Stochastic Modelling: Analysis and Applications

QUEUES WITH CUSTOMER INDUCED INTERRUPTION

Thesis submitted to the
Cochin University of Science and Technology
for the award of the degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science by

VARGHESE JACOB

Department of Mathematics
Cochin University of Science and Technology
Cochin - 682 022

JUNE 2012

Certificate

This is to certify that the thesis entitled ‘ **Queues with Customer Induced Interruption** ’ submitted to the Cochin University of Science and Technology by Mr. Varghese Jacob for the award of the degree of Doctor of Philosophy under the Faculty of Science is a bonafide record of studies carried out by him under our supervision in the Department of Mathematics, Cochin University of Science and Technology. This report has not been submitted previously for considering the award of any degree, fellowship or similar titles elsewhere.

Dr. M. N. Narayanan Namboothiri
(Co-guide)
Assistant Professor
Department of Mathematics
Cochin University of Science
and Technology
Kochi - 682 022, Kerala

Dr. A. Krishnamoorthy
(Research Guide)
Emeritus Professor
Department of Mathematics
Cochin University of Science
and Technology
Kochi - 682 022, Kerala.

Cochin-22
16-6-2012.

Declaration

I, Varghese Jacob, hereby declare that this thesis entitled '**Queues with Customer Induced Interruption**' contains no material which had been accepted for any other Degree, Diploma or similar titles in any University or institution and that to the best of my knowledge and belief, it contains no material previously published by any person except where due references are made in the text of the thesis.

Varghese Jacob
Research Scholar
Registration No. 3376
Department of Mathematics
Cochin University of Science and Technology
Cochin-682 022, Kerala.

Cochin-22
16-06-2012.

To

My Mother

and in loving memory of

my Father

Acknowledgement

'Praise the LORD, my soul! All my being, praise his holy name!
Praise the LORD, my soul, and do not forget how kind he is.'
(Psalm 103:1-2)

With this prayer in mind, I bow my head before the God Almighty for His blessings.

It gives me immense pleasure to thank all my teachers, relatives and friends who directly or indirectly helped me in my research work.

The greatest thanks I owe is to my supervisor and guide Prof. Dr. A. Krishnamoorthy, Emeritus Professor, Department of Mathematics, Cochin University of Science and Technology for the valuable guidance and support extend to me. I am indebted to the bright ideas, his timely help, keen guidance and criticism with love and affection helped me a lot to complete this thesis at the right time. In addition, he is always accessible and willing to help his students with their research. As a result, research life became smooth and rewarding for me.

I owe a special debt of gratitude to Prof. S. R. Chakravarthy, Kettering University, Flint, USA, who have always been a constant source of encouragement during my research study. He offer advice and suggestions whenever I need them. He had inspired me in research and life through our interactions during his visit at Cochin.

I was delighted to interact with Prof. Alexander. N. Dudin, Belarusian State University, Minsk for spending his valuable time for the discussions with us and also for permitting to include our joint work in this thesis.

I am grateful to my co-guide Dr. M. N. Narayanan Namboothiri for his valuable help and advices.

A special word of thanks to Dr. A. Vijayakumar, Head and my teachers, Prof. T. Thrivikraman, Dr. M. Jathavedan, Dr. M. K. Ganapathy, Dr. R. S. Chakravarti, of Department of Mathematics, CUSAT, for their blessings and the valuable help rendered for materializing this attempt. I also thank other faculty members Dr. B. Lakshmi, Dr. P.G. Romeo, Ms. Meena, Mr. Varun Soman and office staff of the Department of Mathematics for their help of various kinds. My gratitude also goes to the authorities of CUSAT for the facilities they provided. I gratefully acknowledge the UGC for providing me Teacher Fellowship under the Faculty Improvement Programme during 11^{th} plan. I am also grateful to the Director of Collegiate Education, Government of Kerala for sanctioning me deputation to avail FIP of UGC.

My sincere thanks to my best friend Dr. Jacob T.M., Asso. Professor, Nirmala College, Muvattupuzha for his inspiration and help during my studies from M.Sc. to research. My special thanks to Dr. Deepak T.G, Asst. Prof., IIST, Trivandrum for his valuable suggestions to start my research work.

I express my feeling of gratitude to my friends Dr. Sasi Gopalan, Mr. Asokan C.K, Dr. Sindhu G. Nair, Dr. G. Indulal, Dr. Vinodkumar P.B., Dr. Sunoj S.M, Dr. Paul Issac, Dr. Seema Varghese. Also I want to express my warmest thanks to my fellow research scholars Mr. Kirankumar V.B, Mr. Pravas K, Mr. Tijo James, Mr. Jayaprasad P.N, Mr. Didimos K.V, Mr. Vivek P.V, K.V, Ms. Anu Varghese, Ms. Pamy Sebastian, Ms. Raji George, Ms. Chithra M. R, Ms. Vineetha, Ms. Seethu Varghese, Mr.

Tonny K.B, Mr. Tibin Thomas, Mr. Gireesan K.K.

I extend my heartfull thanks to friends in my research area Dr. S. Babu, Dr. Varghese C. Joshua, Dr. Viswanath C. N, Dr. Jose K.P, Dr. Pramod P.K, Dr. Ajayakumar C.B, Dr. Lalitha, Mr.C. Sreenivasan, Ms. Deepthi C.P, Mr. Manikandan R, Mr. Manjunath A.S, Ms. Binitha Benny, Ms. Anusha K.K, Mr. Sajeev S Nair, Mr. Gopakumar, Mr. Sathian, Ms. Resmi Varghese, Ms. Jaya S, Ms. Dhanya, Mr. Jaison Jacob, Mr. Vijayagovindan for their interest in my research work and they were always ready to share their bright ideas with me.

I also thank the Principal and staff of Government College, Kottayam for their support and encouraging words. I am also thankful to all my colleagues, especially Prof. V. Jayalal, Prof. Vinayachandran and Prof. Usha K. for their whole hearted support that they kept for me becomes a driving force behind my research work. Also I express my sincere thanks to my former colleagues at St. Joseph's College, Devagri, Calicut and Govt. Polytechnic College, Vennikulam.

It is also not good to forget the love, prayers and blessings of my beloved family members especially my mother Marykutty Jacob, brothers, sister and Parents in-law also have its role to play with in formulating the right spirit behind this move.

And last, but certainly not least, I express my thanks to my wife Ms. Sobha Mary Varkey and my sons Sanjo and Sanay for their prayers, love, understanding and support. But words are not enough to express my feelings towards them.

Varghese Jacob

QUEUES WITH CUSTOMER INDUCED INTERRUPTION

Contents

List of Tables

List of Figures

Notations used

- $\boldsymbol{e}$ denotes column vector of 1's with appropriate dimension;
- $\mathbf{0}$ is a vector consisting of 0's with appropriate dimension;
- $\boldsymbol{e}_j(r)$ denotes column vector of dimension r with 1 in the j^{th} position and 0 elsewhere;
- $\boldsymbol{a}_i = (1, 2, \ldots, i);\ 1 \leqslant i \leqslant K;$
- $\Delta(\boldsymbol{a}_{_i})$ is a diagonal matrix whose diagonal entries are the components of the vector $\boldsymbol{a}_i$;
- $diag\{A_1, \ldots, A_N\}$ is the diagonal matrix having diagonal entries listed in the brackets;
- O is a zero matrix with appropriate dimension;
- I_r denotes identity matrix of dimension r;
- I is an identity matrix of appropriate dimension;

- $A \otimes B$ denotes the Kronecker product of the matrices A and B : $A \otimes B = (a_{ij}B)$

- $A \oplus B$ denotes the Kronecker sum of the matrices A and B (see, e.g., *Graham* [30]) : $A \oplus B = A \otimes I_n + I_m \otimes B$ where A and B are square matrices of order m and n, respectively.

- $sp(R)$ denotes the spectral radius of R;

- $\delta_{i,l}$ is Kronecker delta; that is $\delta_{i,l}$ is equal to 1, if $i = l$ and equal to 0 otherwise;

- $\chi(u)$ is Heavyside function, $\chi(u) = 1$, if $u > 0$ and equal to 0 otherwise;

- $A^{\otimes l} = \underbrace{A \otimes \ldots \otimes A}_{l}$, $A^{\otimes 0} = 1$;

  For the matrix A having M rows,

- $A^{\oplus l} = \sum\limits_{m=0}^{l-1} I_{M^m} \otimes A \otimes I_{M^{l-m-1}}$, $l \geqslant 1$, $\quad A^{\oplus 0} = 0$;

Abbreviation used

PH	:	Phase type;
MAP	:	Markovian Arrival Process;
$CTMC$	:	Continuous-time Markov Chain;
$FIFO$	:	First In First Out;
QBD	:	Quasi-birth-death;
LST	:	Laplace-Stieltjes Transform;
$LIQBD$	:	Level Independent Quasi-Birth-Death;
$LDQBD$	:	Level Dependent Quasi-Birth-Death;
BIP	:	Buffer for Interruption Process;
BIC	:	Buffer for Interruption Completion;
ETP	:	Expected Total Profit;

Chapter 1

Introduction

Congestion is a natural phenomenon in real systems. A service facility gets congested if there are more people than the server can possibly handle. Thus Queueing theory deals with mathematical techniques for analyzing congestion problems. Some examples are : customers arrive at a bank and wait to have certain monetary transactions; queues outside doctor's clinic; super market check out counters; airplanes queue along airport runways; patients have to wait for an operation because either surgeon is busy or there are not enough hospital beds; signals are sent through switches and wait to be transmitted. The main components of a queueing system are the entities that require service, often called customers or jobs, the entities that provide service, usually called servers, and one or more queues. Customers are often external to the service system. Customers who require service are said to arrive at the service facility. At a doctors clinic, the waiting line consists of the patients awaiting their turn to see the doctor; the doctor is the server who dispenses a limited resource, his time.

Our ability to model and analyze systems of queues helps to minimize their inconveniences and maximize the use of the limited resources. An analysis may tell us something about the expected time that a resource will be in use, or the expected time that a customer must wait. This information may then be used to make decisions as to when and how to upgrade the system: for an overloaded doctor to take on an associate.

In this thesis a few queueing models are analyzed by means of continuous time Markov chains in which we use the modelling tools such as Markovian Arrival Process (MAP) and Phase type distributions (PH distributions). Algorithmically tractable tools like these help us to model and analyze the structures so obtained in a very general setup. For example a MAP introduce correlation in the arrival process. The resulting quasi-birth-death processes are solved algorithmically by Matrix Analytic Method.

1.1 Phase Type distribution (Continuous time)

The exponential distribution is widely used in queueing models because of the exceptional mathematical tractability that flows from the memoryless property of this distribution. However, in applications these assumptions are highly restrictive. This lead us to explore ways in which we can model more general distributions while maintaining some of the tractability of the exponential distribution. Thus, M. F. Neuts developed the theory of PH distributions and related point processes. A PH distribution is obtained as the distribution of the time until absorption in a finite state space Markov chain with an absorbing state.

Consider a Markov process $\mathcal{X} = \{X(t) : t \geqslant 0\}$ with finite state space $\{1, 2, \dots, m+1\}$ and the infinitesimal generator matrix

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ m \\ m{+}1 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & \dots & m & m+1 \end{array} \\ \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1m} & T_{1m+1} \\ T_{21} & T_{22} & \dots & T_{2m} & T_{2m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ T_{m1} & T_{m2} & \dots & T_{mm} & T_{mm+1} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \end{array} = \begin{pmatrix} T_{m\times m} & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{pmatrix}$$

where the elements of the matrices T and $\mathbf{T}^0$ satisfy $T_{ii} < 0$ for $1 \leqslant i \leqslant m$, $T_{ij} \geqslant 0$ for $i \neq j$; $T_i^0 \geqslant 0$ and $T_i^0 > 0$ for at least one i, $1 \leqslant i \leqslant m$ and $T\boldsymbol{e} + \mathbf{T}^0 = \mathbf{0}$.

Also the initial distribution of $\mathcal{X}$ is given by $(\boldsymbol{\alpha}, \alpha_{m+1})$ with the property that $\boldsymbol{\alpha}\boldsymbol{e} + \alpha_{m+1} = 1$. Here the states $1, 2, \dots, m$ are called phases and $m+1$ is called absorbing phase.

Let $Z = \inf\{t \geqslant 0 : X(t) = m+1\}$ be the time until absorption in state $m+1$. Then the distribution of Z is called PH distribution with representation $(\boldsymbol{\alpha}, T)$. The dimension m is called the order of the distribution.

(i) The distribution function of Z is given by

$$F(t) = 1 - \boldsymbol{\alpha}\ exp(T.t)\ \boldsymbol{e} \ \text{ for every } t \geqslant 0.$$

It has a jump of magnitude α_{m+1} at $t = 0$ and its density function is given by

$$f(t) = \boldsymbol{\alpha}\ exp(T.t)\ \mathbf{T}^0 \ \text{ for every } t > 0$$

where the function $exp(T.t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} T^i$, the matrix exponential function and

(ii) the Laplace-Stieltjes transform (LST) of $F(.)$ is given by

$$\phi(s) = \alpha_{m+1} + \boldsymbol{\alpha}(sI - T)^{-1}\ \mathbf{T}^0 \ \text{ for } Re(s) \geqslant 0.$$

Theorem 1.1.1 (see, *Latouche and Ramaswami* [43])**.** *Consider a PH distribution* $(\boldsymbol{\alpha}, T)$*. Absorption into state* $m+1$ *occurs with probability 1 from any phase* i *in* $\{1, 2, \ldots, m\}$ *if and only if the matrix T is nonsingular.*

More over, $(-T^{-1})_{i,j}$ *is the expected total time spent in phase* j *during the time until absorption, given that the initial phase is* i*.*

For more information about the PH distribution, see, e.g., *Neuts* [54]. Usefulness of PH distribution as service time distribution in telecommunication networks is elaborated, e.g., in *Pattavina and Parini* [56] and *Riska, Diev and Smirni* [57].

1.2 Quasi-birth-death processes

Consider a Markov Chain with state space $S = \bigcup\limits_{n \geqslant 0} \{(n, i) : 1 \leqslant i \leqslant m\}$. Here the first component n is called level of the Chain and the second component i is called a phase of the n^{th} level. The Markov Chain is called a Quasi-birth-death (QBD) process if the one step transitions from a state is restricted to the same level or to the two adjacent levels. In other words,

$$(i-1, j') \rightleftharpoons (i, j) \rightleftharpoons (i+1, j'') \ \text{ for } i \geqslant 1$$

If the transition rates are level independent, the resulting QBD process is called level independent quasi-birth-death process ($LIQBD$); else it is called level dependent quasi-birth-death process ($LDQBD$).

Arranging the elements of S in lexicographic order, the infinitesimal generator of a $LIQBD$ process has the block tridiagonal matrix form in which three diagonal blocks repeat after some initial levels. We write such a matrix, with modification depending on boundary states, as

$$Q = \begin{bmatrix} B_1 & A_0 & & & \\ B_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \tag{1.1}$$

where the sub matrices A_0, A_1, A_2 are square and have the same dimension; matrix B_1 is also square and need not have the same size as A_1. Also, $B_1\boldsymbol{e} + A_0\boldsymbol{e} = B_2\boldsymbol{e} + A_1\boldsymbol{e} + A_0\boldsymbol{e} = (A_0 + A_1 + A_2)\boldsymbol{e} = \boldsymbol{0}$.

1.3 Matrix Analytic Method

Matrix analytic method, introduced by M.F. Neuts and studied by several researchers in late 1970's, establish a success story, illustrating the enrichment of science and applied probability. Matrix Analytic Method is a tool to construct and analyze a vide class of stochastic models, particularly queueing systems, using a matrix formalism to develop algorithmically tractable solution.

Theorem 1.3.1 (see, Theorem 3.1.1. of *Neuts* [54])**.** *The process Q in (1.1) is positive recurrent if and only if the minimal non-negative solution $\mathcal{R}$ to the matrix-quadratic equation*

$$\mathcal{R}^2 A_2 + \mathcal{R} A_1 + A_0 = O \tag{1.2}$$

has all its eigenvalues inside the unit disk and the finite system of equations

$$\begin{aligned} \boldsymbol{x}_0 \left(B_1 + \mathcal{R} B_2 \right) &= \boldsymbol{0} \\ \boldsymbol{x}_0 (I - \mathcal{R})^{-1} \boldsymbol{e} &= 1 \end{aligned} \tag{1.3}$$

has a unique positive solution $\boldsymbol{x}_0$.

If the matrix $A = A_0 + A_1 + A_2$ is irreducible, then $sp(\mathcal{R}) < 1$ if and only if

$$\boldsymbol{\pi} A_2 \boldsymbol{e} > \boldsymbol{\pi} A_0 \boldsymbol{e} \tag{1.4}$$

where $\boldsymbol{\pi}$ is the stationary probability vector of A.

The stationary probability vector $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots)$ *of* Q *is given by*

$$\boldsymbol{x}_i = \boldsymbol{x}_0 \mathcal{R}^i \quad \textit{for} \ \ i \geqslant 1. \tag{1.5}$$

Once $\mathcal{R}$, the rate matrix, is obtained, the vector $\boldsymbol{x}$ can be computed. We can use an iterative procedure or logarithmic reduction algorithm (see, *Latouche and Ramaswami* [42]) or the cyclic reduction algorithm (see, *Bini and Meini* [7]) for computing $\mathcal{R}$.

1.4 Computation of $\mathcal{R}$ matrix

There are several algorithms for computing rate matrix $\mathcal{R}$. Here we list two of them.

1.4.1 Iterative algorithm

From (1.2), we can evaluate $\mathcal{R}$ in a recursive procedure as follows.

Step 0: $\mathcal{R}(0) = \mathrm{O}$.

Step 1:

$$\mathcal{R}(n+1) = A_0(-A_1)^{-1} + \mathcal{R}^2(n) A_2 (-A_1)^{-1}, \quad n = 0, 1, \ldots$$

Continue **Step 1** until $\mathcal{R}(n+1)$ is close to $\mathcal{R}(n)$.

That is, $||\mathcal{R}(n+1) - \mathcal{R}(n)||_\infty < \epsilon$.

1.4.2 Logarithmic reduction algorithm

Logarithmic reduction algorithm is developed by *Latouche and Ramaswami* [42] which has extremely fast quadratic convergence. This algorithm is considered to be the most efficient one. We will list only the main steps involved in the logarithmic reduction algorithm. For full details on the logarithmic reduction algorithm refer *Latouche and Ramaswami* [42].

Step 0: $H \leftarrow (-A_1)^{-1}A_0$, $L \leftarrow (-A_1)^{-1}A_2$, $G = L$, and $T = H$.

Step 1:

$$U = HL + LH$$

$$M = H^2$$

$$H \leftarrow (I - U)^{-1}M$$

$$M \leftarrow L^2$$

$$L \leftarrow (I - U)^{-1}M$$

$$G \leftarrow G + TL$$

$$T \leftarrow TH$$

Continue **Step 1:** until $||\boldsymbol{e} - G\boldsymbol{e}||_\infty < \epsilon$.

Step 2: $\mathcal{R} = -A_0(A_1 + A_0G)^{-1}$.

1.5 Markovian Arrival Process

Markovian Arrival Process (MAP) was introduced in *Lucantoni* [45]. It is a rich class of point processes that includes many well-known processes such as Poisson, PH-renewal processes and Markov-modulated Poisson process. One of the most significant features of a MAP is the underlying Markovian structure and fits ideally in the context of matrix-analytic solutions to stochastic models. The idea of MAP is to significantly generalize Poisson processes and still keep the tractability for modelling purposes. In [10], *Chakravarthy* provides an extensive survey of the Batch Markovian Arrival Process ($BMAP$) in which arrivals are in batches where as it is in singles in a MAP. Currently, MAP (see, e.g., *Heyman and Lucantoni* [32]), is the most popular mathematical model for the telecommunication networks traffic because it catches the typical features of this traffic such as correlation and burstiness. Furthermore, in many practical applications, notably in communication engineering, production and manufacturing engineering, the arrivals do not usually form a renewal process. So, MAP is a convenient tool to model both renewal and non-renewal arrivals. While MAP is defined for both discrete and continuous time and MAP in continuous time is described as follows.

In MAP the customers arrival is directed by an irreducible continuous time Markov chain ν_t, $t \geqslant 0$, with the state space $\{1, \ldots, r\}$. Let Q^* be the generator of this Markov chain. At the end of a sojourn time in state i, that is exponentially distributed with parameter λ_i, one of the following two events could occur: with probability $p_{ij}(1)$ the transition corresponds to an arrival and the underlying Markov chain is in state j with $1 \leqslant i, j \leqslant r$; with probability $p_{ij}(0)$ the transition corresponds to no arrival and the state of the Markov chain is j, $j \neq i$. Note that the Markov chain can go from state i to state i only through an arrival. Also we have

$$\sum_{j=1}^{r} p_{ij}(1) + \sum_{j=1, j\neq i}^{r} p_{ij}(0) = 1, \quad 1 \leqslant i \leqslant r.$$

Define the matrices $D_0 = (d_{ij}^0)$ and $D_1 = (d_{ij}^1)$ such that $d_{ii}^0 = -\lambda_i, 1 \leqslant i \leqslant r, d_{ij}^0 = \lambda_i p_{ij}(0),$ for $j \neq i$ and $d_{ij}^1 = \lambda_i p_{ij}(1), 1 \leqslant i, j \leqslant r$.

By assuming D_0 to be a nonsingular matrix, the inter-arrival times will be finite with probability one and the arrival process does not terminate. Hence, we see that D_0 is a stable matrix. The generator Q^* is then given by $Q^* = D_0 + D_1$. Thus D_0 governs the transitions corresponding to no arrival and D_1 governs those corresponding to an arrival. Vector $\boldsymbol{\theta}$ of the stationary distribution of the process ν_t is the unique solution to the system $\boldsymbol{\theta}(D_0 + D_1) = \mathbf{0}$, $\boldsymbol{\theta}\mathbf{e} = 1$. Then the constant $\lambda = \boldsymbol{\theta} D_1 \mathbf{e}$ is called fundamental rate of a MAP which gives the expected number of arrivals per unit time in the stationary version of a MAP.

1.6 Motivation of work

In classical queueing models, servers are always available to serve customers. However, in many practical queueing systems, servers may become unavailable for a period of time due to variety of reasons such as

1. services get interrupted due to failure of the servers, (see, e.g., *Avi-Itzhak and Naor* [5], *Takine and Sengupta* [61], *Gaver* [26], *Neuts and Lucantoni* [51], *Krishnamoorthy, Pramod and*

Deepak [37], *Choudhury and Tadj* [17], *Takagi* [59, 60], *Wang* [64], *Atencia and Moreno* [4], *Sherman and Kharoufeh* [49]);

2. waiting for servers to get back from vacation, (see, e.g., *Doshi* [19, 20], *Krishna Kumar and Madheswari* [39], *Krishna Kumar* [40], *Tian and Zhang* [62]);

3. removed from the system due to catastrophic events or negative arrivals, (see, e.g., *Chakravarthy, Dudin and Klimenlok.* [14], *Dudin and Semenova* [21], *Artalejo* [1], *Klimenok and Dudin* [36]);

4. getting pre-empted due to the arrivals of higher priority customers, (see, e.g., *White and Christie* [65], *Jaiswal* [33]).

We refer a recent work by *Krishnamoorthy, Pramod and Chakravarthy* [38] for details on queues with interruptions.

So far all work reported deal with cases in which service interruption are generated by sources other than customers. In this thesis we introduce a new type of service interruption that is induced by the customers while undergoing service. A customer who is currently in service can be (self) interrupted.

The motivation for such interruptions arise from situations seen in practice. A more commonly occurring example is the following : In a doctors clinic, while a patient is being examined, the physician may find that one or more tests are needed for prescription of medicine. Hence he/she is asked to undergo these and return to the clinic. Such patients can be regarded as self-interrupted customer. While trying to send a package in post office, the customer may require a service such as confirmation of the delivery of the package which calls for the customer to fill in certain proforma (if it is not done earlier). This is to be done by moving away from the server to a place that has all relevant forms. Another example is the case of the customer getting interrupted through his/her cell phone which needs immediate attention. Other examples can be found in Online services.

Salient features of this type of interruption as opposed to server interruptions are that the system (a) can have more interrupted customers than the number of servers in the system, and (b) can offer services to other customers while a/some customers is/are undergoing interruptions irrespective of whether the system is run by a single or several servers. Case (a) arises in system induced interruption as well in the retrial setup (see, *Sherman and Kharoufeh* [49])

1.7 Summary of the thesis

As indicated earlier this thesis is on customer induced interruption. This is motivated by real life situations, though in no problem discussed, we go for specifics. From the time we started the work on customer induced interruption more and more examples started revealing the need of such a study.

In this thesis the queueing models are analyzed by identifying $CTMC$ as a consequence of the use of the modelling tools such as a MAP and a PH distributions. Numerically tractable tools like these help us to model and analyze the structures so obtained in a very general setup. For example a MAP introduces correlation in the arrival process. The resulting QBD processes are solved algorithmically by Matrix Analytic Method.

Now we turn to the content of the thesis. The thesis entitled "Queues with Customer Induced Interruption" is divided into 5 chapters including this introductory chapter. In chapters 2 to 5 the systems under study are queueing systems with Customer induced interruption.

In chapter 1, PH distributions, Markovian Arrival Process (MAP), quasi-birth-death (QBD) process are formally defined and their representation and properties discussed. Also we provide a brief description of the computation of rate matrix $\mathcal{R}$ using iterative method and logarithmic reduction algorithm.

Chapter 2 discusses a single server queueing system with Customer induced interruption while in service. We consider an infinite capacity queueing system to which customers arrive according to a Poisson process and the service time follows an exponential distribution. The Customer Induced Interruption (Self Interruption) while in service occurs according to a Poisson process and the interruption duration follows an exponential distribution. The self interrupted customers enter into a finite buffer of size K (referred to as BIP). Any interrupted customer, finding the buffer full, is considered lost. Those interrupted customers who complete their interruptions are placed into another buffer (referred to as BIC) of the same size . The interrupted customers waiting for service are given ' non-preemptive priority ' over new (primary) customers. That is, a customer in service will be completely or until it gets interrupted, served before this BIC customer is taken for service. We investigate the behavior of this queueing system. The mean waiting time in the queue of an arriving customer by deriving its LST and waiting time distribution of a primary customer since taken for service as primary unit, are derived. Several performance measures are evaluated. Numerical illustrations of the system behavior are also provided. An optimization problem of interest are discussed through an illustrative example.

In Chapter 3, we discuss c server queueing system with Customer induced interruption which is an extension of the model described in Chapter 2. All c servers are assumed to be homogeneous and that the service times follows exponential distribution. Under the same assumption as in Chapter 2, service of a primary customer with any of the c servers can get interrupted and it enters into BIP, should there be a space available. Otherwise, such a customer is lost for ever. Customers in the BIP, upon completion of their interruption, enter into BIC. Customers in BIC are given non-preemptive priority over the primary customers. Duration of each interruption and repair/fixing time follows exponential distribution. Here also we assume that at most one interruption is allowed for a primary customer. The mean waiting time in the queue of an arriving customer is obtained through the LST and several performances measure are evaluated. Numerical illustrations of the system behavior are also provided. Optimization problem to maximize the revenue with respect to number of servers to be employed and optimal buffer size for the self-interrupted customers are discussed through two illustrative examples.

In the fourth chapter, a queueing model consisting of two multi-server service systems is considered. Primary customers arrive at a multi-server queueing system-1 having an infinite buffer. The input flow is described by a MAP. The service time of a primary customer has a PH distribution. The service of a primary customer can be interrupted and arrival of interruptions are described by a MAP. An interruption deletes a primary customer from the service if the state (phase) of its PH service process does not belong to a given set of protected phases. The interrupted customer leaves the system permanently with some probability. With complementary probability, the interrupted primary customer moves for service to system-2 (BIP). This system consists of K independent identical servers and has no buffer. If all K servers are busy at the moment of a primary customer's service interruption, this customer will be lost. Otherwise, this primary customer starts the service with an arbitrary idle server of system-2. It is assumed that the service time of a primary customer by a server of system-2 has PH distribution. Upon completion of the service at system-2, the customer becomes priority customer. No more than one knock out of primary customer by negative customers is only allowed. If, at the service completion moment, there are free servers at system-1, the priority customer immediately starts getting the service at system-1. It is assumed that the service time of a priority customer by a server of system-1 has PH distribution and this service can not be interrupted. If, at the moment when the primary customer finished the service at system-2, there are no idle servers at system-1, this customer is placed into the finite buffer for priority customers BIC of capacity K. When a server of system-1 becomes free, it takes for service a priority customer from BIP, if any. Primary customers are picked-up from the infinite buffer only if BIC is empty at the service completion moment at system-1. Behavior of this system is described by a multi-dimensional Markov chain. Algorithms for checking ergodicity condition and computing the stationary distribution are presented. Formulas for computing important performance measures of the system are derived. Some illustrative examples are discussed in the Numerical section.

Chapter 5 analyses an $M/PH/1$ queue with Customer induced interruption and retrial of interrupted customers. The interruption occurs according to a Poisson process. As in earlier chapters we assume that no more than one interruption is allowed for a customer while in service. In this model, when an interruption occurs, the interrupted customer enters into an orbit of finite capacity, K, should there be a space available and from there it retries for service after the interruption is completed. In the case that the orbit is full, an interrupted customer joining the orbit is blocked and is forced to leave the system for ever. On the other hand, when an orbital customer retries and find that the server is still busy, he returns to the orbit. In the models in chapters 2 to 4, we assume a non-preemptive priority for the interrupted customers where as in chapter 5, interruption of service by customers is not encouraged and as a consequence those violating this dictum are punished by way of their being sent to orbit for retrial. Further such customers cannot access the server as long as there are primary customers in the system. This is in sharp contrast to *Sherman and Kharoufeh* [49] where customer service gets interrupted due to system breakdown. Such customers are then to go to an orbit of infinite capacity to retry for service. Thus customers are punished for system fault. The service times of primary and orbital customers are assumed to follow distinct PH distributions. The inter-retrial times of each orbital customer follows an exponential distribution. We derive an explicit expression for the stability condition of the system. We also obtain explicit expression for the rate matrix $\mathcal{R}$ in the $M/M/1$ setup. The long run behavior of the system is studied and several system performance measures are obtained. Numerical illustration are also provided.

