approximate the deterministic time horizon. CoV of this particular distribution decreases with approximately $2/k^2$ which decreases much faster compared to the Erlang-k distribution with increased order k. Therefore, ME-fication will provide a better alternative than Erlangization when the same order is to be used.

- Another important contribution is that we propose a numerical method for the steady-state and transient solution of the M/M/S queue with generally distributed impatience times, namely the M/M/S+G queue. This model is motivated by increasing application of conventional call center models in customer contact centers, health-care systems, and telecommunication networks. We also provide a 66-state MRMFQ used in previous studies for comparative performance evaluation.

1.3 Thesis Outline

The rest of this thesis is organized as follows. Related work is given in Section 2. A preliminary study for Markov fluid queues including SRMFQs, MRMFQs and solution of them is provided in Section 3. The method used for transient and first passage time solutions of multi-regime Markov fluid queues is presented in Section 4. A model for M/M/S+G queues and a model for its transient solution is introduced Section 5. In order to analyze the performance of the proposed models numerical examples are given in Section 6. Finally, conclusions and future work are provided in Section 7.

Chapter 2

Related Work

2.1 Transient Analysis of Markov Fluid Queues

There are a number of studies on how to acquire the transient (time dependent) solution, see [10]. Their approach is based on an approximation of the fluid model by the amounts of work in a sequence of Markov modulated queues of the quasi birth and death (QBD) type that enables the use of matrix-geometric methods for queues. They obtain the Laplace transforms (with respect to time) of the time dependent joint distribution of the fluid level and the phase in a form that lends itself to relatively stable computations. The characterization uses certain kernels that appear in the matrix-geometric method and enable one to “jump” directly to a “last epoch before time t” and thereby avoids building the story over many small intervals up to time t as in a numerical solution of the pdes. They also apply these to a Markov modulated insurance risk model with phase type claims in which a bound b is imposed. They used a numerical SRMFQ example which was previously constructed in [1] for transient analysis results which is also used by us to measure the performance of our transient method and also extended to the MRMFQ case.

2.2 Application of Multi-Regime Markov Fluid Queues in Energy Management

The energy management issue in energy harvesting wireles sensor nodes is approached in [11] by using an MRMFQ model. The work in [11] proposes a risk-theoretic Markov fluid queue model to compute the battery outage probability of a wireless sensor node for a given finite life-horizon. The proposed method enables the performance evaluation of a wide spectrum of energy management policies including those with adaptive sensing rate (or duty cycling). In this model, the node gathers data from the environment according to a Poisson process whose rate is to depend on the instantaneous battery level and/or the state of the energy harvesting process (EHP) which is characterized by a Continuous time Markov Chain (CTMC). They used the solution steps of MRMFQ also used by this thesis such as additive decomposition method which is a useful tool for avoiding individual eigenvector computations and takes care of zero drifts and absorbing states. Also they implemented the block tri-diagonal LU factorization which is a method that reduces the time required to solve boundary problems.

2.3 Multi-Server Queueing System with Generally Distributed Impatient Customers

Generalization of Markovian model is already considered in the literature. In [12] the performance of M/M/S/k+G system is analyzed, where the $+G$ notation indicates that patience is generally distributed, i.i.d. over customers and independently of everything else. In [13] a generalized solution for offered waiting time, probability of missing deadline, and the probability of blocking for M(n)/M/S/k+G queue has been developed. [14] [15] present the derivations for the occupancy distribution for the state dependent M(n)/M(n)/S/k+G queue. Moreover, some empirical results are derived for M/M/S/k+G system in [16]. In [17] an engineering approach to model the M/G/S/k+G queue and derivation of several performance metrics for the customers in the system as well as delay are developed. Many server asymptotics for M/M/S+G model for large values of arrival rate and number of servers are derived in [18].

2.4 The Erlangization Method for Markovian Fluid Flows

The method of Erlangization is used in fluid flow models in [8] as a relatively fast method to approximate t for time dependent observations. This procedure involves approximating the joint distribution of the fluid level and the environment (otherwise referred to as "phase") at time t by the corresponding distribution at an independent, Erlang-distributed horizon with mean t. In addition to the computation of the distribution at time t, it also applies to the evaluation of a variety of first passage time distributions. The method appears to have been first used in [19] to approximate finite-time ruin probabilities in certain insurance risk models and also in [20] it is proved Erlang is the least variable phase type distribution.

Chapter 3

Markov Fluid Queues

Markov fluid queues are stochastic fluid systems in which the net rate of change (called drift) of the fluid at any moment depends on the state of a finite state, continuous time Markov chain (CTMC) which will be called as the background process. These systems are joint processes that consist of two processes: $X(t)$ and $Z(t)$ where $X(t)$ represents the fluid level and $Z(t)$ is the background CTMC that determines the rate of change, i.e., the drift at time t. For every state of CTMC $Z(t) \in \{1, 2, \ldots, N\}$ there exist a drift value r_i and the process $X(t)$ changes by this drift value r_i, if $Z(t) = i$. The buffer capacity B can either be finite or infinite and buffer level $X(t)$ cannot exceed this capacity i.e., $X(t) \in [0, B]$.

MFQs can be categorized by dependence of their parameters to the buffer level. Based on this classification, two categories of MFQs will be covered:

- Single-Regime Markov Fluid Queues (SRMFQ): The infinitesimal generator of background process and drift rate matrix stay the same through all buffer levels i.e., the parameters of the MFQ are independent of the fluid level.
- Multi-Regime Markov Fluid Queues (MRMFQ): In such queues, the parameter of MFQ is piece-wise dependent with respect to the buffer level. The buffer is partitioned into multiple regimes. Each regime has possibly

different infinitesimal generators and drift rate matrices. The buffer level at any moment determines which regime the system is in.

For better understanding, numerical examples for both categories will be given below with their numerical parameters.

Single-Regime Markov Fluid Queue (SRMFQ) Example

Consider a SRMFQ that has a finite buffer capacity $B = 3$ meaning that $X(t) \in [0,3]$. Background process $Z(t)$ is a three-state Markov chain that has the following infinitesimal generator matrix:

$$Q = \begin{bmatrix} -1 & 0.25 & 0.75 \\ 0.3 & -0.65 & 0.35 \\ 0.2 & 0.6 & -0.8 \end{bmatrix},$$

for which; the drift rates for states 1, 2 and 3 are -0.5, -1 and 1 respectively. These rates of states are represented with a drift matrix in the following manner:

$$R = \begin{bmatrix} -0.5 & & \\ & -1 & \\ & & 1 \end{bmatrix}$$

In Fig. 3.1, a sample time evolution of the above SRMFQ example is given. Graphs of the buffer level $X(t)$ and state transition $Z(t)$ is provided and it can be seen how $X(t)$ is correlated with $Z(t)$. When $Z(t)$ is in states 1, 2 or 3, it can be observed that $X(t)$'s rate of change is -0.5, -1, 1 respectively. When $X(t) = 0$ it can be seen that it can not decrease further even though $Z(t) = 2$ which means the drift is equal to -1. Same applies to the case when $X(t) = B = 3$ and the drift is positive. Another observation is that the drift rates and the state transition rates stay the same through every buffer level which shows that MFQ parameters are independent of the buffer level.

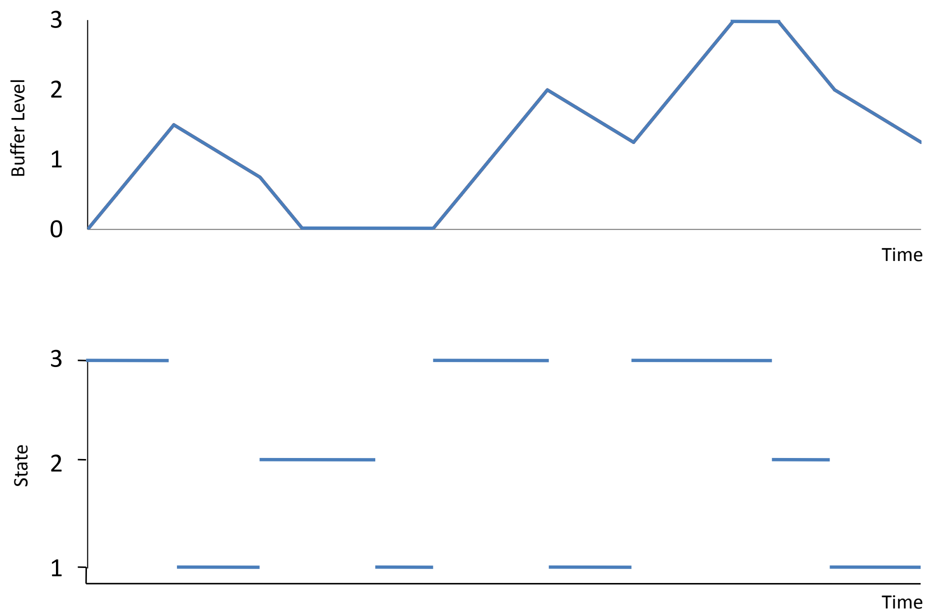

Figure 3.1: A sample path for the buffer level $X(t)$ and state transition $Z(t)$ of the SRMFQ example.

Multi-Regime Markov Fluid Queue (MRMFQ) Example

Let the buffer capacity $B = 3$ again. The capacity is partitioned into two regimes: (0,2) and (2,3). For regimes 1 and 2 the infinitesimal generators of background processes are

$$Q^{(1)} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad Q^{(2)} = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix},$$

respectively. Also, for regimes 1 and 2 the drift matrices are:

$$R^{(1)} = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \quad \text{and} \quad R^{(2)} = \begin{bmatrix} 0.5 & \\ & -0.5 \end{bmatrix}$$

Furthermore, we also need to define the boundary behavior of this fluid at levels 0, 2 and 3. For this example, let us assume that boundary levels 0 and 2 have the same drift rates and background processes with regime 1 and boundary at level 3 has the same parameters with regime 2.

