Stochastic Modelling: Analysis and Applications

ANALYSIS OF SOME PRIORITY QUEUES AND A PROBLEM ON DIAGNOSTICS

Thesis submitted to

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

for the award of the degree of

DOCTOR OF PHILOSOPHY

under the Faculty of Science

by

MANJUNATH A. S.

Department of Mathematics
Cochin University of Science and Technology
Kochi - 682 022, Kerala, India

January 2017

ANALYSIS OF SOME PRIORITY QUEUES AND A PROBLEM ON DIAGNOSTICS

Ph.D. thesis in the field of Stochastic Modelling & Analysis

Author:
Manjunath A. S.
Department of Mathematics
Cochin University of Science and Technology
Kochi - 682 022, Kerala, India
Email: manjunadem@gmail.com

Supervisor:
Dr. A. Krishnamoorthy
Professor (Rtd)
Department of Mathematics
Cochin University of Science and Technology
Kochi - 682 022, Kerala, India.
Email: achyuthacusat@gmail.com

January 2017

Dr. A. Krishnamoorthy
Professor (Rtd)
Department of Mathematics
Cochin University of Science and Technology
Kochi - 682 022, India.

28th January 2017

Certificate

Certified that the work presented in this thesis entitled "**ANALYSIS OF SOME PRIORITY QUEUES AND A PROBLEM ON DIAGNOSTICS**" is based on the authentic record of research carried out by Mr. Manjunath A. S. under my guidance in the Department of Mathematics, Cochin University of Science and Technology, Kochi- 682 022 and has not been included in any other thesis submitted for the award of any degree. Also certified that all the relevant corrections and modifications suggested by the audience during the Pre-synopsis seminar and recommended by the Doctoral Committee of the candidate has been incorporated in the thesis and the work done is adequate and complete for the award of Ph. D. Degree.

Dr. A. Krishnamoorthy
(Supervising Guide)

Phone : 9446428647 Email: achyuthacusat@gmail.com

Declaration

I, Manjunath A S, hereby declare that the work presented in this thesis entitled "**ANALYSIS OF SOME PRIORITY QUEUES AND A PROBLEM ON DIAGNOSTICS**" is based on the original research work carried out by me under the supervision and guidance of Dr. A. Krishnamoorthy, formerly Professor, Department of Mathematics, Cochin University of Science and Technology, Kochi-682 022 and has not been included in any other thesis submitted previously for the award of any degree.

Manjunath A S
Research Scholar (Reg. No. 4121)
Department of Mathematics
Cochin University of Science & Technology
Cochin - 682 022

Cochin-22
28^{th} January 2017

To
All those whom I love,
especially my
father Sivaraman Nair
mother Saraladevi
wife Anitha
daughters Rithvika and Sathvika
and
the memory of my Ammumma

अज्ञानतिमिरान्धस्य ज्ञानाञ्जनशलाकया ।
चक्षुरुन्मीलितं येन तस्मै श्रीगुरवे नमः ॥

Acknowledgments

I would like to offer my humble salutations to my Guru, Guide and inspiration Dr A Krishnamoorthy who has been a force of motivation to my life for a long time. Without him a single word of this thesis would not be possible and I bow my head in front of his masterly mind and paternal affection that helped me to make my dream come true.

My co-guide Dr B Lakshmi has always been a source of knowledge and inspiration and has been a great help to me in my endeavor.

I find myself indebted to Dr Sunil Jacob John for making me realize my path, for giving me an orientation and for being a great friend and inspiration.

A word of thanks to Prof. B D Choi whose genuine interest in my research was a driving force in my journey and Prof. S R Chakravarthy, whose words of encouragement and constant support will always be remembered with deep gratitude from my part.

Every single chapter of this work is a tribute to my teachers- those extraordinary minds I found myself fortunate to get a glimpse of - Dr A Vijayakumar who has been a great motivational spirit giving me his kind guidance, Dr P G Romeo who gave me his helping hand when help was most needed. I extend my heartfelt gratitude to Thrivikraman sir, Ganapathy sir, Namboodiri sir, Chakravarthy sir, Jadhavedan sir and Joseph Mathew sir .