Finally a section of concluding remarks and some suggestions for future study are included.

Chapter 2

On Customer Induced Interruption in a Single Service System

Service interruption in a queueing systems is a common phenomenon; there is an extensive literature on this. For example, in a production and manufacturing set up the machine (offering services to jobs) can fail in the middle of a service due to wear and tear of the tool. We refer a recent work by *Krishnamoorthy, Pramod and Chakravarthy* [38] for details on queues with interruptions. So far all work on interruptions deal with server induced ones. In this chapter and those to follow we introduce a

Some results of this chapter are included in the following paper.
Varghese Jacob, Chakravarthy, S. R. and Krishnamoorthy, A. : On a customer induced interruption in a service systems. Journal of Stochastic Analysis and Applications, 30, 1–13, 2012. Taylor & Francis, USA.

new type of interruption that is induced by the customers while in service. A customer who is currently in service can be self-interrupted. Salient features of this type of interruption as opposed to server interruptions are that the system (a) can have more interrupted customers than the number of servers in the system, and (b) can offer services to other customers while a/some customers is/are undergoing interruptions irrespective of whether the system is manned by a single or more than one servers. Case (a) arises in system induced interruption as well in the retrial setup (see, *Sherman and Kharoufeh* [49])

This chapter is organized as follows. In Section 2.1 the model under study is described. Section 2.2 provides the steady-state analysis of the model, including a few key performance measures and an optimization problem. Two illustrative examples are discussed in section 2.3.

2.1 Model description

We consider an infinite capacity queueing system with a single sever to which customers arrive according to a Poisson process with rate λ. The service times are assumed to follow an exponential distribution with parameter μ. We introduce a customer induced interruption while in service. The interruption occurs according to a Poisson process of rate θ. When an interruption occurs, the customer currently in service will be forced to leave the service facility. The freed server is ready to offer services to other customers. The interrupted customer will enter into a buffer (referred to as BIP) of finite capacity, K, should there be a space available. Otherwise such a customer is lost for ever. The interrupted customers will

spend a random period of time that is independent of other customers and the time is assumed to follow an exponential distribution with parameter η. In this chapter we assume that no more than one interruption is allowed for a customer while in service. That is, an interrupted customer who gets into service again will leave the system with no further interruption. All interrupted customers upon completing their interruptions enter into a finite buffer (referred to as BIC) whose size is K. Customers who are in BIC are given non-preemptive priority over new customers but are served in the order in which they enter into this buffer. Thus, a free server will offer services to those customers waiting in BIC before serving new customers by maintaining the first-in-first-served order. Because of this restriction coupled with the fact that at most one interruption is allowed for any customer, the total number of customers in BIC will never exceed the size of BIP and hence we assume the buffer sizes to be the same.
The model is studied as a QBD process and a matrix-geometric type solution is obtained (see, *Neuts* [54] and *Latouche and Ramaswami*[43]). To obtain the state space of the QBD in the sequel we use the following notations.

- $N(t)$ = Number of primary customers in the queue at time t;
- $N_1(t)$ = Number of customers in BIC at time t;
- $N_2(t)$ = Number of customers in BIP at time t;
- $S(t) = \begin{cases} 0 & \text{if server is idle;} \\ 1 & \text{if server is busy with a customer from primary queue;} \\ 2 & \text{if server is busy with a customer from BIC;} \end{cases}$
- $M = (K+1)(K+2)/2$.

The process $\{(N(t), S(t), N_1(t), N_2(t)) : t \geqslant 0\}$ is a $CTMC$ whose state space

$$\Omega = l^* \cup \bigcup_{n=0}^{\infty} l(n),$$

where

$$l^* = \{\ (0, 0, 0, i_2) :\ i_2 = 0, 1, \ldots, K\},$$

and for $n \geqslant 0$,

$$l(n) = \{(n, j, i_1, i_2) : j = 1, 2;\ i_1, i_2 = 0, \ldots, K; 0 \leqslant i_1 + i_2 \leqslant K\}.$$

The infinitesimal generator matrix Q of this $CTMC$ is a level independent quasi-birth-death process $(LIQBD)$ of the form:

$$Q = \begin{bmatrix} B_1 & B_0 & & & \\ B_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \qquad (2.1)$$

where

$$B_1 = -\left[\lambda I_{K+1} + \eta\ \Delta(0,\ a_{_K})\right],\ B_0 = \begin{bmatrix} B_{01} & B_{02} \end{bmatrix},\ B_{01} = \lambda \begin{bmatrix} I_{K+1} & \mathrm{O} \end{bmatrix};$$

$$B_{02} = \eta \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \Delta(a_K) & \mathrm{O} \end{bmatrix},\ B_2 = \begin{bmatrix} B_{10} \\ B_{20} \end{bmatrix};$$

$$B_{10} = \mu \begin{bmatrix} I_{K+1} \\ \mathrm{O} \end{bmatrix} + \theta \begin{bmatrix} \mathbf{0} & I_K \\ 0 & e_K^{'}(K) \\ \mathbf{0} & \mathrm{O} \end{bmatrix},\ B_{20} = \mu \begin{bmatrix} I_{K+1} \\ \mathrm{O} \end{bmatrix};$$

$$A_1 = \begin{bmatrix} B_{11} & B_{12} \\ \mathrm{O} & B_{22} \end{bmatrix};$$

$$B_{11} = -\left[(\lambda + \mu + \theta) I_M + \eta \Delta(G_K \dots G_0)\right] + \eta \begin{bmatrix} \mathrm{O} & \Delta(F_K \dots F_1) \\ \mathbf{0} & \mathbf{0} \end{bmatrix};$$

$$B_{22} = -\left\{(\lambda + \mu) I_M + \eta \Delta(G_K \dots G_0)\right\} + \eta \begin{bmatrix} \mathrm{O} & \Delta(F_K \dots F_1) \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \mu \begin{bmatrix} \mathrm{O} & \mathbf{0} \\ \Delta(J_K \dots J_1) & \mathbf{0} \end{bmatrix};$$

$$B_{12} = \begin{bmatrix} \mathrm{O} & \mathbf{0} \\ \Delta(H_K \dots H_1) & \mathbf{0} \end{bmatrix};$$

where

$$G_i = \Delta(0 \;\; a_i),\; G_0 = [\,0\,],\; F_i = \begin{bmatrix} \mathbf{0} \\ \Delta(a_i) \end{bmatrix},\; J_i = \begin{bmatrix} I_i & \mathbf{0} \end{bmatrix};$$

$$H_i = \mu \begin{bmatrix} I_i & \mathbf{0} \end{bmatrix} + \theta \begin{bmatrix} \mathbf{0} & I_i \end{bmatrix},\; 1 \leqslant i \leqslant K;$$

$$A_2 = \begin{bmatrix} A_{11} & \mathrm{O} \\ A_{21} & \mathrm{O} \end{bmatrix};$$

$$A_{11} = \mu \begin{bmatrix} I_{K+1} & \mathrm{O} \\ \mathbf{0} & \mathrm{O} \end{bmatrix} + \theta \begin{bmatrix} \mathbf{0} & I_K & \mathrm{O} \\ 0 & e_K^{'}(K) & \mathbf{0} \\ \mathbf{0} & \mathrm{O} & \mathrm{O} \end{bmatrix},\; A_{21} = \mu \begin{bmatrix} I_{K+1} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix};$$

$$A_0 = \lambda I.$$

The dimensions of the matrices B_0, B_1 and B_2 are, respectively, $(K+1) \times (K+1)$, $(K+1) \times 2M$ and $2M \times (K+1)$ respectively. The matrices A_0,

A_1 and A_2 are square matrices of order $(K+1)(K+2)$.

2.2 Analysis of the system

In this section we perform the steady-state analysis of the queueing model under study by first establishing the stability condition of the queueing system.

2.2.1 Stability condition

Let $\boldsymbol{\pi}$ denote the steady-state probability vector of the generator $A_0 + A_1 + A_2$. That is, $\boldsymbol{\pi}(A_0 + A_1 + A_2) = \mathbf{0},\ \boldsymbol{\pi}\boldsymbol{e} = 1$. The *LIQBD* description of the model indicates that the queueing system is stable (see, *Neuts* [54]) if and only if

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}. \tag{2.2}$$

That is, the rate of drift to the left has to be higher than that to the right. The vector, $\boldsymbol{\pi}$, cannot be obtained explicitly in terms of the parameters of the model, and hence the stability condition is known only implicitly. If we partition the vector $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi} = (\boldsymbol{\pi}_{1,1}, \ldots, \boldsymbol{\pi}_{1,M}, \boldsymbol{\pi}_{2,1}, \ldots, \boldsymbol{\pi}_{2,M})$$

and then using the structure of A_0 and A_2 matrices, equation (2.2) is given by

$$\lambda < (\mu + \theta) \sum_{i=1}^{K+1} \boldsymbol{\pi}_{1,i} + \mu \sum_{i=1}^{K+1} \boldsymbol{\pi}_{2,i}. \tag{2.3}$$

For future reference, we define the traffic intensity, ρ as

$$\rho = \frac{\boldsymbol{\pi} A_0 \boldsymbol{e}}{\boldsymbol{\pi} A_2 \boldsymbol{e}}. \tag{2.4}$$

Note that the stability condition in (2.2) is equivalent to $\rho < 1$. We will discuss the impact of the input parameters of the model on the traffic intensity in Section 2.3.

2.2.2 Steady-state distribution

Since the model studied as a QBD process, its steady-state distribution has a matrix-geometric solution under the stability condition. Assume that the stability condition (2.2) holds. Let $\boldsymbol{x}$ denote the steady-state probability vector of the generator Q given in (2.1). That is,

$$\boldsymbol{x}Q = \mathbf{0}, \quad \boldsymbol{x}\boldsymbol{e} = 1. \tag{2.5}$$

Partitioning $\boldsymbol{x}$ as

$$\boldsymbol{x} = (\boldsymbol{x}^*, \boldsymbol{x}(0), \boldsymbol{x}(1), \ldots) \tag{2.6}$$

we see that $\boldsymbol{x}$, under the assumption that the stability condition (2.2) holds, is obtained as (see, *Neuts* [54])

$$\boldsymbol{x}(n) = \boldsymbol{x}(0)\mathcal{R}^n, \quad n \geqslant 1, \tag{2.7}$$

where $\mathcal{R}$ is the minimal non-negative solution to the matrix quadratic equation:

$$\mathcal{R}^2 A_2 + \mathcal{R} A_1 + A_0 = \mathrm{O}, \tag{2.8}$$

and the boundary equations are given by

$$(\boldsymbol{x}^*, \boldsymbol{x}(0)) \begin{bmatrix} B_1 & B_0 \\ B_2 & A_1 + \mathcal{R}A_2 \end{bmatrix} = \mathbf{0}. \tag{2.9}$$

The normalizing condition of (2.5) results in

$$\boldsymbol{x}^* \boldsymbol{e} + \boldsymbol{x}(0)(I - \mathcal{R})^{-1} \boldsymbol{e} = 1. \tag{2.10}$$

Once the rate matrix $\mathcal{R}$ matrix is obtained, the vector $\boldsymbol{x}$ can be computed by exploiting the special structure of the coefficient matrices. We can use logarithmic reduction algorithm *Latouche and Ramaswami*[43] for computing $\mathcal{R}$.

2.2.3 Stationary waiting time distribution in the queue

In this section the *LST* of waiting time distribution and mean waiting time of a customer in the queue will be discussed.

First note that an arriving customer will enter into service immediately with probability $w_0 = \boldsymbol{x}^* \boldsymbol{e}$. With probability $1 - w_0$ the arriving customer has to wait before getting into service. The waiting time may be viewed as the time until absorption in a Markov chain with a highly sparse structure. The state space (that includes the arriving customer in its count) of the Markov chain is given by

$$\tilde{\Omega} = \{*\} \bigcup \{(n, j, i_1, i_2) : j = 1, 2;\ 0 \leqslant i_1, i_2 \leqslant K; 0 \leqslant i_1 + i_2 \leqslant K; n \geqslant 1\}.$$

The state * is obtained by lumping together the states that correspond to the server being idle. That is, * is obtained by lumping $\{(0,0,0,i_2) : 0 \leqslant i_2 \leqslant K\}$. Its generator matrix $\tilde{Q}$ is given by

$$\tilde{Q} = \begin{pmatrix} 0 & \mathbf{0} & & & \\ \boldsymbol{a} & \tilde{A}_1 & & & \\ & A_2 & \tilde{A}_1 & & \\ & & A_2 & \tilde{A}_1 & \\ & & & \ddots & \ddots \end{pmatrix}, \tag{2.11}$$

where

$$\tilde{A}_1 = A_1 + \lambda I, \quad \boldsymbol{a} = A_2\boldsymbol{e} = \begin{bmatrix} (\mu+\theta)\boldsymbol{e} \\ \mathbf{0} \\ \mu\boldsymbol{e} \\ \mathbf{0} \end{bmatrix}. \tag{2.12}$$

The initial probability vector of $\tilde{Q}$ is denoted by $\boldsymbol{z}$ and in partitioned form is given by

$$\boldsymbol{z} = (w_0, \boldsymbol{x}(0), \boldsymbol{x}(1), \cdots).$$

Define $\widetilde{\mathbf{W}}(t), t \geqslant 0$ to be the probability that an arriving customer will enter into service no later than time t. We will now derive the LST, $\tilde{w}(s)$, of $\widetilde{\mathbf{W}}(t)$. This transform is useful in deriving an expression for the mean waiting time. Using the structure of $\tilde{Q}$, it can readily be verified that

Theorem 2.2.1. *The LST, $\tilde{w}(s)$, of $\widetilde{\mathbf{W}}(t)$ is given by*

$$\tilde{w}(s) = w_0 + \sum_{i=0}^{\infty} \boldsymbol{x}(i) \left[(sI - \tilde{A}_1)^{-1}A_2\right]^i (sI - \tilde{A}_1)^{-1}\boldsymbol{a}. \tag{2.13}$$

Corollary 1. The mean waiting time, $\mu'_{\widetilde{W}}$, in the queue of an arriving customer is given by $\mu'_{\widetilde{W}} =$

$$\left[\boldsymbol{x}(0)(I-\mathcal{R})^{-1}-\boldsymbol{x}(0)\sum_{k=0}^{\infty}\mathcal{R}^kP^{k+1}+\boldsymbol{x}(0)(I-\mathcal{R})^{-2}\tilde{P}\right](I-P+\tilde{P})^{-1}(-\tilde{A}_1)^{-1}\boldsymbol{e}, \tag{2.14}$$

where

$$P=(-\tilde{A}_1)^{-1}A_2, \quad \tilde{P}=\boldsymbol{ep}, \tag{2.15}$$

and $\boldsymbol{p}$ is the invariant probability vector of P. That is,

$$\boldsymbol{p}P=\boldsymbol{p}, \quad \boldsymbol{pe}=1. \tag{2.16}$$

Note: In the computation of the mean waiting time, $\mu'_{\widetilde{W}}$, we need to evaluate the infinite sum $\sum_{k=0}^{\infty}\mathcal{R}^kP^{k+1}$. On noting that P is a stochastic matrix, we get

$$\boldsymbol{x}(0)\sum_{k=0}^{\infty}\mathcal{R}^kP^{k+1}\boldsymbol{e}=1-\boldsymbol{x}^*\boldsymbol{e},$$

and hence in truncating the infinite sum we find N^* such that

$$|\boldsymbol{x}(0)\sum_{k=0}^{N^*}\mathcal{R}^kP^{k+1}\boldsymbol{e}-(1-\boldsymbol{x}^*\boldsymbol{e})|<\epsilon,$$

where ϵ is a pre-determined quantity such as 10^{-7}.

2.2.4 Waiting time distribution of a primary customer since taken for service as primary unit

1. Waiting time distribution of a primary customer who completes his service without interruption

$$\hat{\mathbf{W_1}}(t) = \frac{\mu}{\mu+\theta}\left(1-e^{-(\mu+\theta)t}\right); \tag{2.17}$$

2. Waiting time distribution of a primary customer who leaves the system after the interruption without completing service due to BIP full

$$\hat{\mathbf{W_2}}(t) = \frac{\theta}{\mu+\theta}P_{BPFL}\left(1-e^{-(\mu+\theta)t}\right); \tag{2.18}$$

3. Waiting time distribution of a customer who leaves the system completing service after interruption

$$\begin{aligned}\hat{\mathbf{W_3}}(t) \;=\;& \tfrac{\theta}{\mu+\theta}(1-P_{BPFL})\left[(1-e^{-(\mu+\theta)t})+(1-e^{-\eta t})+P_{idle}(1-e^{-\mu t})\right]\\ &+\tfrac{\theta}{\mu+\theta}(1-P_{BPFL})(1-P_{idle})\left\{\sum_{n=1}^{K}A(n)\int_0^t\frac{\mu e^{-\mu x}(\mu x)^{n-1}}{(n-1)!}dx\right.\\ &\left.+P_{BSYP}\left(1-e^{-(\mu+\theta)t}\right)+P_{BSYI}\left(1-e^{-\mu t}\right)\right\};\end{aligned} \tag{2.19}$$

where P_{BPFL}, P_{idle} ,$A(n)$, P_{BSYP} , P_{BSYI} respectively denote the probabilities that BIP is full, server is idle, BIC have n customers, server is busy with primary customers and server is busy with interrupted customer.

Hence waiting time distribution of a primary customer since taken for service as primary unit, $\hat{W}(t)$ can obtained as

$$\hat{\mathbf{W}}(t) = \hat{\mathbf{W}}_{\mathbf{1}}(t) + \hat{\mathbf{W}}_{\mathbf{2}}(t) + \hat{\mathbf{W}}_{\mathbf{3}}(t). \tag{2.20}$$

Also distribution of the time spent by a customer in the system can be obtained as

$$\mathbf{W}(t) = (\widetilde{\mathbf{W}} * \hat{\mathbf{W}})(t), \tag{2.21}$$

where $*$ denote convolution.

2.2.5 Expected waiting time of a customer after taken for service

We can obtain the expected waiting time as

$$\begin{aligned} \mu'_{\hat{W}} \quad = \quad & \tfrac{\mu}{(\mu+\theta)^2} + \tfrac{\theta}{(\mu+\theta)^2} P_{BPFL} + \tfrac{\theta}{\mu+\theta}(1 - P_{BPFL}) \left[\tfrac{1}{\mu+\theta} + \tfrac{1}{\eta} + \tfrac{P_{idle}}{\mu} \right] \\ & + \tfrac{\theta}{\mu+\theta}(1 - P_{BPFL})(1 - P_{idle}) \left[\sum_{n=1}^{K} \tfrac{nA(n)}{\mu} + \tfrac{P_{BSYP}}{\mu+\theta} + \tfrac{P_{BSYI}}{\mu}) \right]. \end{aligned} \tag{2.22}$$

2.2.6 System performance measures

In this section we list a number of key system performance measures to bring out the qualitative aspects of the model under study. These are listed below along with their formula for computation. Towards this end, we further partition the vectors $\boldsymbol{x}^*$, $\boldsymbol{x}(n), n \geqslant 0$, into $\boldsymbol{x}(n) = (\boldsymbol{x}_1(n), \boldsymbol{x}_2(n)), n \geqslant$

0, where

$$\begin{aligned} \boldsymbol{x}^* &= (x_0^*, \cdots, x_K^*) \\ \boldsymbol{x}_1(n) &= (\boldsymbol{x}_{1,0}(n), \boldsymbol{x}_{1,1}(n), \cdots, \boldsymbol{x}_{1,K}(n)), n \geqslant 0, \\ \boldsymbol{x}_2(n) &= (\boldsymbol{x}_{2,0}(n), \boldsymbol{x}_{2,1}(n), \cdots, \boldsymbol{x}_{2,K}(n)), n \geqslant 0. \end{aligned}$$

Note that $\boldsymbol{x}_{j,r}(n), j = 1, 2; 0 \leqslant r \leqslant K; n \geqslant 0$, is of dimension $K + 1 - r$.

1. **The probability that the server is idle**:

$$P_{idle} = \boldsymbol{x}^* \boldsymbol{e}.$$

2. **The probability that the server is busy with a primary customer**:

$$P_{BSYP} = \sum_{n=0}^{\infty} \boldsymbol{x}_1(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - \mathcal{R})^{-1}(\boldsymbol{e}_1(2) \otimes \boldsymbol{e}).$$

3. **The probability that the server is busy with an interrupted customer**:

$$P_{BSYI} = \sum_{n=0}^{\infty} \boldsymbol{x}_2(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - \mathcal{R})^{-1}(\boldsymbol{e}_2(2) \otimes \boldsymbol{e}).$$

4. **The probability that an interrupted customer is lost**:

$$P_{loss} = \frac{\theta}{\theta + \mu} \sum_{n=0}^{\infty} x_{n,1,0,K}.$$

5. **Mean number of primary customers in the queue**:

$$\mu_{PQ} = \boldsymbol{x}(0)\mathcal{R}(I - \mathcal{R})^{-2}\boldsymbol{e}.$$

6. **The mean number of interrupted customers in the** BIP:

$$\mu_{BIP} = \sum_{i_2=0}^{K} i_2 x^*_{i_2} + \sum_{n=0}^{\infty}\sum_{j=1}^{2}\sum_{i_2=0}^{K}\sum_{i_1=0}^{K-i_2} i_2 x_{n,j,i_1,i_2}.$$

7. **The mean number of interrupted customers in the** BIC:

$$\mu_{BIC} = \sum_{n=0}^{\infty}\sum_{j=1}^{2}\sum_{i_1=0}^{K}\sum_{i_2=0}^{K-i_1} i_1 x_{n,j,i_1,i_2}.$$

8. **The mean waiting time in the queue,** $\mu'_{\widetilde{W}}$**, is as given in** (2.14).

We conclude this section by showing that the server is busy with primary customers is independent of K.

Theorem 2.2.2. *We have*

$$P_{BSYP} = \frac{\lambda}{\theta + \mu}.$$

Proof. The steady-state equations given in (2.5) can be written as

$$\boldsymbol{x}^* B_1 + \boldsymbol{x}(0)B_2 = \mathbf{0}, \tag{2.23}$$

$$\boldsymbol{x}^* B_0 + \boldsymbol{x}(0)A_1 + \boldsymbol{x}(1)A_2 = \mathbf{0}, \tag{2.24}$$

and

$$\boldsymbol{x}(i-1)A_0 + \boldsymbol{x}(i)A_1 + \boldsymbol{x}(i+1)A_2 = \mathbf{0}, \quad i \geqslant 1. \tag{2.25}$$

Post-multiplying equations (2.23) through (2.25) by $\boldsymbol{e}$ of appropriate dimensions, we get

$$\lambda \boldsymbol{x}^* \boldsymbol{e} + \eta \sum_{r=0}^{K} r x_r^* = (\mu + \theta) \boldsymbol{x}_{1,0}(0) \boldsymbol{e} + \mu \boldsymbol{x}_{2,0}(0) \boldsymbol{e}, \tag{2.26}$$

and

$$\lambda (\boldsymbol{x}_1(i) \boldsymbol{e} + \boldsymbol{x}_2(i) \boldsymbol{e}) = (\mu + \theta) \boldsymbol{x}_{1,0}(i+1) \boldsymbol{e} + \mu \boldsymbol{x}_{2,0}(i+1) \boldsymbol{e}, \; i \geqslant 0. \tag{2.27}$$

Now post-multiplying equations (2.24) and (2.25) by $(\boldsymbol{e}_1(2) \otimes \boldsymbol{e})$ and adding over $i = 1$ to ∞, we get

$$\lambda \boldsymbol{x}^* \boldsymbol{e} = (\mu + \theta) \sum_{i=0}^{\infty} \boldsymbol{x}_1(i) \boldsymbol{e} - (\mu + \theta) \sum_{i=1}^{\infty} \boldsymbol{x}_{1,0}(i) \boldsymbol{e} - \mu \sum_{i=1}^{\infty} \boldsymbol{x}_{2,0}(i) \boldsymbol{e}. \tag{2.28}$$

The stated result follows by immediately by adding (2.27) over i and (2.28). □

Note: It is interesting to note that the measure P_{BSYP} does not depend on K. This seems to be counter intuitive as one would expect a larger K to increase the server to be busy with serving interrupted customers and hence has a bearing on P_{BSYP}. However, it appears that the server's idle probability is reduced when K is increased to off-set the increase in P_{BSYI} and to keep P_{BSYP} the same.

2.2.7 An Optimization Problem

In this section we propose an optimization problem and discuss it through an illustrative example. Given

- a revenue of r_1 monetary units for each customer leaving the system with an uninterrupted service,
- revenue of r_2 monetary units for each customer leaving the system with an interrupted service,
- a holding cost of c_1 monetary units for each unit of time that a customer has to wait in the primary queue,
- a holding cost of c_2 monetary units for each unit of time that a customer has to wait in the BIC,
- a cost of c_3 monetary units for each customer that is lost due to BIP being full at the time an interruption occurs,

the goal is to find an optimum value for K (when all other parameters are fixed) that maximizes the expected total profit, ***ETP***, as given in the following objective function.

$$\boldsymbol{ETP} = \mu[r_1 P_{BSYP} + r_2 P_{BSYI}] - c_1\mu_{PQ} - c_2\mu_{BIC} - c_3(\theta + \mu)P_{loss}. \quad (2.29)$$

We will discuss this optimization problem in the next section.

2.3 Numerical Examples

In order to bring out the qualitative nature of the model under study, we present a few representative examples.

Example 2.3.1. In this example, by fixing $\lambda = 1, \mu = 1.1$ and $\eta = 1$, we look at the effect of varying θ and K on some selected measures. These are displayed in Figure 2.1 and 2.2. Looking at these figures, we summarize the following observations.

- Noting that increasing K leads to an increase in the customers leaving the system with services, the traffic intensity, ρ, has to increase. This is the case for all values of θ as can be seen in Figure 2.1(a).
- As one would expect increasing θ should result in a decrease in ρ (for all values of K). This is due to the fact that an increase in θ will cause more customers to be interrupted leading to an increase in the number of customers leaving the system without getting services.
- The measures, μ_{PQ} and $\mu'_{\widetilde{W}}$, behave similar to ρ as functions of K and θ (see Figure 2.1(b) and (e)). This is again as expected.
- Comparing the two measures, μ_{BIC} and μ_{BIP}, as K and θ vary, we observe some interesting trends. Recall that at any given time the total number of customers in the BIC and BIP cannot exceed K.

    1. For all values of K, $\mu_{BIC} < \mu_{BIP}$ when θ increases from 0 to 1 (since $\eta = 1$). This is probably due to the fact that not too

many customers are seen waiting in BIP buffer and hence the rate of interrupted customers getting back to service through waiting in BIC buffer will be smaller leading to less customers (on the average) in BIC buffer.

2. However, when θ is varied from 1 (note that $\eta = 1$) to higher values, there appears to be cut-off points, say, $\hat{\theta}$ and $\hat{K}$ such that for $\theta > \hat{\theta}$, $\mu_{BIC} < \mu_{BIP}$, for all $K \leqslant \hat{K}$ and for all $K > \hat{K}$, $\mu_{BIP} < \mu_{BIC}$. [For our example, $\hat{\theta} > 3$ and $\hat{K} = 5$.] A possible explanation for this is as follows. For sufficiently large θ and K, the higher interruption rate causes more interruptions leading to more interrupted customers filling BIP buffer, which in turn increases the rate of self-interrupted customers to get into BIC buffer.

- As is to be expected the measure P_{BSYI} is a non-decreasing function of θ when all other parameters are fixed. This measure is also a non-decreasing function of K when all other parameters are fixed.

- For fixed K, the measure P_{loss} increases initially and then decreases as θ increases. At first one may expect this measure to be a non-decreasing function of θ when all other parameters are fixed. However, after carefully looking into the model and the fact that ρ is a non-increasing function of θ, we see that beyond a certain point for θ, any further increase in θ will only result in the server being busy with interrupted customers the phenomenon of increasing-decreasing of this measure can be justified.

- If θ^*_K denotes the point where P_{loss} attains its maximum as a function of θ for fixed K, then it can be noticed that θ^*_K is a non-decreasing function of K.

Example 2.3.2. In this example we will discuss the optimization problem of section 2.2.7. By fixing $\lambda = 2, \mu = 2.5, r_1 = 60, r_2 = 70, c_1 = 20, c_2 = 30$, and $c_3 = 40$, we look at the optimum K that maximizes the expected total profit, $\boldsymbol{ETP}$ by varying θ and η. The optimum K^* and the corresponding $\boldsymbol{ETP}$ are given in Tab. 2.1 below.