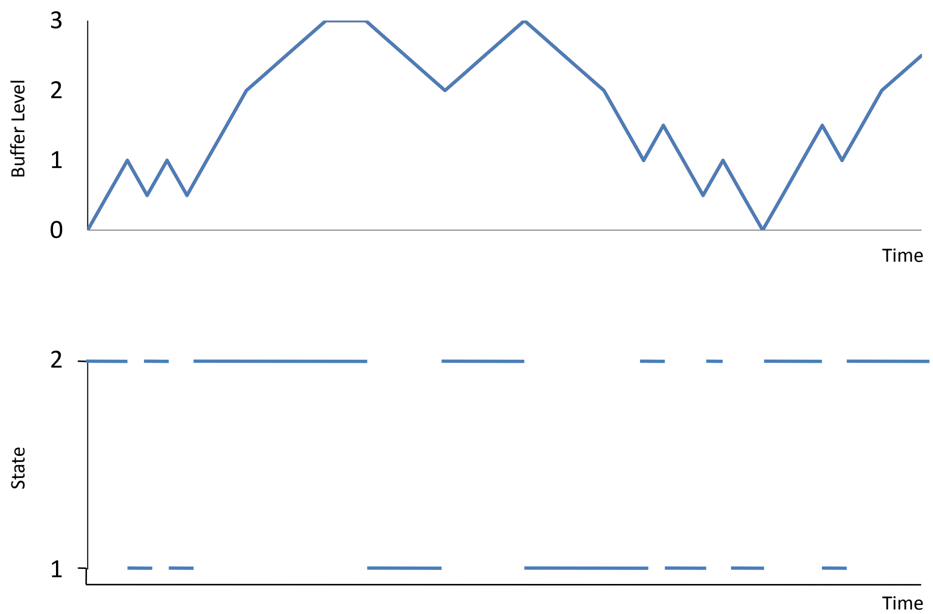

Figure 3.2: A sample path for the buffer level $X(t)$ and state transition $Z(t)$ of the MRMFQ example.

The evolution of the MRMFQ example across time is given in Fig. 3.2. In this sample path, the first thing to notice is that, unlike the SRMFQ case, the drift and transition rates dependent on the buffer level. The drift rate and the transition rates decrease as fluid enters the second regime. This behavior happens because the values in $Q^{(2)}$ and $R^{(2)}$ are smaller than of those in $Q^{(1)}$ and $R^{(1)}$.

In the folowing section, we will proceed with formal definitions of SRMFQs and MRMFQs. Also, the solution steps will also be described. The following sections are based on [21], [1].

3.1 Single-Regime Markov Fluid Queues (SRMFQ)

A SRMFQ is a Markovian joint process which consists of a background CTMC process $Z(t)$ and process that indicates buffer level $X(t)$. The buffer level $X(t)$ is bounded by the buffer capacity value which can either be infinity $X(t) \in [0, \infty)$ or bounded by a finite value B meaning $X(t) \in [0, B]$ For every state i of the process $Z(t)$ there is a drift value r_i. These drift values can be represented in a diagonal drift matrix R in the following manner:

$$R = \begin{bmatrix} r_1 & & & & \\ & r_2 & & & \\ & & \ddots & & \\ & & & r_{N-1} & \\ & & & & r_N \end{bmatrix}.$$

At a given time t, if $Z(t) = i$, then the change rate of $X(t)$ is r_i. Of course, that is the case when $0 < X(t) < \infty$ for infinite buffer capacity and $0 < X(t) < B$ for the finite buffer capacity. When $X(t)$ is equal to the buffer capacity levels 0 or B (for finite case), it can not decrease (increase) further than that level. Therefore, negative (positive) drifts act as zero drift. Given that $Z(t) = i$:

$$\frac{dX(t)}{dt} = \begin{cases} \max\{0, r_i\}, & X(t) = 0, \\ r_i, & 0 < X(t) < B, \\ \min\{0, r_i\}, & X(t) = B. \end{cases} \tag{3.1}$$

for the infinite buffer capacity case. For the infinity buffer capacity case:

$$\frac{dX(t)}{dt} = \begin{cases} \max\{0, r_i\}, & X(t) = 0, \\ r_i, & 0 < X(t). \end{cases} \tag{3.2}$$

From this point, we will continue with the finite buffer capacity case. In order to obtain the solution of SRMFQs, the joint pdf vector $f(x)$ and the probability mass accumulation vector at boundary points k must be specified:

$$f(x) = \begin{bmatrix} f_1(x) & f_2(x) & \dots & f_N(x) \end{bmatrix} \quad and \quad c^{(k)} = \begin{bmatrix} c_1^{(k)} & c_2^{(k)} & \dots & c_N^{(k)} \end{bmatrix}. \tag{3.3}$$

Each element $f_i(x)$ and $c_i^{(k)}$ are defined as:

$$f_i(x) = \lim_{t\to\infty} \frac{d}{dx} \Pr\{X(t) \leq x, Z(t) = i\}, \quad and \quad c_i^{(k)} = \lim_{t\to\infty} \Pr\{X(t) = k,\ Z(t) = i\}, \tag{3.4}$$

where $k \in \{0, B\}$. For the joint pdf vector $f(x)$, the following differential equation holds [21]:

$$\frac{d}{dx} f(x) R = f(x) Q. \tag{3.5}$$

In addition to Eqn. (3.5), the following set of boundary equations must also hold [21]. First, consider S_-, S_0 and S_+ as the set of states with negative, zero and positive drifts, respectively. Then,

$$c_i^{(0)} = 0, \quad \forall i \in S_+ \tag{3.6}$$

$$c_i^{(B)} = 0, \quad \forall i \in S_- \tag{3.7}$$

$$f(0+)R = c^{(0)}Q \tag{3.8}$$

$$f(B-)R = c^{(B)}Q \tag{3.9}$$

$$\left(\int_0^B f(x)dx + c^{(0)} + c^{(B)} \right) \mathbf{1} = 1 \tag{3.10}$$

must hold. After providing the boundary conditions (3.5)-(3.10), we continue with the additive decomposition method.

3.1.1 Additive Decomposition Method

This method which is proposed in [1], is a useful tool for avoiding individual eigenvector computations and takes care of zero drifts and absorbing states.

First of all after a treatment for the states with zero drifts, we may assume that there are no states with zero drifts, i.e., $S_0 = \emptyset$ that will be used as an input to the additive decomposition method [21].

If we define $A = QR^{-1}$, Eqn. (3.5) can be rewritten as

$$\frac{d}{dx} f(x) = f(x)A. \tag{3.11}$$

Then, there exists a Y such that defining $\bar{f}(x) = f(x)Y$ Eqn. (3.11) becomes

$$\frac{d}{dx} \bar{f}(x) = \bar{f}(x) \begin{bmatrix} A_0 & & \\ & A_- & \\ & & A_+ \end{bmatrix}. \tag{3.12}$$

where A_0, A_- and A_+ are upper triangular matrices and have the eigenvalues that are 0, and that have negative and positive real parts respectively on their diagonals. By doing this the LTI system is divided into constant, forward (stable) and backward (antistable) parts. Such Y can be found by Schur decomposition.

Assuming that the background process is irreducible and the mean drift is nonzero, the matrix QR^{-1} has exactly one eigenvalue at 0, meaning that $A_0 = 0$. Therefore, Eqn. (3.12) becomes

$$\frac{d}{dx}\bar{f}(x) = \bar{f}(x)\begin{bmatrix} 0 & & \\ & A_- & \\ & & A_+ \end{bmatrix}. \tag{3.13}$$

At this point, it is easy to see that $\bar{f}(x) = a_0 + a_- e^{A_- x} + a_+ e^{A_+ x}$. Assume that the buffer capacity is B, we can rewrite growing exponential term $a_+ e^{A_+ x} = a_+ e^{A_+ B} e^{-A_+(B-x)}$ and $e^{A_+ B}$ can be absorbed into a_+. Therefore, the solution becomes

$$\bar{f}(x) = a_0 + a_- e^{A_- x} + a_+ e^{-A_+(B-x)}. \tag{3.14}$$

It is known that $f(x) = \bar{f}(x) Y^{-1}$. If Y^{-1} is partitioned as

$$Y^{-1} = \begin{bmatrix} L_0 \\ L_- \\ L_+ \end{bmatrix},$$

then the overall solution for joint pdf $f(x)$ becomes

$$f(x) = \begin{bmatrix} a_0 & a_- & a_+ \end{bmatrix} \begin{bmatrix} 1 & & \\ & e^{A_- x} & \\ & & e^{-A_+(B-x)} \end{bmatrix} \begin{bmatrix} L_0 \\ L_- \\ L_+ \end{bmatrix}. \tag{3.15}$$

From this point, the boundary equations (3.6)-(3.10) must be applied to find the coefficients a_0, a_- and a_+ and probability masses at 0 and B, i.e., $c^{(0)}$ and $c^{(B)}$.

3.2 Multi-Regime Markov Fluid Queues (MRMFQ)

MRMFQs are specific type of MFQs such that the system parameters depend on the fluid level at any given time t [22–25]. Consider buffer of a Markov fluid queue partitioned into K regimes by using the thresholds $0 = T^{(0)} < T^{(1)} < \cdots < T^{(K-1)} < \infty$ for the infinite buffer capacity case, i.e., if the capacity is finite then $T^{(K)} = B$. At any given time t if the fluid level is in the interval $T^{(k-1)} < X(t) < T^{(k)}$, then system is said to be in regime k. While the system is in k^{th} regime the drift matrix is $R^{(k)}$ and infinitesimal generator of the background process $Z(t)$ is $Q^{(k)}$. Each regime possibly has different $R^{(k)}$ and $Q^{(k)}$.

In addition to the regime parameters $R^{(k)}$ and $Q^{(k)}$, for each boundary level $T^{(k)}$ there exist specific drift and infinitesimal generator matrices $\tilde{R}^{(k)}$ and $\tilde{Q}^{(k)}$, respectively.

Similar to SRMFQs system the joint pdf vectors $f^{(k)}(x)$ and the probability mass accumulations at boundary points $c^{(k)}$ must be specified to obtain the steady-state solution

$$f^{(k)}(x) = \begin{bmatrix} f_1^{(k)}(x) & f_2^{(k)}(x) & \dots & f_N^{(k)}(x) \end{bmatrix} \qquad 1 \le k \le K \tag{3.16}$$

and

$$c^{(k)} = \begin{bmatrix} c_0^{(k)} & c_1^{(k)} & \dots & c_N^{(k)} \end{bmatrix} \qquad 0 \le k \le K, \tag{3.17}$$

Each element $f_i(x)^{(k)}$ and $c_i^{(k)}$ are defined as

$$f_i^{(k)}(x) = \lim_{t\to\infty} \frac{d}{dx} \Pr\{X(t) \le x,\ Z(t) = i\}, \tag{3.18}$$

$$c_i^{(k)} = \lim_{t\to\infty} \Pr\{X(t) = T^{(k)},\ Z(t) = i\}. \tag{3.19}$$

Similar to SRMFQ for the joint pdf vectors $f(x)^{(k)}$ the following differential equation holds [21].