A special note of thanks to my friend and colleague Varghese Jacob for being my pillar of strength in hard times.

Dhanya for her immense help during my research will always be remembered with warm gratitude.

I consider this work as a result of prayers from many well wishers- Joshua sir, Jayaprasad, Jayalal sir, Vinayan sir, Sudin, Jose K P sir and Madhavan Namboodiri.

I consider it a blessing to get friends like Gireesan, Kiran, Tijo, Sajeesh, Didimos, Satheesh, Noufal, Ambily, Jaya, Savitha Chechi and Bineetha during my research tenure and I thank you for giving me a new outlook and perception.

I would like to thank Dr. C Sreenivasan, Dr. R Manikandan, Dr. Viswanath C N, Dr. Reshmi T, Dr. Sajeev S. Nair, Dr. Gopakumar B., Dr. Sathian, Ms. Divya, Mr. Shajeb, Mr. Prince, Mr. Abdul Rof, Prof. Sindhu, Ms. Anu Nutan for their help and support. .

My fellow research scholars Mr. Pravas K., Ms. Pamy Sebastian, Mr. Shyam Sunder, Ms. Anu Varghese, Mr. Balesh K., Ms. Seethu Varghese, Ms. Akhila R., Ms Susan Mathew, Ms. Elisabeth Reshma, Dr. Raji George, Ms. Savitha K.N., Ms. Smisha, Mr. Arun Kumar C S, and Mr. Rahul will always be remembered with affection for the support they have given me.

Mr Renjith Mohan, my colleague at Govt College Kottayam deserves a special mention for helping me through administrative paper works related to FIP.

Prof. Santhakumari, former principal of Government Engineering College Idukki has always been a friend, a guiding spirit and well-wisher. Same goes true with Prof. Leelamma, Dr. Syamala kumari, Dr. P V Sasi former Principals of Govt College Kottayam and Dr. Babu Sebastian, Principal of Govt. College Kottayam.

My heartfelt thanks to Manoj Kovur for well-timed advice and Fasila for being a friend who always motivated and gave me courage to carry on through difficult situations.

Ramkumar, Pramada, Teena, Murali, Nishamol and Lekshmi will be remembered with love and affection for being those constant faces who kept up with me and my life always.

Office staff at CUSAT and GCK were always a support and a special thanks is due to the librarians at Maths, CUSAT for their great help.

My sincere thanks to Sreelatha chechi for bringing me food while I was not well.

I place my gratitude to UGC for my fellowship through which I could work for my lifetime dream.

To my family- my lifeline …first of all I bow my head to the memory of my grand mother Rajarajeswari Amma who made me the person I am, to Ammachi Saraladevi and Achan Sivaraman Nair for always being there, my in-laws Kochukuttan Pillai and Geetha for those considerate actions, Manu and Vinu for sheer support and prayer, Maniyan chittappan to whom this thesis owes a lot, Vishnu for his help, Vineeth, Anjali, Priya and Anju for support.

To Anitha, my wife and Ammu and Sachu, my daughters for love.

To God for giving me a tough situation, for helping me to wade through and I believe every moment that you gave me for this Thesis was a gift which will always be remembered with love and I believe without You this would just be a dream not a deed.

Manjunath A S

Contents

Notations and Abbreviations used

$\boldsymbol{e}$	:	column vector of 1's with appropriate dimension
$\mathbf{0}$	:	vector consisting of 0's with appropriate dimension
$\mathbf{O}$	:	zero matrix with appropriate dimension
$\boldsymbol{e}_j$	:	column vector of appropriate dimension with 1 in the j^{th} position and 0 elsewhere
$\boldsymbol{e}'_j$	:	row vector of appropriate dimension with 1 in the j^{th} position and 0 elsewhere
I	:	identity matrix of appropriate dimension
I_r	:	identity matrix of dimension r
PH	:	Phase type
$CTMC$	:	Continuous Time Markov Chain
QBD	:	Quasi-birth-and-death
$LIQBD$	:	Level Independent Quasi-Birth-and-Death process
$\otimes$	:	Kronecker product
$\oplus$	:	Kronecker sum