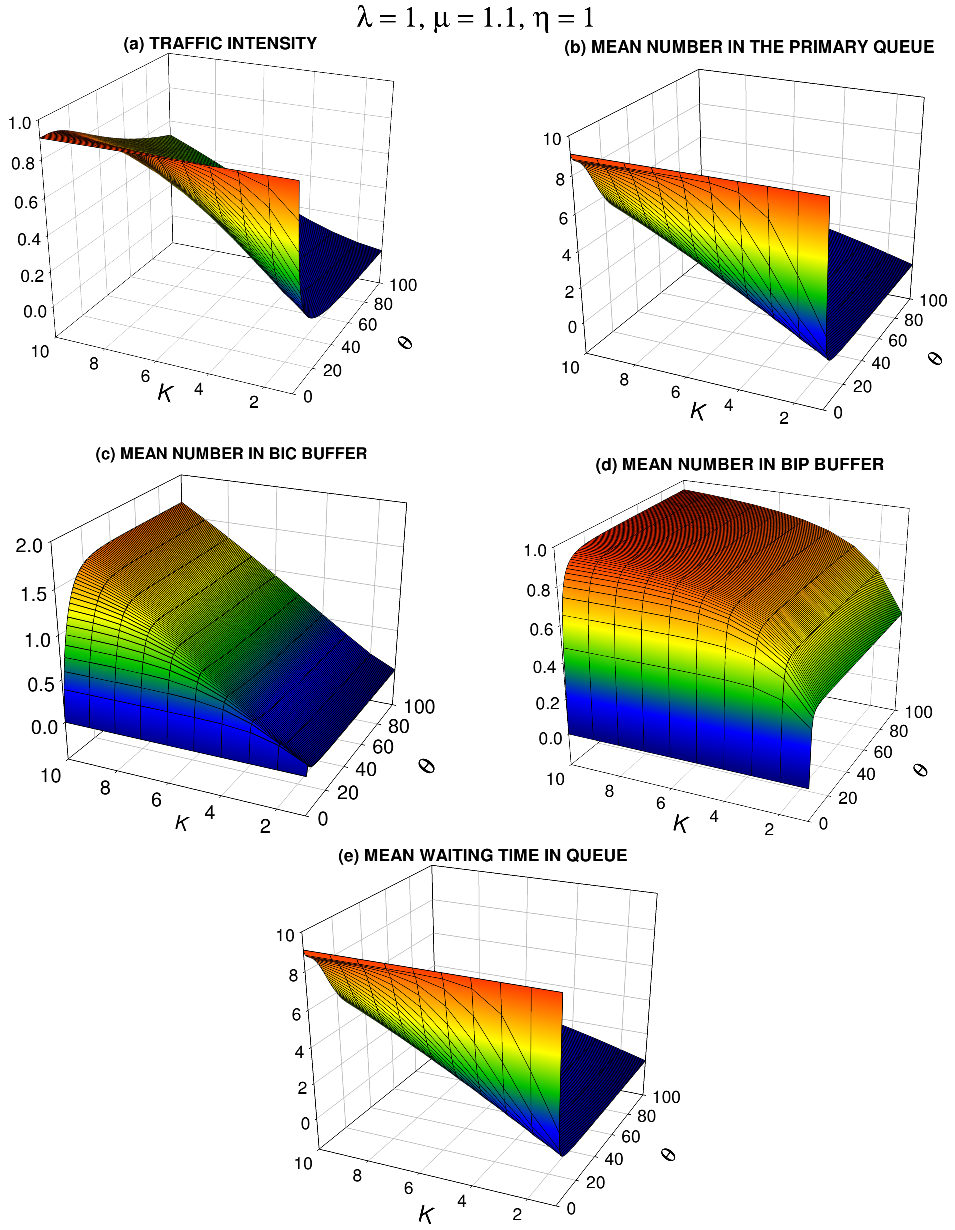
- For fixed η, we notice that K^* is a nondecreasing function of θ. This is to be expected since an increase in θ results in more customers being interrupted and hence requires more buffer space to avoid the loss of (interrupted) customers. However, it is interesting to see that K^* approaches a limiting value as θ increases. The limiting value depends on η.

- For fixed θ, we notice that K^* is a non-increasing function of η. This is to be expected since an increase in η results in a faster clearance of BIC buffer leading to lower values for K^*.

- For fixed but small values of η, we notice that the $\boldsymbol{ETP}$ at optimum K^* increases initially and then decreases as θ increases. A possible explanation is that as the buffer size increases, the cost associated with the customers waiting in BIC buffer increases resulting in a reduction in the $\boldsymbol{ETP}$.

Table 2.1: K^* and $\boldsymbol{ETP}$ for selected η and θ.

					η			
θ	Measure	0.1	0.2	1	2	5	10	20
0.1	K^*	1	1	1	1	1	1	1
	$\boldsymbol{ETP}$	59.145	58.418	57.136	56.861	56.671	56.603	56.568
0.2	K^*	1	1	1	1	1	1	1
	$\boldsymbol{ETP}$	62.117	61.402	58.973	58.166	57.537	57.298	57.172
1	K^*	6	3	1	1	1	1	1
	$\boldsymbol{ETP}$	65.654	66.166	66.760	67.514	65.431	63.479	62.081
2	K^*	9	6	2	2	1	1	1
	$\boldsymbol{ETP}$	65.259	67.068	70.040	68.856	70.538	69.116	66.953
5	K^*	14	9	3	2	2	1	1
	$\boldsymbol{ETP}$	59.448	63.820	71.903	73.332	73.351	74.390	74.411
10	K^*	17	10	4	3	2	2	1
	$\boldsymbol{ETP}$	51.648	58.322	72.226	75.802	78.046	76.242	77.350
20	K^*	18	11	5	4	3	2	2
	$\boldsymbol{ETP}$	44.294	52.648	71.254	75.922	78.802	80.968	78.426
50	K^*	19	12	5	4	3	3	2
	$\boldsymbol{ETP}$	37.745	47.419	69.564	76.444	81.772	82.289	83.872
100	K^*	19	12	6	4	3	3	2
	$\boldsymbol{ETP}$	35.106	45.190	68.871	76.034	82.037	84.757	84.823



(a) TRAFFIC INTENSITY

(b) MEAN NUMBER IN THE PRIMARY QUEUE

(c) MEAN NUMBER IN BIC BUFFER

(d) MEAN NUMBER IN BIP BUFFER

(e) MEAN WAITING TIME IN QUEUE

Figure 1

Figure 2.1: Effect of K and θ on ρ, μ_{PQ}, μ_{BIC}, μ_{BIP} and $\mu'_{\widetilde{W}}$

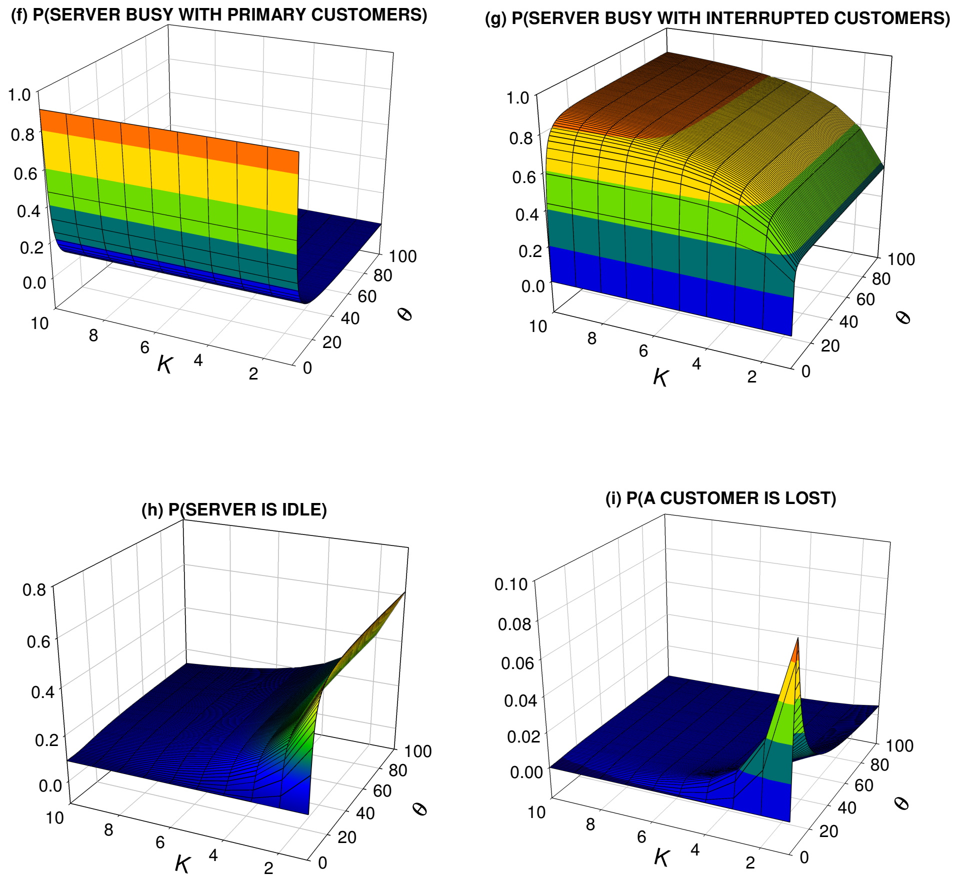


Figure 1

Figure 2.2: Effect of K and θ on P_{BSYP}, P_{BSYI}, P_{idle} and P_{loss}

Chapter 3

Customer Induced Interruption in a Multi-server System

In this chapter we extend the model discussed in chapter 2 to a multi-server system with customer induced interruption. We attempt to obtain compact expressions for stability of the system. This will be missing in the chapter to follow since quite general there. Further the arrival process in it follows a MAP where correlation is inbuilt.

This chapter is arranged as follows. In Section 3.1 the problem is mathematically formulated and analyzed. Then the steady-state analy-

Some results of this chapter are included in the following paper.
Krishnamoorthy, A. and Varghese Jacob : Analysis of customer induced interruption in a multi-server system. Neural, Parallel and Scientific Computations, 20, 153–172, 2012. Dynamic publishers, Inc, USA.

sis of the model, including a few key performance measures are obtained. The mean waiting time in the queue of an arriving customer is obtained through the LST and an optimization problem of the model under study is described. These are done in section 3.2. Finally, in section 3.3 some illustrative examples including optimization problem to maximize the revenue with respect to number of servers to be employed and optimal buffer size for the self-interrupted customers are discussed.

3.1 Model description

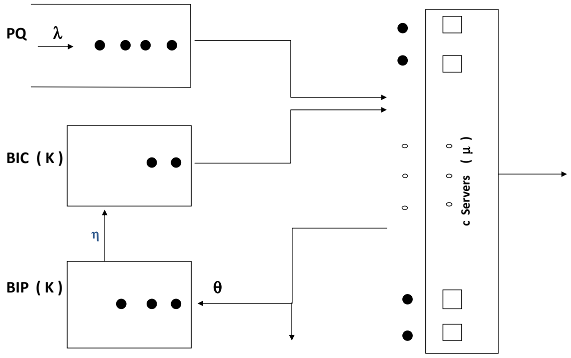


Figure 3.1: Customer Induced Interruption in an $M/M/c$ queueing system

We consider an infinite capacity multi server queuing model to which customers arrive according to a Poisson process with rate λ (Figure 3.1). The service facility consists of c servers. All c servers are assumed to be homogeneous and that the service times are exponentially distributed with parameter μ. An arriving customer, finding a free server, enters into service immediately; otherwise the customer is placed into the buffer of infinite capacity and he/she will be picked up for service according to the order of their arrival. We consider customer induced interruption while his/her service is going on. The interruption occurs according to a Poisson process of rate θ. When an interruption occurs, the customer currently in service will be forced to leave the service facility. The freed server is ready to offer services to other customers. The interrupted customer enters into a buffer (referred to as BIP) of finite capacity, K, should there be a space available. Otherwise, such a customer is lost for ever. An interrupted customer spends a random period of time for completion of interruption, independent of other customers. The duration of an interruption follows an exponential distribution with parameter η. In this chapter we assume that no more than one interruption is allowed for a customer while in service. That is, an interrupted customer who gets into service again will leave the system with no further interruption. All interrupted customers, upon completing their interruptions enter into a finite buffer (referred to as BIC) whose size is K. Customers who are in BIC are given non-preemptive priority over new customers but are served in the order in which they enter into this buffer. Thus, a free server will offer services to those customers waiting in BIC before serving new customers by maintaining the order of their arrival. Because of this restriction coupled, with the fact that at most one interruption is allowed for a customer, the total number of customers in BIC and BIP will never exceed the size of BIP and hence we assume the buffer sizes to be the same.

In the sequel we use the following notations.

- $N(t)$= Number of primary customers in the queue at time t;
- $N_1(t)$= Number of busy servers at time t;
- $N_2(t)$= Number of servers busy with primary customers at time t;
- $N_3(t)$ = Number of customers in BIC at time t;
- $N_4(t)$ = Number of customers in BIP at time t;
- $LC = L * (c+1)$.

The process $\{(N(t), N_1(t), N_2(t), N_3(t), N_4(t)) : t \geqslant 0\}$ is a $CTMC$ whose state space is given by

$$\begin{aligned}\Omega &= \{(0,0,0,0,i_2) : 0 \leqslant i_2 \leqslant K\}\\ &\bigcup\{(0,j,m,0,i_2) : 1 \leqslant j \leqslant c-1; 0 \leqslant m \leqslant j; 0 \leqslant i_2 \leqslant K\}\\ &\bigcup\{(n,c,m,i_1,i_2) : n \geqslant 0; 0 \leqslant m \leqslant c; 0 \leqslant i_1, i_2, i_1+i_2 \leqslant K\}\end{aligned}$$

A brief description of the above states are given below.

- $(\underline{0,0}) = (0,0,0,0,i_2) : -$ the system has no customers in the primary queue, all servers including primary servers are idle, no customers in the BIC and BIP has i_2 customers.
- $(\underline{j,m}) = (0,j,m,0,i_2) : -$ the system has no customer in the primary queue, there are j servers are busy of which m servers are busy with primary queue customers ($1 \leqslant j \leqslant c-1$ and $0 \leqslant m \leqslant j$), no customer in BIC and BIP has i_2 customers.
- $(\underline{c,m}) = (n,c,m,i_1,i_2) : -$ there are $n(n \geqslant 0)$ customers in the primary queue, all c servers are busy of which m servers are busy

with primary queue customers, $(0 \leqslant m \leqslant c)$, BIC has i_1 customers and BIP has i_2 customers.

Level $l(0,j)$ denotes the union of $(j+1)(K+1)$ states given by

$$l(0,j) = \bigcup_{m=0}^{j} \{(0,j,m,0,i_2) : \ 0 \leqslant i_2 \leqslant K\}; \ 0 \leqslant j \leqslant c-1.$$

Level $l(n,c)$ denotes the union of LC states given by

$$l(n,c) = \bigcup_{m=0}^{c} \{(n,c,m,i_1,i_2) : 0 \leqslant i_1, i_2, i_1+i_2 \leqslant K\} \ ; \ n \geqslant 0.$$

To write down the infinitesimal generator Q, we introduce additionally the following notations:

- $I^* = \begin{bmatrix} \mu & \theta & & & \\ & \mu & \theta & & \\ & & \ddots & \ddots & \\ & & & \mu & \theta \\ & & & & \mu+\theta \end{bmatrix}_{(K+1)\times(K+1)}$;

- $\widetilde{I}^* = \begin{bmatrix} I^* \\ \mathrm{O} \end{bmatrix}_{L\times(K+1)}$;

- $\widetilde{I}^{**} = \begin{bmatrix} I^* & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}_{L\times L}$;

- $F^* = \eta \begin{bmatrix} \mathbf{0} & 0 \\ \Delta(\mathbf{a_K}) & \mathbf{0} \end{bmatrix}_{(K+1)\times(K+1)}$;

- $\widehat{F}^* = \begin{bmatrix} F^* & \mathrm{O} \end{bmatrix}_{(K+1)\times L}$;

- $\widetilde{I}_{K+1} = \begin{bmatrix} I_{K+1} \\ \mathrm{O} \end{bmatrix}_{L\times(K+1)}$;

- $\widehat{I}_{K+1} = \begin{bmatrix} I_{K+1} & \mathrm{O} \end{bmatrix}_{(K+1)\times L}$;

For $1 \leqslant p \leqslant K$,

- $F_p = \begin{bmatrix} \mathbf{0} \\ \Delta(\boldsymbol{a}_p) \end{bmatrix}_{(p+1)\times p}$, $J_p = \begin{bmatrix} I_p & \mathbf{0} \end{bmatrix}$, $G_p = \Delta(0 \ \ \boldsymbol{a}_p),\ G_0 = \ 0;$

- $H_p = \begin{bmatrix} \mu & \theta & & & \\ & \mu & \theta & & \\ & & \ddots & \ddots & \\ & & & \mu & \theta \end{bmatrix}_{p\times(p+1)}$;

If the states in Ω are listed in lexicographical order then the infinitesimal generator of the $CTMC$ governing the system is given by $Q =$

$$\begin{array}{c|cccccccc|}
 & l(0,0) & l(0,1) & l(0,2) & \dots & l(0,c{-}1) & l(0,c) & l(1,c) & \dots \\
\hline
l(0,0) & E_0 & C_0 & & & & & & \\
l(0,1) & B_1 & E_1 & C_1 & & & & & \\
l(0,2) & & B_2 & E_2 & C_2 & & & & \\
\vdots & & & \ddots & \ddots & \ddots & & & \\
l(0,c{-}1) & & & & B_{c-1} & E_{c-1} & C_{c-1} & & \\
l(0,c) & & & & & B_c & A_1 & A_0 & \\
l(1,c) & & & & & & A_2 & A_1 & A_0 \\
\vdots & & & & & & & \ddots & \ddots \quad \ddots
\end{array}, \tag{3.1}$$

where the coefficient matrices appearing in (3.1) are given by

$$E_j = \begin{array}{c|ccccc|}
 & (\underline{j,0}) & (\underline{j,1}) & \dots & (\underline{j,j-1}) & (\underline{j,j}) \\
\hline
(\underline{j,0}) & D_{j,0} & & & & \\
(\underline{j,1}) & & D_{j,1} & & & \\
\vdots & & & \ddots & & \\
(\underline{j,j-1}) & & & & D_{j,j-1} & \\
(\underline{j,j}) & & & & & D_{j,j}
\end{array}, \tag{3.2}$$

where

$$D_{j,i} = -\Delta\left(\lambda + j\mu + i\theta, \lambda + j\mu + i\theta + \eta, \dots, \lambda + j\mu + i\theta + K\eta\right),$$

$$j = 0, 1, \dots, (c-1), \quad i = 0, \dots, j;$$

$$B_j = \begin{array}{c} \\ (\underline{j,0}) \\ (\underline{j,1}) \\ (\underline{j,2}) \\ \vdots \\ (\underline{j,j-1}) \\ (\underline{j,j}) \end{array} \begin{array}{ccccc} (\underline{j-1,0}) & (\underline{j-1,1}) & \dots & (\underline{j-1,j-2}) & (\underline{j-1,j-1}) \\ \left(\begin{array}{c} j\mu I_{K+1} \\ I^* \\ \\ \\ \\ \\ \end{array}\right. & \begin{array}{c} \\ (j-1)\mu I_{K+1} \\ 2I^* \\ \\ \\ \\ \end{array} & \begin{array}{c} \\ \\ \ddots \\ \ddots \\ \\ \\ \end{array} & \begin{array}{c} \\ \\ \\ \ddots \\ (j-1)I^* \\ \\ \end{array} & \left.\begin{array}{c} \\ \\ \\ \\ \mu I_{K+1} \\ jI^* \end{array}\right) \end{array}, \tag{3.3}$$

$j = 0, \dots, c-1;$

$$C_j = \begin{array}{c} \\ (\underline{j,0}) \\ (\underline{j,1}) \\ \vdots \\ (\underline{j,j-1}) \\ (\underline{j,j}) \end{array} \begin{array}{cccccc} (\underline{j+1,0}) & (\underline{j+1,1}) & (\underline{j+1,2}) & \dots & (\underline{j+1,j}) & (\underline{j+1,j+1}) \\ \left(\begin{array}{c} F^* \\ \\ \\ \\ \\ \end{array}\right. & \begin{array}{c} \lambda I_{K+1} \\ F^* \\ \\ \\ \\ \end{array} & \begin{array}{c} \\ \lambda I_{K+1} \\ \ddots \\ \\ \\ \end{array} & \begin{array}{c} \\ \\ \ddots \\ \ddots \\ \\ \end{array} & \begin{array}{c} \\ \\ \\ \lambda I_{K+1} \\ F^* \end{array} & \left.\begin{array}{c} \\ \\ \\ \\ \lambda I_{K+1} \end{array}\right) \end{array}, \tag{3.4}$$

$j = 0, \dots, c-2;$

$$B_c = \begin{array}{c|ccccc}
 & (\underline{c-1,0}) & (\underline{c-1,1}) & \dots & (\underline{c-1,c-2}) & (\underline{c-1,c-1}) \\ \hline
(\underline{c,0}) & c\mu\widetilde{I}_{K+1} & & & & \\
(\underline{c,1}) & \widetilde{I}^* & (c-1)\mu\widetilde{I}_{K+1} & & & \\
(\underline{c,2}) & & 2\widetilde{I}^* & \ddots & & \\
\vdots & & & \ddots & 2\mu\widetilde{I}_{K+1} & \\
(\underline{c,c-1}) & & & & (c-1)\widetilde{I}^* & \mu\widetilde{I}_{K+1} \\
(\underline{c,c}) & & & & & c\widetilde{I}^*
\end{array} ; \tag{3.5}$$

$$C_{c-1} = \begin{array}{c|cccccc}
 & (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \dots & (\underline{c,c-1}) & (\underline{c,c}) \\ \hline
(\underline{c-1,0}) & \widehat{F}^* & \lambda\widehat{I}_{K+1} & & & & \\
(\underline{c-1,1}) & & \widehat{F}^* & \lambda\widehat{I}_{K+1} & & & \\
\vdots & & & \ddots & \ddots & & \\
(\underline{c-1,c-2}) & & & & \ddots & \lambda\widehat{I}_{K+1} & \\
(\underline{c-1,c-1}) & & & & & \widehat{F}^* & \lambda\widehat{I}_{K+1}
\end{array} ; \tag{3.6}$$

$$A_1 = \begin{array}{c|cccccc}
 & (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \dots & (\underline{c,c-1}) & (\underline{c,c}) \\ \hline
(\underline{c,0}) & A^{(1)}_{0,0} & & & & & \\
(\underline{c,1}) & A^{(1)}_{1,0} & A^{(1)}_{1,1} & & & & \\
(\underline{c,2}) & & A^{(1)}_{2,1} & A^{(1)}_{2,2} & & & \\
\vdots & & & \ddots & \ddots & & \\
(\underline{c,c-1}) & & & & \ddots & A^{(1)}_{(c-1),(c-1)} & \\
(\underline{c,c}) & & & & & A^{(1)}_{c,(c-1)} & A^{(1)}_{c,c}
\end{array} ; \tag{3.7}$$

where

$$A_{i,i}^{(1)} = -\left\{(\lambda + c\mu + i\theta)\; I_L + \eta\; \Delta(G_K \ldots G_0)\right\} + \eta \begin{bmatrix} \mathrm{O} & \Delta(F_K \ldots F_1) \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$+\; (c-i)\; \mu \begin{bmatrix} \mathrm{O} & \mathbf{0} \\ \Delta(J_K \ldots J_1) & \mathbf{0} \end{bmatrix}, \quad i = 0, \ldots, c;$$

$$A_{i(i-1)}^{(1)} = i \begin{bmatrix} \mathrm{O} & \mathbf{0} \\ \Delta(H_K \ldots H_1) & \mathbf{0} \end{bmatrix}, \quad i = 1, \ldots, c;$$

$$A_2 = \begin{array}{c} \\ (\underline{c,0}) \\ (\underline{c,1}) \\ (\underline{c,2}) \\ \vdots \\ (\underline{c,c-1}) \\ (\underline{c,c}) \end{array} \!\!\! \begin{array}{c} \begin{array}{cccccc} (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \ldots & (\underline{c,c-1}) & (\underline{c,c}) \end{array} \\ \left(\begin{array}{cccccc} \mathrm{O} & A_0^{(2)} & & & & \\ & \widetilde{I}^{**} & A_1^{(2)} & & & \\ & & 2\widetilde{I}^{**} & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & (c-1)\widetilde{I}^{**} & A_{c-1}^{(2)} \\ & & & & & c\widetilde{I}^{**} \end{array} \right) \end{array}, \qquad (3.8)$$

where

$$A_j^{(2)} = (c-j)\; \mu \begin{bmatrix} I_{K+1} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}_{L \times L}, \quad j = 0, \ldots, c-1;$$

$$A_0 = \lambda\; I_{LC}; \qquad (3.9)$$

3.2 Steady-state analysis

In this section we perform the steady-state analysis of the queueing model under study by first establishing the stability condition of the queueing system.

3.2.1 Stability condition

Let $\boldsymbol{\pi}$ denote the steady-state probability vector of the generator $A_0 + A_1 + A_2$. That is, $\boldsymbol{\pi}(A_0 + A_1 + A_2) = \mathbf{0},\ \boldsymbol{\pi}\boldsymbol{e} = 1$. The *LIQBD* description of the model indicates that the queueing system is stable (see, *Neuts* [54]) if and only if

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}. \tag{3.10}$$

The vector, $\boldsymbol{\pi}$, cannot be obtained explicitly in terms of the parameters of the model, and hence the stability condition is known only implicitly. If we partition the vector $\boldsymbol{\pi}$ as $\boldsymbol{\pi} = (\boldsymbol{\pi}_{0,1}, \ldots, \boldsymbol{\pi}_{0,L}, \boldsymbol{\pi}_{1,1}, \ldots, \boldsymbol{\pi}_{1,L}, \boldsymbol{\pi}_{c,1}, \ldots, \boldsymbol{\pi}_{c,L})$ and then using the structure of A_0 and A_2 matrices, equation (3.10) is given by

$$\lambda < c\mu \sum_{i=0}^{c} \sum_{j=1}^{K+1} \boldsymbol{\pi}_{i,j} + \theta \sum_{i=0}^{c} \sum_{j=1}^{K+1} i\boldsymbol{\pi}_{i,j}. \tag{3.11}$$

For future reference, we define the traffic intensity, ρ as

$$\rho = \frac{\boldsymbol{\pi} A_0 \boldsymbol{e}}{\boldsymbol{\pi} A_2 \boldsymbol{e}}. \tag{3.12}$$

Note that the stability condition in (3.10) is equivalent to $\rho < 1$. We will discuss the impact of the input parameters of the model on the traffic intensity in Section 3.3.

3.2.2 Steady-state probability vector

Let $\boldsymbol{x}$ denote the steady-state probability vector of the generator Q given in (3.1). That is,

$$\boldsymbol{x}\, Q = \mathbf{0}, \quad \boldsymbol{x}\boldsymbol{e} = 1. \tag{3.13}$$

Let $\boldsymbol{x}$ be partitioned as

$$\boldsymbol{x} = (\boldsymbol{x}^*(0), \boldsymbol{x}^*(1), \ldots, \boldsymbol{x}^*(c-1), \boldsymbol{x}(0), \boldsymbol{x}(1), \ldots) \tag{3.14}$$

we see that $\boldsymbol{x}$, under the assumption that the stability condition holds, the steady-state probability vector is obtained as

$$\boldsymbol{x}(n) = \boldsymbol{x}(0)\mathcal{R}^n, \quad n \geqslant 1, \tag{3.15}$$

where $\mathcal{R}$ is the minimal non-negative solution to the matrix quadratic equation:

$$\mathcal{R}^2 A_2 + \mathcal{R} A_1 + A_0 = \mathrm{O},$$

and the vectors $\boldsymbol{x}^*(0), \boldsymbol{x}^*(1), \ldots, \boldsymbol{x}^*(c-1), \boldsymbol{x}(0)$ are obtained from boundary equations

$$\boldsymbol{x}^*(0)E_0 + \boldsymbol{x}^*(1)B_1 = \mathbf{0},$$

$$\boldsymbol{x}^*(i-1)C_{i-1} + \boldsymbol{x}^*(i)E_i + \boldsymbol{x}^*(i+1)B_{i+1} = \mathbf{0}, \quad 1 \leqslant i \leqslant c-2 \tag{3.16}$$

$$\boldsymbol{x}^*(c-2)C_{c-2} + \boldsymbol{x}^*(c-1)E_{c-1} + \boldsymbol{x}(0)B_c = \mathbf{0},$$

$$\boldsymbol{x}^*(c-1)C_{c-1} + \boldsymbol{x}(0)\left[A_1 + \mathcal{R}A_2\right] = \mathbf{0},$$

Once $\mathcal{R}$ matrix is obtained, from the boundary equation we obtain

$$\boldsymbol{x}(0) = \boldsymbol{x}^*(c-1)\mathcal{R}_{c-1},$$

$$\boldsymbol{x}^*(i) = \boldsymbol{x}^*(i-1)\mathcal{R}_{i-1}, \quad 1 \leqslant i \leqslant (c-1),$$

which gives $\boldsymbol{x}(0) = \boldsymbol{x}^*(0)\prod\limits_{i=0}^{c-1}\mathcal{R}_i$ where $\mathcal{R}_{c-1} = C_{c-1}\left[-(A_1 + \mathcal{R}A_2)\right]^{-1}$ and $\mathcal{R}_{i-1} = C_{i-1}\left[-(E_i + \mathcal{R}_i B_{i+1})\right]^{-1}$. The component $\boldsymbol{x}^*(0)$ is the steady-state distribution of the Markov chain with generator matrix $E_0 + \mathcal{R}_0 B_1$ subject to the normalizing equation

$$\boldsymbol{x}^*(0)\left(I + \sum_{i=0}^{c-2}\prod_{j=0}^{i}\mathcal{R}_j + \prod_{j=0}^{c-1}\mathcal{R}_j(I-\mathcal{R})^{-1}\right)\boldsymbol{e} = 1. \tag{3.17}$$

Thus, the vector $\boldsymbol{x}$ can be computed by exploiting the special structure of the coefficient matrices.