$$\frac{d}{dx} f^{(k)}(x) R^{(k)} = f^{(k)}(x) Q^{(k)}. \tag{3.20}$$

Also there exist some set of boundary equations to be used to find the solution of MRMFQs. First, consider $S_-^{(k)}$, $S_0^{(k)}$ and $S_+^{(k)}$ as the set of states in k^{th} regime with negative, zero and positive drifts, respectively. Similarly, $\tilde{S}_-^{(k)}$, $\tilde{S}_0^{(k)}$ and $\tilde{S}_+^{(k)}$ are the set of states in k^{th} boundary level with negative, zero and positive drifts, respectively. Then, the boundary conditions are

$$c_i^{(0)} = 0, \quad \forall i \in S_+^{(1)}, \tag{3.21}$$

$$c_i^{(k)} = 0, \quad \forall i \in \left(S_+^{(k)} \cap S_+^{(k+1)}\right) \cup \left(S_-^{(k)} \cap S_-^{(k+1)}\right), \tag{3.22}$$

$$c_i^{(k)} = 0, \quad \forall i \in \left(S_-^{(k)} \cap S_+^{(k+1)}\right) \cap \left(\tilde{S}_+^{(k)} \cup S_-^{(k)}\right), \tag{3.23}$$

$$c_i^{(K)} = 0, \quad \forall i \in S_-^{(K)}, \tag{3.24}$$

$$f^{(1)}(0+) R^{(1)} = c^{(0)} \tilde{Q}^{(0)}, \tag{3.25}$$

$$f^{(k+1)}(T^{(k)}+) R^{(k+1)} - f^{(k)}(T^{(k)}-) R^{(k)} = c^{(k)} \tilde{Q}^{(k)}, \tag{3.26}$$

$$f_i^{(k)}(T^{(k)}-) = 0 \quad \forall i \in S_-^{(k)} \cup \left(\tilde{S}_0^{(k)} \cap \tilde{S}_+^{(k)}\right), \tag{3.27}$$

$$f^{(K)}(B-) R^{(K)} = -c^{(K)} \tilde{Q}^{(K)}, \tag{3.28}$$

$$f_i^{(k+1)}(T^{(k)}+) = 0 \quad \forall i \in \left(\tilde{S}_0^{(k)} \cap \tilde{S}_-^{(k)}\right) \cup S_+^{(k+1)}, \tag{3.29}$$

$$\left(\sum_{k=1}^{K} \int_{T^{(k-1)}+}^{T^{(k)}-} f^{(k)}(x) dx + \sum_{k=0}^{K-1} c^{(k)} \right) \mathbf{1} = 1. \tag{3.30}$$

For the remainder of this section, steady-state solution technique is described by utilizing the equations (3.20)-(3.28).

3.2.1 Solution of MRMFQs

The solution starts by using the additive decomposition method described in section 3.1.1. The method is applied to Eqn. (3.20).

By defining $A^{(k)} = Q(k)(R^{(k)})^{-1}$, Eqn. (3.20) becomes

$$\frac{d}{dx} f^{(k)}(x) = f(x)^{(k)} A^{(k)}. \tag{3.31}$$

Notice that Eqn. (3.31) is ready for additive decomposition method. As an output of this method, we are able to acquire the pdf of each regime

$$f^{(k)}(x) = \begin{bmatrix} a_0^{(k)} & a_-^{(k)} & a_+^{(k)} \end{bmatrix} \begin{bmatrix} 1 & & \\ & e^{A_-^{(k)}(x-T^{(k-1)})} & \\ & & e^{-A_+^{(k)}(T^{(k)}-x)} \end{bmatrix} \begin{bmatrix} L_0^{(k)} \\ L_-^{(k)} \\ L_+^{(k)} \end{bmatrix} \tag{3.32}$$

where $1 \leq k \leq K$. At this point, to find the pdf coefficients $a^{(k)}$ and probability masses at boundary points $c^{(k)}$, we need to apply boundary equations (3.20)-(3.28). From equations (3.21)-(3.24) we already know some $c_i^{(k)} = 0$. By excluding the $c_i^{(k)}$ that are equal to zero, we construct new vectors $\tilde{c}^{(k)}$ from $c^{(k)}$. Then by using $a^{(k)}$ and unknown vectors $\tilde{c}^{(k)}$ a new z vector is constructed:

$$z = \begin{bmatrix} \tilde{c}^{(0)} & a^{(1)} & \tilde{c}^{(1)} & a^{(2)} & \tilde{c}^{(2)} & \dots & a^{(K-1)} & \tilde{c}^{(K-1)} & a^{(K)} & \tilde{c}^{(K)} \end{bmatrix}. \tag{3.33}$$

Then, by using the boundary conditions a H matrix is constructed such that

$$zH = 0 \tag{3.34}$$

which represents the system of linear equations that is constructed from equations (3.25)-(3.30).

The constructed H matrix will have the following form as described in [9].

$$H = \begin{bmatrix} H_1^M & H_1^U & & & \\ H_1^L & H_2^M & H_2^U & & \\ & H_2^L & H_3^M & \ddots & \\ & & \ddots & \ddots & H_{K-1}^U \\ & & & H_{K-1}^L & H_K^M \end{bmatrix},$$

where H_k^M for $0 \leq k \leq K-1$ are n_k by n_k square matrices and

$$n_k = \begin{cases} \text{length}(\tilde{c}^{(0)}) & \text{if } k = 1, \\ \text{length}(\tilde{c}^{(k-1)}) + \text{length}(\tilde{a}^{(k-1)}) & \text{if } 2 \leq k \leq K. \end{cases}$$

With this information it is easy to find the sizes of H_k^U and H_k^L for $1 \leq k \leq K-1$.

With this structure of H, the matrix is suitable for block tri-diagonal LU factorization which is a method that reduces the time required to solve Eqn. (3.34) dramatically, i.e., the computational complexity can be brought down to $\mathcal{O}(N^3K)$ [9].

Chapter 4

Transient and First Passage Time Solutions of Multi-Regime Markov Fluid Queues

In this section, the system model is explained which will be used for the transient and first passage time solutions of multi-regime Markov fluid queues. An auxiliary fluid queue will be presented for each. Lastly, the results in question will be obtained from the steady-state solution of the auxiliary MRMFQ.

These models can be applied to both multi and single-regime Markov fluid queues; but, for the rest of this thesis the method will be described on multi-regime Markov fluid queues due to the fact that the solution for MRMFQs covers the SRSMFQ case.

4.1 Transient Solution of Multi-Regime Markov Fluid Queues

Consider a MRMFQ process $X(t) = (X_f(t), X_m(t))$ with the given fluid level process $0 \leq X_f(t) \leq B$ and the modulating background Markov process $X_m(t) \in \{1, 2, ..., n\}$. The MRMFQ process is characterized by the pair of piece-wise constant drift and infinitesimal generator matrices $\{Q_x(x), R_x(x)\}$, i.e., $Q_x(x)$ and $R_x(x)$ stay constant inside a regime ($Q^{(k)}$ and $R^{(k)}$ in k^{th} regime) and may possibly have different values for each regime. The steady-state solution of such MFQs can be obtained as described in section 3.2.; but, for the transient (time dependent) case, we must construct an auxiliary background Markov process from the original one to give problem a time horizon T. Then steady-state solution of the auxiliary MFQ with time horizon will include the results for transient solution of the original one.

In order to produce the MFQ mentioned above, let us define a PH-type distribution that is characterized by (α,S) with $S^0 = -Se$ where e is a column vector of ones of appropriate length. The fluid level process starts from a specified level $X_f(0) = a$ and background process from state j with probability π_j. So;

$$\pi = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \end{bmatrix}, \quad \pi_j = Pr\{X_m(0) = j\},\ j = 1, 2, \ldots, n. \tag{4.1}$$

In order to refer to Eqn. (4.1) in other equations, the notation $X_m(0) \sim \pi$ is used. After a random time horizon T expires we are trying to obtain the following joint cumulative distribution function

$$F_T^{a,\pi}(j, x) = \Pr\{X_f(T) \leq x, X_m(T) = j | X_f(0) = a, X_m(0) \sim \pi\} \tag{4.2}$$

and the corresponding joint pdf is denoted by $f_T^{a,\pi}(j, x)$. The joint cdf vector which covers all states is defined as

$$F_T^{a,\pi}(x) = \begin{bmatrix} F_T^{a,\pi}(1, x) & F_T^{a,\pi}(2, x) & \cdots & F_T^{a,\pi}(n, x) \end{bmatrix}, \tag{4.3}$$

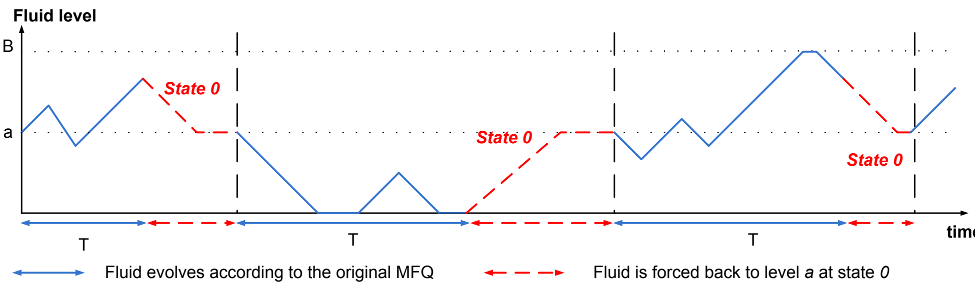

Figure 4.1: A sample path of the auxiliary MFQ for the transient distribution of the original MFQ.

and the corresponding joint pdf vector is denoted by $f_T^{a,\pi}(x)$. If the random time horizon T replaced with a deterministic time value t, then Eqn. (4.2) provides an expression for the joint cdf of the transient behavior of the MFQ at time t.

After these definitions, the auxiliary MRMFQ can be constructed in order to obtain Eqn. (4.2) with deterministic time value t. Consider a MRMFQ $Y(t) = (Y_f(t), Y_m(t))$ which is to be constructed from the original one, $X(t)$ and time horizon $T \sim PH(\alpha, S)$. This MRMFQ has $1+mn$ states. mn of these corresponds to the pairs $(i, j), 1 \leq i \leq m, 1 \leq j \leq n$ where j is keeping track of state of original modulating process $X_m(t)$ and i denotes the state of PH-type distributed time horizon T, and the final state is an absorbing state which will be called as state 0. This fluid process starts from level a and evolves according to level dependent parameters of the original MRMFQ $\{Q_x(x), R_x(x)\}$ until it reaches to the time horizon $T \sim PH(\alpha, S)$. After timer expires, the fluid level process is forced back to level a while the modulating process is at state 0. This is achieved by a negative drift (positive drift) while $X_f(t) > a$ ($X_f(t) < a$) at state 0. When the fluid level reaches a drift of state 0 becomes zero and fluid level is forced to stay at level a in state 0 for an exponentially distributed duration. Then the modulating process escapes from state 0 with probability $\alpha_i \pi_j$ to state $(i, j), 1 \leq i \leq m, 1 \leq j \leq n$ and starts to evolve according to original MRMFQ again. The described process repeats itself forever while $t \to \infty$. A sample path evolution with respect to time of $Y_f(t)$ is given in Fig. 4.1. In this model, the state 0 is introduced to push the system to start from level a over and over again at the beginning of every epoch.