Chapter 1

Introduction

Most of us experience queueing systems directly or indirectly; directly through waiting in line ourselves or indirectly through some of our items waiting in line, such as a print job waiting in the printer buffer queue, or a packet waiting at a router node for processing. In all the cases, we want the delay to be minimum and also not to be turned away by the system due to the buffer space being not available. These possibilities of delay and denial are the major issues in a queuing system and how to minimize them is the concern. A trade off between the cost to the system due to customers denied admission as a consequence of overflow and profit due to large number of customers in the system is what is needed. Stochastic modelling of the system along with construction of a suitable cost function provides answers to most of the questions. Thus the performance of a queuing system can be evaluated and information can be generated for making decisions as to when and how to upgrade the system to improve its future performance.

In queueing systems, and all systems that operate over time with uncertainty being model characteristic, we need a sequence or a family of random variables to represent such a phenomenon over time. A stochastic process is a family or a sequence of random variables indexed by a parameter, usually time. A continuous time Markov chain is a continuous time stochastic process that enjoy memoryless property which means that no matter what the past was, the current state is all that is needed to predict the future. This memoryless property allows flexibility in modelling and produces tractable models. This thesis analyzes a few queueing models by means of continuous time Markov chains. Modelling tools such as Markovian Arrival Process (MAP) and Phase-type service distributions (PH-distributions) are used in this regard.

Queueing systems with phase-type arrival or service mechanisms give rise to transition matrices that are block tridiagonal and are referred to as quasi-birth-death(QBD) processes. The matrix geometric method developed by Neuts is employed in solving such queueing models. In this thesis we have extensively made use of Phase type distributions for service time and Markovian arrival process for arrival of customers. Hence it is apt to give a brief description of these.

1.1 Phase Type distribution (Continuous time)

Phase type(PH) distributions and related point processes provide a versatile set of tractable models in applied probability. They are based on the method of stages, a technique introduced by A. K. Erlang and generalized by M. F. Neuts. The key idea is to model random time intervals as being made up of a (possibly random) number of exponentially distributed segments and to exploit the resulting Markovian structure without losing computational tractability.

The continuous PH distributions are introduced as a natural generalization of the exponential and Erlang distributions. A $PH-$distribution is obtained as the distribution of the time until absorption in a finite state space Markov chain with an absorbing state. Phase-type distributions have matrix representations that are not unique. Furthermore, phase-type distributions constitute a versatile class of distributions that can approximate arbitrarily closely any probability distribution defined on the nonnegative real line.

A non-negative random variable X has a Phase-Type (PH) distribution if its distribution function is given by

$$F(t) = P(X \leq t) = 1 - \boldsymbol{\alpha} \exp(Tt)e \equiv 1 - \boldsymbol{\alpha} \left(\sum_{r=0}^{\infty} \frac{t^r T^r}{r!} \right) e, \quad t \geq 0$$

where,

- $\boldsymbol{\alpha}$ is row vector of non-negative elements of order $m(> 0)$ satisfying $\boldsymbol{\alpha} e \leq 1$.
- T is an $m \times m$ matrix such that i) all off-diagonal elements are nonnegative ii) all main diagonal elements are negative iii) all row sums are non-positive and iv) T is invertible

The 2- tuple $(\boldsymbol{\alpha}, T)$ is called a phase-type representation of order m for the PH distribution and T is called a generator of the PH distribution..

Let $\mathcal{X} = \{X(t) : t \geq 0\}$ be a homogeneous Markov chain with finite state space $\{1, ..., m, m+1\}$ and generator

$$\mathcal{Q} = \begin{pmatrix} \mathcal{T}_{m \times m} & \mathcal{T}^0 \\ \mathbf{0} & 0 \end{pmatrix}$$

where the elements of the matrices $\mathcal{T}$ and $\mathcal{T}^0$ satisfy $\mathcal{T}_{ii} < 0$ for $1 \leq i \leq m$, $\mathcal{T}_{ij} \geq 0$ for $i \neq j$; $\mathcal{T}_i^0 \geq 0$ and $\mathcal{T}_i^0 > 0$ for at least one i, $1 \leq i \leq m$ and $\mathcal{T}\boldsymbol{e} + \mathcal{T}^0 = \mathbf{0}$.