3.2.3 Stationary waiting time distribution in the queue

The stationary waiting time distribution for this queueing model, in general, is analytically intractable. However we will obtain the *LST* of the waiting time of a customer in the queue and derive an expression for its mean. First note that an arriving customer will enter into service immediately with probability $w_0 = \sum_{i=0}^{c-1}\boldsymbol{x}^*(i)\boldsymbol{e}$. With probability $1 - w_0$ the arriving customer has to wait before getting into service. The waiting time may be viewed as the time until absorption in a Markov chain with a highly sparse structure. The state space (that includes the arriving customer in its count) of the Markov chain is given by

$$\tilde{\Omega} = \{*\} \bigcup \{(n, c, j, i_1, i_2) : 0 \leqslant j \leqslant c;\ 0 \leqslant i_1, i_2, i_1 + i_2 \leqslant K; n \geqslant 0\}.$$

The state * is obtained by lumping together the states that correspond to at least one of the server being idle. That is, * is obtained by lumping $\{(0, j, m, 0, i_2) :\ 0 \leqslant j \leqslant c - 1;\ 0 \leqslant m \leqslant j;\ 0 \leqslant i_2 \leqslant K\}$. Its generator matrix $\tilde{Q}$ is given by

$$\tilde{Q} = \begin{pmatrix} 0 & \mathbf{0} & & & & \\ \boldsymbol{a} & \tilde{A}_1 & & & & \\ & A_2 & \tilde{A}_1 & & & \\ & & A_2 & \tilde{A}_1 & & \\ & & & \ddots & \ddots & \end{pmatrix}, \tag{3.18}$$

where

$$\tilde{A}_1 = A_1 + \lambda I, \quad \boldsymbol{a} = A_2 \boldsymbol{e}.$$

The initial probability vector of $\tilde{Q}$ is denoted by $\boldsymbol{z}$ and in partitioned form is given by

$$\boldsymbol{z} = (w_0, \boldsymbol{x}(0), \boldsymbol{x}(1), \cdots).$$

Define $W(t), t \geqslant 0$ to be the probability that an arriving customer will enter into service no later than time t. We will now derive the LST, $\tilde{w}(s)$, of $W(t)$. This transform is useful in deriving an expression for the mean waiting time. Using the structure of $\tilde{Q}$, it can readily be verified that

Theorem 3.2.1. *The LST, $\tilde{w}(s)$, of $W(t)$ is given by*

$$\tilde{w}(s) = w_0 + \sum_{i=0}^{\infty} \boldsymbol{x}(i) \left[(sI - \tilde{A}_1)^{-1} A_2 \right]^i (sI - \tilde{A}_1)^{-1} \boldsymbol{a}. \tag{3.19}$$

Corollary 2. The mean waiting time E_W^Q, in the queue of an arriving customer is given by $E_W^Q =$

$$\left[\boldsymbol{x}(0)(I - \mathcal{R})^{-1} - \boldsymbol{x}(0) \sum_{k=0}^{\infty} \mathcal{R}^k P^{k+1} + \boldsymbol{x}(0)(I - \mathcal{R})^{-2} \tilde{P} \right] (I - P + \tilde{P})^{-1} (-\tilde{A}_1)^{-1} \boldsymbol{e}, \tag{3.20}$$

where

$$P = (-\tilde{A}_1)^{-1} A_2, \quad \tilde{P} = \boldsymbol{e}\boldsymbol{p}, \tag{3.21}$$

and $\boldsymbol{p}$ is the invariant probability vector of P. That is,

$$\boldsymbol{p}P = \boldsymbol{p}, \quad \boldsymbol{p}\boldsymbol{e} = 1. \tag{3.22}$$

Note: As in chapter 2, to evaluate the infinite sum $\sum_{k=0}^{\infty} \mathcal{R}^k P^{k+1}$ in the expression (3.20), we need to find N^* such that

$$|\boldsymbol{x}(0) \sum_{k=0}^{N^*} \mathcal{R}^k P^{k+1} \boldsymbol{e} - (1 - w_0)| < \epsilon,$$

where ϵ is a pre-determined quantity such as 10^{-7}.

3.2.4 System performance measures

In this section we list a number of key system performance measures to bring out the qualitative aspects of the model under study. These are listed below along with their formula for computation. Towards this end, we further partition the vectors $\boldsymbol{x}^*(i)$ and $\boldsymbol{x}(n)$ into smaller vectors as follows:

$$\boldsymbol{x}^*(i) = (\ \boldsymbol{x}^*_{j,0,i_2}(i)\ ),\ i = 0, \ldots, c-1; j = 0, \ldots, i;$$

$$\boldsymbol{x}(n) = (\ x_{c,m,i_1,i_2}(n)\ ),\ n \geqslant 0;\ 0 \leqslant m \leqslant c;\ 0 \leqslant i_1,\ i_2,\ i_1 + i_2 \leqslant\ K;$$

Note that $\boldsymbol{x}^*(i)$ are of dimension $(i+1)(K+1)$ and $\boldsymbol{x}(n)$ are of dimension LC,

1. **The probability that all servers are idle** :

$$P_{idle} = \boldsymbol{x}^*(0)\boldsymbol{e}.$$

2. **The probability that an interrupted customer is lost**:

$$P_{loss} = \frac{\theta}{\theta+\mu}\sum_{i=1}^{c-1}\sum_{j=1}^{i} x^*_{j,0,K}(i) + \frac{\theta}{\theta+\mu}\sum_{n=0}^{\infty}\sum_{m=1}^{c} x_{c,m,0,K}(n).$$

3. **Mean number of idle servers** :

$$\mu_{IDS} = \sum_{i=0}^{c-1}(c-i)\ \boldsymbol{x}^*(i)\boldsymbol{e}.$$

4. **Mean number of busy servers** :

$$\mu_{BYS} = \sum_{i=1}^{c-1} i\ \boldsymbol{x}^*(i)\boldsymbol{e} + c\ \boldsymbol{x}(0)(I-\mathcal{R})^{-1}\boldsymbol{e}.$$

5. **Mean number of servers busy with primary customers** :

$$\mu_{SBYP} = \sum_{i=1}^{c-1}\sum_{j=1}^{i}\sum_{i_2=0}^{K} j\ x^*_{j,0,i_2}(i) + \sum_{n=0}^{\infty}\sum_{m=1}^{c}\sum_{i_1=0}^{K}\sum_{i_2=0}^{K-i_1} m\ x_{c,m,i_1,i_2}(n).$$

6. **Mean number of servers busy with** BIC **customers** :

$$\mu_{SBYI} = \sum_{i=1}^{c-1}\sum_{j=0}^{i-1}\sum_{i_2=0}^{K}(i-j)x^*_{j,0,i_2}(i) + \sum_{n=0}^{\infty}\sum_{m=0}^{c-1}\sum_{i_1=0}^{K}\sum_{i_2=0}^{K-i_1}(c-m)x_{c,m,i_1,i_2}(n).$$

7. **Mean number of primary customers in the queue**:

$$\mu_{PQ} = \boldsymbol{x}(0)\mathcal{R}(I-\mathcal{R})^{-2}\boldsymbol{e}.$$

8. **The mean number of interrupted customers in the** BIP:

$$\mu_{BIP} = \sum_{i=0}^{c-1}\sum_{j=0}^{i}\sum_{i_2=0}^{K} i_2\ x^*_{j,0,i_2}(i) + \sum_{n=0}^{\infty}\sum_{m=0}^{c}\sum_{i_2=0}^{K}\sum_{i_1=0}^{K-i_2} i_2\ x_{c,m,i_1,i_2}(n).$$

9. **The mean number of interrupted customers in the** BIC:

$$\mu_{BIC} = \sum_{n=0}^{\infty}\sum_{m=0}^{c}\sum_{i_1=0}^{K}\sum_{i_2=0}^{K-i_1} i_1\ x_{c,m,i_1,i_2}(n).$$

10. **The mean waiting time in the queue** E_W^Q**, is as given in** (3.20).

3.2.5 An Optimization Problem

In this section we propose an optimization problem and discuss it through illustrative examples 3.3.2 and 3.3.3 in section 3.3. To construct an objective function we assume that customer induced interruptions produce revenue to the system in contrast to server induced interruptions . Interrupted customers have to pay more cost than those without interruption. Also idle servers, loss of customers and waiting spaces in primary queue and BIC involve expenditure to the system. Thus we introduce per unit time revenue and cost as follows.

- revenue r_1 monetary units per customer leaving the system with an uninterrupted service,
- revenue $r_2(> r_1)$ monetary units per customer leaving the system on completion of service after an interruption,
- holding cost c_1 monetary units per unit time that a customer has to wait in the primary queue,
- holding cost c_2 monetary units per unit time that a customer has to wait in the BIC buffer,
- cost c_3 monetary units per unit time that each customer lost due to BIP buffer being full at the time an interruption occurs.
- cost c_4 monetary units per unit time for each idle servers,

The problem of interest is to find an optimum value the number of servers c to be employed and optimum value for K (when all other parameters are fixed) that maximizes the expected total profit $\boldsymbol{ETP}$, as given in the following objective function.

$$\boldsymbol{ETP} = r_1\mu_{SBYP} + r_2\mu_{SBYI} - c_1\mu_{PQ} - c_2\mu_{BIC} - c_3(\theta+\mu)P_{loss} - c_4\mu_{IDS}. \tag{3.23}$$

3.3 Numerical Illustrations

Now we present numerical results for implementing the qualitative nature of the model under study. The correctness and the accuracy of the code are verified by a number of accuracy checks. We consider a few representative examples.

Example 3.3.1. The purpose of this example is to see the impact of parameter θ for the case when $c = K = 2, 4, 6, 8$ on some measures. In this example, by fixing $\lambda = 15, \mu = 8$ and $\eta = 2$, we look at the effect of varying θ on some selected measures. These are displayed in Figure 3.2 and Figure 3.3. Looking at these figures, we summarize the following observations.

- As θ increases, the traffic intensity ρ, appears to decrease for all values of c and K. The rate of decrease is small for higher values of c and K. ρ is largest for the case when $c = K = 2$. This is as expected since increasing θ will cause an increase in the customers getting lost due to BIP being full for small values of c and K and for higher values of c and K, that is with more servers and more waiting space in BIP, help to clear the customers at a faster rate. When θ is progressively decreased and comes closer and closer to zero,our model converges to the classical queueing problem without interruption. Thus the ratio $\frac{\pi A_0 e}{\pi A_2 e}$ converges to the traffic intensity ρ of the classical situation.

- As is to be expected the measure P_{idle} is a non-decreasing function of θ when all other parameters are fixed.
- From Figure 3.3 we see that P_{loss} increases with increase in the interruption rate θ and the rate of increase is small for higher values of c and K, of course this is as expected.
- From Figure 3.3 it is seen that P_{loss} decrease with increase in the BIP size K for every c fixed, this is as expected. Also for each fixed K, this measure increases as c increases. This is as expected, since for a fixed K, as c increases, more customers may get interrupted from different servers and as a consequence the BIP gets filled (note that $\eta = 2$).
- We notice from Figure 3.4 that the measure E_W^Q decreases with increase in θ. This measure is largest for the case when $c = K = 2$ and for higher values of c and K, it is quite negligible as to be expected.

Now we discuss two optimization problems associate with Section 3.2.5.

Example 3.3.2. In this example, we fix $K = 5, \lambda = 20, \mu = 11, \eta = 5, r_1 = \$300, r_2 = \$400, c_1 = \$10, c_2 = \$20, c_3 = \$30, c_4 = \$5$. The optimal number of servers, c, that maximizes the expected total profit $\boldsymbol{ETP}$, for various combinations of θ are displayed in Figure 3.5.

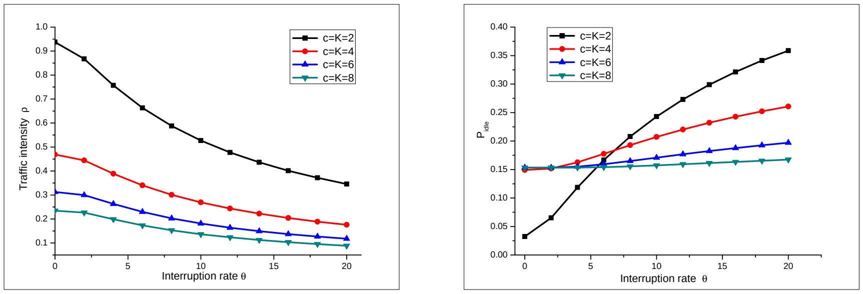


Figure 3.2: θ versus ρ and θ versus P_{idle}

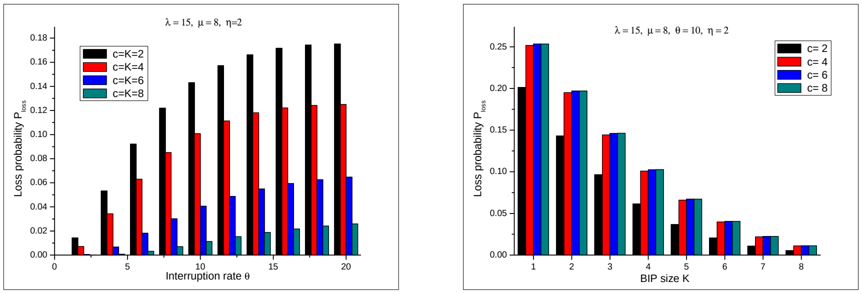


Figure 3.3: θ versus P_{loss} and K versus P_{loss}

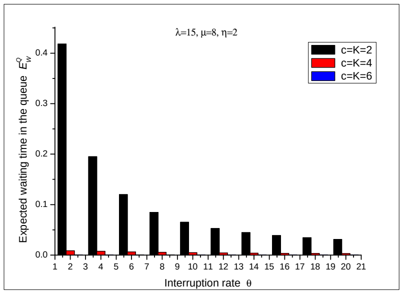


Figure 3.4: θ versus Expected waiting time in the queue

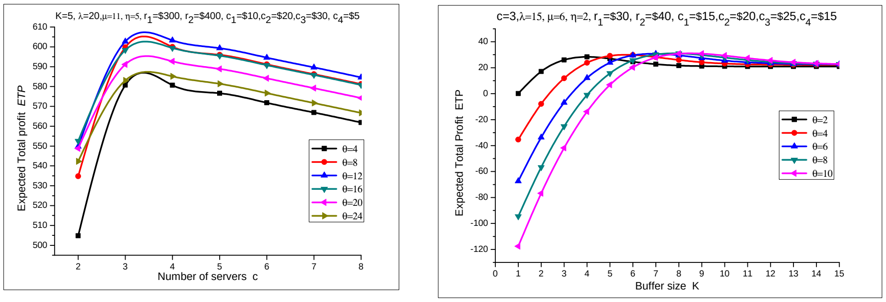


Figure 3.5: Optimum values of c and K for different θ

It is seen from the numerical experiments that $\boldsymbol{ETP}$ increases first and then decreases with increasing θ. The optimum c and the corresponding $\boldsymbol{ETP}$ are given in Tab. 3.1.

Table 3.1: Optimum c and $\boldsymbol{ETP}$ for selected θ

θ	4	8	12	16	20	24
Optimum c	3	4	4	4	4	4
$\boldsymbol{ETP}$	580.762	599.969	603.297	599.414	592.693	585.104

Example 3.3.3. Here we fix $c = 3, \lambda = 15, \mu = 6, \eta = 2, r_1 = \$30, r_2 = \$40, c_1 = \$15, c_2 = \$20, c_3 = \$25, c_4 = \$15$. The optimum value of K that maximizes the expected total profit $\boldsymbol{ETP}$, for various combinations of θ are displayed in Figure 3.5. The optimum K and the corresponding $\boldsymbol{ETP}$ are given in Tab. 3.2.

Table 3.2: Optimum K and $\boldsymbol{ETP}$ for selected θ

θ	2	4	6	8	10
Optimum K	3	5	6	7	8
$\boldsymbol{ETP}$	26.90732	29.35727	30.31224	30.78644	31.08506

Example 3.3.4. In this example we fix $\lambda = 15, \mu = 7.5, \eta = 2$ and vary the parameters c, K and θ. In Tab. 3.3 and Tab. 3.4, we display the measures $\mu_{IDS}, \mu_{BYS}, \mu_{SBYP}, \mu_{SBYI}, \mu_{PQ}, \mu_{BIC}$ and μ_{BIP}. A look at these tables reveal some notable observations.

- For each fixed pair c and K, μ_{IDS} increases and μ_{BYS} decreases as θ increases. This is due to the fact that an increase in θ will cause more customers to be interrupted from different servers leading to an increase in the number of customers leaving the system without getting service.

- For each fixed pair c and K, μ_{SBYI} is a non-decreasing function of θ whereas μ_{SBYP} is a non-increasing function of θ. This is again as expected.

- The measure μ_{PQ} is a non-increasing function of θ for all values of c and K and is largest for the case when $c = 2$ and $K = 5$. This is to be expected since increase in θ results in interrupted customers, for lower values of K, getting lost and for higher values of K they get back to service through BIC buffer.

- Finally, looking at the measures μ_{BIC} and μ_{BIP}, we see some interesting trends. Recall that at any given time the total number of customers in the BIC and BIP buffers cannot exceed K. For all values of c and K, $\mu_{BIC} < \mu_{BIP}$ when θ increases. For higher interruption rate causes more interruption leading to more interrupted customers filling BIP buffer (note that η=2) and hence the rate of interrupted customers getting back to service through BIC buffer will be smaller leading to less customers (on the average) in BIC

buffer. Also we notice that for each values of c, μ_{BIC} increases initially and then decreases as θ increases further, for higher value K. This is probably due to the fact that as θ reaches a certain value, any further increase in θ will only result in the server being busy with customers in BIC, for higher values of K.

We conclude this section by showing that the mean number of servers busy with primary customers, μ_{SBYP}, is independent of K and c. We are able to prove this only for the case when $c = 1$. Even though the result appear to be true in general, which we verified through numerical computation as we can see in Tab. 3.3 and Tab. 3.4.

Theorem 3.3.1. *The server is busy with primary customers is given by*

$$P_{BSYP} = \frac{\lambda}{\theta + \mu}.$$

Proof. The steady-state equations given in (3.13) can be written as

$$\boldsymbol{x}^*(0)E_0 + \boldsymbol{x}(0)B_1 = 0, \tag{3.24}$$

$$\boldsymbol{x}^*(0)C_0 + \boldsymbol{x}(0)A_1 + \boldsymbol{x}(1)A_2 = 0, \tag{3.25}$$

and

$$\boldsymbol{x}(i-1)A_0 + \boldsymbol{x}(i)A_1 + \boldsymbol{x}(i+1)A_2 = 0, i \geqslant 1. \tag{3.26}$$

Post-multiplying equations (3.24) through (3.26) by $\boldsymbol{e}$ of appropriate dimensions, we get

$$\lambda \boldsymbol{x}^*(0)\boldsymbol{e} + \eta \sum_{r=0}^{K} r x^*_{0,0,r}(0) = \mu \boldsymbol{x}_{1,0,0}(0)\boldsymbol{e} + (\mu + \theta)\boldsymbol{x}_{1,1,0}(0)\boldsymbol{e}, \tag{3.27}$$

and

$$\lambda(\boldsymbol{x}_{1,0}(i)\boldsymbol{e}+\boldsymbol{x}_{1,1}(i)\boldsymbol{e}) = \mu\boldsymbol{x}_{1,0,0}(i+1)\boldsymbol{e}+(\mu+\theta)\boldsymbol{x}_{1,1,0}(i+1)\boldsymbol{e},\ i \geqslant 0. \quad (3.28)$$

Now post-multiplying equations (3.25) and (3.26) by $(\boldsymbol{e}_1(2)\otimes\boldsymbol{e})$ and adding over $i = 1$ to ∞, we get

$$\lambda\boldsymbol{x}^*(0)\boldsymbol{e} = (\mu+\theta)\sum_{i=0}^{\infty}\boldsymbol{x}_{1,1}(i)\boldsymbol{e} - \mu\sum_{i=1}^{\infty}\boldsymbol{x}_{1,0,0}(i)\boldsymbol{e} - (\mu+\theta)\sum_{i=1}^{\infty}\boldsymbol{x}_{1,1,0}(i)\boldsymbol{e}. \quad (3.29)$$

The stated result follows by immediately by adding (3.28) over i and (3.29). □

Table 3.3: Effect of K and θ on some selected measures when $c = 2, 4$

K	θ	μ_{IDS}	μ_{BYS}	μ_{SBYP}	μ_{SBYI}	μ_{PQ}	μ_{BIC}	μ_{BIP}
$c = 2$								
1	2	0.2613	1.7387	1.5789	0.1597	5.2553	0.05010	0.5989
	5	0.6037	1.3963	1.2000	0.1963	1.3087	0.0329	0.7362
	20	1.2313	0.7687	0.5455	0.2232	0.1479	0.0068	0.8371
2	2	0.1438	1.8562	1.5789	0.2773	11.1043	0.1053	1.0399
	5	0.4341	1.5659	1.2000	0.3659	2.3990	0.0835	1.3719
	20	1.0236	0.9764	0.5455	0.4309	0.3521	0.0268	1.6161
3	2	0.0682	1.9318	1.5789	0.3528	25.8552	0.1528	1.3232
	5	0.2958	1.7042	1.2000	0.5042	4.2736	0.1477	1.8906
	20	0.8355	1.1645	0.5455	0.6190	0.6785	0.0655	2.3213
4	2	0.0272	1.9728	1.5789	0.3938	69.0971	0.1846	1.4769
	5	0.1898	1.8102	1.2000	0.6102	7.6242	0.2176	2.2883
	20	0.6704	1.3296	0.5455	0.7842	1.1503	0.1250	2.9406
5	2	0.0091	1.9909	1.5789	0.4120	215.1687	0.2011	1.5450
	5	0.1135	1.8865	1.2000	0.6865	14.0691	0.2844	2.5745
	20	0.5298	1.4702	0.5455	0.9248	1.7933	0.2036	3.4678
$c = 4$								
2	2	2.1376	1.8624	1.5789	0.2835	0.1116	0.0074	1.0631
	5	2.4240	1.5759	1.2000	0.3759	0.0480	0.0044	1.4099
	20	3.0147	0.9854	0.5455	0.4399	0.0049	0.0006	1.6499
3	2	2.0621	1.9380	1.5789	0.3591	0.1339	0.0113	1.3466
	5	2.2778	1.7223	1.2000	0.5223	0.0706	0.0088	1.9584
	20	2.8137	1.1863	0.5455	0.6408	0.0115	0.0018	2.4031
4	2	2.0232	1.9768	1.5789	0.3979	0.1483	0.0141	1.4919
	5	2.1659	1.8341	1.2000	0.6341	0.0945	0.0140	2.3779
	20	2.6304	1.3696	0.5455	0.8242	0.0227	0.0044	3.0907
5	2	2.0072	1.9928	1.5789	0.4138	0.1556	0.0156	1.5519
	5	2.0890	1.9110	1.2000	0.7109	0.1163	0.0192	2.6662
	20	2.4684	1.5316	0.5455	0.9862	0.0387	0.0089	3.6982
6	2	2.0019	1.9981	1.5789	0.4192	0.1585	0.0162	1.5718
	5	2.0424	1.9576	1.2000	0.7576	0.1331	0.0235	2.8409
	20	2.3312	1.6689	0.5455	1.1234	0.0592	0.0149	4.2128

Table 3.4: Effect of K and θ on some selected measures when $c = 6, 8$

K	θ	μ_{IDS}	μ_{BYS}	μ_{SBYP}	μ_{SBYI}	μ_{PQ}	μ_{BIC}	μ_{BIP}
$c = 6$								
2	2	4.1372	1.8628	1.5789	0.2838	0.00501	0.00040	1.0644
	5	4.4236	1.5765	1.2000	0.3765	0.00161	0.00017	1.4117
	20	5.0144	0.9856	0.5455	0.4401	0.00006	0.00001	1.6505
4	2	4.0230	1.9769	1.5789	0.3980	0.00721	0.00089	1.4926
	5	4.1649	1.8351	1.2000	0.6351	0.00384	0.00069	2.3815
	20	4.6292	1.3708	0.5455	0.8253	0.00045	0.00010	3.0949
5	2	4.0072	1.9929	1.5789	0.4139	0.00768	0.00102	1.5522
	5	4.0881	1.9119	1.2000	0.7119	0.00502	0.00105	2.6697
	20	4.4664	1.5336	0.5455	0.9882	0.00092	0.00024	3.7056
6	2	4.0019	1.9981	1.5789	0.4192	0.00788	0.00108	1.5719
	5	4.0418	1.9582	1.2000	0.7582	0.00600	0.00137	2.8434
	20	4.3284	1.6716	0.5455	1.1262	0.00165	0.00050	4.2232
8	2	4.0001	1.9999	1.5789	0.4210	0.00796	0.00111	1.5786
	5	4.0065	1.9935	1.2000	0.7935	0.00708	0.00178	2.9756
	20	4.1347	1.8653	0.5455	1.3198	0.00372	0.00135	4.9493
$c = 8$								
4	2	6.0230	1.9770	1.5789	0.3981	0.00029	0.00004	1.4927
	5	6.1649	1.8351	1.2000	0.6351	0.00012	0.00002	2.3817
	20	6.6292	1.3708	0.5455	0.8253	0.00001	0.00000	3.0949
5	2	6.0071	1.9929	1.5789	0.4139	0.00031	0.00005	1.5522
	5	6.0881	1.9119	1.2000	0.7119	0.00017	0.00004	2.6698
	20	6.4664	1.5337	0.5455	0.9882	0.00002	0.00000	3.7058
7	2	6.0004	1.9996	1.5789	0.4206	0.00033	0.00005	1.5774
	5	6.0175	1.9825	1.2000	0.7825	0.00026	0.00007	2.9344
	20	6.2175	1.7825	0.5455	1.2370	0.00006	0.00002	4.6388
8	2	6.0001	1.9999	1.5789	0.4209	0.00033	0.00005	1.5786
	5	6.0065	1.9935	1.2000	0.7935	0.00028	0.00008	2.9756
	20	6.1346	1.8654	0.5455	1.3199	0.00010	0.00004	4.9498
10	2	6.0000	2.0000	1.5789	0.4211	0.00033	0.00005	1.5789
	5	6.0007	1.9994	1.2000	0.7994	0.00030	0.00009	2.9976
	20	6.0409	1.9591	0.5455	1.4136	0.00018	0.00008	5.3010

Chapter 4

A Multi-server Queueing System with Service Interruption, Partial Protection and Repetition of Service

In this chapter we propose to study a very general queueing model with customer induced interruption. The customer arrival process is assumed to be a MAP thereby bringing in an inbuilt correlation structure. Further

Some results of this chapter are included in the following paper.
Dudin, A. N., Varghese Jacob and Krishnamoorthy, A. : A Multi-Server Queueing System with Service Interruption, Partial Protection and Repetition of service. To appear in Annals of Operations Research, Springer, 2013.

a stream of interruption arrivals is also introduced that removes the customer in service if the service is not in a protected phase and if he is not already a once interrupted customer. We retain the priority of interrupted customers over fresh arrivals for service. In the next chapter we consider a problem in which self-interruption is not encouraged; such customers are sent to an orbit with limited capacity to retry in case they are able to get into it.

Motivated by a few real life situations we consider a queueing system with customer induced service interruption, partial protection and repetition of service. For instance, the model fit in are : (i). Production inventory : A batch of items (row materials) is being processed. During the processing this batch is seen not to satisfy certain specifications. Then its further processing is deferred. If it is not rectifiable the entire batch is discarded. During the intervening time the production process goes on without any hindrance. (ii). Patients admitted to hospitals for surgery form is another typical example for customer oriented interruptions. While a patient is being prepared for surgery, his blood pressure may shoot up resulting in postponing the surgery. (iii). Administration of antibiotics : A patient is asked to take antibiotics for a specified period for an ailment. During the course of the treatment if he develops some other health problems for which the antibiotics has severe reaction, then the course of the antibiotics is terminated temporarily until he relapses completely from the newly developed problem. (iv). Ayurveda - ancient Indian system of medicine: A number of treatment procedures in this system comes under services (treatment) with customer (patient) induced interruption, see the book *Chopra* [16]. (v). In production process, especially of expensive commodities, it is essential to give some sort of protection at the time of manufacture. There are certain stages in the manufacturing process wherein the equipment used itself is a very expensive item (as in case of surgery) which has to be protected from variations in power supply for example: The voltage fluctuation can be considered as arrival of interruption; this can affect the customer being served or even the server. Thus protection from breakdown of service/collapse of customer has to be ensured.

The probability p of rejection of customer being served in unprotected phase can be explained as follows: when service is in an unprotected stage, the customer (an item in production) can be badly affected with probability p. This item may not be worth further processing. Thus it is rejected for which we have to have p as the rejection probability.

This chapter is arranged as follows. In Section 4.1 the model under study is mathematically formulated. In Section 4.2 we obtain the system-state and the infinitesimal generator matrix. Section 4.3 provides the steady-state analysis of the model. In Section 4.4 we provide a number of system performance measures of interest. Numerical results are discussed in section 4.5.

4.1 Model description

We consider a queueing model consisting of two queueing systems. One system (we refer it as primary system or system-1) is a c-server queueing system with an infinite buffer. The servers are identical and independent of each other. Ordinary (primary, or type-1) customers arrive to this

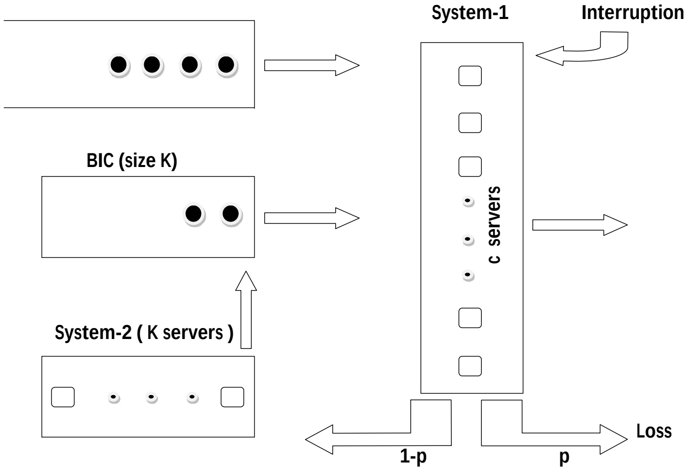


Figure 4.1: A multi-server queueing system with service interruption

system according to a MAP. In a MAP, the customers arrival is directed by an irreducible $CTMC$ ν_t, $t \geqslant 0$, with the state space $\{0, 1, \ldots, W\}$. The intensities of transition of the Markov chain ν_t, $t \geqslant 0$, which are accompanied by arrival of k customers, are described by the matrices D_k, $k = 0, 1$. Vector $\boldsymbol{\theta}$ of the stationary distribution of the process ν_t is the unique solution to the system $\boldsymbol{\theta}(D_0 + D_1) = \mathbf{0}$, $\boldsymbol{\theta}\boldsymbol{e} = 1$. Fundamental rate λ of MAP is given by $\lambda = \boldsymbol{\theta} D_1 \boldsymbol{e}$.