Based on the above description and sample path at Fig. 4.1 we can proceed to give mathematical expressions for $Y(t)$ using $X(t)$ and time horizon $T \sim PH(\alpha, S)$. The infinitesimal generator and drift of $Q_Y(x)$ and $R_Y(x)$ must be found to find the steady-state solution of $Y(t)$. First of all, the states are enumerated as $(0,(1,1),(1,2),\ldots,(1,n),(2,1),\ldots,(m,n))$. Then the following pair of $Q_Y(x)$ and $R_Y(x)$ can characterize $Y(t)$:

$$Q_Y(x) = \left[\begin{array}{c|c} 0 & 0 \\ \hline S^0 \otimes e & I_m \otimes Q_X(x) + S \otimes I_n \end{array}\right] \text{ for } x \neq a, \tag{4.4}$$

$$Q_Y(a) = \left[\begin{array}{c|c} -1 & \alpha \otimes \pi \\ \hline S^0 \otimes e & I_m \otimes Q_X(a) + S \otimes I_n \end{array}\right], \tag{4.5}$$

and for $0 < a < B$

$$R_Y(x) = \begin{cases} \mathbf{diag}\{1, I \otimes R_X(x)\} & \text{if } x < a, \\ \mathbf{diag}\{-1, I \otimes R_X(x)\} & \text{if } x > a, \\ \mathbf{diag}\{0, I \otimes R_X(a)\} & \text{if } x = a. \end{cases} \tag{4.6}$$

For the specific case $a = 0$, $Q_Y(x)$ is the same as in Eqn. (4.4), but Eqn. (4.6) needs a slight modification:

$$R_Y(x) = \begin{cases} \mathbf{diag}\{-1, I \otimes R_X(x)\} & \text{if } x > 0, \\ \mathbf{diag}\{0, I \otimes R_X(0)\} & \text{if } x = 0. \end{cases} \tag{4.7}$$

On the other hand, when $a = B$,

$$R_Y(x) = \begin{cases} \mathbf{diag}\{1, I \otimes R_X(x)\} & \text{if } x < B, \\ \mathbf{diag}\{0, I \otimes R_X(0)\} & \text{if } x = B. \end{cases} \tag{4.8}$$

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

$$F_Y(s,x) = \lim_{t\to\infty} \Pr\{Y_f(t) \leq x, Y_m(t) = s\}, s = 0 \text{ or } s = (i,j), 1 \leq i \leq m, 1 \leq j \leq n. \tag{4.9}$$

Also let $f_Y(s,x)$ denote the corresponding joint pdf allowing impulses corresponding to the probability masses at the boundaries of the original MFQ. From sample path arguments, it is not difficult to show that

$$f_T^{a,\pi}(j,x) = \frac{\sum_i f_Y((i,j),x) S_i^0}{\sum_{i,j} f_Y((i,j),x) S_i^0}, \tag{4.10}$$

where S_i^0 denotes the ith entry of S^0.

4.2 First Passage Time Solution of Multi-Regime Markov Fluid Queues

The same MRMFQ process $\mathbf{X}(t) = (X_f(t), X_m(t))$ described in the previous section characterized with the matrix pair $(Q_X(x), R_X(x))$ will be used as the original MRMFQ which the first passage time solution will be applied. Similar to the previous section, for the first passage time solution the MRMFQ process starts at level $X_f(0) = a$ with initial probability vector π, i.e., $X_m(0) \sim \pi$. Let $\tau^{a,b,\pi}$ denote the first passage time from level a to level b:

$$\tau^{a,b,\pi} = \inf_t \{X_f(t) = b | X_f(0) = a, X_m(0) \sim \pi\}. \tag{4.11}$$

We are interested in the following probability

$$F_\tau^{a,b,\pi}(T) = \Pr\{\tau^{a,b,\pi} < T\} \tag{4.12}$$

for PH-distributed T characterized with the pair (α, S). When T is deterministic with $T = t$, then Eqn. (4.12) provides an expression for the cdf of the first passage time from level a to level b at time t.

We propose to construct an auxiliary MRMFQ denoted by $\mathbf{Z}(t)$ whose steady-state solution provides a solution for the probability given in Eqn. (4.12) for the original MRMFQ $\mathbf{X}(t)$ similar to the transient solution case. This process

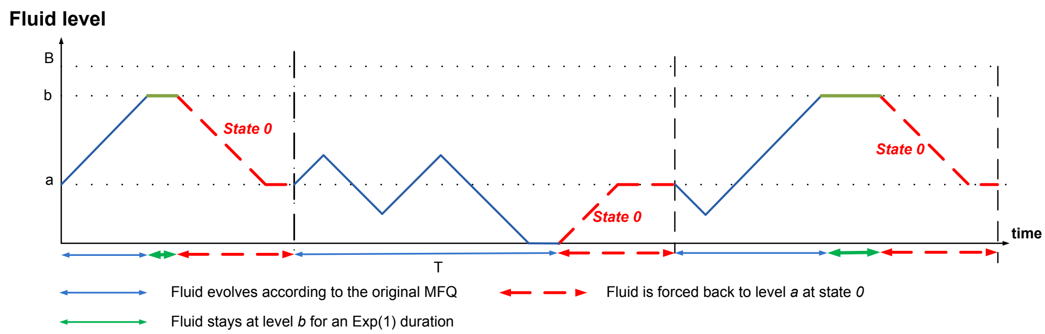

Figure 4.2: A sample path of the auxiliary MRMFQ $\mathbf{Z(t)}$ constructed for the first passage times of the original MRMFQ $\mathbf{X(t)}$.

starts from fluid level a and behaves according to the original MRMFQ parameters $(Q_X(x), R_X(x))$ until either the threshold b is reached or random time T is expired. When the former situation occurs the fluid stays at level b for some random exponential time whose mean is set to unity then escapes from level b with state 0. In the latter situation, the fluid is forced back to level a directly with state 0. When fluid level hits a it waits there for a random exponential time whose mean is set to unity again at state 0. Then process escapes from state 0 with probability $\alpha_i \pi_j$ to state $(i, j), 1 \leq i \leq m, 1 \leq j \leq n$ and starts to act like original MRMFQ once again until it is pushed back to state 0 again. The described process repeats itself forever while $t \to \infty$. Consider the sample path of the auxiliary fluid process $\mathbf{Z}(t)$ in Fig. 4.2. First and third episodes are the ones that hit level b and in the second one timer expires.

The fluid process given in Fig. 4.2 is actually an MRMFQ process denoted by $\mathbf{Z}(t) = (Z_f(t), Z_m(t))$ with the same state-space as that of the MRMFQ process $\mathbf{Y}(t)$ of the previous section transient solution case). Actually, $(Q_Z(x), R_Z(x))$ is the same as the pair $(Q_Y(x), R_Y(x))$ for $0 \leq x < b$. Moreover, $Z_f(t)$ will not take a value in the interval $(b, B]$ (for the case which $b < a$, $Z_f(t)$ will not take a value in the interval $[0, b)$). On the other hand, 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,

$$Q_Z(x) = \left[\begin{array}{c|c} 0 & 0 \\ \hline S^0 \otimes e & I_m \otimes Q_X(x) + S \otimes I_n \end{array}\right] \text{ for } x \neq a, b, \tag{4.13}$$

$$Q_Z(a) = \left[\begin{array}{c|c} -1 & \alpha \otimes \pi \\ \hline S^0 \otimes e & I_m \otimes Q_X(a) + S \otimes I_n \end{array}\right]. \tag{4.14}$$

For level $Z_f(t) = b$, the system is forced to state 0 after some exponential time spent on level b. Therefore, all states can only go to state 0 after some exponential time whose mean is set to unity and state 0 must an absorbing state. To achieve this

$$Q_Z(b) = \left[\begin{array}{c|c} 0 & 0 \\ \hline e & -I_{mn} \end{array}\right] \tag{4.15}$$

and for $0 < a < B$

$$R_Z(x) = \begin{cases} \textbf{diag}\{1, I \otimes R_X(x)\} & \text{if } x < a, x \neq b, \\ \textbf{diag}\{-1, I \otimes R_X(x)\} & \text{if } x > a, x \neq b, \\ \textbf{diag}\{0, I \otimes R_X(a)\} & \text{if } x = a, x \neq b. \end{cases} \tag{4.16}$$

For the specific case $a = 0$, $Q_Z(x)$ is the same as in equations (4.13)-(4.15), but Eqn. (4.16) needs a slight modification:

$$R_Z(x) = \begin{cases} \textbf{diag}\{-1, I \otimes R_X(x)\} & \text{if } x > 0, x \neq b, \\ \textbf{diag}\{0, I \otimes R_X(0)\} & \text{if } x = 0, x \neq b. \end{cases} \tag{4.17}$$

On the other hand, when $a = B$,

$$R_Z(x) = \begin{cases} \textbf{diag}\{1, I \otimes R_X(x)\} & \text{if } x < B, x \neq b, \\ \textbf{diag}\{0, I \otimes R_X(0)\} & \text{if } x = B, x \neq b. \end{cases} \tag{4.18}$$

For level $Z_f(t) = b$ the system is forced to state 0 after some exponential time spent on level b (to achieve this the drift should be zero for all states except state

0):

$$R_Z(b) = \begin{cases} \mathbf{diag}\{-1, \mathbf{0}_{mn}\} & \text{if } a < b, \\ \mathbf{diag}\{1, \mathbf{0}_{mn}\} & \text{if } a > b, \end{cases} \tag{4.19}$$

where $\mathbf{0}_{mn}$ is a zero row vector of length mn. As observed in Fig. 4.2, the overall trajectory is episodic. All episodes terminate with the fluid level staying at level a for an exponentially distributed duration of time with unity mean. In episodes when the fluid level reaches level b before the timer expires, the fluid process visits level b for an exponentially distributed duration of time with unity mean. Let $c(b)$ denote the overall steady-state probability mass at level b for all states $(i, j), 1 \le i \le m, 1 \le j \le n$ for the auxiliary MFQ $\mathbf{Z}(t)$. Also let $c_0(a)$ be the steady-state probability mass at level a for state 0. Based on this notation, we have

$$F_{\tau}^{a,b,\pi}(T) = \frac{c(b)}{c_0(a)}. \tag{4.20}$$

4.3 Phase Type and Matrix Exponential Distributions

In order to describe a phase type (PH-type) distribution, a continuous-time Markov chain is defined on the state space $\{1, \ldots, k, k+1\}$ with state $k+1$ being absorbing, other states being transient, initial probability vector $(\alpha, 0)$, and an infinitesimal generator

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

Here, α is a row probability vector of size k, the sub-generator S is a $k \times k$ sub-stochastic matrix, e denotes a column vector of ones of appropriate size and $S^0 = -Se$ is a column vector of size k. Let X denote the time till absorption into the absorbing state $k+1$. Then, the distribution of X is called phase type (PH-type), i.e., $X \sim PH(\alpha, S)$ with order k. For a detailed study of PH-type distributions, we refer the reader to [26]. The probability density function (pdf) of $X \sim PH(\alpha, S)$ denoted by $f_X(x)$ is then given as:

$$f_X(x) = -\alpha e^{Sx} S e, x \geq 0. \tag{4.21}$$

A superset of PH-type distributions is the so-called Matrix Exponential (ME) distributions; see [27] and [28] for a detailed description of ME distributions and their properties. We say $X \sim ME(\alpha, S, s)$ with order k if the pdf of random variable X is in the following form:

$$f_X(x) = \alpha e^{Sx} s, x \geq 0 \tag{4.22}$$

for a $k \times k$ matrix S. If $s = -Se$ then Eqn. (4.22) reduces to the algebraic form (4.21) given for the pdf of PH-type distribution and s can always be forced to satisfy $s = -Se$ using similarity transformations. When so, we simply say $X \sim ME(\alpha, S)$. ME distributions do not necessarily possess the stochastic interpretation of that of PH-type distributions.