Let the initial distribution of $\mathcal{X}$ be the row vector $(\boldsymbol{\alpha}, \alpha_{m+1})$, $\boldsymbol{\alpha}$ being a row vector of dimension m with the property that $\boldsymbol{\alpha e} + \alpha_{m+1} = 1$. The states $1, 2, \ldots, m$ shall be transient, while the state $m+1$ is absorbing.

Let $\mathcal{Z} = \inf\{t \geq 0 : X(t) = m+1\}$ be the random variable representing the time until absorption in state $m+1$. Then the distribution of $\mathcal{Z}$ is Phase type distribution (or shortly PH distribution) with representation $(\boldsymbol{\alpha}, \mathcal{T})$. The dimension m of $\mathcal{T}$ is called the order of the distribution. The states $1, 2, ..., m$ are also called phases.

- The density function is

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

- $E[X^n] = (-1)^n n! \boldsymbol{\alpha} \mathcal{T}^{-n} \mathbf{e},\ n \geq 1.$

- The Laplace-Stieltjes transform of $F(.)$ is

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

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

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

For further information about the PH distribution, see, *Neuts*, [52], *Breuer and Baum*, [9], *Latouche and Ramaswami* , [44] and *Qi-Ming He*, [55]. Usefulness of PH distribution as service time distribution in telecommunication networks is elaborated, e.g., in *Pattavina and Parini* [53] and *Riska, Diev and Smirni* [54].

1.2 Markovian Arrival Process

Markovian Arrival Processes (MAP) are introduced in *Neuts* [50]. 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. A salient feature of the MAP is the underlying Markovian structure that fits ideally in the context of matrix-analytic solutions to stochastic models. MAP significantly generalizes the Poisson processes and still keep the tractability for modelling purposes. Currently, the MAP 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. 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 MAP.

A continuous time Markovian arrival process is a counting process that is defined on a finite state continuous time Markov chain. However, unlike PH-distributions an underlying Markov chain for a Markovian arrival process has no absorption state (phase). A Markovian arrival process counts the number of arrivals, which can be associated with changes of state in the underlying Markov chain. The arrivals can also occur during the stay in each state of the underlying Markov chain. For a MAP, the transitions of state with arrival, transitions of state without arrival, and arrivals without a transition of state, are all referred to as events. Arrival rates of events can be customized for different states, demonstrating the versatility inherent to MAPs.

In a MAP, the customers arrival is directed by an irreducible continuous time Markov chain $\{\boldsymbol{\phi}_t, t \geq 0\}$ with the state space $\{1, 2, \ldots, m\}$. Let $\mathcal{D}$ 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 $1 \leq i, j \leq m$; with probability $p_{ij}(0)$ the transition corresponds to no arrival and the state of the Markov chain is j, $j \neq i$. The Markov chain can go from state i to state i only through an arrival. Also we have

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

The transition intensities of the Markov chain $\{\boldsymbol{\phi}_t, t \geq 0\}$ which are accompanied by arrival of k customers are described by the matrices $D_k, k = 0, 1$. Define $D_0 =$

$(d_{ij}^{(0)})$ and $D_1 = (d_{ij}^{(1)})$ such that $d_{ii}^{(0)} = -\lambda_i, 1 \leq i \leq m, d_{ij}^{(0)} = \lambda_i p_{ij}(0)$, for $j \neq i$ and $d_{ij}^{(1)} = \lambda_i p_{ij}(1), 1 \leq i, j \leq m$.

By assuming D_0 to be a nonsingular matrix, the inter-arrival time is finite with probability one and the arrival process does not terminate. Hence, we see that D_0 is a stable matrix. The generator $\mathcal{D}$ is then given by $\mathcal{D} = D_0 + D_1$. Thus D_0 governs the transitions corresponding to no arrival and D_1 governs those corresponding to an arrival. Vector $\boldsymbol{\eta}$ of the stationary distribution of the process $\{\boldsymbol{\phi}_t, t \geq 0\}$ is the unique solution to the system

$$\boldsymbol{\eta}(D_0 + D_1) = \boldsymbol{\eta}\mathcal{D} = \mathbf{0} \text{ and } \boldsymbol{\eta}\mathbf{e} = 1. \tag{1.1}$$

Fundamental rate λ of the MAP is given by $\lambda = \boldsymbol{\eta} D_1 e$ which gives the expected number of arrivals per unit time in the stationary version of the MAP.