If an arriving primary customer meets available servers at station-1, it immediately occupies one free server and starts getting service. Otherwise, this customer is placed into the buffer of infinite capacity and he/she will be picked up for service according to the $FIFO$ discipline. It is assumed that the service time of a primary customer by a server has a

PH distribution with an irreducible representation $(\boldsymbol{\beta}^{(1)}, S^{(1)})$. It means the following. Service time is interpreted as the time until the $CTMC$ $\eta_t^{(1)}$, $t \geqslant 0$, with the state space (set of phases) $\{1, \dots, M_1+1\}$ reaches the single absorbing state (phase) M_1+1. Transitions of the chain $\eta_t^{(1)}$, $t \geqslant 0$, within the state space $\{1, \dots, M_1\}$ are defined by the sub-generator $S^{(1)}$ while the intensities of transition into the absorbing state are defined by the vector $\mathbf{S}_0^{(1)} = -S^{(1)}\boldsymbol{e}$. At the service beginning epoch, the state of the process $\eta_t^{(1)}$, $t \geqslant 0$, is chosen within the state space $\{1, \dots, M_1\}$ according to the row vector of probabilities $\boldsymbol{\beta}^{(1)} = (\beta_1^{(1)}, \dots, \beta_{M_1}^{(1)})$. It is assumed that the matrix $S^{(1)} + \mathbf{S}_0^{(1)}\boldsymbol{\beta}^{(1)}$ is an irreducible one. The mean service time is computed as $b_1^{(1)} = \boldsymbol{\beta}^{(1)}(-S^{(1)})^{-1}\boldsymbol{e}$. The mean intensity μ_1 of the service is given by $\mu_1 = (b_1^{(1)})^{-1}$.

Service of a primary customer can be interrupted. Interruptions (or negative customers) arrive to the system according to a MAP. A MAP is defined by the state space $\{0, 1, \dots, Z\}$ of underlying process $\zeta_t,\ t \geqslant 0$, and by the matrices H_0 and H_1. Fundamental rate h of MAP is given by $h = \boldsymbol{\sigma} H_1 \boldsymbol{e}$ where the row vector $\boldsymbol{\sigma}$ is the unique solution to the system $\boldsymbol{\sigma}(H_0 + H_1) = \mathbf{0},\ \boldsymbol{\sigma}\boldsymbol{e} = 1$.

An arriving interruption with equal probability is directed to any server of system-1. If the selected server is idle upon the interruption arrival, this interruption has no effect on the system. If the selected server is providing a service to a primary customer, the service can be interrupted or not interrupted. Following to *Klimenok and Dudin* [36], we suppose that there exists a set of phases of service process that are protected from the effect of arriving interruptions. Without loss of generality we define this set as $\{m_1+1, \dots, M_1\}$. If the state of PH service process of a primary customer, which occupies the server, belongs to the set $\{1, \dots, m_1\}$ the negative customer interrupts the service of such a primary customer. In the opposite case, the primary customer is considered to be protected and the interruption leaves the system without any effect.

A primary customer, service of which is interrupted, leaves the system permanently with probability p, and he/she will considered as lost customer. With complementary probability $1 - p$, the primary customer moves for the service to system-2. This system consists of K independent identical servers and has no buffer (we refer it as BIP). So, if all K servers are busy at the moment of a primary customer interruption, this customer will be lost. Otherwise, the primary customer starts the service at an arbitrary idle server of BIP.

It is assumed that the service time of a primary customer by a server of system-2 has PH distribution with an irreducible representation $(\boldsymbol{\alpha}, T)$. It is directed by the $CTMC$ ϕ_t, $t \geqslant 0$, with the set of phases $\{1, \dots, R, R+1\}$ where $R+1$ is the single absorbing state. We denote $\mathbf{T}_0 = -T\boldsymbol{e}$. Intensity μ of service at system-2 is defined by $\mu^{-1} = \boldsymbol{\alpha}(-T)^{-1}\boldsymbol{e}$.

Upon completion of the service at system-2, the customer becomes priority or type-2 customer. If, at this service completion moment, there are free servers at system-1, the type-2 customer immediately starts the service at system-1. It is assumed that the service time of a priority customer by a server of system-1 has a PH type distribution with an irreducible representation $(\boldsymbol{\beta}^{(2)}, S^{(2)})$. It is directed by the $CTMC$ $\eta_t^{(2)}$, $t \geqslant 0$, with the state space (set of phases) $\{1, \dots, M_2+1\}$ and is interpreted as the time until this chain reaches the single absorbing state (phase) $M_2 + 1$. We denote $\mathbf{S}_0^{(2)} = -S^{(2)}\boldsymbol{e}$. The mean service time is computed as $b_1^{(2)} = \boldsymbol{\beta}^{(2)}(-S^{(2)})^{-1}\boldsymbol{e}$. The mean intensity μ_2 of the service is given by

$\mu_2 = (b_1^{(2)})^{-1}$. Service of a priority customer can not be interrupted. So, if an arriving interruption is directed to the server providing a service to priority customer, this interruption leaves the system without any effect.

If, at the moment when the primary customer finished the service at BIP and becomes priority customer, there are no available servers at system-1, the type-2 customer is placed into the finite buffer for priority customers of capacity K (we refer it as BIC). He will be picked up for the service according to the $FIFO$ discipline. If some server of system-1 becomes free, it takes for the service a priority customer if any from BIC. So, type-1 customers are picked-up from the infinite buffer only if BIC is empty at the service completion moment at system-1.

We analyze performance measures of the described queueing model.

4.2 The Process of the System States

Let

- $i_t,\ i_t \geqslant 0$, be the number of primary customers in system-1;
- $n_t,\ n_t \in \{0, \dots, c\}$, be the number of priority customers in service at system-1;
- $k_t \in \{0, \dots, K\}$, be the number of priority customers in the BIC;
- $j_t \in \{0, \dots, K - k_t\}$, be the number of customers in BIP;
- $\eta_n^{(2)}(t) \in \{1, \dots, M_2\}$, be the state of PH underlying process in the

nth server among the servers of system-1 providing the service to type-2 customer, $n \in \{1, \dots n_t\}$;

- $\eta_n^{(1)}(t) \in \{1, \dots, M_1\}$, be the state of PH service process by the nth server among the servers of system-1 providing the service to type-1 customer, $n \in \{1, \dots, \min\{i_t, c - n_t\}\}$;
- $\phi_j(t),\ \phi_j(t) \in \{1, \dots, R\},\ j \in \{1, \dots, j_t\}$, be the state of PH service process by the jth server of BIP at the epoch $t,\ t \geqslant 0$.
- ν_t, $\nu_t \in \{0, \dots, W\}$, be the state of MAP of primary customer arrival;
- ζ_t, $\zeta_t \in \{0, \dots, Z\}$, be the state of the interruptions process according to a MAP.

It is obvious that the process

$$\xi_t = \{i_t, n_t, k_t, j_t, \eta_1^{(2)}(t), \dots, \eta_{n_t}^{(2)}(t), \eta_1^{(1)}(t), \dots, \eta_{\min\{i_t, c-n_t\}}^{(1)}(t),$$
$$\phi_1(t), \dots, \phi_{j_t}(t), \nu_t, \zeta_t\},$$

$t \geqslant 0$, is an irreducible regular $CTMC$.

Note that $k_t = 0$ if $i_t + n_t < c$.

Let us enumerate the states of the components

$$\{\eta_1^{(2)}(t), \dots, \eta_{n_t}^{(2)}(t), \eta_1^{(1)}(t), \dots, \eta_{\min\{i, c-n_t\}}^{(1)}(t), \phi_1(t), \dots, \phi_{j_t}(t), \nu_t, \zeta_t\}$$

of this Markov chain in lexicographical order and form the so called macro-states $\{i_t, n_t, k_t, j_t\}$ from the corresponding states of the Markov chain ξ_t.

In the sequel we will use the following notations:

- $\bar{I}_{M_1}$ is a diagonal matrix of size M_1 having zeros as the first m_1 diagonal entries and 1's as the rest of diagonal;
- $\hat{\boldsymbol{e}}$ is a column vector of size M_1, having 1's as the first m_1 entries and zeros as the rest entries;
- $\bar{W} = W + 1,\ \ \bar{Z} = Z + 1.$

Analyzing transitions of the Markov chain ξ_t during an interval having an infinitesimal length, we can compute the matrices defining transition rates of this chain.

The transition rates of the Markov chain ξ_t without the change of the macro-state (i, n, k, j) are defined by formula

$$(S^{(2)})^{\oplus n} \oplus (S^{(1)})^{\oplus \min\{i,c-n\}} \oplus T^{\oplus j} \oplus D_0 \oplus H_0$$

$$+\frac{1}{c} I_{M_2^n} \otimes \left[\bar{I}_{M_1}^{\oplus \min\{i,c-n\}} + (c-i-n)\chi(c-i-n) I_{M_1}^{\otimes i} + n I_{M_1}^{\otimes \min\{i,c-n\}} \right] \otimes I_{R^j} \otimes I_{\bar{W}} \otimes H_1. \tag{4.1}$$

The first term in (4.1) corresponds to the possible transitions of the service phase of priority and non-priority customers by system-1, service by system-2, and the underlying processes of arrival of customers and interruptions that do not lead to service completion or arrival. The second term in (4.1) corresponds to the case when interruption occurs but it is ignored by the system (because the interruption selected the server which provides the service in protected phase for a primary customer or an idle server or one who is providing service for a priority customer). Note that the diagonal entries of the matrix defined by formula (4.1) are negative and define, up to the sign, the intensity of leaving the corresponding state of the Markov chain ξ_t.

The transition rates from the macro-state (i, n, k, j) to other macro-states are given in following tables Tab. 4.1 to Tab. 4.3. The first column defines the state, to which a transition can occur, the second column explains condition when this transition occurs. The third column contains the block matrix defining the rate of the corresponding transition.

Let us enumerate the macro-states (i, n, k, j) in the lexicographical order of components (k, j) and form the macro-states (i, n).

Let Q be the infinitesimal generator of the Markov chain $\xi_t,\ t \geqslant 0$, consisting of blocks $Q_{i,l}$, which consists of the matrices $(Q_{i,l})_{n,n'}$ defining the intensity of transition in this Markov chain from the macro-state (i, n) to the macro-state $(l, n'),\ n, n' \in \{0, \dots, c\}$.

To write down the generator Q, we introduce additionally the following notation:

- $\tilde{R}_k = \frac{R^{k+1}-1}{R-1}$, if $R \neq 1$ and $\tilde{R}_k = k+1$ if $R = 1,\ k \in \{0, \dots, K\}$,
- $\tilde{R} = \sum\limits_{k=0}^{K} \tilde{R}_k;$
- $\tilde{N} = \sum\limits_{n=0}^{c} M_1^{c-n} M_2^n \tilde{R} \bar{W} \bar{Z}, \quad \tilde{N}_i = \sum\limits_{n=0}^{c} M_1^i M_2^n \tilde{R}_K \bar{W} \bar{Z}, \quad i \in \{0, \dots, c-1\};$
- $\mathcal{T} = diag\{\{T^{\oplus 0}, T^{\oplus 1}, \dots, T^{\oplus (K-k)}\},\ k \in \{0, \dots, K\}\};$
- $\mathcal{T}_0 = diag\{T^{\oplus 0}, T^{\oplus 1}, \dots, T^{\oplus K}\};$

$(i{+}1,n,0,j)$	arrival of type-1 customer	$I_{M_2^n}\otimes I_{M_1^{c-n}}\otimes I_{R^j}\otimes D_1\otimes I_{\bar{Z}}$
$(i,n{-}1,0,j)$	service completion of type-2 customer	$(\mathbf{S}_0^{(2)})^{\oplus n}\otimes I_{M_1^{c-n}}\otimes\boldsymbol{\beta}^{(1)}\otimes I_{R^j}\otimes I_{\bar{W}}\otimes I_{\bar{Z}}$
$(i{-}1,n,0,j)$	service completion of type-1 customer	$I_{M_2^n}\otimes(\mathbf{S}_0^{(1)}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)}\otimes I_{R^j}\otimes I_{\bar{W}}\otimes I_{\bar{Z}},\ i>c{-}n,$ $I_{M_2^n}\otimes(\mathbf{S}_0^{(1)})^{\oplus(c-n)}\otimes I_{R^j}\otimes I_{\bar{W}}\otimes I_{\bar{Z}},\ i=c-n,$
$(i,n,1,j{-}1)$	service completion in system-2	$I_{M_2^n}\otimes I_{M_1^{c-n}}\otimes\mathbf{T}_0^{\oplus j}\otimes I_{\bar{W}}\otimes I_{\bar{Z}}$
$(i{-}1,n,0,j)$	interruption arrival to the server at unprotected phase and loss of a customer	$\frac{p}{c}I_{M_2^n}\otimes(\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)}\otimes I_{R^j}\otimes I_{\bar{W}}\otimes H_1$
$(i{-}1,n,0,j{+}1)$ $j<K$	interruption arrival to the server at unprotected phase and move of a customer to system-2	$\frac{1-p}{c}I_{M_2^n}\otimes(\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)}\otimes I_{R^j}\otimes\boldsymbol{\alpha}\otimes I_{\bar{W}}\otimes H_1$
$(i{-}1,n,0,K)$ $j=K$	interruption arrival to the server at unprotected phase and loss of a customer due to all servers in system-2 are busy	$\frac{1-p}{c}I_{M_2^n}\otimes(\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)}\otimes I_{R^K}\otimes I_{\bar{W}}\otimes H_1$

Table 4.1: The intensities of transitions from macro-state $(i,n,0,j),\ i\geqslant c-n$

$(i{+}1, n, k, j)$	arrival of type-1 customer	$I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes I_{R^j} \otimes D_1 \otimes I_{\bar{Z}}$
$(i, n, k{-}1, j)$	service completion of type-2 customer	$(\mathbf{S}_0^{(2)}\boldsymbol{\beta}^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i{-}1, n{+}1, k{-}1, j)$	service completion of type-1 customer	$(I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)}) \otimes (\mathbf{S}_0^{(1)})^{\oplus(c-n)} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i, n, k{+}1, j{-}1)$	service completion in system-2	$I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes \mathbf{T}_0^{\oplus j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i{-}1, n{+}1, k{-}1, j)$	interruption arrival to the server at unprotected phase and customer loss	$\frac{p}{c}(I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)}) \otimes \hat{\boldsymbol{e}}^{\oplus(c-n)} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes H_1$
$(i{-}1, n{+}1, k{-}1, j{+}1)$	interruption arrival to the server at unprotected phase and move of a customer to system-2	$\frac{1-p}{c}(I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)}) \otimes \hat{\boldsymbol{e}}^{\oplus(c-n)} \otimes I_{R^j} \otimes \boldsymbol{\alpha} \otimes I_{\bar{W}} \otimes H_1$

Table 4.2: The intensities of transitions from macro-state (i, n, k, j), $i \geqslant c-n$, $k \geqslant 1$

$(i{+}1, n, 0, j)$	arrival of type-1 customer	$I_{M_2^n} \otimes I_{M_1^i} \otimes \boldsymbol{\beta}^{(1)} \otimes I_{R^j} \otimes D_1 \otimes I_{\bar{Z}}$
$(i, n{-}1, 0, j)$	service completion of type-2 customer	$(\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^i} \otimes \boldsymbol{\beta}^{(1)} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i{-}1, n, 0, j)$	service completion of type-1 customer	$I_{M_2^n} \otimes (\mathbf{S}_0^{(1)})^{\oplus i} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i, n{+}1, 0, j{-}1)$	service completion in system-2	$I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes I_{M_1^i} \otimes \mathbf{T}_0^{\oplus j} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$
$(i{-}1, n, 0, j)$	interruption arrival to the server at unprotected phase and loss of a customer	$\frac{1}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes I_{R^j} \otimes I_{\bar{W}} \otimes H_1(\delta_{j,K} + p(1 - \delta_{j,K}))$
$(i{-}1, n, 0, j{+}1)$ $j < K$	interruption arrival to the server at unprotected phase and move of a customer to system-2	$\frac{1-p}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes I_{R^j} \otimes \boldsymbol{\alpha} \otimes I_{\bar{W}} \otimes H_1$

Table 4.3: The intensities of transitions from macro-state $(i, n, 0, j),\ i < c - n$

- $$F_k = \begin{pmatrix} \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \mathbf{T}_0^{\oplus 1} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & \mathbf{T}_0^{\oplus 2} & \dots & \mathrm{O} & \mathrm{O} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{O} & \mathrm{O} & \dots & \mathbf{T}_0^{\oplus k} & \mathrm{O} \end{pmatrix}, \ k \in \{1, \dots, K\};$$

- $$\mathcal{F} = \begin{pmatrix} \mathrm{O} & F_K & \mathrm{O} & \dots & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & F_{K-1} & \dots & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & F_1 \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} \end{pmatrix}, \ \ \mathcal{F}_0 = F_K;$$

- $$G_k = \begin{pmatrix} \mathrm{O} & \boldsymbol{\alpha} & \mathrm{O} & \dots & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & I_{R^1} \otimes \boldsymbol{\alpha} & \dots & \mathrm{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & I_{R^k} \otimes \boldsymbol{\alpha} \end{pmatrix}, \ k \in \{0, \dots, K-1\} \ ;$$

- $$G_K = \begin{pmatrix} \mathrm{O} & \boldsymbol{\alpha} & \mathrm{O} & \dots & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & I_{R^1} \otimes \boldsymbol{\alpha} & \dots & \mathrm{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & I_{R^{K-1}} \otimes \boldsymbol{\alpha} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & I_{R^K} \end{pmatrix};$$

- $$\mathcal{G} = \begin{pmatrix} \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ G_{K-1} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & G_{K-2} & \dots & \mathrm{O} & \mathrm{O} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{O} & \mathrm{O} & \dots & G_0 & \mathrm{O} \\ & & & & \end{pmatrix}, \;\; \mathcal{G}_0 = diag\{G_K, \mathrm{O}, \dots, \mathrm{O}\};$$

- $$\hat{I}_k = \left(I_{\tilde{R}_k} | \;\; \mathrm{O}_{\tilde{R}_k \times R^{k+1}} \right), \; k \in \{0, \dots, K-1\};$$

- $$I^- = \begin{pmatrix} \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \hat{I}_{K-1} & \mathrm{O} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \mathrm{O} & \hat{I}_{K-2} & \mathrm{O} & \dots & \mathrm{O} & \mathrm{O} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathrm{O} & \mathrm{O} & \mathrm{O} & \dots & \hat{I}_0 & \mathrm{O} \end{pmatrix};$$

- $$\tilde{I} = diag\{I_{\tilde{R}_K}, \mathrm{O}, \dots, \mathrm{O}\};$$

- $$E^- = \begin{pmatrix} I_{\tilde{R}_K} \\ \mathrm{O}_{\tilde{R}_{K-1} \times \tilde{R}_K} \\ \vdots \\ \mathrm{O}_{\tilde{R}_0 \times \tilde{R}_K} \end{pmatrix}, \;\; E^+ = (I_{\tilde{R}_K} | \mathrm{O}_{\tilde{R}_K \times \tilde{R}_{K-1}} | \dots | \mathrm{O}_{\tilde{R}_K \times \tilde{R}_0}).$$

Recall that, as it was mentioned above, the BIC is empty (the component k of the Markov chain is equal to 0) if the sum of the components i an n is less then c which means that there are free servers at station 1. When this sum is greater than or equal to c, the component k takes values from the set $\{0, \dots, K\}$. So, the macro-state (i, n) consists of $M_1^{c-n} M_2^n \tilde{R} \bar{W} \bar{Z}$ states if $i + n \geqslant c$ and it consists of only $M_1^i M_2^n \tilde{R}_K \bar{W} \bar{Z}$ states if $i + n < c$.

Theorem 4.2.1. *Generator Q has tri-diagonal block structure:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & \dots & O & O & O & O & \dots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \dots & O & O & O & O & \dots \\ O & Q_{2,1} & Q_{2,2} & \dots & O & O & O & O & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ O & O & O & \dots & Q_{c-1,c-1} & Q_{c-1,c} & O & O & \dots \\ O & O & O & \dots & Q_{c,c-1} & Q_{c,c} & Q_{c,c+1} & O & \dots \\ O & O & O & \dots & O & Q_{c+1,c} & Q_{c,c} & Q_{c,c+1} & \dots \\ O & O & O & \dots & O & O & Q_{c+1,c} & Q_{c,c} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix} \tag{4.2}$$

where the blocks $Q_{i,j}$ are defined as follows:

- *the block $Q_{c,c+1}$ is the diagonal matrix of order $\tilde{N}$ with diagonal blocks*

$$(Q_{c,c+1})_{n,n} = I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes I_{\tilde{R}} \otimes D_1 \otimes I_{\bar{Z}}, \quad n \in \{0, \dots, c\},$$

- *the block $Q_{c,c}$ is lower bidiagonal matrix of order $\tilde{N}$ having non-zero* *diagonal and sub-diagonal blocks defined by*

$$(Q_{c,c})_{n,n} = (\mathbf{S}_0^{(2)}\boldsymbol{\beta}^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I^{-} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes \mathcal{F} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$$

$$+ (S^{(2)})^{\oplus n} \oplus (S^{(1)})^{\oplus (c-n)} \oplus \mathcal{T} \oplus D_0 \oplus H_0 + \frac{1}{c} I_{M_2^n} \otimes \left[\bar{I}_{M_1}^{\oplus (c-n)} + n I_{M_1}^{\otimes (c-n)} \right]$$

$$\otimes I_{\tilde{R}} \otimes I_{\bar{W}} \otimes H_1, \quad n \in \{0, \ldots, c\},$$

$$(Q_{c,c})_{n,n-1} = (\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes \boldsymbol{\beta}^{(1)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}, \ \ n \in \{1, \ldots, c\};$$

- *the block $Q_{c+1,c}$ is upper bidiagonal matrix of order $\tilde{N}$ having non-zero diagonal and up-diagonal blocks defined by*

$$(Q_{c+1,c})_{n,n} = I_{M_2^n} \otimes (\mathbf{S}_0^{(1)}\boldsymbol{\beta}^{(1)})^{\oplus (c-n)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus (c-n)}$$

$$\otimes \tilde{I} \otimes I_{\bar{W}} \otimes H_1 + \frac{1-p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus (c-n)} \otimes \mathcal{G}_0 \otimes I_{\bar{W}} \otimes H_1, \ \ n \in \{0, \ldots, c\},$$

$$(Q_{c+1,c})_{n,n+1} = I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes (\mathbf{S}_0^{(1)})^{\oplus (c-n)} \otimes I^{-} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes \hat{\boldsymbol{e}}^{\oplus (c-n)}$$

$$\otimes I^{-} \otimes I_{\bar{W}} \otimes H_1 + \frac{1-p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes \hat{\boldsymbol{e}}^{\oplus (c-n)} \otimes \mathcal{G} \otimes I_{\bar{W}} \otimes H_1, \ \ n \in \{0, \ldots, c-1\};$$

- *the blocks $Q_{i,i}$, $Q_{i,i+1}$ and $Q_{i,i-1}$ are respectively of dimension N_i, $N_i \times N_{i+1}$ and $N_i \times N_{i-1}$. Also the non-zero entries $(Q_{i,i})_{n,n'}$ of the blocks $Q_{i,i}$, entries $(Q_{i,i+1})_{n,n}$ of the blocks $Q_{i,i+1}$, entries $(Q_{i,i-1})_{n,n}$ of the blocks $Q_{i,i-1}$ are defined as follows:*

$$(Q_{i,i})_{n,n} = (S^{(2)})^{\oplus n} \oplus (S^{(1)})^{\oplus i} \otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + I_{M_2^n} \otimes I_{M_1^i} \otimes \mathcal{T}_0 \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$$

$$+ I_{M_2^n} \otimes I_{M_1^i} \otimes I_{\tilde{R}_K} \otimes (D_0 \oplus H_0) + \frac{1}{c} I_{M_2^n} \otimes \left[\bar{I}_{M_1}^{\oplus i} + (c-i) I_{M_1}^{\otimes i} \right]$$

$$\otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes H_1, \quad n \in \{0, \dots, c-i-1\},$$

$$(Q_{i,i})_{n,n} = (\mathbf{S}_0^{(2)} \boldsymbol{\beta}^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I^- \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes \mathcal{F} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}$$

$$+ (S^{(2)})^{\oplus n} \oplus (S^{(1)})^{\oplus (c-n)} \oplus \mathcal{T} \oplus D_0 \oplus H_0 + \frac{1}{c} I_{M_2^n} \otimes \left[\bar{I}_{M_1}^{\oplus (c-n)} + n I_{M_1}^{\otimes (c-n)} \right]$$

$$\otimes I_{\tilde{R}} \otimes I_{\bar{W}} \otimes H_1, \quad n \in \{c-i, \dots, c\},$$

$$(Q_{i,i})_{n,n-1} = (\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^i} \otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}, \ \ n \in \{1, \dots, c-i-1\},$$

$$(Q_{i,i})_{n,n-1} = (\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^i} \otimes E^- \otimes I_{\bar{W}} \otimes I_{\bar{Z}}, \ n = c-i,$$

$$(Q_{i,i})_{n,n-1} = (\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes \boldsymbol{\beta}^{(1)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}, \ \ n \in \{c-i+1, \dots, c\},$$

$$(Q_{i,i})_{n,n+1} = I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes I_{M_1^i} \otimes \mathcal{F}_0 \otimes I_{\bar{W}} \otimes I_{\bar{Z}}, \ \ n \in \{0, \dots, c-i-2\},$$

$$(Q_{i,i})_{n,n+1} = I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes I_{M_1^i} \otimes (\mathcal{F}_0 | O_{\tilde{R}_K \times \tilde{R}_{K-1}} | \dots | O_{\tilde{R}_K \times \tilde{R}_0}) \otimes I_{\bar{W}} \otimes I_{\bar{Z}},$$

$$n \ = \ c-i-1,$$

$$(Q_{i,i+1})_{n,n} = I_{M_2^n} \otimes I_{M_1^i} \otimes \boldsymbol{\beta}^{(1)} \otimes I_{\tilde{R}_K} \otimes D_1 \otimes I_{\bar{Z}}, \ \ n \in \{0, \dots, c-i-2\},$$

$$(Q_{i,i+1})_{n,n} = I_{M_2^n} \otimes I_{M_1^i} \otimes \boldsymbol{\beta}^{(1)} \otimes E^+ \otimes D_1 \otimes I_{\bar{Z}}, \ n = c-i-1,$$

$$(Q_{i,i+1})_{n,n} = I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes I_{\tilde{R}} \otimes D_1 \otimes I_{\bar{Z}}, \quad n \in \{c-i, \dots, c\},$$

$$(Q_{i,i-1})_{n,n} = I_{M_2^n} \otimes (\mathbf{S}_0^{(1)})^{\oplus i} \otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes H_1$$

$$+ \frac{1-p}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes G_K \otimes I_{\bar{W}} \otimes H_1,$$

$$n \in \{0, \dots, c-i-1\},$$

$$(Q_{i,i-1})_{n,n} = I_{M_2^n} \otimes (\mathbf{S}_0^{(1)})^{\oplus i} \otimes E^- \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes E^- \otimes I_{\bar{W}} \otimes H_1$$

$$+\frac{1-p}{c} I_{M_2^n} \otimes \hat{\boldsymbol{e}}^{\oplus i} \otimes \begin{pmatrix} G_K \\ O_{\tilde{R}_{K-1}\times \tilde{R}_K} \\ \vdots \\ O_{\tilde{R}_0 \times \tilde{R}_K} \end{pmatrix} \otimes I_{\bar{W}} \otimes H_1, n = c-i,$$

$$(Q_{i,i-1})_{n,n} = I_{M_2^n} \otimes (\mathbf{S}_0^{(1)} \boldsymbol{\beta}^{(1)})^{\oplus(c-n)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)}$$

$$\otimes \tilde{I} \otimes I_{\bar{W}} \otimes H_1 + \frac{1-p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)} \otimes \mathcal{G}_0 \otimes I_{\bar{W}} \otimes H_1, \; n \in \{c-i+1, \dots, c\},$$

$$(Q_{i,i-1})_{n,n+1} = I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes (\mathbf{S}_0^{(1)})^{\oplus(c-n)} \otimes I^- \otimes I_{\bar{W}} \otimes I_{\bar{Z}} + \frac{p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)}$$

$$\otimes \hat{\boldsymbol{e}}^{\oplus(c-n)} \otimes I^- \otimes I_{\bar{W}} \otimes H_1 + \frac{1-p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes \hat{\boldsymbol{e}}^{\oplus(c-n)} \otimes \mathcal{G} \otimes I_{\bar{W}} \otimes H_1,$$

$$n \in \{c-i, \dots, c-1\}.$$

Proof. Proof of the theorem consists of careful packing of the transition rates presented by tables Tab. 4.1 to Tab. 4.3 into the block matrix form.

We shall explain the method to obtain some selected block matrices. Now consider the matrix $Q_{c,c}$, which describes all transitions in which the level, i.e., the value of the component i_t, $i_t \geqslant c$, of the Markov chain does not change (that is transitions within a level). Here the possible transitions are from the macro-state (c, n) to macro-states (c, n), $n = 0, \dots, c$ and to $(c, n-1)$, $n = 1, \dots, c$. Transition from (c, n) to (c, n) occurs in the following cases.

1. Service completion of a priority customers in one of busy servers among n servers of system-1 and movement of BIC customer to system-1. So the number of customers in BIC decreased by one.