The specific case of X being deterministic, i.e., $X = t$ is key to the current paper. Clearly, the pdf of X is a dirac delta function located at t and X is neither PH-type nor ME-type. However, $X = t$ can arbitrarily well be approximated by PH-type distributions. In the Erlangization alternative (see [19], [29]), X is approximated by $\tilde{X}_k \sim PH(\alpha_k, S_k)$ (also called Erlang-k with order k) where

$$\alpha_k = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, S_k = \frac{k}{t} \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \\ & & & -1 \end{bmatrix}.$$

As the order k increases, $\tilde{X}_k$ converges to X in distribution but with relatively slow convergence rates since the corresponding coefficient of variation (CoV) of $\tilde{X}_k = 1/k$. Another alternative we propose to use in this paper is the concentrated ME distribution introduced in [5] with parameter l denoted by CME-l with order $k = 2l + 1, l \geq 1$ whose CoV behaves approximately as $2/k^2$ and decays to zero much faster than the Erlang distribution with increased order k. The CME-l

distribution is characterized by a pair of matrices (α, S) with pdf in the same algebraic form as in (4.21) but without the stochastic interpretation. For CME-l distribution, there is no structure like Erlang-k; but, for CME with order $l = 5$, α and S values are given below as an example:

$$\alpha = \begin{bmatrix} 5.4024415 & -2.9882265 & -2.3449328 & 0.3749164 & 0.5558015 \end{bmatrix},$$

$$S = \frac{1}{t}\begin{bmatrix} -3.1922581 & 0 & 0 & 0 & 0 \\ 0 & -3.1922581 & -3.0266150 & 0 & 0 \\ 0 & 3.0266150 & -3.1922581 & 0 & 0 \\ 0 & 0 & 0 & -3.1922581 & -6.0532301 \\ 0 & 0 & 0 & 6.0532301 & -3.1922581 \end{bmatrix}.$$

In the numerical examples chapter, both alternatives will be comparatively studied.

4.4 Numerical Application of Transient Methods on a 3-State MRMFQ

Consider a MRMFQ that has a finite buffer capacity $B = 3$ meaning that $X(t) \in [0, 3]$. The capacity is partitioned into two regimes: (0,2) and (2,3). For regime 1 and 2 the Background process is a three-state Markov chain that has the following infinitesimal generator matrices

$$Q^{(1)} = \begin{bmatrix} -1 & 0.25 & 0.75 \\ 0.5 & -1 & 0.5 \\ 0.7 & 0.3 & -1 \end{bmatrix} \quad , \quad Q^{(2)} = \begin{bmatrix} -1 & 0.8 & 0.2 \\ 0.3 & -1 & 0.7 \\ 0.5 & 0.5 & -1 \end{bmatrix}$$

and drift matrices

$$R^{(1)} = \begin{bmatrix} 0 & & \\ & -1 & \\ & & 1 \end{bmatrix}, \quad R^{(2)} = \begin{bmatrix} 0 & & \\ & -0.5 & \\ & & 0.5 \end{bmatrix}.$$

The boundary infinitesimal generators for $X(t) = 0, 2, 3$ respectively are

$$\tilde{Q}^{(1)} = \begin{bmatrix} -1 & 0.25 & 0.75 \\ 0.5 & -1 & 0.5 \\ 0.7 & 0.3 & -1 \end{bmatrix}, \quad \tilde{Q}^{(2)} = \begin{bmatrix} -1 & 0.25 & 0.75 \\ 0.5 & -1 & 0.5 \\ 0.7 & 0.3 & -1 \end{bmatrix}, \quad \tilde{Q}^{(3)} = \begin{bmatrix} -1 & 0.8 & 0.2 \\ 0.3 & -1 & 0.7 \\ 0.5 & 0.5 & -1 \end{bmatrix},$$

and drift matrices

$$\tilde{R}^{(1)} = \begin{bmatrix} 0 & & \\ & -1 & \\ & & 1 \end{bmatrix}, \quad \tilde{R}^{(2)} = \begin{bmatrix} 0 & & \\ & -1 & \\ & & 1 \end{bmatrix}, \quad \tilde{R}^{(3)} = \begin{bmatrix} 0 & & \\ & -0.5 & \\ & & 0.5 \end{bmatrix}.$$

Therefore, MRMFQ $X(t)$ is characterized by the pair $(Q(x), R(x))$ which is described above. We intend to find the transient distribution and first passage time distributions of $X(t)$, namely

$$F_T^{a,\pi}(j, x) = \Pr\{X_f(T) \leq x, X_m(T) = j | X_f(0) = a, X_m(0) \sim \pi\}$$

and

$$F_\tau^{a,b,\pi}(T) = \Pr\{\tau^{a,b,\pi} < T\}$$

where

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

respectively. In this case the variable T is a deterministic time horizon and for this example $T = 100$. The fluid level process starts from a specified level $X_f(0) = a = 0$ and background process from state 3 with probability 1, i.e., $\pi = [0 \quad 0 \quad 1]$. For the case of first passage time distribution first passage time to level $b = 2.5$ need to be found.

Having specified all the unknowns we need to specify the characteristics of PH-type to introduce the time horizon to the problem. For better understanding of the calculations, we choose Erlangization with order $k = 3$ which is characterized by

$$\alpha_3 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \qquad S_3 = \frac{3}{100} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

We now proceed to construct the new MRMFQ process $Y(t)$ which will be used for the transient solution of the original MRMFQ.

4.4.1 Transient Solution

By implementing the method which was described in Section 4.1, we aim to find the transient solution of $X(t)$. First, we start with constructing the auxiliary MRMFQ process $Y(t)$ which is characterized by the pair $Q_Y(x)$ and $R_Y(x)$. $Q_Y(x)$ and $R_Y(x)$ matrices can be constructed according to the equations (4.4)-(4.8) using $X(t)$ and Erlang-3 that has been defined above. There will be $1 + mn$ states in the new MRMFQ which is equal to 10. The infinitesimal generator matrices of $Y(t)$ are the following

$$Q_Y^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & -1.03 & 0.50 & 0 & 0.03 & 0 & 0 & 0 & 0 \\ 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.700 & 0.30 & -1.03 \end{bmatrix},$$

$$Q_Y^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.03 & 0.80 & 0.20 & 0.03 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.30 & -1.030 & 0.70 & 0 & 0.03 & 0 & 0 & 0 & 0 \\ 0 & 0.50 & 0.50 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.03 & 0.80 & 0.20 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.30 & -1.03 & 0.70 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & 0.50 & -1.03 & 0 & 0 & 0.03 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.80 & 0.20 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.30 & -1.03 & 0.70 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0.50 & -1.03 \end{bmatrix}$$

and for the boundary conditions, infinitesimal generator at level $a = 0$ has a different distribution since it is the beginning level a of fluid process

$$
Q_Y(a) = \tilde{Q}_Y^{(1)} = \begin{bmatrix}
-1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.5 & -1.03 & 0.50 & 0 & 0.03 & 0 & 0 & 0 & 0 \\
0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 & 0 & 0.03 & 0 \\
0 & 0 & 0 & 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.700 & 0.30 & -1.03
\end{bmatrix}.
$$

For the other boundary levels, $\tilde{Q}_Y^{(2)} = Q_Y^{(1)}$ and $\tilde{Q}_Y^{(3)} = Q_Y^{(2)}$

$$
R_Y(x) = \begin{cases}
\mathbf{diag}\{0, 0, -1, 1, 0, -1, 1, 0, -1, 1\} & \text{if } x = 0, \\
\mathbf{diag}\{-1, 0, -1, 1, 0, -1, 1, 0, -1, 1\} & \text{if } 0 < x \leq 2, \\
\mathbf{diag}\{-1, 0, -0.5, 0.5, 0, -0.5, 0.5, 0, -0.5, 0.5\} & \text{if } 2 < x \leq 3.
\end{cases}
$$

At this point, we constructed the MRMFQ which is necessary for transient solution. We should apply the steady-state solution for MRMFQs to $Y(t)$ which is described in section 3.2 and then find the result based on Eqn. (4.10).

4.4.2 First Passage Time Distribution Solution

In order to obtain the first passage time distribution, a new MRMFQ $Z(t)$ must be constructed which is characterized by the pair $Q_Z(x)$ and $R_Z(x)$. It is similar to previously constructed $Y(t)$ but the fluid level is bounded by $b = 2.5$ and $Q_Z(b)$ and $R_Z(b)$ have different values:

$$
Q_Z^{(1)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.5 & -1.03 & 0.50 & 0 & 0.03 & 0 & 0 & 0 & 0 \\
0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 & 0 & 0.03 & 0 \\
0 & 0 & 0 & 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.700 & 0.30 & -1.03
\end{bmatrix},
$$

$$
Q_Z^{(2)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1.03 & 0.80 & 0.20 & 0.03 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.30 & -1.030 & 0.70 & 0 & 0.03 & 0 & 0 & 0 & 0 \\
0 & 0.50 & 0.50 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1.03 & 0.80 & 0.20 & 0.03 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.30 & -1.03 & 0.70 & 0 & 0.03 & 0 \\
0 & 0 & 0 & 0 & 0.50 & 0.50 & -1.03 & 0 & 0 & 0.03 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.80 & 0.20 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.30 & -1.03 & 0.70 \\
0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & 0.50 & -1.03
\end{bmatrix},
$$

$$Q_Z(a) = \tilde{Q}_Z^{(1)} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & -1.03 & 0.50 & 0 & 0.03 & 0 & 0 & 0 & 0 \\ 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 & 0.03 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.70 & 0.30 & -1.03 & 0 & 0 & 0.03 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & -1.03 & 0.25 & 0.75 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.50 & -1.03 & 0.50 \\ 0.03 & 0 & 0 & 0 & 0 & 0 & 0 & 0.700 & 0.30 & -1.03 \end{bmatrix},$$

for the other boundary levels, $\tilde{Q}_Y^{(2)} = Q_Y^{(1)}$ and the boundary level at $b = 2.5$

$$Q_Z(b) = \tilde{Q}_Z^{(3)} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

$$R_Y(x) = \begin{cases} \mathbf{diag}\{0, 0, -1, 1, 0, -1, 1, 0, -1, 1\} & \text{if } x = 0, \\ \mathbf{diag}\{-1, 0, -1, 1, 0, -1, 1, 0, -1, 1\} & \text{if } 0 < x \leq 2, \\ \mathbf{diag}\{-1, 0, -0.5, 0.5, 0, -0.5, 0.5, 0, -0.5, 0.5\} & \text{if } 2 < x < 2.5, \\ \mathbf{diag}\{-1, 0, 0, 0, 0, 0, 0, 0, 0, 0\} & \text{if } x = 2.5. \end{cases}$$

At this point, we constructed the MRMFQ which is necessary for first passage time solution. We should apply the steady state solution for MRMFQs to $Y(t)$ which is described in Section 3.2 and then find the first passage time distributions based on Eqn. (4.20).