1.3 Quasi-birth-death processes

Quasi-birth-death processes (QBDs) are matrix generalizations of simple birth-and-death processes on the nonnegative integers in the same way as PH distributions are matrix generalization of the exponential distribution. Consider a Markov Chain $\left\{X_t,\ t \in \mathbf{R}^+\right\}$ on the two dimensional state space $\Omega = \bigcup_{n \geq 0} \{(n, j) : 1 \leq j \leq m\}$. The first coordinate n is called the level, and the second coordinate j is called a phase of the n^{th} level. The number of phases in each level may be either finite or infinite. The Markov chain is called a QBD process if one-step transitions from a state are restricted to phases in the same level or to the two adjacent levels. In other words,

$$(n-1, j') \rightleftharpoons (n, j) \rightleftharpoons (n+1, j'') \text{ for } n \geq 1.$$

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 Ω in lexicographic order, the infinitesimal generator of a *LIQBD* process is block tridiagonal and has the following form:

$$\boldsymbol{Q} = \begin{pmatrix} B_1 & A_0 & & & \\ B_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} \tag{1.2}$$

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, the matrices B_2, A_2 and A_0 are nonnegative and the matrices B_1 and A_1 have nonnegative off-diagonal elements and strictly negative diagonals. The row sums of Q are equal to zero, so that we have $B_1\boldsymbol{e} + A_0\boldsymbol{e} = B_2\boldsymbol{e} + A_1\boldsymbol{e} + A_0\boldsymbol{e} = (A_0 + A_1 + A_2)\boldsymbol{e} = \mathbf{0}$.

Among the several tools that we employed in this thesis Matrix geometric method plays a key role. A brief description of this is given below.

1.4 Matrix Geometric Method

Matrix Geometric Method introduced by M. F. Neuts is a tool to construct and analyze a wide class of stochastic models, particularly queueing systems, using a matrix formalism to develop algorithmically tractable solution. The transform techniques employed in solving QBD processes are replaced largely by the matrix geometric approach with the advent of high speed computers and efficient algorithms. In the matrix geometric method the distribution of a random variable is defined through a matrix; its density function, moments, etc. are expressed with this matrix. The modelling tools such as Phase type distributions, Markovian Arrival Processes, Batch Markovian Arrival Processes, Markovian Service Processes etc. are well suited for Matrix Geometric Methods.

Theorem 1.4.1 (see Theorem 3.1.1. of *Neuts* [52]). *The process* $\boldsymbol{Q}$ *in (1.2) is positive recurrent if and only if the minimal non-negative solution* R *to the* *matrix-quadratic equation*

$$R^2A_2 + RA_1 + A_0 = O \tag{1.3}$$

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

$$\begin{aligned} \boldsymbol{x}_0 \, (B_1 + RB_2) &= \boldsymbol{0} \\ \boldsymbol{x}_0(I - R)^{-1}\boldsymbol{e} &= 1 \end{aligned} \tag{1.4}$$

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

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

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e} \tag{1.5}$$

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

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

$$\boldsymbol{x}_i = \boldsymbol{x}_0 R^i \quad \text{for} \;\; i \geq 1. \tag{1.6}$$

Once 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* [43]) or the cyclic reduction algorithm (see *Bini and Meini* [4]) for computing R.

1.5 Computation of R matrix

There are several algorithms for computing rate matrix R. Here we list two of them.

Iterative algorithm

From (1.3), we can evaluate R in a recursive procedure as follows.

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

Step 1:

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

Continue **Step 1** until $R(n+1)$ is close to $R(n)$.

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

Logarithmic reduction algorithm

Logarithmic reduction algorithm is developed by *Latouche and Ramaswami* [43] which has extremely fast quadratic convergence. This algorithm is considered to be the most efficient one. The main steps involved in the logarithmic reduction algorithm are listed below. For further details on the logarithmic reduction algorithm refer *Latouche and Ramaswami* [43].