The corresponding intensities of transition are given by the matrix

$$(\mathbf{S}_0^{(2)}\boldsymbol{\beta}^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I^{-} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}.$$

2. Service completion of an interrupted customer in BIP. So the number of customers in BIP decreased by one and the number in BIC increased by one. This can be recorded by one step to left in the matrix F_k and one step to right in $\mathcal{F}$ respectively. In this case the transitions are given by the matrix

$$I_{M_2^n} \otimes I_{M_1^{c-n}} \otimes \mathcal{F} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}.$$

3. Phase changes that do not lead to service completion or interruption arrival or primary arrival. This transitions are given by the matrix

$$(S^{(2)})^{\oplus n} \oplus (S^{(1)})^{\oplus(c-n)} \oplus \mathcal{T} \oplus D_0 \oplus H_0.$$

4. Interruption arrival to a server which provides the service in protected phase for primary customers or one who is providing service for priority customers. The corresponding matrix is given by

$$\frac{1}{c} I_{M_2^n} \otimes \left[\bar{I}_{M_1}^{\oplus(c-n)} + n I_{M_1}^{\otimes(c-n)}\right] \otimes I_{\tilde{R}} \otimes I_{\bar{W}} \otimes H_1.$$

Transition from (c, n) to $(c, n-1)$ occurs if service completion of a priority customer in system-1 and a primary customer is taken for service due to BIC is empty. This can be expressed as

$$(\boldsymbol{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes \boldsymbol{\beta}^{(1)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}.$$

Now we consider the matrix $Q_{c+1,c}$, which describes the transitions from level i, $i \geqslant c+1$, to $i-1$. This can be occur only when service completion of primary customers in one of busy servers in system-1 or interruption arrival to one of the busy servers serving primary customer in unprotected phase of service. Here the possible transitions are from the macro-state $(c+1,n)$ to macro-states (c,n), $n = 0,\ldots,c$ and to $(c,n+1)$, $n = 1,\ldots,c-1$. We consider these transition in two cases when BIC is empty and BIC is non-empty. If BIC is empty then transition from $(c+1,n)$ to (c,n) occurs in the following ways.

1. Service completion of primary customers in one of busy servers of system-1 and a primary customer is taken for service. The corresponding intensities of transition are given by the matrix

$$I_{M_2^n} \otimes (\mathbf{S}_0^{(1)}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes I_{\bar{Z}}.$$

2. Interruption arrival to a primary server in unprotected phase and with probability p the current primary customer is lost and a new primary customer is taken for service. These transitions are given by the matrix

$$\frac{p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)} \otimes \tilde{I} \otimes I_{\bar{W}} \otimes H_1.$$

3. With complimentary probability $1-p$, the interrupted customer moves to BIP which results in increase of number in BIP by one (this is recorded in $\mathcal{G}_0$). If all servers in BIP is busy then the interrupted customer is lost and this transition is recorded in the

diagonal block of G_K. So we obtain the transition matrix as

$$\frac{1-p}{c} I_{M_2^n} \otimes (\hat{\boldsymbol{e}}\boldsymbol{\beta}^{(1)})^{\oplus(c-n)} \otimes \mathcal{G}_0 \otimes I_{\bar{W}} \otimes H_1.$$

Again, if BIC is non-empty then transition from $(c+1, n)$ to $(c, n-1)$ occurs in the following ways.

1. Service completion of primary customer in system-1 and a BIC customer is taken for service. So the number in BIC is decreased by one. In this case transition matrix is given by

$$I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes (\mathbf{S}_0^{(1)})^{\oplus(c-n)} \otimes I^- \otimes I_{\bar{W}} \otimes I_{\bar{Z}}.$$

2. Interruption arrival to a primary server in unprotected phase and with probability p, the current primary customer is lost and a BIC customer is taken for service. In this case no change of number in BIP and the number in BIC is decreased by one. These transitions are given by the matrix

$$\frac{p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes \hat{\mathbf{e}}^{\oplus(c-n)} \otimes I^- \otimes I_{\bar{W}} \otimes H_1.$$

3. With complimentary probability $1-p$, the interrupted customer moves to BIP and a BIC customer is taken for service. In this case the number in BIP increased by one (that is, one step right in G_k) and number in BIC reduced by one (that is, one step to left in $\mathcal{G}$). These transitions are given by

$$\frac{1-p}{c} I_{M_2^n} \otimes \boldsymbol{\beta}^{(2)} \otimes \hat{\mathbf{e}}^{\oplus(c-n)} \otimes \mathcal{G} \otimes I_{\bar{W}} \otimes H_1.$$

In a similar way we can prove the the expressions for other blocks of Q. □

4.3 Steady-state distribution

Ergodicity condition can be expressed in terms of the matrices $Q_{c+1,c}$, $Q_{c,c}$, $Q_{c,c+1}$ as follows. Defining $\mathcal{A} = Q_{c+1,c} + Q_{c,c} + Q_{c,c+1}$ and $\boldsymbol{\pi}$ to be the steady-state probability vector of the irreducible matrix $\mathcal{A}$. That is, $\boldsymbol{\pi}\mathcal{A} = \mathbf{0}$, $\boldsymbol{\pi}\boldsymbol{e} = 1$. The *LIQBD* description of the model indicates that the queueing system is stable (see, *Neuts* [54]) if and only if

$$\boldsymbol{\pi} Q_{c,c+1}\boldsymbol{e} < \boldsymbol{\pi} Q_{c+1,c}\boldsymbol{e}. \tag{4.3}$$

The vector $\boldsymbol{\pi}$, cannot be obtained explicitly in terms of the parameters of the model, and hence the stability condition is known only implicitly.
Let $\boldsymbol{x}$ denote the steady state probability vector of the generator Q. That is,

$$\boldsymbol{x}Q = \mathbf{0}, \quad \boldsymbol{x}\boldsymbol{e} = \mathbf{1}. \tag{4.4}$$

Let $\boldsymbol{x}$ be partitioned as

$$\boldsymbol{x} = (\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(c-1), \boldsymbol{x}(c), \ldots)$$

Algorithm for computing the stationary probabilities

$$\boldsymbol{x}(i, n, k, j, \eta_1^{(2)}, \ldots, \eta_n^{(2)}, \eta_1^{(1)}, \ldots, \eta_{\min\{i,c-n\}}^{(1)}, \phi_1, \ldots, \phi_j, \nu, \zeta) =$$

$$\lim_{t\to\infty} P\{i_t = i, n_t = n, k_t = k, j_t = j, \eta_1^{(2)}(t) = \eta_1^{(2)}, \ldots, \eta_{n_t}^{(2)}(t) = \eta_n^{(2)}, \eta_1^{(1)}(t) = \eta_1^{(1)},$$

$$\ldots, \eta_{\min\{i,c-n_t\}}^{(1)}(t) = \eta_{\min\{i,c-n\}}^{(1)}, \phi_1(t) = \phi_1, \ldots, \phi_{j_t}(t) = \phi_j, \nu_t = \nu, \zeta_t = \zeta\},$$

corresponding vectors of these probabilities listed in lexicographic order $\boldsymbol{x}(i,n,k,j)$, $\boldsymbol{x}(i,n)$, and $\boldsymbol{x}(i)$, is as follows.
We see under the assumption that the stability condition (4.4) holds, the steady-state probability vector $\boldsymbol{x}$, is obtained (see, e.g., *Neuts* [54]) as

$$\boldsymbol{x}(i+c) = \boldsymbol{x}(c)\mathcal{R}^i, \quad i \geqslant 1, \tag{4.5}$$

where $\mathcal{R}$ is the minimal non-negative solution to the matrix quadratic equation:

$$\mathcal{R}^2 Q_{c+1,c} + \mathcal{R}Q_{c,c} + Q_{c,c+1} = \mathbf{0}, \tag{4.6}$$

and the vectors $\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(c)$ are obtained from the boundary equations

$$\begin{gathered}\boldsymbol{x}(0)Q_{0,0} + \boldsymbol{x}(1)Q_{1,0} = \mathbf{0},\\ \boldsymbol{x}(i-1)Q_{i-1,i} + \boldsymbol{x}(i)Q_{i,i} + \boldsymbol{x}(i+1)Q_{i+1,i} = \mathbf{0}, \quad 1 \leqslant i \leqslant c-1,\\ \boldsymbol{x}(c-1)Q_{c-1,c} + \boldsymbol{x}(c)\left(Q_{c,c} + \mathcal{R}Q_{c+1,c}\right) = \mathbf{0},\end{gathered} \tag{4.7}$$

subject to the normalizing equation

$$\sum_{i=0}^{c-1} \boldsymbol{x}(i)\boldsymbol{e} + \boldsymbol{x}(c)(I-\mathcal{R})^{-1}\boldsymbol{e} = 1. \tag{4.8}$$

Once $\mathcal{R}$ matrix is obtained, from the boundary equation we obtain

$$\boldsymbol{x}(c) = \boldsymbol{x}(c-1)\mathcal{R}_{c-1},$$

$$\boldsymbol{x}(i) = \boldsymbol{x}(i-1)\mathcal{R}_{i-1}, \quad 1 \leqslant i \leqslant c-1,$$

where

$$\mathcal{R}_{c-1} = Q_{c-1,c}\left(-(Q_{c,c} + \mathcal{R}Q_{c+1,c})\right)^{-1}$$

and

$$\mathcal{R}_{i-1} = Q_{i-1,i}\left(-(Q_{i,i}+\mathcal{R}_i Q_{i+1,i})\right)^{-1}, \quad 1 \leqslant i \leqslant c-1.$$

The component $\boldsymbol{x}(0)$ is the steady-state distribution of the Markov Chain with generator matrix $Q_{0,0} + \mathcal{R}_0 Q_{1,0}$.

Thus, the vector $\boldsymbol{x}$ can be computed by exploiting the special structure of the coefficient matrices. We can use logarithmic reduction algorithm for computing $\mathcal{R}$. For full details on the logarithmic reduction algorithm we refer *Latouche and Ramaswami* [42].

4.4 Performance measures of the system

Once the stationary probability vectors $\boldsymbol{x}(i,n,k,j)$, $\boldsymbol{x}(i,n)$, and $\boldsymbol{x}(i), i \geqslant 0,\ n \in \{0,\ldots,c\}, j \in \{0,\ldots,K-k\},\ k \in \{0,\ldots,K\},$ have been computed, we can calculate various performance measures of the system. The most essential measures are as follows.

1. The average number of primary customers in the system-1

$$L_1 = \sum_{i=1}^{\infty} i\boldsymbol{x}(i)\boldsymbol{e}.$$

2. The average number of priority customers in the system-1

$$L_2 = \sum_{i=0}^{\infty}\sum_{n=0}^{c} n\boldsymbol{x}(i,n)\boldsymbol{e}.$$

3. The average number of priority customers in the buffer (after service at system-2)

$$L_2^{BIC} = \sum_{i=0}^{\infty}\sum_{n=0}^{c}\sum_{k=0}^{K}\sum_{j=0}^{K-k} k\boldsymbol{x}(i,n,k,j)\boldsymbol{e}.$$

4. The average number of customers in system-2

$$L_2^{BIP} = \sum_{i=0}^{\infty}\sum_{n=0}^{c}\sum_{k=0}^{K}\sum_{j=0}^{K-k} j\boldsymbol{x}(i,n,k,j)\boldsymbol{e}.$$

5. The probability $P^{interrupt}$ that the service of an arbitrary primary customer will be interrupted, but the customer will not be lost is computed by

$$P^{interrupt} = \frac{1}{\lambda}\frac{1-p}{c}\Big\{\sum_{i=1}^{c-1}\sum_{n=0}^{c-i-1}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus i}\otimes I^{**}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=1}^{c-1}\sum_{n=c-i}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes I^{*}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=c}^{\infty}\sum_{n=0}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes I^{*}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}\Big\},$$

where the diagonal matrices I^{**} and I^{*} of size $\tilde{R}_K$ and $\tilde{R}$, respectively, are defined by

$$I^{**} = diag\{I_{R^0}, I_{R^1}, \dots, I_{R^{K-1}}, \mathrm{O}_{R^K}\}$$

and

$$I^* = diag\{I_{R^0}, I_{R^1}, \ldots, I_{R^{K-1}}, \mathrm{O}_{R^K}, I_{R^0}, I_{R^1}, \ldots, I_{R^{K-1}}, \ldots, I_{R^0}\}.$$

6. The probability P_1^{loss} of an arbitrary customer loss due to interruption arrival is computed by

$$P_1^{loss} = \frac{1}{\lambda}\frac{p}{c}\Bigg\{\sum_{i=1}^{c-1}\sum_{n=0}^{c-i-1}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus i}\otimes I_{\tilde{R}_K}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=1}^{c-1}\sum_{n=c-i}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes I_{\tilde{R}}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=c}^{\infty}\sum_{n=0}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes I_{\tilde{R}}\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}\Bigg\}.$$

7. The probability P_2^{loss} that the service of an arbitrary primary customer will be lost due to interruption and, then, rejection in system-2 is computed by

$$P_2^{loss} = \frac{1}{\lambda}\frac{1-p}{c}\Bigg\{\sum_{i=1}^{c-1}\sum_{n=0}^{c-i-1}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus i}\otimes(I_{\tilde{R}_K}-I^{**})\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=1}^{c-1}\sum_{n=c-i}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes(I_{\tilde{R}}-I^*)\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}$$

$$+\sum_{i=c}^{\infty}\sum_{n=0}^{c}\boldsymbol{x}(i,n)(I_{M_2^n}\otimes\hat{\boldsymbol{e}}^{\oplus(c-n)}\otimes(I_{\tilde{R}}-I^*)\otimes I_{\bar{W}}\otimes H_1)\boldsymbol{e}\Bigg\},$$

8. The probability P^{loss} that an arbitrary customer will be lost in the

system is computed by

$$P^{loss} = P_1^{loss} + P_2^{loss}.$$

9. Intensity of priority customer's service completion at system-1 is computed by

$$\mu^{priority} = \sum_{i=0}^{c-1} \sum_{n=1}^{c-i-1} \boldsymbol{x}(i,n)((\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^i} \otimes I_{\tilde{R}_K} \otimes I_{\bar{W}} \otimes I_{\bar{Z}})\boldsymbol{e}$$

$$+\sum_{i=0}^{c-1} \sum_{n=c-i}^{c} \boldsymbol{x}(i,n)((\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I_{\tilde{R}} \otimes I_{\bar{W}} \otimes I_{\bar{Z}})\boldsymbol{e}$$

$$+\sum_{i=c}^{\infty} \sum_{n=1}^{c} \boldsymbol{x}(i,n)((\mathbf{S}_0^{(2)})^{\oplus n} \otimes I_{M_1^{c-n}} \otimes I_{\tilde{R}} \otimes I_{\bar{W}} \otimes I_{\bar{Z}})\boldsymbol{e}.$$

4.5 Numerical Illustrations

Next we present numerical experiments to illustrate the behavior of the queueing model under study. The correctness and the accuracy of the code are verified by a number of accuracy checks. In the first example we look into the impact of fundamental arrival rate λ on the main performance measures.

Example 4.5.1. Here we fix MAP arrival of type-1 customers is described by the matrices

$$D_0 = \begin{bmatrix} -7.6 & 0.30 \\ 0.25 & -0.75 \end{bmatrix} \text{ and } D_1 = \begin{bmatrix} 6 & 1.3 \\ 0.25 & 0.25 \end{bmatrix}$$

and then normalized to get the fundamental rate $\lambda = 2, 2.5, 3, 3.5, 4, 4.5, 5$ with coefficient of correlation of successive inter-arrival time $corr_\lambda = 0.0762$ and MAP interruption arrival is described by the matrices

$$H_0 = \begin{bmatrix} -2 & 0.00001 \\ 0.001 & -0.01 \end{bmatrix} \text{ and } H_1 = \begin{bmatrix} 1.99998 & 0.00001 \\ 0.002 & 0.007 \end{bmatrix}$$

then normalized to get the fundamental rate $h = 2$ with coefficient of correlation of successive inter-interruption arrival time $corr_h = 0.3421$. We analyze the system with $c = 2$ servers in system-1, $p = 0.3$, the probability that interrupted customers leaves the system-1 without service. We take PH service process in system-1 for type-1 and type-2 customers are by the vectors $\boldsymbol{\beta^{(i)}} = (1, 0), i = 1, 2$ and $S^{(1)} = \begin{bmatrix} -12 & 5 \\ 0 & -12 \end{bmatrix}$ and $S^{(2)} = \begin{bmatrix} -12 & 4 \\ 3 & -12 \end{bmatrix}$ and normalized with service rates $\mu_1 = 6$, $\mu_2 = 5$ respectively. That is the service time of type-1 customer has Coxian distribution of order 2. Here we assume that the first phase is not protected while the second one is protected. The PH service process of interrupted customers in system-2 are characterized by the vector $\boldsymbol{\alpha} = (1, 0)$ and $T = \begin{bmatrix} -12 & 7 \\ 8 & -12 \end{bmatrix}$ normalized with service rate $\mu = 3$. Here we vary the number of servers in system-2 by $K = 1, 2, 3$.

From figure 4.2, as λ increases, the measure L_1 increases for all values K. For small values of λ, the measure is independent of the number of servers in system-2 and for higher values of λ the measure shows only a slight variation for different values of K. These are as expected. From figure 4.2 when λ increases, the measure P_2^{loss} increases for all values of K

and the rate of increase in P_2^{loss} is negligible for higher values of K. This is due to the fact that if we increase the number of servers in system-2 this helps to accommodate more interrupted customers in BIP and thereby reducing the loss probability.

From figure 4.3, $P^{interrupt}$ decreases with increase of λ for every K. This is also as expected because increase in fundamental arrival rate results in increase of loss probability of customers due to all servers in system-2 busy. Also the rate of decrease in the measure is small for higher values of K, as expected. From figure 4.3, the measure $\mu^{priority}$ is a non-decreasing function of λ. This is to be expected since increase in λ results in increase of interrupted customers get back to service in system-1 through system-2 (BIP buffer). Also from Tab.4.4 we observe that for each K, the measures L_2, L_2^{BIC}, L_2^{BIP} and P^{loss} are non-decreasing function of λ. Note that the increase in the measure P_1^{loss} is quite negligible and a possible explanation is that for a fixed p and h, the loss of type-1 customer from system-1 has no effect on the arrival rate of such customers.

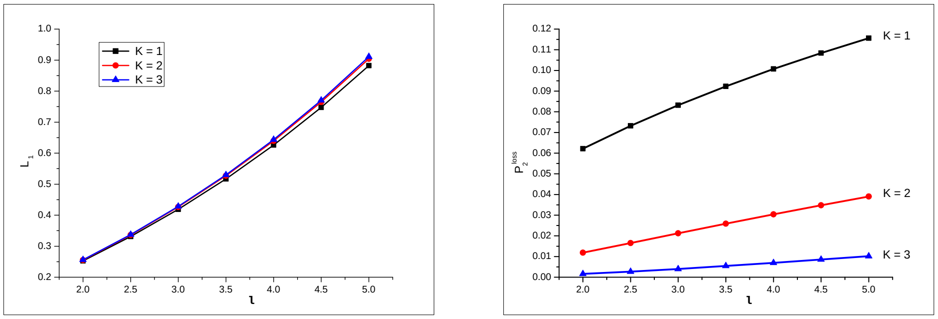


Figure 4.2: Effect of fundamental arrival rate λ on L_1 and P_2^{loss}.

Example 4.5.2. In this example we investigate the influence of the loss probability p on the performance measures for various values of service

Table 4.4: Effect of Fundamental arrival rate λ.

λ	L_2	L_2^{BIC}	L_2^{BIP}	P_1^{loss}	P^{loss}
$K = 1$					
2.0	0.06623814	0.00520949	0.15777769	0.09761938	0.15980259
2.5	0.07729384	0.00724554	0.18512034	0.09761945	0.17081049
3.0	0.08675321	0.00939641	0.21001713	0.09761954	0.18080977
3.5	0.09482057	0.01163043	0.23285999	0.09761964	0.18994084
4.0	0.10166306	0.01392802	0.25390216	0.09761976	0.19832036
4.5	0.10741199	0.01627686	0.27331454	0.09761989	0.20605298
5.0	0.11216673	0.01866898	0.29121815	0.09762004	0.21323339
$K = 2$					
2.0	0.08370817	0.00574557	0.25644991	0.09761937	0.10951737
2.5	0.10051027	0.00803395	0.30863611	0.09761944	0.11413869
3.0	0.11542034	0.01046270	0.35782440	0.09761952	0.11883695
3.5	0.12854037	0.01299537	0.40432779	0.09761962	0.12348401
4.0	0.13999854	0.01560953	0.44838166	0.09761973	0.1280129
4.5	0.14991498	0.01829123	0.49016820	0.09761986	0.13239759
5.0	0.15838693	0.02103147	0.52982908	0.09762001	0.13663638
$K = 3$					
2.0	0.08672734	0.00581331	0.34172068	0.09761937	0.09924439
2.5	0.10524121	0.00817475	0.41475839	0.09761943	0.10032917
3.0	0.1220468	0.01071652	0.48452078	0.09761951	0.10161769
3.5	0.1371536	0.01340585	0.55125481	0.09761961	0.10305119
4.0	0.15061929	0.01622115	0.61515389	0.09761972	0.10458131
4.5	0.16251100	0.01914782	0.67637507	0.09761985	0.10617337
5.0	0.17288551	0.02217583	0.73504352	0.09761999	0.10780518

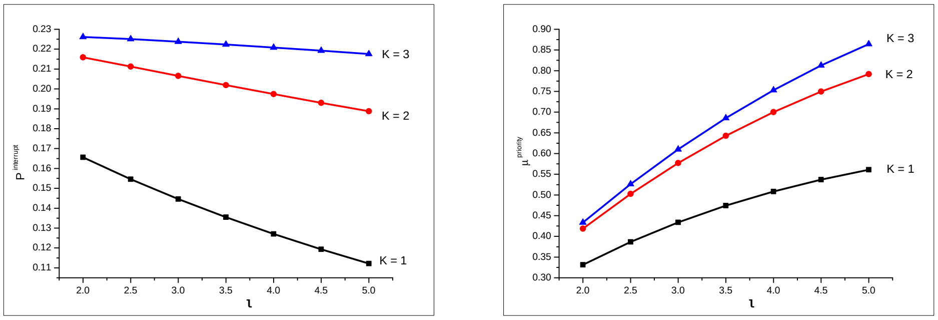


Figure 4.3: Effect of fundamental arrival rate λ on $P^{interrupt}$ and $\mu^{priority}$.

intensity, μ at system-2. We fix the number of servers in system-1 and in system-2 as $c = 2$, $K = 2$ and the matrices for MAP arrival of type-1 customer are

$$D_0 = \begin{bmatrix} -2 & 0.00001 \\ 0.001 & -0.01 \end{bmatrix} \text{ and } D_1 = \begin{bmatrix} 1.99998 & 0.00001 \\ 0.002 & 0.007 \end{bmatrix}$$

and then normalized to get the fundamental rate $\lambda = 6$ with $corr_\lambda = 0.253$ and the matrices for MAP interruption arrival are

$$H_0 = \begin{bmatrix} -1.3526 & 0 \\ 0 & -0.04391 \end{bmatrix} \text{ and } H_1 = \begin{bmatrix} 1.3436 & 0.009 \\ 0.02446 & 0.01945 \end{bmatrix}$$

This MAP has normalized fundamental rate $h = 2$ with $corr_h = 0.20023$. We take PH service process in system-1 for type-1 and type-2 customers are by the vectors $\boldsymbol{\beta}^{(i)} = (1, 0), i = 1, 2$ and $S^{(1)} = \begin{bmatrix} -4 & 4 \\ 0 & -4 \end{bmatrix}$ and $S^{(2)} = \begin{bmatrix} -6 & 4 \\ 3 & -6 \end{bmatrix}$ with normalized intensity of service rates $\mu_1 = 6.5$, $\mu_2 = 5$ respectively. Here we assume that the first phase is not protected and second phase is protected as in the example 4.5.1. The PH service process of interrupted customers in system-2 are characterized by the vector $\boldsymbol{\alpha} = (1, 0)$ and $T = \begin{bmatrix} -9 & 7 \\ 8 & -9 \end{bmatrix}$ with normalized service rates $\mu = 1, 3, 10$. Variation of μ is performed by multiplication of T by a scalar.

From figure 4.4 it is seen that the measure L_2^{BIP} decreases with increase in the lost probability p for every μ, this is due to the fact that as p increases interrupted customers leaving the system-1 without entering in to system-2. Also we observe that for every p, this measure is smaller for higher values of μ. This is to be expected since an increase in μ results in a faster clearance of customers from BIP.

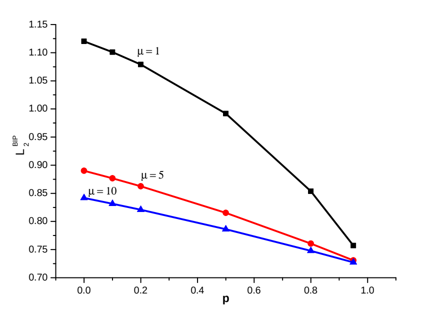


Figure 4.4: Effect of interruption probability p on L_2^{BIP}.

Example 4.5.3. In this example we show the impact of interruption arrival to the system-1 by varying λ. We fix the matrices D_0, D_1, H_0, H_1, $S^{(1)}, S^{(2)}$, T with normalized intensity of service rates $\mu_1 = 6.5$, $\mu_2 = 5$, $\mu = 3$ same as in example 4.5.2. Take the number of servers in system-1 and in system-2 as $c = K = 2$, $\lambda = 6, 8$,10, 12 and the lost probability $p = 0.3$.

We notice from figure 4.5 is that the measures P_1^{loss} and $\mu^{priority}$ increases as interruption arrival rate h increases for every λ, of course this is as expected. Also from figure 4.5 we observe some interesting observation is that for each fixed h, the measure $\mu^{priority}$ initially increases and then decreases when λ increases. This is due to the fact that more primary customers leaving the system-1 with service completion or rejection by BIP due to all servers in BIP are busy (note that $\mu = 3$).

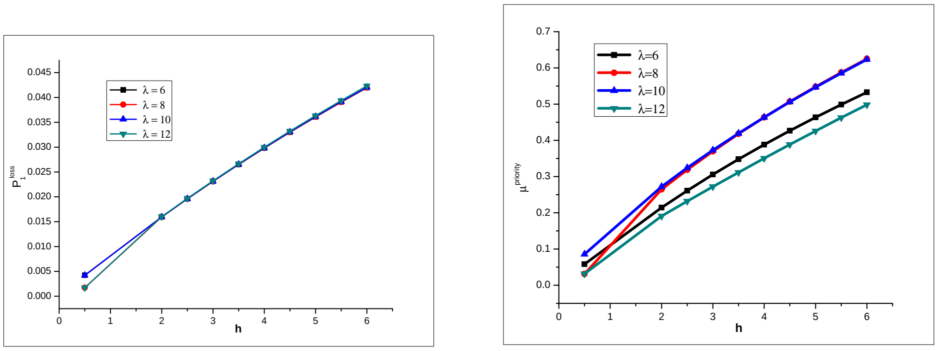


Figure 4.5: Effect of interruption arrival rate h on P_1^{loss} and $\mu^{priority}$.

Tab.4.5 shows that the measures L_1, L_2^{BIC} decreases when h increases. This is to be expected since increase in the interruption rate results in more type-1 customers getting lost from system-1 due to all the servers in system-2 are busy (note that $K = 2$ and $\mu = 3$). Also we observe that the measures L_2,L_2^{BIP}, P_2^{loss}, P^{loss}, $P^{interrupt}$ increases when h increases. This is again as expected.

Example 4.5.4. In this example we are interested in the impact of intensity of service rates μ_1 in system-1 and μ in system-2. We fix the matrices D_0, D_1, H_0, H_1, $S^{(1)}, S^{(2)}$, T with normalized fundamental rates

Table 4.5: Effect of Fundamental interruption arrival rate h when $\lambda = 6$;

h	L_1	L_2	L_2^{BIC}	L_2^{BIP}	P_2^{loss}	P^{loss}	$P^{interrupt}$
0.5	1.10216	0.01167	0.07289	0.66064	0.00001	0.00422	0.00983
2.0	1.05707	0.04290	0.07126	0.69162	0.00035	0.01631	0.03689
3.0	1.02975	0.06112	0.07022	0.70906	0.00095	0.02408	0.05300
4.0	1.00437	0.07764	0.06921	0.72453	0.00186	0.03168	0.06771
5.0	0.98072	0.09271	0.06825	0.73836	0.00304	0.03912	0.08115
6.0	0.95865	0.10652	0.06732	0.75082	0.00443	0.04640	0.09348

$\lambda = 6$, $h = 4$ and normalized intensity of service rate of type-2 customer $\mu_2 = 5$, same as in example 4.5.2. Take the number of servers in system-1 and in system-2 as $c = K = 2$, the lost probability $p = 0.3$ and the first phase is not protected.

In Tab. 4.6 we can see that the increase of μ_1 implies the decrease of all the measures. This is as expected because more type-1 customers are leaving the system-1 after completing the service with out any interruption. Also we observe that the measure L_1 increases when μ grows. This is because as increase of service rate μ in system-2 results in a faster clearance of customers from BIP and so system-1 is being busy with type-2 customers. This also results in decrease of the measures L_2^{BIP}, P_1^{loss}, P_2^{loss}. Other measures L_2, L_2^{BIC}, $P^{interrupt}$, $\mu^{priority}$ are increases as μ increases. This is again as expected.

Example 4.5.5. The purpose of this example is to see the influence of number of servers in system-2 (BIP) for various h. Here we fix the matrices D_0, D_1, $S^{(1)}, S^{(2)}$, T with fundamental rate $\lambda = 6$ and the normalized intensity of service rates $\mu_1 = 6.5$, $\mu_2 = 5$, $\mu = 3$ same as in example 4.5.2 and the interruption arrival follow Poisson processes with rate $h = 2, 6, 10$. Take the number of servers in system-1 as $c = 2$ and

μ_1	L_1	L_2	L_2^{BIC}	L_2^{BIP}	P_1^{loss}	P_2^{loss}	$P^{interrupt}$	$\mu^{priority}$
6.1	1.08949	0.08085	0.07611	0.75377	0.03150	0.00229	0.07120	0.40427
6.5	1.00437	0.07764	0.06921	0.72453	0.02982	0.00186	0.06771	0.38821
7.0	0.91645	0.07378	0.06185	0.69099	0.02795	0.00146	0.06375	0.36891
7.5	0.84377	0.07015	0.05560	0.66040	0.02630	0.00117	0.06020	0.35076
8.0	0.78252	0.06678	0.05027	0.63241	0.02484	0.00095	0.05701	0.33387
8.5	0.73010	0.06365	0.04567	0.60669	0.02353	0.00078	0.05412	0.31824
9.0	0.68465	0.06076	0.04169	0.58299	0.02235	0.00065	0.05150	0.30380
μ	L_1	L_2	L_2^{BIC}	L_2^{BIP}	P_1^{loss}	P_2^{loss}	$P^{interrupt}$	$\mu^{priority}$
1	1.00028	0.06797	0.06932	0.86990	0.02983	0.00857	0.06100	0.33983
2	1.00329	0.07526	0.06920	0.76510	0.02982	0.00334	0.06623	0.37632
3	1.00437	0.07764	0.06921	0.72453	0.02982	0.00186	0.06771	0.38821
4	1.00492	0.07877	0.06924	0.70316	0.02982	0.00123	0.06834	0.39383
5	1.00528	0.07941	0.06927	0.68999	0.02981	0.00090	0.06867	0.39704
6	1.00552	0.07982	0.06930	0.68109	0.02981	0.00070	0.06887	0.39910
7	1.00571	0.08010	0.06932	0.67466	0.02981	0.00057	0.06899	0.40052

Table 4.6: Effect of μ_1 and μ on some selected measures.

the lost probability $p = 0.3$. Here we assume that the first phase is not protected.