Chapter 5

Transient Solution of Multi-Server Queueing System with Stochastic Impatient Customers

In this chapter, a multi-server queueing system with stochastic impatient customers is considered, which is motivated by increasing application of conventional call center models in customer contact centers, health-care systems and telecommunication networks. In this system, a customer leaves the queue without getting served if its waiting time in the queue exceeds a random variable.

In order to obtain the model of such systems, we consider M/M/S queue with generally distributed impatience time, i.e., M/M/S+G. Multi-Regime Markov Fluid queue (MRMFQ) is deployed to derive the distribution of virtual waiting time and several other metrics of interest.

In this chapter, we intend to give detailed description of this system and the MRMFQ process $X(t)$ which will be used to model these systems. Later we will describe how to obtain the metrics such as expected waiting time, probability of abandonement etc. After these definitions, the transient solution of this model will be introduced using the ME-fication technique which was described in chapter 4, but, with slight modification due to a state in which time does not evolve.

5.1 Model Description

A stationary M/M/S queue with stochastic impatience times is considered in this section. The queueing system has S homogeneous servers and a waiting room of infinite capacity. Customers arrive at the system according to a Poisson arrival process with rate λ. The arriving customers are served according to First-Come First-Served (FCFS) policy. The service times of customers are independent and identically distributed (i.i.d.) according to an exponential distribution with mean S/μ, where $\mu > 0$. Customers are impatient and if their elapsed waiting times exceed a generally distributed random threshold, which is drawn i.i.d. for each customer, they will leave the system without receiving their service.

In order to model the M/M/S+G queueing system as a Markov fluid queue, virtual waiting time process is deployed. The virtual waiting time, which is denoted by $V(t)$, is denoted as the overall time a fictitious customer who arrives at time t would spend in the queue until his/her service begins. If there are less than S customers in the system, $V(t) = 0$. If there are exactly S customers in the system, $V(t)$ will indicate the time required until the first departure. When $V(t) > 0$, each new arrival decides whether it is going to wait in the queue or not based on the value of its stochastic abandonement time (patience). The cumulative distribution function (CDF) of customer abandonment process is denoted by $g(X)$. If the value of the stochastic abandonment time is smaller than the virtual waiting time, the customer is not patient enough and will leave the system immediately. Hence, any packets that decide to wait in the queue will eventually get served. To be precise, each new arrival will abandon the system with probability $g(V(t))$, while it will wait in line to get served with probability $1 - g(V(t))$. The sample path for the process $V(t)$ is given in Fig. 5.1 for an example scenario with three servers. At each event (customer arrival or departure) instant, all of the customers in the system are demonstrated as a column where the three topmost ones denote the customers in service. Each customer is illustrated by two digits: one showing the sequence number of the customer, while the other showing the remaining service time. For instance, (4)8 indicates the fourth customer who requires eight time slots to get served. Obviously, the remaining service time of the customers in the queue is fixed until they start receiving the service. The packet arrival and departure instants are indicated by the small upside and downside arrows, respectively. In this example scenario, the customer number 5 decides to abandon the system due to excessive expected waiting time, while the rest of the customers wait in line and eventually get served.

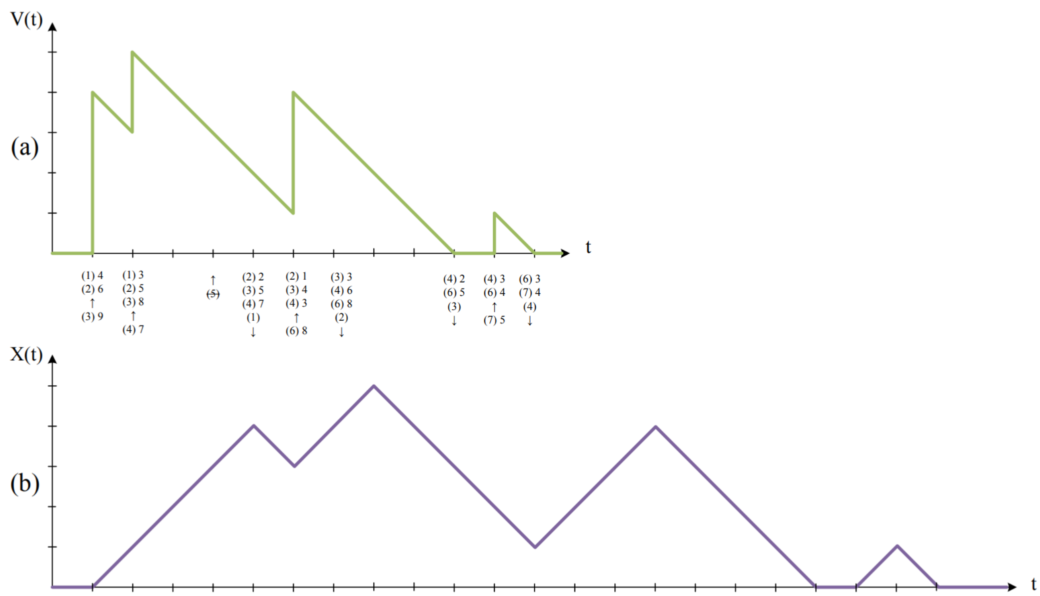

Figure 5.1: Sample paths of (a) the virtual waiting time process $V(t)$ and (b) auxiliary random process $X(t)$.

Due to abrupt jumps, the virtual waiting time process is not suitable to be modeled as a fluid queue [30]. Therefore, $X(t)$ auxiliary random process is introduced by replacing the abrupt jumps in the virtual waiting time process with linear increments corresponding to a drift of one (see Fig. 5.1 (b)). Moreover, it is clear from the sample path arguments that the steady-state distribution of the process $V(t)$ can be derived from that of $X(t)$ by censoring out the states corresponding to positive drifts.

We first focus on the fluid model for $X(t)$. We define S service states indicating the number of busy servers denoted by $\{0, 1, ..., S-2, S-1(D)\}$ and one increment state I. In this model, states $i \in \{0, 1, ..., S-2, S-1(D)\}$ when $X(t) = 0$ represents i busy servers and $X(t) > 0$ means there are S busy servers. During state $i \in \{0, 1, ..., S-2, S-1(D)\}$ and $X(t) = 0$, the customers are served with rate $i\mu$, because there is at least one empty server the service of which start immediately with a new arrival. However, when in state $S-1(D)$ if a new arrival occurs waiting time must increase, which is achieved by state I. In this case, customers are served with rate $S\mu$ and now that all servers are full $X(t)$ starts to increase in order to introduce the waiting time of the next customer in the line. Once state I increases the fluid level to the virtual waiting time which is an exponentially distributed time with rate $S\mu$, system goes back to state $S-1(D)$, but when $X(t) > 0$ state $S-1(D)$ no longer represents $S-1$ busy servers exist. It represents the decrement of the waiting time when S servers are busy and there is no new arrival. In state $S-1(D)$ waiting time starts to decrease until $X(t) = 0$ is reached. When $X(t) > 0$ if a new customer decides to wait in the queue, which happens with probability $1 - g(X)$, the system transitions to state I again and increases the fluid level to the virtual waiting time which is an exponentially distributed time with rate $S\mu$ again, then system goes back to state $S-1(D)$ again. In state I, level of the fluid increases in order to take into account the impact of the new arrival on the virtual waiting time. Whenever, the next customer decides to abandon, which happens with probability $g(X)$, the system stays in state $S-1(D)$. During state I, time does not evolve, since it is only an auxiliary state that is added to system in order to facilitate the modeling of the system via fluid queues. Note that at the boundary $X(t) = 0$, no abandonment occurs and $g(X) = 0$.

In essence, the M/M/S+G queueing system is modeled via a Continuous Feedback Markov Fluid Queue (CFMFQ) which is denoted by the random process $X(t)$ and controlled by the underlying Markov chain. The underlying Markov chain is determined by $X_m(t)$ which determines the number of busy servers. The analytical treatment of such CFMFQ system requires the solution where $Q(x)$ and $R(x)$ are infinitesimal generator matrix of the underlying Markov chain and drift matrix of the fluid, respectively, which both depend continuously on the fluid level. Our proposed call center model is characterized by an infinitesimal generator matrix for $X(t) > 0$ denoted by $Q(x)$, an infinitesimal generator matrix for the boundary $X(t) = 0$ denoted by $\tilde{Q}(x)$, and their corresponding rate matrices denoted by $R(x)$ and $\tilde{R}(x)$, respectively. In order to implement CFMFQ, we will quantize the fluid level and introduce model as a MRMFQ problem.

On the basis of the above descriptions, the infinitesimal generator matrices for the $X(t) > 0$ and $X(t) = 0$ for states $\{0, 1, ..., S-2, S-1(D), I\}$ respectively can be obtained as

$$Q(x) = \begin{bmatrix} 0 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & -\lambda(1-g(x)) & \lambda(1-g(x)) \\ 0 & \dots & S\mu & -S\mu \end{bmatrix}, \tag{5.1}$$

$$\tilde{Q}(x) = \begin{bmatrix} -\lambda & \lambda & & & & \\ \mu & -\lambda-\mu & \lambda & & & \\ & 2\mu & -\lambda-2\mu & \lambda & & \\ & & \ddots & \ddots & \ddots & \\ & & & (S-1)\mu & -\lambda-(S-1)\mu & \lambda \\ & & & & S\mu & -S\mu \end{bmatrix} \tag{5.2}$$

and for the drift matrices

$$\tilde{R}(X) = R(X) = \mathbf{diag}\{-1, -1, \dots, -1, 1\} \tag{5.3}$$

where diag denotes a diagonal matrix. From this model, we are able to find various metrics of interest. The key point of finding them is to neglect the time spent at state I. Some metrics and how to calculate them are listed below.