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

1.6 Supplementary variable technique

In most practical queueing systems inclusion of one or more supplementary variable could make the system Markovian. The use of the supplementary variable technique in queueing dates back to 1942 when it was introduced by Kosten [36]. In this method to get a Markov Process, we keep track of some additional information together with the underlying random variable. Consider an $M/G/1$ queue, where G denotes the distribution of service time. The process $\{N_t,\ t \in R^+\}$, where N_t gives the state of the system or the system size at an arbitrary time t is then non-Markovian. This process is not Markov. However such a process could be made Markovian by the inclusion of variable x_t defined as the amount of time spent/remaining for the customer in service at time t, if any. In other words the collection $\{(N_t, x_t)\ ; t \geq 0, x_t \geq 0\}$ is a Markov process. For the $GI/M/1$ the supplementary information on time elapsed until t since last arrival, in addition to the number of customers at the pre-arrival epoch prior to time t, provides a two dimensional Markv chain. For details of the supplementary variable techniques applied to $M/G/1$ queue see Cox [19], Keilson and Kooharain [34] and Cohen [18].

This thesis provides analysis of priority queues. These priorities need not be on the basis of source of external (primary) customers. We have deviated from the classical priority queue by bringing in 'internal' priority generation. This root of internal priority generation is considered in Krishnamoorthy et al.[38] and Gomez-Corral et al.[21]. Krishnamoorthy et al. [38] also analyze multi priority queues with 'internal priority generation'. Very often internal priority generation may be t higher priorities (like patients waiting in a queue to consult a physician). In contrast the internal priority generation discussed in this thesis takes the customer to lower priority queue; such queues get generated internally. An example of such situation is a customer interrupting his service to attend a phone call.

1.7 Review of related work

In queueing literature, priority queues stand for customers belonging to different classes joining distinct waiting lines (one for each class) to receive service. The highest classes of customers have priority (preemptive or non-preemptive) over the rest; the next in the order gets priority over all lower class customers and so on.

Priority queues are first considered by White and Christie [64] as a queue with interruption of service of low priority customers to provide service to higher priority customers. A priority queue with preemptive service can be regarded as a queue with service interruption for e.g. a doctor renders his service to a causality patient urgently by interrupting his other consultation. Jaiswal [27] is on preemptive priority queue with resumption of service of the low priority customer and Jaiswal [29] discusses time dependent solution in priority queues. Cobham [17] considers a non-preemptive priority queue and derived equilibrium expected waiting time. A detailed discussion of development in priority queues until 1968 is given in Jaiswal [30]. More recent developments on priority queues could be found in Takagi [61] and in Brodal [8].

Concept of interruption in service is introduced in the context of the failure of service system (see the recent survey paper by A. Krishnamoorthy et al. [39]). Customer induced service interruption as coined by Jacob et al. [26] is a contrast situation to that of interruption due to server failure. This is done for the single server case, where service interrupted customers are given priority over primary customers. Here self-interrupted customer takes an exponentially distributed time to get out of interruption. This is extended to the multi-server case in Krishnamoorthy and Jacob [37]. All underlying distributions (inter-arrival time, service time, inter-interruption time, interruption fixation time) are assumed to be independent exponential random variables. Dudin et al. [20] extend the above case to Markovian Arrival Process and Phase type service with c servers and negative customers with a few protected service phases.

The priority queuing system considered in the second chapter differs from those discussed above as follows: We assume that the interrupted customers are allotted low priority. As an e.g. a person who applies for credit card with bad credit history. Also in the previous models discussed, the interrupted customers are entered in a buffer space of finite capacity where as in this model the interrupted customers join a waiting line with an infinite capacity. The interruption for a customer can occur a finite number of times, say N, resulting in $N + 1$ queues. Each waiting line is generated by the customers in the immediately preceding queue, except the highest priority customers who form the primary queue (external source). Thus the low priority queues are dependent even in its evolution. Both preemptive and non-preemptive service disciplines are analyzed.