Figure 4.6 shows that the measure P_2^{loss} decreases whereas L_2^{BIP} increases as the number of servers in BIP increases. Explanation of this decreasing and increasing is the following. If we increase the number of servers in BIP more and more interrupted type-1 customers can enter into BIP. So the probability that the interrupted customers are lost due to BIP being full is getting reduced and the average number of customers in BIP is increased.

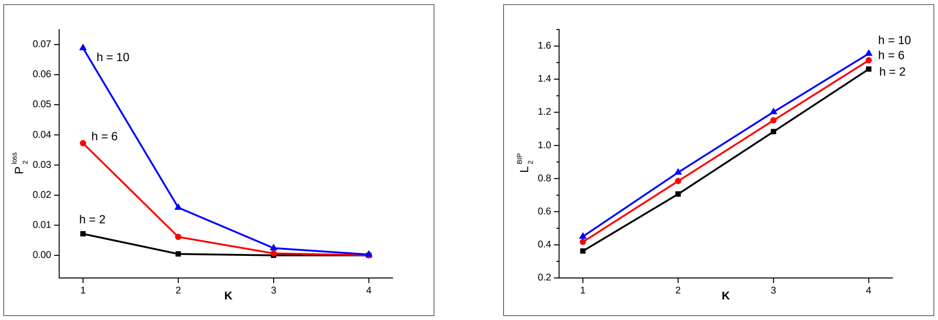


Figure 4.6: Effect of number of servers in system-2 on P_2^{loss} and L_2^{BIP}.

From Tab. 4.7, looking at measures L_2^{BIC} and P_1^{loss} we see some interesting trends. The measure L_2^{BIC} increases initially and then gradually decreases as K increases. This is probably due to the fact that after carefully looking into the model, we see that beyond certain value of K, any further increase in K will only result in the servers in system-1 being busy with type-2 customers. So the rate of interrupted customers getting back to service through waiting in the BIC buffer will be smaller leading to less customers (on the average) in BIC buffer. Also the measure P_1^{loss} does not depending on K.

Table 4.7: Effect of number of servers in system-2.

K	L_1	L_2	L_2^{BIC}	P_1^{loss}	P^{loss}	$P^{interrupt}$	$\mu^{priority}$
1	1.03309	0.05141	0.06973	0.02143	0.02858	0.04285	0.25707
2	1.03604	0.05757	0.07050	0.02143	0.02190	0.04953	0.28782
3	1.03621	0.05789	0.07041	0.02143	0.02145	0.04998	0.28947
4	1.03621	0.05790	0.07040	0.02143	0.02143	0.05001	0.28952

Chapter 5

Analysis of Customer Induced Interruption and Retrial of Interrupted Customers

In chapters 2 to 4 we gave priority to self-interrupted customers for service upon the completion of interruption. In many real life situations self-interruption is not encouraged. Thus we reverse the role of such customers to a lower level. They are proceed to an orbit on interruption from where they continuously keep trying to access the servers whenever found idle. Thus primary customers are given priority over interrupted customers. This is in sharp contrast to *Sherman and Kharoufeh* [49] wherein they punish customers whose service get interrupted due to the fault of the

Some results of this chapter are included in the following paper.
Krishnamoorthy, A. and Varghese Jacob : Analysis of customer induced Interruption and retrial of interrupted customers. (communicated).

systems.

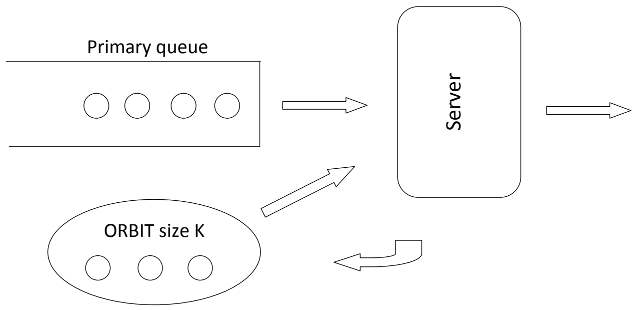


Figure 5.1: An $M/(PH,PH)/1$ queue with retrial of interrupted customers

5.1 Model description

We consider a single sever queueing system to which primary customers arrive according to a Poisson process of rate λ. The service times of primary customers are assumed to follow a PH distribution with representation $(\boldsymbol{\alpha}, T)$ of order m_1. Then $\boldsymbol{T^0} = -T\boldsymbol{e}$ is the vector of absorption rates. An arriving primary customer who finds the server idle, obtains service immediately. Otherwise, this customer is placed into the buffer of infinite capacity and will be picked up for service according to the order of arrival.

We consider customer induced interruption while his/her service is going on. In this chapter we assume as done in earlier chapters that no more than one interruption is allowed for a customer while in service. That is, an interrupted customer who gets into service again will leave the system with no further interruption. Interruptions occur according to a Poisson process of rate θ. When an interruption occurs, the customer currently in service will be forced to leave the service facility. The server thereby becoming free, is ready to offer service to other customers. Notice that in the system oriented interruption, mainly system breakdown, the customer in service while interruption strikes will be the one to be taken for service on removal of interruption. Further no service is possible while the server is under repair consequent to system breakdown; in contrast to the present model. The interrupted customer enters into an orbit of finite capacity K, should there be a space available and from there he/she retries for service after the interruption is completed. In the case that the orbit is full, an interrupted customer is blocked from joining the orbit and is forced to leave the system for ever. On the other hand, when an orbital customer retries and finds that the server is busy, he returns to the orbit. The service time of orbital customers are assumed to follow PH distribution with representation $(\boldsymbol{\beta}, S)$ of order m_2. We denote $\boldsymbol{S^0} = -S\boldsymbol{e}$, the vector of absorption rates from transient phases. The inter-retrial times are distributed exponentially with parameter γ. We also assume that inter-arrival time of primary customers, inter-occurrence period of interruptions, service time and inter-retrial times are mutually independent. The orbital customer in service is not preempted due to arrival of a primary customer.

In the sequel, $\hat{I}_n$ denotes a square matrix of order n with 1 in the first row and first column and all other entries are zeros. The average service rate μ_1 of primary customers is given by $\mu_1 = [\boldsymbol{\alpha}(-T)^{-1}\boldsymbol{e}]^{-1}$ and that of orbital customers is given by $\mu_2 = [\boldsymbol{\beta}(-S)^{-1}\boldsymbol{e}]^{-1}$.

5.2 The System State process

Let

- $N_1(t)$ be the number of primary customers in the system at time t;
- $N_2(t)$ be the number of interrupted customers in the orbit plus one in service (if any) at time t;
- $S(t)$ be the status of the server.
  That is, $S(t) = \begin{cases} 0 & \text{if server is idle at time } t; \\ 1 & \text{if server is serving a primary customer at time } t; \\ 2 & \text{if server is serving an orbital customer at time } t; \end{cases}$
- $S_1(t)$ be the phase of service at time t.

The process $\mathbf{X} = \{(N_1(t), S(t), N_2(t), S_1(t)) : t \geqslant 0\}$ is a $CTMC$ whose state space

$$\begin{aligned}\boldsymbol{\Psi} &= \{(0,0,k) : k = 0, \ldots K\} \\ &\cup\{(i,1,k,l) : i \geqslant 1; k = 0, \ldots K; l = 1, \ldots, m_1\} \\ &\cup\{(i,2,k,l) : i \geqslant 0; k = 1, \ldots K; l = 1, \ldots, m_2\}\end{aligned}$$

If the states in $\mathbf{\Psi}$ are listed in lexicographical order then the infinitesimal generator of the $CTMC$ governing the system, is given by

$$Q = \begin{bmatrix} B_1 & B_0 & & & \\ B_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{5.1}$$

where each entry is described as below:

$$B_1 = \begin{bmatrix} B_{00} & B_{02} \\ B_{20} & B_{22} \end{bmatrix}_{(m_2+1)K+1\times(m_2+1)K+1} \quad \text{with}$$

$$B_{00} = -\left(\lambda+\gamma\right) I_{K+1} + \gamma \hat{I}_{K+1};\ B_{02} = \gamma \begin{bmatrix} \mathbf{0} \\ I_K \otimes \boldsymbol{\beta} \end{bmatrix}_{(K+1)\times Km_2};$$

$$B_{20} = \begin{bmatrix} I_K \otimes \boldsymbol{S^0} & \mathbf{0} \end{bmatrix}_{Km_2\times(K+1)};\ B_{22} = I_K \otimes \left(S - \lambda I_{m_2}\right);$$

$$B_0 = \lambda \begin{bmatrix} I_{K+1} \otimes \boldsymbol{\alpha} & \mathrm{O} \\ \mathrm{O} & I_{Km_2} \end{bmatrix}_{(m_2+1)K+1\times(m_1+m_2)K+m_1};$$

$$B_2 = \begin{bmatrix} B_{11} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}_{(m_1+m_2)K+m_1\times(m_2+1)K+1};$$

$$B_{11} = I_{K+1} \otimes \boldsymbol{T^0} + B_\theta \text{ where } B_\theta = \begin{bmatrix} \mathbf{0} & \theta\boldsymbol{e} & & \\ & \ddots & \ddots & \\ & & \mathbf{0} & \theta\boldsymbol{e} \\ & & & \theta\boldsymbol{e} \end{bmatrix}$$

$$A_1 = \begin{bmatrix} A_{11}^{(1)} & \mathrm{O} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{bmatrix}_{(m_1+m_2)K+m_1 \times (m_1+m_2)K+m_1};$$

$$A_{11}^{(1)} = I_{K+1} \otimes (T - (\lambda+\theta)I_{m_1})\,; \; A_{21}^{(1)} = \begin{bmatrix} I_K \otimes \boldsymbol{\alpha} & \mathrm{O} \end{bmatrix};$$

$$A_{22}^{(1)} = I_K \otimes (S - \lambda I_{m_2})\,;$$

$$A_2 = \begin{bmatrix} A_{11}^{(2)} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix}_{(m_1+m_2)K+m_1 \times (m_1+m_2)K+m_1};$$

$$A_{11}^{(2)} = I_{K+1} \otimes \boldsymbol{T^0}\boldsymbol{\alpha} + A_{\theta}^{(2)}; \quad \text{where } A_{\theta}^{(2)} = \begin{bmatrix} \mathrm{O} & \theta\boldsymbol{e}\boldsymbol{\alpha} & & \\ & \ddots & \ddots & \\ & & \mathrm{O} & \theta\boldsymbol{e}\boldsymbol{\alpha} \\ & & & \theta\boldsymbol{e}\boldsymbol{\alpha} \end{bmatrix}$$

$$A_0 = \begin{bmatrix} A_{11}^{(0)} & \mathrm{O} \\ \mathrm{O} & A_{22}^{(0)} \end{bmatrix}_{(m_1+m_2)K+m_1 \times (m_1+m_2)K+m_1};$$

$$A_{11}^{(0)} = \lambda I_{(K+1)m_1}, \quad A_{22}^{(0)} = \lambda I_{Km_2};$$

5.3 Steady-state analysis

In this section we perform the steady-state analysis of the queueing model under study by first establishing the stability condition of the queueing system.

5.3.1 Stability condition

Denote by $\boldsymbol{\pi}$ the steady-state probability vector of the generator $A = A_0 + A_1 + A_2$. That is, $\boldsymbol{\pi} A = \mathbf{0},\ \boldsymbol{\pi}\boldsymbol{e} = 1$. The following theorem gives the stability of the queueing system under study.

Theorem 5.3.1. *The Markov Chain* $\mathbf{X}$ *is stable if and only if*

$$\lambda\boldsymbol{\alpha}\left(\theta I - T\right)^{-1}\boldsymbol{e} < 1. \tag{5.2}$$

Proof. The *LIQBD* description of the model indicates that the queueing system is stable (see, *Neuts* [54]) if and only if

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}. \tag{5.3}$$

Let

$$\boldsymbol{\pi} = \left(\boldsymbol{\pi}_1^1, \ldots, \boldsymbol{\pi}_{K+1}^1, \boldsymbol{\pi}_1^2, \ldots, \boldsymbol{\pi}_K^2\right) = \left(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2\right)$$

The matrix A is given by

$$A = \left[\begin{array}{ccccccccc}
D & \theta\boldsymbol{e}\boldsymbol{\alpha} & & & & & & & \\
 & D & \theta\boldsymbol{e}\boldsymbol{\alpha} & & & & & & \\
 & & \ddots & \ddots & & & & & \\
 & & & D & \theta\boldsymbol{e}\boldsymbol{\alpha} & & & & \\
 & & & & \widetilde{D} & & & & \\
\boldsymbol{S}^{\mathbf{0}}\boldsymbol{\alpha} & & & & & S & & & \\
 & \boldsymbol{S}^{\mathbf{0}}\boldsymbol{\alpha} & & & & & S & & \\
 & & \ddots & & & & & \ddots & \\
 & & & \boldsymbol{S}^{\mathbf{0}}\boldsymbol{\alpha} & \mathrm{O} & & & & S
\end{array}\right] \tag{5.4}$$

where $D = T + \boldsymbol{T^0}\boldsymbol{\alpha} - \theta I$ and $\widetilde{D} = D + \theta \boldsymbol{e}\boldsymbol{\alpha}$
It is easy to verify that

$$\boldsymbol{\pi}_i^1 = \mathbf{0},\ i = 1, \dots, K$$

$$\boldsymbol{\pi}_{K+1}^1 \left(T + \boldsymbol{T^0}\boldsymbol{\alpha} - \theta I + \theta \boldsymbol{e}\boldsymbol{\alpha}\right) = \mathbf{0} \tag{5.5}$$

$$\boldsymbol{\pi}_2 = \mathbf{0}$$

From equation (5.5) and using normalizing condition, it follows that

$$\boldsymbol{\pi}_{K+1}^1 \left(\boldsymbol{T^0} + \theta \boldsymbol{e}\right) = \left(\boldsymbol{\alpha} \left(\theta I - T\right)^{-1} \boldsymbol{e}\right)^{-1} \tag{5.6}$$

Then stability condition (5.3) implies that

$$\lambda < \boldsymbol{\pi}_{K+1}^1 \left(\boldsymbol{T^0} + \theta \boldsymbol{e}\right)$$

Then by using (5.6), the stated result follows immediately. □

Note: When $\theta = 0$, the traffic intensity $\rho = \lambda \boldsymbol{\alpha} \left(-T\right)^{-1} \boldsymbol{e}$ is the traffic intensity for the classical $M/PH/1$ model. Also note that traffic intensity is independent of the orbit size K.

5.3.2 Steady-state probability vector

Algorithm for computing the stationary probabilities

$$\boldsymbol{x}_{j,k,l}(i) = \lim_{t\to\infty} P\{N_1(t) = i, S(t) = j, N_2(t) = k, S_1(t) = l\},\quad (i,j,k,l) \in \boldsymbol{\Psi}$$

as follows.

Since the *CTMC*, $\mathbf{X}$ is a *LIQBD* process, its stationary distribution (if it exits) has a matrix geometric solution. We refer to *Neuts* [54] and *Latouche and Ramaswami* [43] for details about the matrix geometric solution of the *QBD* processes.

We assume that $\rho < 1$, then there exists the steady-state probability vector $\boldsymbol{x}$ of the generator Q given in (5.1). That is,

$$\boldsymbol{x}Q = \mathbf{0}, \quad \boldsymbol{x}\boldsymbol{e} = 1. \tag{5.7}$$

Partitioning $\boldsymbol{x}$ as

$$\boldsymbol{x} = (\boldsymbol{x}(0), \boldsymbol{x}(1), \boldsymbol{x}(2), \ldots) \tag{5.8}$$

we see that $\boldsymbol{x}$, is obtained as

$$\boldsymbol{x}(n) = \boldsymbol{x}(1)\mathcal{R}^{n-1}, \quad n \geqslant 2, \tag{5.9}$$

where the vector $\boldsymbol{x}(0)$ and $\boldsymbol{x}(1)$ are the unique solution of the boundary equations and the normalizing condition in (5.7):

$$\begin{gathered} \boldsymbol{x}(0)B_1 + \boldsymbol{x}(1)B_2 = \mathbf{0}; \\ \boldsymbol{x}(0)B_0 + \boldsymbol{x}(1)\left(A_1 + \mathcal{R}A_2\right) = \mathbf{0}; \\ \boldsymbol{x}(0)\boldsymbol{e} + \boldsymbol{x}(1)(I - \mathcal{R})^{-1}\boldsymbol{e} = 1. \end{gathered} \tag{5.10}$$

Here $\mathcal{R}$, the rate matrix, is minimal non-negative solution of the matrix quadratic equation

$$\mathcal{R}^2 A_2 + \mathcal{R}A_1 + A_0 = \mathrm{O}, \tag{5.11}$$

and the following relation

$$\mathcal{R}A_2\boldsymbol{e} = A_0\boldsymbol{e}, \tag{5.12}$$

[see, *Neuts* [54], p.82–83]. The above equation implies that the rate of transition from a state where there are i customers, to a state with $i-1$ customers, is equal to the transition rate from i to $i+1$. One can use logarithmic reduction algorithm *Latouche and Ramaswami* [42] for computing $\mathcal{R}$ directly.

5.3.3 Special case of $M/M/1$ queueing model

In this section we will obtain an explicit expression of $\mathcal{R}$ in the special case when $T = -\mu_1$ and $S = -\mu_2$. Then our model reduces to the $M/M/1$ queueing system with customer induced interruption and finite orbit. In this case the coefficient matrices A_0, A_1 and A_2 are given by

$$A_2 = \begin{bmatrix} A_{11}^{(2)} & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{bmatrix};$$

$$A_{11}^{(2)} = \begin{bmatrix} \mu_1 & \theta & & \\ & \ddots & \ddots & \\ & & \mu_1 & \theta \\ & & & \mu_1+\theta \end{bmatrix}_{(K+1)\times(K+1)};$$

$$A_1 = \begin{bmatrix} A_{11}^{(1)} & \mathrm{O} \\ A_{21}^{(1)} & A_{22}^{(1)} \end{bmatrix};$$

$A_{11}^{(1)} = d_1 I_{K+1}$; $A_{22}^{(1)} = d_2 I_K$; where $d_1 = -(\lambda+\mu_1+\theta)$, $d_2 = -(\lambda+\mu_2)$;

$$A_{21}^{(1)} = \begin{bmatrix} \mu_2 & 0 & & \\ & \ddots & \ddots & \\ & & \mu_2 & 0 \end{bmatrix}_{K\times(K+1)};$$

$$A_0 = \begin{bmatrix} A_{11}^{(0)} & \mathrm{O} \\ \mathrm{O} & A_{22}^{(0)} \end{bmatrix};$$

$$A_{11}^{(0)} = \lambda I_{K+1},\ A_{22}^{(0)} = \lambda I_K;$$

Theorem 5.3.2. *If $\frac{\lambda}{\mu_1+\theta} < 1$, the matrix equation (5.11) has the minimal non-negative solution*

$$\mathcal{R} = \begin{bmatrix} R_{11} & O \\ R_{21} & R_{22} \end{bmatrix} \tag{5.13}$$

where

$$R_{11} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_{K-1} & \alpha_K & \beta_{K+1} \\ & \alpha_1 & \alpha_2 & \dots & \alpha_{K-2} & \alpha_{K-1} & \beta_K \\ & & \alpha_1 & \dots & \alpha_{K-3} & \alpha_{K-2} & \beta_{K-1} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & \alpha_1 & \alpha_2 & \beta_3 \\ & & & & & \alpha_1 & \beta_2 \\ & & & & & & \beta_1 \end{bmatrix}_{(K+1)\times(K+1)}; \tag{5.14}$$

$$R_{21} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_2 & \bar{\alpha}_3 & \dots & \bar{\alpha}_{K-1} & \bar{\alpha}_K & \bar{\beta}_{K+1} \\ & \bar{\alpha}_1 & \bar{\alpha}_2 & \dots & \bar{\alpha}_{K-2} & \bar{\alpha}_{K-1} & \bar{\beta}_K \\ & & \bar{\alpha}_1 & \dots & \bar{\alpha}_{K-3} & \bar{\alpha}_{K-2} & \bar{\beta}_{K-1} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & \bar{\alpha}_1 & \bar{\alpha}_2 & \bar{\beta}_3 \\ & & & & & \bar{\alpha}_1 & \bar{\beta}_2 \end{bmatrix}_{K\times(K+1)} ; \tag{5.15}$$

$$R_{22} = \gamma_1 I_K \ with\ \gamma_1 = \frac{\lambda}{(\lambda+\mu_2)}; \tag{5.16}$$

$$\alpha_1 = \frac{-d_1 - \sqrt{d_1^2 - 4\mu_1\lambda}}{2\mu_1};\ \alpha_2 = \frac{\theta\alpha_1^2}{-d_1 - 2\mu_1\alpha_1};$$

$$\alpha_i = \frac{\theta \sum_{j=1}^{i-1} \alpha_j\alpha_{i-j} + \mu_1 \sum_{j=2}^{i-1} \alpha_j\alpha_{i-j+1}}{-d_1 - 2\mu_1\alpha_1};\ i = 3, \dots, K. \tag{5.17}$$

$$\beta_1 = \frac{\lambda}{\mu_1+\theta};\ \beta_i = \frac{\lambda}{\mu_1+\theta} - \sum_{j=1}^{i-1} \alpha_j;\ i = 2, \dots, K+1.$$

$$\bar{\alpha}_1 = \frac{\gamma_1\mu_2}{-d_1 - (\alpha_1+\gamma_1)\mu_1};\ \bar{\alpha}_2 = \frac{(\theta(\alpha_1+\gamma_1) + \mu_1\alpha_2)\,\bar{\alpha}_1}{-d_1 - (\alpha_1+\gamma_1)\mu_1};$$

$$\bar{\alpha}_i = \frac{\sum_{j=1}^{i-2} (\theta\alpha_{i-j} + \mu_1\alpha_{i-j+1})\,\bar{\alpha}_j + (\theta(\alpha_1+\gamma_1) + \mu_1\alpha_2)\,\bar{\alpha}_{i-1}}{-d_1 - (\alpha_1+\gamma_1)\mu_1};\ i = 3, \dots, K. \tag{5.18}$$

$$\bar{\beta}_i = \frac{\lambda}{\mu_1+\theta} - \sum_{j=1}^{i-1} \bar{\alpha}_j;\ i = 2, \dots, K+1.$$

Proof. Since the coefficient matrices A_0, A_1 and A_2 are all lower triangular, we observe that the rate matrix $\mathcal{R}$ has the same structure :

$$\mathcal{R} = \begin{bmatrix} R_{11} & \mathrm{O} \\ R_{21} & R_{22} \end{bmatrix}$$

Substituting $\mathcal{R}^2$ and $\mathcal{R}$ into equation(5.11) we obtain the following matrix equations

$$R_{11}^2 A_{11}^{(2)} + R_{11} A_{11}^{(1)} + A_{11}^{(0)} = \mathrm{O} \tag{5.19}$$

$$\left(R_{21} R_{11} + R_{22} R_{21}\right) A_{11}^{(2)} + R_{21} A_{11}^{(1)} + R_{22} A_{21}^{(1)} = O \tag{5.20}$$

$$R_{22} A_{22}^{(1)} + A_{22}^{(0)} = \mathrm{O} \tag{5.21}$$

Due to the structure of each blocks in the coefficient matrices A_0, A_1 and A_2 and from equation (5.19), we see that the block R_{11} is an upper triangular matrix of order $K+1$. From equation(5.12), it follows that

$$R_{11}\boldsymbol{e} = \frac{\lambda}{\mu_1 + \theta}\, \boldsymbol{e} \text{ and } R_{21}\boldsymbol{e} = \frac{\lambda}{\mu_1 + \theta}\, \boldsymbol{e} \tag{5.22}$$

Suppose $R_{11} = \left(r_{ij}\right)_{i \leqslant j}$. Using equation(5.19) we get the following set of equations:

$$\mu_1 r_{ii}^2 + d_1 r_{ii} + \lambda = 0,\ i = 1, \ldots, K;$$

$$(\mu_1 + \theta) r_{(K+1)(K+1)}^2 + d_1 r_{(K+1)(K+1)} + \lambda = 0;$$

$$\theta r_{ii}^2 + \mu_1 \sum_{j=i}^{i+1} r_{ij} r_{j(i+1)} + d_1 r_{i(i+1)} = 0,\ i = 1, \ldots, K-1;$$

$$\theta \sum_{j=i}^{i+1} r_{ij} r_{j(i+1)} + \mu_1 \sum_{j=i}^{i+2} r_{ij} r_{j(i+2)} + d_1 r_{i(i+2)} = 0,\ i = 1, \ldots, K-2;$$

$$\vdots$$

$$\theta\sum_{j=1}^{K-1} r_{1j}r_{j(K-1)} + \mu_1\sum_{j=1}^{K} r_{1j}r_{jK} + d_1 r_{1K} = 0;$$

To obtain the minimal non-negative solution of (5.19), from the first set of quadratic equations we obtain $\alpha_1 = r_{ii} = \frac{-d_1-\sqrt{d_1^2-4\mu_1\lambda}}{2\mu_1}$. Thus by solving the above set of equations and using equation(5.22) we obtain the matrix R_{11} stated in equation(5.14).

Suppose $R_{21} = (s_{ij})_{i\leqslant j}$ and defining $B = R_{11}A_{11}^{(2)} + \gamma_1 A_{11}^{(2)} + d_2 I_{K+1}$. Evidently, B is an upper triangular matrix and equation(5.20) gives $R_{21}B + \gamma_1 A_{21}^{(1)} = \mathrm{O}$, we obtain the following system of equations

$$((\alpha_1+\gamma_1)\mu_1 + d_1)\, s_{ii} + \gamma_1\mu_2 = 0,\ i = 1,\ldots,K;$$

$$(\theta(\alpha_1+\gamma_1) + \mu_1\alpha_2)\, s_{ii} + (\mu_1(\alpha_1+\gamma_1) + d_1)s_{i(i+1)} = 0,\ i = 1,\ldots,K-1;$$

$$(\theta\alpha_2 + \mu_1\alpha_3)s_{ii} + (\theta(\alpha_1+\gamma_1) + \mu_1\alpha_2)s_{i(i+1)} + (\mu_1(\alpha_1+\gamma_1) + d_1)s_{i(i+2)} = 0,$$

$$i = 1,\ldots,K-2;$$

$$\vdots$$

$$\sum_{j=1}^{K-1}(\theta\alpha_{K-j} + \mu_1\alpha_{K-j+1})s_{1j} + ((\alpha_1+\gamma_1)\mu_1 + d_1)\, s_{1K} = 0;$$

In a similar way we can obtain the matrix R_{21} by solving the above set of equations and equation(5.22).
From equation(5.21), $R_{22} = A_{22}^{(0)}\left(-A_{22}^{(1)}\right)^{-1} = \gamma_1 I_{K+1}$ where $\gamma_1 = \frac{\lambda}{\lambda+\mu_2}$.

□

In this special case we obtain the spectral radius of $\mathcal{R}$, $sp(\mathcal{R})$. We know from (5.13) that,

$$|\mathcal{R}| = |R_{11}||R_{22}|$$

So the eigenvalues of the rate matrix $\mathcal{R}$ are α_1, β_1 and γ_1 with multiplicities K, 1 and $K+1$ respectively. That is,

$$sp(\mathcal{R}) = \ \max\ \{\alpha_1, \beta_1, \gamma_1\}.$$

Evidently, $0 < \gamma_1 < 1$ for any $\theta > 0$ and $\mu_2 > 0$. Also, from equation (5.22) it follows that $\alpha_1 < \beta_1$.

Hence, $sp(\mathcal{R}) = \beta_1 = \frac{\lambda}{\mu_1+\theta}$

Then by Theorem 3.1.1 in *Neuts* [54] the QBD process is positive recurrent if and only if $\frac{\lambda}{\mu_1+\theta} < 1$, which is the stability condition of the system derived in (5.2).

5.3.4 Performance characteristics

In this section we list some useful descriptors to bring out the qualitative aspects of the model under study. These are listed below along with their formula for computation. Towards this end, we further partition the vectors $\boldsymbol{x}(i)$ into smaller vectors as follows:

$$\boldsymbol{x}(0) = (\boldsymbol{x}_0(0), \boldsymbol{x}_2(0))\ ;$$

$$\boldsymbol{x}_0(0) = (x_{0,0}(0), \ldots, x_{0,K}(0)),\ \boldsymbol{x}_2(0) = (x_{2,1}(0), \ldots, x_{2,K}(0))\ ;$$

$$\boldsymbol{x}(i) = (\boldsymbol{x}_1(i), \boldsymbol{x}_2(i));$$

$$\boldsymbol{x}_1(i) = (\boldsymbol{x}_{1,0}(i), \ldots, \boldsymbol{x}_{1,K}(i)),\ \boldsymbol{x}_2(i) = (\boldsymbol{x}_{2,1}(i), \ldots, \boldsymbol{x}_{2,K}(i)),\ i \geqslant 1;$$

Note that $\boldsymbol{x}_1(i), i \geqslant 1$ and $\boldsymbol{x}_2(i), i \geqslant 0$ are of dimensions $(K+1)m_1$ and Km_2 respectively.