Probability that waiting time at queue $W = 0$

$$P(W = 0) = \frac{c(0)}{1 - P(X_m(t) = S + 1)}. \tag{5.4}$$

Probability that an arrival result with abandonement

$$P(A) = \frac{\int_0^\infty f_S(x)g(x)dx}{1 - PX_m(t) = S + 1)}. \tag{5.5}$$

Expected waiting time in the queue of a customer that didn't abandon

$$E[W|S] = \frac{\int_0^\infty f_S(x)xdx}{(1 - P(X_m(t) = S + 1))(1 - P(A))}. \tag{5.6}$$

5.2 Transient Model

For the transient model of multi-server queueing systems with stochastic impatient customers the model in section 4.1 will be applied to the model constructed in section 5.1. However, a slight modification will be needed since there is a state called I in which time does not evolve; so, in order to make our time approximations precise the time spent in state I must not be included to the approximation.

We start by defining a PH-type distribution of order k that is characterized by (α;S) with $S^0 = -Se$ where e is a column vector of ones of appropriate length. We will use this PH-type distribution for our time approximations. The fluid level process starts from a specified level $X_f(0) = a$ and background process from state j with probability π_j. We can use arbitrary a and π but for this model the most logical ones would be $a = 0$ and $\pi = [1 \quad 0 \quad 0 \quad 0 \quad \ldots \quad 0]$.

After the definitions above, we deploy the equations (4.4)-(4.8) to obtain the infinitesimal generator and drift matrices of $Y(t) = (Y_f(t), Y_m(t))$ which is the MRMFQ process for the transient approximation. The state space size of $Y_m(t)$ is $1 + k(S+1)$ where S is the number of servers. For this purpose we use $X(t)$, M/M/S+G model, which is characterized with the pair of matrices $\{Q_x(x), R_x(x)\}$ constructed from the models (5.1)-(5.3).

We mentioned before that for I which is the $S+1$th state of $X(t)$ a slight modification is needed to neglect the time spent at state I. The only thing we need to do is to block the state transitions from $(i, S+1)$ to $(j, S+1)$ where $i \neq j$, i.e., no transitions of PH-type distribution must happen when modulating process of $X(t)$ is in state I.

In order to achieve this we define a new diagonal matrix of size $S+1$

$$\tilde{I}_{S+1} = \textbf{diag}\{1, 1, \ldots, 1, 0\} \tag{5.7}$$

and a new vector of length $S+1$

$$\tilde{e}_{S+1} = \{1, 1, \ldots, 1, 0\} \tag{5.8}$$

$\tilde{I}_{S+1}$ will replace I_{S+1} which is an identity matrix of size $S+1$ and $\tilde{e}_{S+1}$ will replace e_{S+1} which is a vector of ones. Therefore, the infinitesimal generator of $Y(t)$ is

$$Q_Y(x) = \left[\begin{array}{c|c} 0 & 0 \\ \hline S^0 \otimes \tilde{e}_{S+1} & I_k \otimes Q_X(x) + S \otimes \tilde{I}_{S+1} \end{array}\right] \text{for } x \neq 0, \tag{5.9}$$

$$Q_Y(0) = \left[\begin{array}{c|c} -1 & \alpha \otimes \pi \\ \hline S^0 \otimes \tilde{e}_{S+1} & I_k \otimes Q_X(0) + S \otimes \tilde{I}_{S+1} \end{array}\right]. \tag{5.10}$$

And the drift matrix is

$$R_Y(x) = \begin{cases} \mathbf{diag}\{-1, I_k \otimes R_X(x)\} & \text{if } x > 0, \\ \mathbf{diag}\{0, I_k \otimes R_X(0)\} & \text{if } x = 0. \end{cases} \tag{5.11}$$

After obtaining the steady state solution of $Y(t)$ it is not difficult to modify equations (5.4)-(5.6) to find the metrics of interest. The numerical examples for this model will given in section 6 along with the simulation results.

Chapter 6

Numerical Examples

In this chapter, two numerical examples will be given in order to verify the transient and first passage time distribution method which we introduced in chapter 4 and transient solution of the M/M/S+G queue we obtained in chapter 5.

6.1 Numerical Example for the Transient and First Passage Distribution Method

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 [1] and [3]. 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 [31]. The probability of the source being in the ON state is 0.4 and each ON source emits traffic at a unity rate. 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 $\mathbf{X}_1(t)$ with $n = 66$ states. As in [3], 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. Erlang-k and CME-l distributions are used for approximating the deterministic time horizon in all the numerical examples. The complementary cdf (ccdf) of the buffer content

$$G_t^{0,\pi}(x) = \Pr\{X_f(t) > x | X_f(0) = 0, X_m(0) \sim \pi\} \tag{6.1}$$

for various values of x and for two values of $t = 100, 1000$ are tabulated in Tables 6.1 and 6.2, respectively, along with the numerical results reported in the work by [3]. The convergence with CME-l is rapid when compared to that obtained by Erlang-k and the results obtained with the particular CME-50 are in line with the results of [3] up to four digits.

We also modify the original MFQ in such a way that each ON source emits traffic at a rate of r_1=1.5 when the queue occupancy is less than a threshold $B_T = 50$ and it is kept unchanged at the rate of one above this threshold (r_2=1). This modification gives rise to an MFQ called $\mathbf{X}_2(t)$ with two regimes. Then to get more meaningful results for the first passage time distribution we applied the same modification with r_1=1.25 and called the MFQ as $\mathbf{X}_3(t)$. The ccdf $G_t^{0,\pi}(x)$ for the MFQ $\mathbf{X}_2(t)$ is presented in Tables 6.3 and 6.4 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. The results obtained with the particular CME-50 approximation are in line with simulation results up to four digits.

Table 6.1: Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 100$ with respect to varying x.

		Erlang-k		CME-l				
x	Ahn et. al.	$k = 50$	$k = 100$	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
0	0.8005439	0.8006440	0.8005932	0.8005748	0.8005507	0.8005466	0.8005453	0.8005447
10	0.5146490	0.5144087	0.5145341	0.5145469	0.5146283	0.5146410	0.5146448	0.5146465
20	0.4086868	0.4080080	0.4083574	0.4084831	0.4086421	0.4086687	0.4086772	0.4086809
30	0.3285634	0.3274791	0.3280335	0.3282759	0.3284980	0.3285364	0.3285490	0.3285546
40	0.2634434	0.2620412	0.2627537	0.2630950	0.2633622	0.2634095	0.2634252	0.2634322
50	0.2095419	0.2079325	0.2087444	0.2091570	0.2094505	0.2095034	0.2095212	0.2095291
60	0.1647741	0.1630696	0.1639226	0.1643756	0.1646780	0.1647334	0.1647521	0.1647605
70	0.1275597	0.1258657	0.1267065	0.1271688	0.1274644	0.1275191	0.1275378	0.1275461
80	0.0963395	0.0947629	0.0955402	0.0959784	0.0962507	0.0963015	0.0963189	0.0963267
90	0.0689034	0.0675842	0.0682333	0.0686029	0.0688292	0.0688717	0.0688862	0.0688927
100-	0.0261030	0.0256553	0.0258796	0.0260011	0.0260782	0.0260925	0.0260973	0.0260995

Table 6.2: Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 1000$ with respect to varying x.

		Erlang-k		CME-l				
x	Ahn et. al.	$k = 50$	$k = 100$	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
0	0.8292083	0.8292038	0.8292064	0.8292036	0.8292075	0.8292080	0.8292082	0.8292082
10	0.5859580	0.5859471	0.5859535	0.5859286	0.5859530	0.5859563	0.5859571	0.5859575
20	0.4930416	0.4930291	0.4930364	0.4930091	0.4930359	0.4930396	0.4930406	0.4930410
30	0.4195613	0.4195483	0.4195559	0.4195282	0.4195556	0.4195594	0.4195604	0.4195608
40	0.3562558	0.3562429	0.3562504	0.3562239	0.3562504	0.3562539	0.3562549	0.3562553
50	0.2999119	0.2998997	0.2999068	0.2998826	0.2999069	0.2999102	0.2999111	0.2999114
60	0.2488533	0.2488422	0.2488487	0.2488274	0.2488489	0.2488518	0.2488525	0.2488529
70	0.2018935	0.2018838	0.2018895	0.2018714	0.2018897	0.2018922	0.2018928	0.2018931
80	0.1578151	0.1578072	0.1578118	0.1577973	0.1578120	0.1578140	0.1578146	0.1578148
90	0.1142349	0.1142291	0.1142325	0.1142219	0.1142327	0.1142342	0.1142346	0.1142347
100-	0.0397938	0.0397920	0.0397931	0.0397895	0.0397931	0.0397936	0.0397937	0.0397938

Table 6.3: 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 a rate of 1.5 when $x < 50$.

		CME-l				
x	Simulation	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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

Table 6.4: Ccdf of the buffer content $G_t^{0,\pi}(x)$ for $t = 1000$ with respect to varying values of x when each ON source sends traffic at a rate of 1.5 when $x < 50$.

		CME-l				
x	Simulation	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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

We also tabulate the ccdf of first passage time

$$G_{\tau}^{0,b,\pi}(t) = \Pr\{\tau^{0,b,\pi} > t\} \tag{6.2}$$

for the MFQ $\mathbf{X}_1(t)$ for $t = 100$ and $t = 1000$, respectively, in Tables 6.5 and 6.6, with respect to varying values of the parameter b. The same ccdf of the first passage time for the MFQ $\mathbf{X}_3(t)$ is presented in Tables 6.7 and 6.8 for $t = 100$ and $t = 1000$, respectively.

Table 6.5: Ccdf of the first passage time $G_{\tau}^{0,b,\pi}(100)$ for varying values of b for the MFQ $\mathbf{X}_1(t)$.

		CME-l				
b	Simulation	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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 6.6: Ccdf of the first passage time $G_{\tau}^{0,b,\pi}(1000)$ for varying values of b for $\mathbf{X}_1(t)$.

		CME-l				
b	Simulation	$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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 6.7: Ccdf of the first passage time $G_{\tau}^{0,b,\pi}(100)$ for varying values of b for the MFQ $\mathbf{X}_3(t)$.

	Simulation	CME-l				
b		$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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

Table 6.8: Ccdf of the first passage time $G_{\tau}^{0,b,\pi}(1000)$ for varying values of b for the MFQ $\mathbf{X}_3(t)$.