Compared to the second chapter, the third chapter analyzes a priority queueing model where low priority lines consists of customers who come back for service getting repeated. Thus the third chapter focuses on priority queues with feedback customers. Various feedback policies on different queuing models are studied in detail in literature. Some of the works are reported in [7], [13], [14], [15], [32], [59] and [60]. Krishnakumar et al. [40] consider M/G/1 retrial queue with feedback, the feedback customer goes to the tail end of the queue. Krishnakumar et al. [41] analyze a multiserver feedback system in which also feedback is to the tail end of the queue. A single server retrial queue with collisions and feedback is analyzed in Krishnakumar et al. [42]. The feedback considered in literature fall mainly in two categories. Either the customer joins the tail end of queue on completion of service to get his service repeated or he occupies the server immediately on completion of service without joining the queue. In the latter case, service at the head of the queue is paused for a while to provide service to the immediately fed back customer. In both cases there is no separate queue for feedback customers and there is no way of identifying a feedback customer in the first case.

In the third chapter, we introduce feedback queue in a different setting. Even though the customer feedback is instantaneous, it is assigned a lower priority in our system, added to that we assume there is external entry also. For instance, a company providing annual maintenance contract with certain number of free services. Here the waiting lines are not as dependent as discussed in the second chapter. Yet, if we block the external entry to the low priority lines, then the queues will be formed only by the feedback customers(so that analysis of feedback customers will be made easy). We restrict our attention to the case of a single feedback.

From here the work proceeds to a different priority queuing model, which contains a virtual queue of infinite capacity and a finite queue of physically arriving customers. For example, a store may have two types deliveries-one direct and other over phone. Crowdsourcing happens when the store decides to serve indirect customers through willing direct customers, the store being main server and willing customers being servers for the store. Crowdsourcing coined from 'crowd' and 'outsourcing' according to Howe [25] is the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. For a discussion on the crowdsourcing queueing system one may refer to Chakravarthy and Dudin [11]. They discuss the problem as a priority queue with non-preemption. Motivated from this we analyze a preemptive priority crowdsourcing model.

In all the three models discussed above, it is assumed that the server is completely aware of the service requirement of a customer (see Gross and Harris [22], though there is no mention about exact requirement). In fact this is the case with all models discussed so far in queueing theory. Quite often only one type of service is offered by the system and so conflict does not occur. It is also true that the customers arriving to such system know the type of service needed. Thus there is no conflict on the service provided to the customer. However, there are several real life situations where the customer (or server or even both) is not knowledgeable about the exact service requirement. This is especially the case when several types of services are available at a service station. As a concrete example we have vehicles for repair at service stations, patients consulting physicians for diagnosis and medication and a specific call for a service at a customer care center. If the right service required is not identified and instead the diagnosis turned out to be wrong the result could be disastrous. A wrong diagnosis and consequent service provided may sometimes turn out to be even fatal or may result in the equipment being rendered unusable or the loss of a customer altogether. These types of diagnostic problems are analyzed in the last two chapters of this thesis.

This thesis analyzes models providing explicit solution for system state distribution and also those that need algorithmic analysis. The matrix-geometric structure of the steady-state distributions introduced by Neuts and an extended version by Miller [49] for doubly infinite queues are used in the models for obtaining solutions.

1.8 Summary of the thesis

This thesis is basically analysis of some priority queues and a problem on diagnostics which we face in many real life situations. In this thesis a few queueing models are studied by means of continuous time Markov chains. The modelling tools such as Markovian Arrival Process (MAP) and Phase type distributions (PH-distributions) are used. We analyze the resulting systems as quasi-birth-death processes, mainly using matrix geometric method.

Now we turn to the content of the thesis. The thesis entitled "Analysis of some priority queues and a problem on diagnostics" is divided into 6 chapters including the introductory chapter. The chapters 2 and 3 discuss doubly infinite queues, chapter 4 discuss a crowdsourcing model and chapters 5 and 6 are on diagnostic problems. All models discussed in this thesis involve interruption in service in some form or other.