1. **The probability that the server is idle**:

$$P_{idle} = \boldsymbol{x}(0)(\boldsymbol{e}_1(2) \otimes \boldsymbol{e}).$$

2. **The probability that the server is busy**:

$$P_{busy} = 1 - P_{idle}.$$

3. **The probability that the server is busy with primary customers**:

$$P_{bsyp} = \sum_{i=1}^{\infty} \boldsymbol{x}_1(i)\boldsymbol{e}.$$

4. **The probability that the server is busy with orbital customers** :

$$P_{bsyo} = \sum_{i=0}^{\infty} \boldsymbol{x}_2(i)\boldsymbol{e}.$$

5. **Mean number of primary customers in the system**:
The mean number $\mu_{primary}$, of primary customers in the system is given by

$$\mu_{primary} = \sum_{i=0}^{\infty} i\boldsymbol{x}(i)\boldsymbol{e} = \boldsymbol{x}(1)(I-\mathcal{R})^{-2}\boldsymbol{e}.$$

6. **Mean number of primary customers in the queue**:

   The mean μ_{PQ}, number of primary customers in the queue is given by

$$\mu_{PQ} = \sum_{i=1}^{\infty}(i-1)\boldsymbol{x}(i)\boldsymbol{e} = \boldsymbol{x}(1)\mathcal{R}(I-\mathcal{R})^{-2}\boldsymbol{e}.$$

7. **Mean number of orbital customers in the system**:

   The mean μ_{orbit}, number of orbital customers is given by

$$\mu_{orbit} = \sum_{k=1}^{K} k\left(\boldsymbol{x}_{0,k}(0) + \boldsymbol{x}_{2,k}(0)\boldsymbol{e}\right) + \sum_{i=1}^{\infty}\sum_{k=1}^{K} k\left(\boldsymbol{x}_{1,k}(i) + \boldsymbol{x}_{2,k}(i)\right)\boldsymbol{e}.$$

8. **The probability that an interrupted primary customer is lost due to the orbit being full**:

$$P_{loss} = \frac{\theta}{\theta+\mu_1}\sum_{i=1}^{\infty}\boldsymbol{x}_{1,K}\boldsymbol{e}.$$

9. **The successful rate of retrials**:

   The rate at which the orbiting customer successfully reach a free server is given by

$$\gamma_1^* = \gamma\boldsymbol{x}_0(0)\boldsymbol{e}.$$

10. **The overall rate of retrials**:

    The overall rate of retrials at which the orbiting customers request service is given by

$$\gamma_2^* = \gamma\left(\sum_{k=1}^{K} k\left(\boldsymbol{x}_{0,k}(0) + \boldsymbol{x}_{2,k}(0)\right)\boldsymbol{e} + \sum_{i=1}^{\infty}\sum_{k=1}^{K} k\left(\boldsymbol{x}_{1,k}(i)\boldsymbol{e} + \boldsymbol{x}_{2,k}(i)\boldsymbol{e}\right)\right).$$

11. **The fraction of successful rate of retrials**:
The fraction, FSR, of successful rate of retrial is given by

$$FSR = \frac{\gamma_1^*}{\gamma_2^*}.$$

5.4 Numerical Results

In this section, we present some numerical examples that describe the performance characteristics of the queueing model under study. For service process, we consider respectively the following PH distributions for primary customers and orbital customers.

$$\boldsymbol{\alpha} = (1,0),\ T = \begin{bmatrix} -6 & 4 \\ 2 & -7 \end{bmatrix} \text{ and } \boldsymbol{\beta} = (1,0),\ S = \begin{bmatrix} -9 & 7 \\ 8 & -10 \end{bmatrix}$$

These PH distributions will be normalized according when the service rates are changed in our numerical examples.

Example 5.4.1. The purpose of this example is to study the effect of arrival of primary customers to the system for various interruption rate θ. Consider the case when $K = 5, \mu_1 = 4$, $\mu_2 = 4.5, \gamma = 2$. The values of the measures P_{bsyp} and P_{bsyo} are graphed in Figures 5.2.

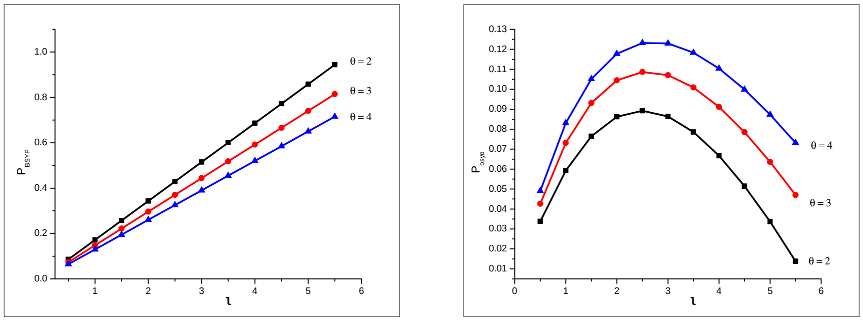


Figure 5.2: λ versus P_{bsyp} and λ versus P_{bsyo}

We notice the following from these figures.

- As is to be expected when λ increases P_{bsyp} also increases for all values of θ when all other parameters are fixed. This is because with the arrival of primary customers, the server will be active for longer time with primary customers.

- Also from Figure 5.2 we observe that for fixed θ, the measure P_{bsyo} increases initially and then decreases as λ increases. This is due to fact that after carefully looking into the model, we see that beyond a certain value of λ, any further increase in its value will only result in the server being busy with primary customers thereby justifying the phenomenon of increasing and then decreasing of this measure.

Example 5.4.2. In this example we show the impact of orbit size K for various values of θ by fixing $\lambda = 3, \mu_1 = 3, \mu_2 = 4, \gamma = 2$. In Figure 5.3 we display the graph of the measures P_{loss} and μ_{orbit} as functions of K for various θ. Examining these graphs and the Table 5.1 we note the following points :

Table 5.1: Some selected measures when $\lambda = 3, \mu_1 = 3, \mu_2 = 4, \gamma = 2$

θ	ρ	P_{idle}	P_{bsyp}	P_{bsyo}	$\mu_{primary}$
1	0.7674	0.1876	0.767397	0.045032	3.234449
2	0.6207	0.2970	0.620739	0.082287	1.727963
5	0.3915	0.4624	0.391528	0.146088	0.811940
10	0.2402	0.5688	0.240232	0.190996	0.502724

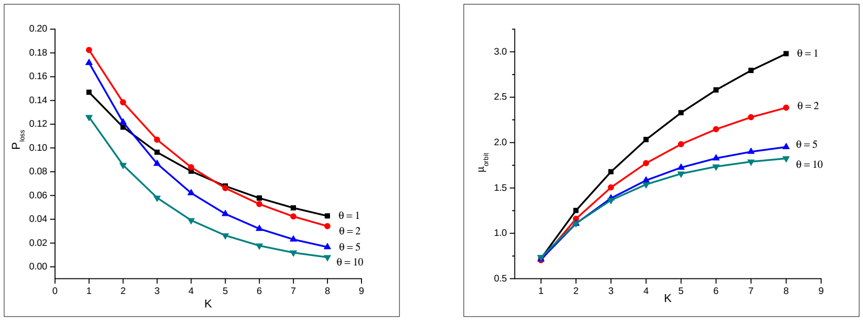


Figure 5.3: K versus P_{loss} and K versus μ_{orbit}

- As is to be expected, with K increasing P_{loss} decreases and μ_{orbit} increase for all values of θ when all other parameters are fixed.

- Looking at the graph in Figure 5.3, it is interesting to note that for larger K, P_{loss} appears to decrease when θ increases and for smaller K, P_{loss} increases initially and then deceases. This seems to be counter intuitive as one would expect for a larger K, as θ increases more primary customers get interrupted and moves to orbit resulting in a successful retrial. Also for smaller K, it is observed that as θ is varied from 1 to higher values, there appears to be cut-off points such that P_{loss} decreases. A possible explanation for this is as follows. Higher interruption rate causes more interruptions leading large number of interrupted customers in the orbit, which in

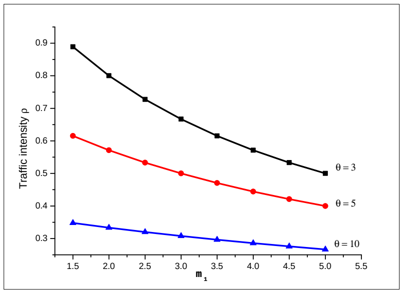


Figure 5.4: μ_1 versus traffic intensity ρ

turn increases the rate of self-interrupted customers getting back to service (note that $\lambda = 3, \mu_2 = 4$).

- From Figure 5.3, with respect to the measure μ_{orbit}, we see some interesting trends. For a fixed K, the measure μ_{orbit} decreases as θ increases. At first one may expect this measure to be non-decreasing function of θ when all other parameters are fixed. This is due to the fact that increase in θ will only result in more customers being interrupted and so the server stays busy with orbital customers thereby clearing the orbital customers faster (note that $\lambda = 3, \mu_2 = 4$).

Example 5.4.3. In this example we investigate the influence of μ_1 on the performance measures for various values of θ. Take $K = 5, \lambda = 4, \mu_2 = 4, \gamma = 2$. These are displayed in the Figure 5.4 and 5.5. We summarize the following observations.

- Looking at the Figure 5.4, it is seen that the measure ρ is a non-increasing function of μ_1 for every θ. This is expected since increase

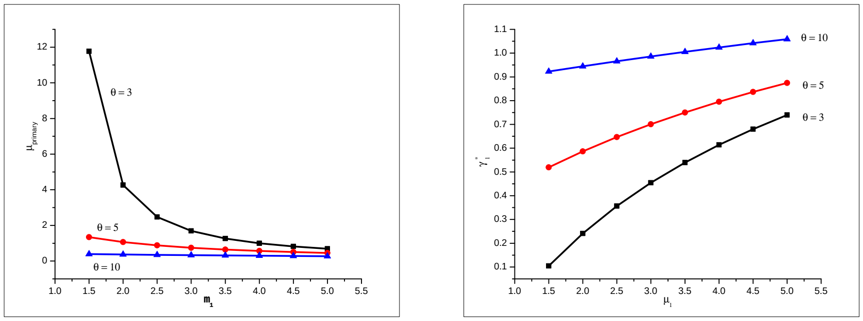


Figure 5.5: μ_1 versus $\mu_{primary}$ and μ_1 versus γ_1^*

in μ_1 helps to clear the primary customers at a faster rate so that traffic intensity get reduced.

- Figure 5.5 shows that the measure $\mu_{primary}$ decreases with increasing values of μ_1 and that it decreases more rapidly as θ increases as is to be expected. Also the measure γ_1^* is a non-decreasing function of μ_1 for all fixed values of θ. This is due to the fact that as μ_1 increases more primary customers complete their service without interruption resulting in a faster clearance of primary customers leading to increase in servers's idle probability and hence increase in successful retrial.

Example 5.4.4. Through this example we aim at checking the influence of θ for various values of μ_2. Here we fix $K = 5, \lambda = 3, \mu_1 = 3, \gamma = 2$. In Figure 5.6, we display the graph of the measures P_{loss} and μ_{orbit} as functions of θ for various μ_2 values.

- From Figure 5.6 and Table 5.2 and due to the observations in Example 5.4.3, we note that the measure P_{loss} increases initially and

Table 5.2: P_{bsyp} and P_{bsyo} for selected θ and μ_2 values

					θ			
μ_2		0.5	1	1.5	2	2.5	3	3.5
2	P_{bsyp}	0.8689	0.7674	0.6866	0.6207	0.5661	0.5201	0.4809
	P_{bsyo}	0.0405	0.0813	0.1158	0.1449	0.1695	0.1905	0.2086
4	P_{bsyp}	0.8689	0.7674	0.6866	0.6207	0.5661	0.5201	0.4809
	P_{bsyo}	0.0219	0.0450	0.0651	0.0823	0.0969	0.1096	0.1206
6	P_{bsyp}	0.8689	0.7674	0.6866	0.6207	0.5661	0.5201	0.4809
	P_{bsyo}	0.0148	0.0308	0.0448	0.0567	0.0670	0.0759	0.0836

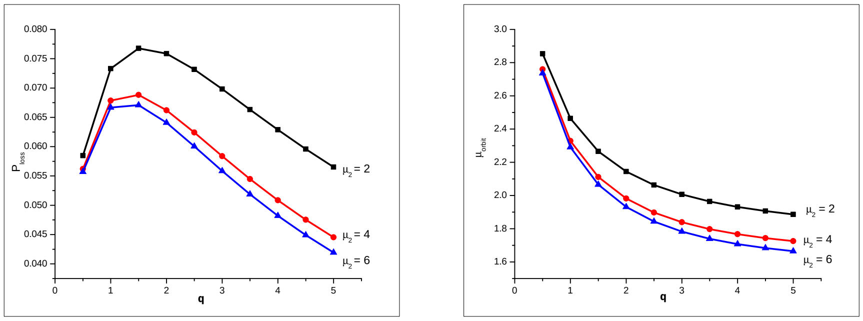


Figure 5.6: θ versus P_{loss} and θ versus μ_{orbit}

then decreases as θ increases.

- As θ increases, μ_{orbit} appears to decrease for all μ_2 values. This is also is explained by the observation in Example 5.4.3.

Example 5.4.5. The purpose here is to see the effect of γ for various μ_1. Here we fix $K = 5, \lambda = 4, \mu_2 = 4, \theta = 3$.

- Figure 5.7 shows that an increase in γ results in decrease in P_{loss}.

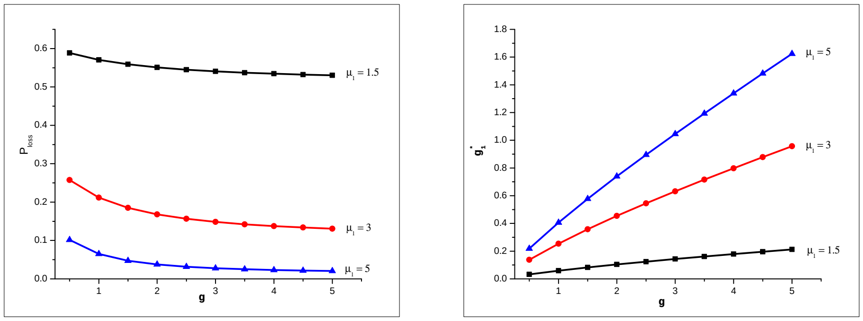


Figure 5.7: γ versus P_{loss} and γ versus γ_1^*

This is because if we increase the retrial rate γ, effective successful rate increases and hence more orbital customers leave the system after completing service. So the loss probability, P_{loss} gets reduced.

- From figure 5.7 it is seen that the measure γ_1^* is a non-decreasing function of γ for every μ_1 value, which is as expected. Further, it can be noticed that for a fixed γ, the measure γ_1^* is larger for higher values of μ_1. This is to be expected since an increase in μ_1 helps in a faster clearance of primary customers leading to increase in the idle probability.

Concluding remarks and suggestions for further study

In this thesis we have introduced and studied the notion of self interruption of service by customers. Service interruption in queueing systems have been extensively discussed in literature (see, *Krishnamoorthy, Pramod and Chakravarthy* [38]) for the most recent survey. So far all work reported deal with cases in which service interruptions are generated by sources other than customers. However, there are situations where interruptions are due to the customers rather than the system. Such situations are especially arise at doctors clinic, banks, reservation counter etc. Our attempt is to quantify a few of such problems. Systematically we have proceed from single server queue (in Chapter 2) to multi-server queues (Chapter 3). In Chapte 4, we have studied a very general multi-server queueing model with service interruption and protection of service phases. We also introduced customer interruption in a retrial setup (in Chapter 5). All models (from Chapter 2 to Chapter 4) that were analyzed involve 'non-preemptive priority' for interrupted customers where as in the model discussed in Chapter 5 interruption of service by customers is not encouraged. So the interrupted customers cannot access the server as long as there are primary customers in the system. In Chapter 5 we have obtained an explicit expression for the stability condition of the system. In all models analyzed in this thesis, we have assumed that no more than one interruption is allowed for a customer while in service. Since the models are not analytically tractable, a large number of numerical illustrations were given in each chapter it illustrate the working of the systems.

We can extend the models discussed in this thesis to several directions. For example some of the models can be analyzed with both server induced and customer induced interruptions the results for which are not available till date. Another possible extension of work is to the case where there is no bound on the number of interruptions a customer is permitted to have before service completion. More complex is the case where a customer is permitted to have a finite number $(K \geqslant 2)$ of interruptions during service.

Bibliography

[1] *Artalejo, J.* (2000) : G-networks: A versatile approach for work removal in queueing networks. European Journal of Operational Research. 126, 233–249.

[2] *Artalejo, J.R. and Gomez-Corral, A.* (2008) : Retrial Queueing Systems: A Computational Approach: Springer, Berlin.

[3] *Artalejo, J.R.* (2010) : Accessible bibliography on retrial queues: Progress in 2000-2009. Mathematical and computer modelling, 51, 1071–1081.

[4] *Atencia, I. and Moreno, P.* (2006) : A Discrete-Time Geo/G/1 retrial queue with the server subject to starting failures. Annals Oper. Res., 141, 85–107.

[5] *Avi-Itzhak, B. and Naor, P.* (1963) : Some queueing problems with the service station subject to breakdowns. Operations Research, 11, 303–320.

[6] *Bellman, R.* (1960) : Introduction to Matrix Analysis. McGraw Hill book Co.,New York.

[7] *Bini, D. and Meini, B.* (1995) : On cyclic reduction applied to a class of Toeplitz matrices arising in queueing problems. In Computations with Markov Chains, Ed., W. J. Stewart, Kluwer Academic Publisher, 21–38.

[8] *Breuer, L. and Baum, D.* (2005) : An introduction to queueing theory and matrix analytic methods, Springer, The Netherlands.

[9] *Chakravarthy, S.R. and Alfa, A.S.* (1994) : A Finite Capacity Queue With Markovian Arrivals and two Servers with group services. Journal of Applied Mathematics and Stochastic Analysis 7, Number 2, 161–178.

[10] *Chakravarthy, S.R.* (2001) : The batch Markovian arrival process: a review and future work: A. Krishnamoorthy, et al. (Eds.), Advances in Probability Theory and Stochastic Process: Proc., Notable Publications, NJ, pp. 21–49.

[11] *Chakravarthy, S. R., Krishnamoorthy, A. and Joshua, V. C.* (2006) : Analysis of a multi-server retrial queue with search of customers from the orbit; Performance Evaluation; Vol. 63, Issue 8, pp. 776–798.

[12] *Chakravarthy, S.R. and Saligrama Agnihothri, R.* (2008) : A server backup model with Markovian arrivals and phase type services. European Journal of Operational Research, 184, 584-609.

[13] *Chakravarthy, S.R.* (2009) : Analysis of a Multi-server queue with Markovian arrivals and synchronous phase type vacations. Asia-Pacific Journal of Operational Research. Vol. 26, No.1, 85–113.

[14] *Chakravarthy, S.R., Dudin, A. N. and Klimenlok, V.I.* (2010) : A retrial queueing model with MAP arrivals, catastrophic failures with

repairs and customer impatience. Asia-Pacific Journal of Operational Research, 27, 727–752.

[15] *Choi, B.D. and Chang, Y.* (1999) : $MAP_1, MAP_2/M/c$ retrial queue with the re-trial group of finite capacity and geometric loss. Mathematical and Computer Modelling, 30(3-4).

[16] *Chopra, A.S.* (2003) : Ayurveda In Medicine Across Cultures: History and Practice of Medicine in Non-Western Cultures", Edited by H. Selin. Kluwer Academic Publishers, Norwell, MA, 75–83.

[17] *Choudhury, G. and Tadj, L.* (2009) : An M/G/1 queue with two phases of service subject to the server breakdowns and delayed repair. Applied Mathematical Modelling, 33(6), 2699–2709.

[18] *Chung Kai Lai and Aitsahlia Farid.* (2003) : Elementary probability theory-with Stochastic Processes and an introduction to Mathematical finance, Springer.

[19] *Doshi, B.T.* (1986) : Queueing systems with vacations-a survey. Queueing Systems, Theory and Applications, 1(1), 29–66.

[20] *Doshi, B.T.* (1990) : Single server queues with vacations. In Stochastic Analysis of Computer and Communication Systems, H. Takagi (editor), 217–265, Elsevier Science Publishers B.V. (North-Holland), Amsterdam.

[21] *Dudin, A.N. and Semenova, O.V.* (2004): Stable algorithm for stationary distribution calculation for a $BMAP/SM/1$ queueing system with Markovian arrival input of disasters. Journal of Applied Probability. 42, 547-556

[22] *Edward Kao, P.C. and Marison Spokony Smith* (1992) : On Excess, Current and Total-Life distributions of phase type renewal processes, Naval Research Logistics, vol.32, p.p 789–799.

[23] *Erhan Cinlar* (1975) : Introduction to Stochastic Processes, New Jersy: Prentice-Hall.

[24] *Falin, G.I. and Templeton, J.G.C.* (1997) : Retrial Queues, Chapman & Hall.

[25] *Fisher, W. and Meier-Hellstern, K.S.* (1993) : The Markov-modulated Poisson process (MMPP) cookbook. Performance Evaluation. 18, 149–171.

[26] *Gaver, D.P.* (1962) : A waiting line with interrupted service including priority. Journal of Royal Statistical Society, B24, 73–90.

[27] *Gaver, D.P., Jacobs, P.A. and Lathouche, G.* (1984) : Finite birth and death models in randomly changing environments. Adv. in Appl.Probab., 16: 715–731.

[28] *Gomez-corral, A., Krishnamoorthy, A. and Narayan, V.C.* (2005) : The Impact of Self Generation of Priorities on multi-server queues with finite capacity Stochastic models 21:427–447.

[29] *Gomez-Corral, A.* (2006) : A biblographical guide to the analysis of retrial queues through matrix analytic techniques. Annals of Operations Research, 141:163–191.

[30] *Graham, A.* (1981) : Kronecker Products and Matrix Calculus with Applications, Ellis Horwood, Cichester.

[31] *Gross, D. and Harris, C.M.* (1988) : Fundamentals of Queueing Theory, John Wiley and Sons, New York.

[32] *Heyman, D.P. and Lucantoni, D.* (2003) : Modelling multiple IP traffic streams with rate limits. IEEE/ACM Transactions on Networking. 11, 948–958.

[33] *Jaiswal, N.K.* (1961) : Preemptive resume priority queue. Operations Research, 732–770.

[34] *Karlin, S. and Taylor, H.M.* (1975) : A first course in Stochastic Processes, Academic press, Newyork.

[35] *Karlin, S. and Taylor, H.M.* (1981) : A second course in Stochastic Processes, Academic press, Newyork.

[36] *Klimenok, V.I. and Dudin, A.N.* (2012) : A $BMAP/PH/N$ queue with negative customers and partial protection of service. Communications in Statistics - Simulation and Computation. 41, No 1062–1082.

[37] *Krishnamoorthy, A., Pramod, P.K. and Deepak, T.G.* (2009) : On a queue with interruptions and repeat/resumption of service, Non-linear Analysis, Theory, Methods and Applications, 71, e1673-e1683, Elsevier.

[38] *Krishnamoorthy, A., Pramod, P.K. and Chakravarthy, S.R.* (2012) : Queues with interruption : A Survey. TOP-Spanish journal of Statistics & Operations Research, DOI 10.1007/s11750-012-0256-6, 2012.

[39] *Krishna Kumar, B. and Madheswari, S.* (2005) : An $M/M/2$ queueing system with heterogeneous servers and multiple vacations. Mathematical and Computer Modelling, 41(13), 1415–1429.

[40] *Krishna Kumar, B.* (2007) : Transient Solution of an $M/M/2$ Queue with Heterogeneous Servers Subject to Catastrophes. Information and Management Sciences-New York, 18(1), 63–80.

[41] *Kulkarni, V. G., Nicola, V. F. and Trivedi, K. S.* (1990) : Effects Of Checkpointing And Queueing on Program Performance. Commun. Statist.- Stochastic Models, 6(4), 615-648.

[42] *Latouche, G. and Ramaswami, V.* (1993) : A logarithmic reduction algorithm for quasi-birth-and-death processes. Journal of Applied Probability, 30, 650–674.

[43] *Latouche, G. and Ramaswami, V.* (1999) : Introduction to Matrix Analytic Methods in Stochastic Modeling. SIAM., Philadelphia, PA.

[44] *Li Jihong and Tian Naishuo.* (2007) : The $M/M/1$ queue with working vacations and vacation interruptions, Systems Engineering Society of China and Springer-Verlag.

[45] *Lucantoni, D.M.* (1991) : New results on the single server queue with a batch Markovian arrival process. Communications in Statistics-Stochastic Models. 7, 1–46.

[46] *Medhi, J.* (1984) : Stochastic Processes, New age international, New Delhi.

[47] *Medhi, J.* (2003) : Stochastic models in queueing theory, Academic press, An imprint of Elsevier, USA.

[48] *Miaomiao Yu, Yinghui Tang, Yonghong Fu and Lemeng Pan.* (2011) : An $M/E_k/1$ queueing system with no damage service interruption. Mathematical and Computer Modelling, 54, 1262–1272.

[49] *Nathan P. Sherman and Jeffrey P. Kharoufeh.* (2006) : An $M/M/1$ retrial queue with unreliable server. Operations Research Letters, 34, 697–705.

[50] *Neuts, M.F.* (1979) : A versatile Markovian point process. J. Appl. Prob., 16:764–779.

[51] *Neuts, M. F. and Lucantoni, D. M.* (1979) : A Markovian Queue With N Servers Subject To Breakdowns And Repairs. Management Science, 25(9), 849–861.

[52] *Neuts, M.F.* (1989) : Structured Stochastic Matrices of $M/G/1$ type and their Applications. Marcel Dekker, NewYork.

[53] *Neuts, M.F. and Rao, B.M.* (1990) : Numerical investigations of a multi-server retrial model. Queueing Systems, 7:169–190.

[54] *Neuts, M.F.* (1994) : Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach, 2nd ed., Dover Publications, Inc., New York.

[55] *Neuts, M.F.* (1995) : Algorithmic Probability: A collection of problems. Chapman and Hall, NewYork.

[56] *Pattavina, A. and Parini, A.* (2005) : Modelling voice call interarrival and holding time distributions in mobile networks, in: Performance Challenges for Efficient Next Generation Networks - Proc. of 19th International Teletraffic Congress, pp. 729-738.

[57] *Riska, A., Diev, V. and Smirni, E.* (2002) : Efficient fitting of long-tailed data sets into hyperexponential distributions, Global Telecommunications Conference (GLOBALCOM'02, IEEE), pp. 2513-2517.

[58] *Sapna Isotupa, K. P. and David A. Stanford* (2002) : An Infinite-Phase Quasi-Birth-And-Death Model For The Non-Preemptive Priority $M/PH/1$ Queue. Stochastic Models, Vol. 18, No. 3, 387–424.

[59] *Takagi, H.* (1991) : Queueing Analysis: A Foundation of Performance Evaluation, Vol.1: Vacation and Priority Systems, Part1, Elsevier Science Publications B.V, Amsterdam.

[60] *Takagi, H.* (1993) : Queueing Analysis - A foundation of Performance Evaluation. Vol.III, Elsevier Science Publishers B.V.

[61] *Takine, T. and Sengupta, B.* (1998) : A single server queue with service interruptions, Queueing System 26, 285-300.

[62] *Tian, N. and Zhang, Z.G.* (2006) : Vacation Queueing Models - Theory and Applications. Springer International Series.

[63] *Tijms Henk, C.* (2003) : A First Course in Stochastic Models, John Wiley and sons Ltd. Chichester, England.

[64] *Wang, J.T.* (2004) : An M/G/1 queue with second optional service and server breakdowns, Computers and Mathematics with Applications 47, 1713-1723.

[65] *White, H. and Christie, L.* (1958) : Queuing with preemptive priorities or with breakdown, Operations Research 6, 79-95.

[66] *William J. Stewart.* (2009) : Probability, Markov Chains, Queues and Simulation, the Mathematical Basics of Performance Modelling. Princeton University Press, Princeton and Oxford.

List of Papers accepted/communicated

- ***Varghese Jacob**, Chakravarthy, S.R. and Krishnamoorthy, A.* : On a Customer Induced Interruption in a service system. Journal of Stochastic Analysis and Applications, 30, 1–13, 2012. Taylor & Francis, USA.

- *Krishnamoorthy, A. and **Varghese Jacob**.* : Analysis of Customer Induced Interruption in a multi server system. Neural, Parallel and Scientific Computations, 20, 153–172, 2012. Dynamic publishers, Inc, USA.

- *Dudin, A.N., **Varghese Jacob** and Krishnamoorthy, A.*: A multi-server queueing system with service interruption, partial protection and repetition of service. To appear in Annals of Operations Research, Springer, 2013.

- *Krishnamoorthy, A. and **Varghese Jacob*** : Analysis of customer induced interruption and retrial of interrupted customers. (Submitted for publication).

Papers presented/accepted for presentation

- ***Varghese Jacob**, Chakravarthy, S.R. and Krishnamoorthy, A.* : On a Customer Induced Interruption in a service system. International Workshop on Retrial Queues (8^{th} WRQ),July 27-29,2010, Beijing, China.

- *Krishnamoorthy, A. and **Varghese Jacob*** : Analysis of customer induced interruption and retrial of interrupted customers. International Workshop on Retrial Queues (9^{th} WRQ),June 28-30, 2012, Seville, Spain. (Accepted for presentation).

CURRICULUM VITAE

Name : Varghese Jacob

Present Address :
Department of Mathematics,
Cochin University of Science
and Technology, Cochin,
Kerala, India – 682 022.

Official Address :
Associate Professor,
Department of Mathematics,
Government College, Kottayam,
Kerala, India – 686 013.

Permanent Address :
Thomackal House,
Chenneerkara P.O.
Pathanamthitta,
Kerala, India – 689 517.

Email : v.varghesejacob@gmail.com

Qualifications :
B.Sc. (Mathematics), 1991,
M. G. University, Kottayam,
Kerala, India.

M.Sc. (Mathematics), 1993,
M.Phil.(Mathematics), 1996,

Cochin University of Science
and Technology, Cochin,
Kerala, India – 682 022.

Research Interest : Stochastic Modelling– Queueing Theory.