	Simulation	CME-l				
b		$l = 10$	$l = 20$	$l = 30$	$l = 40$	$l = 50$
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

6.2 Numerical Example for the Transient M/M/S+G Method

In this section, the analytical results are compared to simulation results, in order to validate our analytical approach in Chapter 5. The packet arrival rate is 10 packets per second, the service rate is one, and the number of servers is 10. The packets abandon with rate 1 for the stochastic abandonement systems, In order to numerically derive the results for systems with stochastic abandonments, 1500 quantization levels are used for steady-state method and 100 used for transient method due to the increased complexity of the problem. This example leads to an original MFQ $\mathbf{X}_4(t)$ with $n = 11$ states. As defined before, 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 start from the first state where no customers are getting served at that moment. CME-10 distributions are used for the approximation of the deterministic time horizon in all the numerical examples. CME-10 was sufficient to approximate the time horizon accurately enough to obtain the first four digits of the metrics of interest converged. The zero wait probability $P(W = 0)$, probability of abandonment $P(A)$ and expected waiting time in the queue of a customer that didn't abandon $E[W|S]$ results are compared to those of simulations for performance evaluation and for the steady-state case exact results are also presented [18]. Steady-state result is tabulated in Table 6.9 and transient results at $t = 4, 8, 12, 16, 10000$ are tabulated in Table 6.10.

Table 6.9: Zero wait probability $P(W=0)$, probability of abandonment $P(A)$ and expected waiting time in the queue of a customer that didn't abandon $E[W|S]$ results for steady-state solution of $\mathbf{X}_4(t)$.

	Proposed Method	Exact Method [18]	Simulation Results
$P(W=0)$	0.4579	0.4579	0.4583 ± 0.0021
$P(A)$	0.1251	0.1251	0.1250 ± 0.0008
$E[W\|S]$	0.1149	0.1149	0.1149 ± 0.0008

Table 6.10: Zero wait probability $P(W=0)$, probability of abandonment $P(A)$ and expected waiting time in the queue of a customer that didn't abandon $E[W|S]$ results for transient solution of $\mathbf{X}_4(t)$ at $t=4,8,12,16$.

		$P(W=0)$	$P(A)$	$E[W\|S]$
Proposed Method	$t=4$	0.4820	0.1173	0.1072
	$t=8$	0.4585	0.1249	0.1148
	$t=12$	0.4580	0.1251	0.1149
	$t=16$	0.4580	0.1251	0.1149
Simulation	$t=4$	0.4854 ± 0.0005	0.1141 ± 0.0003	0.1046 ± 0.0002
	$t=8$	0.4585 ± 0.0004	0.1246 ± 0.0004	0.1147 ± 0.0002
	$t=12$	0.4581 ± 0.0007	0.1250 ± 0.0003	0.1148 ± 0.0002
	$t=16$	0.4583 ± 0.0005	0.1247 ± 0.0004	0.1148 ± 0.0002

Chapter 7

Conclusions

In this thesis, we studied on the transient analysis of multi-regime Markov fluid queues which is quite a useful tool for modeling queueing systems. In order to achieve the transient measures of interest, a numerical method is proposed which constructs a new auxiliary MRMFQ from the original one by introducing approximated time horizon with a PH-type distribution. Then, the transient solution is included in the steady-state distribution of the new MRMFQ. By using this method both transient and first passage time distributions are obtainable given the initial conditions of the fluid level and state distribution. We deployed additive decomposition to divide the system into constant, stable, unstable parts and block tri-diagonal LU factorization method in order to reduce the solution time. By following all of the steps described above, our method is able to solve transient MRMFQ problems with non-zero boundary probability masses, absorbing states and zero drifts. Also the fact that it is a treatment for not only SRMFQs but also MRMFQ, makes it applicable to wide variety of queuing problems.

In order to introduce the time horizon to the problem, first the Erlangization method is used where the PH-type distribution is an Erlang-k distribution with order k. As the order k increases, the model starts to make better approximations; but, the coefficient of variation of Erlang-k distribution is equal to $1/k$

which İS ineffective since the size of the model increases vastly for accurate approximations. Then, it is replaced by ME-fication where the Erlang is replaced by the Concentrated ME distribution denoted by CME-l whose coefficient of variation is approximately $1/2l^2$. We also compared these two types of distributions in our numerical examples.

In addition to the transient method, we also proposed a numerical method for the transient solution of multi-server queueing problem with generally distributed impatient customers with a M/M/S+G queueing model. This problem is modeled by using MRMFQ where the fluid level is the queue waiting time; but, an auxiliary state is introduced that does not evolve with time in order to introduce the abrupt changes at queue time with arrivals. Then the proposed transient method is modified for this MRMFQ to eliminate the auxiliary state from the transient measures.

We tried to prove our proposed models' validity by giving two numerical examples. First example is a 66 state MRMFQ/SRMFQ to prove the validity of the transient method which was previously studied in [1], [3]. The second one is an M/M/S+M model for which both transient and steady-state measures of interest are found. Both, numerical examples proved the validity and effectiveness of the proposed method.

For future work of this thesis, the proposed transient method can be applied to model practical and actual scenarios. As it is modified for the M/M/S+G case it can be implemented to other problems and its structure can be modified if necessary.

Bibliography

[1] N. Akar and K. Sohraby, "Infinite- and finite-buffer Markov fluid queues: a unified analysis," *Journal of Applied Probability*, vol. 41, no. 02, p. 557–569, 2004.

[2] H. Kobayashi and Q. Ren, "A mathematical theory for transient analysis of communication networks," *IEICE Trans. Commun*, vol. E-75B, no. 12, p. 1266–1276, 1992.

[3] S. Ahn, A. L. Badescu, and V. Ramaswami, "Time dependent analysis of finite buffer fluid flows and risk models with a dividend barrier," *Queueing Syst. Theory Appl.*, vol. 55, pp. 207–222, apr 2007.

[4] S. Asmussen, F. Avram, and M. Usabel, "Erlangian approximations for finite-horizon ruin probabilities," *Astin Bulletin*, vol. 32, no. 2, pp. 267–282, 2002.

[5] I. Horváth, O. Sáfár, M. Telek, and B. Zámbó, "Concentrated matrix exponential distributions," in *Computer Performance Engineering* (D. Fiems, M. Paolieri, and A. N. Platis, eds.), (Cham), pp. 18–31, Springer International Publishing, 2016.

[6] D. Stanford, F. Avram, A. Badescu, L. Breuer, A. D. S. Soares, and G. Latouche, "Phase-type approximations to finite-time ruin possibilities in the Sparre-Andersen and stationary renewal risk models," *Astin Bulletin*, vol. 35, no. 1, pp. 131–144, 2005.

[7] M. A. Yazici and N. Akar, "The workload-dependent MAP/PH/1 queue with infinite/finite workload capacity," *Perform. Eval.*, vol. 70, pp. 1047–1058, December 2013.

[8] V. Ramaswami, D. G. Woolford, and D. A. Stanford, "The Erlangization method for Markovian fluid flows," *Annals of Operations Research*, vol. 160, no. 1, p. 215–225, 2008.

[9] M. A. Yazici and N. Akar, "The finite/infinite horizon ruin problem with multi-threshold premiums: a Markov fluid queue approach," *Annals of Operations Research*, vol. 252, pp. 85–99, 2017.

[10] S. Ahn and V. Ramaswami, "Transient analysis of fluid flow models via stochastic coupling to a queue," *Stochastic Models*, vol. 20, no. 1, p. 71–101, 2004.

[11] C. Tunc, "Energy management in energy harvesting wireless sensor nodes with lifetime constraints," Master's thesis, Ihsan Doğramacı Bilkent University, 2016.

[12] F. Baccelli and G. Hebuterne, "On queues with impatient customers," *in: Performance '81*, vol. 32, 1981.

[13] A. Movaghar, "On queueing with customer impatience until the beginning of service," *Queueing Systems*, vol. 29, no. 2, pp. 337–350, 1998.

[14] A. Brandt and M. Brandt, "Asymptotic results and a markovian approximation for the M(n)/M(n)/S+GI system," *Queueing Systems*, vol. 41, no. 1/2, p. 73–94, 2002.

[15] A. Brandt and M. Brandt, "On the M(n)/M(n)/S queue with impatient calls," *Performance Evaluation*, vol. 35, no. 1-2, p. 1–18, 1999.

[16] A. Mandelbaum and S. Zeltyn, "The impact of customers? patience on delay and abandonment: some empirically-driven experiments with the M/M/ N G queue," *OR Spectrum*, vol. 26, p. 377–411, Jan 2004.

[17] W. Whitt, "Engineering solution of a basic call-center model," *Management Science*, vol. 51, no. 2, p. 221–235, 2005.

[18] S. Zeltyn and A. Mandelbaum, "Call centers with impatient customers: Many-server asymptotics of the M/M/N G queue," *Queueing Systems*, vol. 51, no. 3-4, p. 361–402, 2005.

[19] S. Asmussen, F. Avram, and M. Usabel, "Erlangian approximations for finite-horizon ruin probabilities," *ASTIN Bulletin*, vol. 32, no. 02, p. 267–281, 2002.

[20] D. Aldous and L. Shepp, "The least variable phase type distribution is Erlang," *Communications in Statistics. Stochastic Models*, vol. 3, no. 3, p. 467–473, 1987.

[21] M. A. Yazici, *Stochastic Modelling with Continious Feedback Markov Fluid Queues*. PhD thesis, Ihsan Doğramacı Bilkent University, 2014.

[22] H. E. Kankaya and N. Akar, "Solving multi-regime feedback fluid queues," *Stochastic Models*, vol. 24, pp. 425–450, July 2008.

[23] A. Badescu, S. Drekic, and D. Landriault, "On the analysis of a multi-threshold Markovian risk model," *Scandinavian Actuarial Journal*, vol. 2007, no. 4, pp. 248–260, 2007.

[24] A. da Silva Soares and G. Latouche, "Fluid queues with level dependent evolution," *European Journal of Operational Research*, vol. 196, no. 3, pp. 1041 – 1048, 2009.

[25] M. Mandjes, D. Mitra, and W. Scheinhardt, "Models of network access using feedback fluid queues," *Queueing Syst. Theory Appl.*, vol. 44, no. 4, pp. 2989–3002, 2003.

[26] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*. Baltimore, MD: Johns Hopkins University Press, 1981.

[27] M. W. Fackrell, *Characterization of Matrix-exponential Distributions*. PhD thesis, The University of Adelaide, 2003.

[28] Q.-M. He and H. Zhang, "On matrix exponential distributions," *Advances in Applied Probability*, vol. 39, no. 1, pp. pp. 271–292, 2007.

[29] V. Ramaswami, D. G. Woolford, and D. A. Stanford, "The Erlangization method for markovian fluid flows," *Annals of Operations Research*, vol. 160, no. 1, pp. 215–225, 2008.

[30] C. Tunc and N. Akar, "Performance modeling of delay-based dynamic speed scaling," in *The Ninth International Conference on Matrix-Analytic Methods in Stochastic Models (MAM9), Budapest, Hungary*, June 2016.

[31] H. Tijms, *Stochastic models: an algorithmic approach.* Wiley series in probability and mathematical statistics, John Wiley & Sons, 1994.