In chapter 2 we consider a priority queueing system where low priority customers are generated by self-interruption while at service. Customers arrive to a single service station from a Poisson stream and form a queue ($\mathcal{P}_1$ line) of infinite capacity, if the server is found busy. They are served one at a time according to FIFO discipline. Customers may have a tendency to interrupt their own service while availing the same due to various reasons. Self interrupted customers are pushed to an infinite capacity low priority ($\mathcal{P}_2$) queue. If the customer at P2 line interrupts his service again, he is sent to a further lower priority queue ($\mathcal{P}_3$ -line) and this may go on a finite number (say N) of times. When at a service completion epoch of a $\mathcal{P}_i$ customer, if there is none left behind in $\mathcal{P}_1$ line, then the server goes to serve customers in $\mathcal{P}_{i+1}$ line. The service time for each category is assumed to follow exponential distribution, but at different service rates. The interruptions that happen are also according to exponential distributions with different parameters. We consider both preemptive and non-preemptive service discipline. We analyze a two priority system in detail where we assume that $\mathcal{P}_2$ customers are not allowed to interrupt their service. The joint system state distribution is obtained from which the marginals are computed. Waiting time distribution of both type of customers are derived. We extend the results to three priority non-preemptive case and the case of $N+1$ priorities is briefly discussed.

Chapter 3 is a modification of the hitherto notion of feedback in queueing theory (see page no.3). Here we analyze a two priority queueing system where high priority ($\mathcal{P}_1$) customers may feedback according to a Bernoulli process if they are not satisfied with the service provided. but they will have to join the tail end of the low priority($\mathcal{P}_2$) line. Arrival of both type of customers are according to independent Poisson processes. Both waiting rooms have infinite capacity. Customers are served one at a time according to FIFO discipline on priority basis: those in $\mathcal{P}_1$ are given priority over the ones in the waiting line $\mathcal{P}_2$. The service time is class dependent phase type. $\mathcal{P}_2$ line customers will be serviced only when, if there is none left behind in $\mathcal{P}_1$ line, at the service completion epoch of a high priority customer. Being a two priority system we assume that $\mathcal{P}_2$ customers are not allowed an additional feedback. Thus the system consists of a primary waiting line and a second waiting line which is generated from the first as well as by customers from outside. We consider both preemptive and non-preemptive service discipline. The joint steady state probability distribution is derived and the corresponding marginal probabilities are computed. The distribution of waiting time of each type of customers is derived. We also point out a situation where there is no external entry to the $\mathcal{P}_2$ line which makes the $\mathcal{P}_2$ line exclusively for feedback customers. Even this special case does not boil down to the main problem discussed in chapter 2, since there, the self-interruption is during service.

Going on, we analyze a crowdsourcing queueing model in chapter 4. We consider a $c-$server queueing system providing service to two types of customers, $\mathcal{P}_1$ and $\mathcal{P}_2$. Customers arrive according to two distinct Poisson processes. A $\mathcal{P}_1$ customer has to receive service by one of the c servers while a Type 2 customer may be served by a $\mathcal{P}_1$ customer who is available to act as a server soon after getting own service or by one of c servers. A $\mathcal{P}_1$ customer will be available for serving a $\mathcal{P}_2$ customer with certain probability provided there is at least one $\mathcal{P}_2$ customer waiting in the queue at the time of the service completion of that $\mathcal{P}_1$ customer. With complementary probability, a $\mathcal{P}_1$ customer will opt out of serving a $\mathcal{P}_2$ customer, if any, waiting in the system. A free server offers service to a $\mathcal{P}_1$ customer on a FCFS basis. However, if there is no $\mathcal{P}_1$ customer waiting in the system, that server will serve a $\mathcal{P}_2$ customer if one of that type is present in the queue. The service time is exponentially distributed for each category. $\mathcal{P}_1$ customers have priority over those of $\mathcal{P}_2$. We consider preemptive service discipline. Condition for system stability is established. Important system characteristics including the average number of busy servers, the loss probability and the expected waiting time of each type in the system are computed. Some examples are numerically illustrated. Finally the characteristics of this model are compared with that of Chakravarthy and Dudin [11].

Now we turn to the diagnostic problem. In real life, there are several service providing systems offering a multitude of service. Neither the server nor the customer may be fully aware of the exact service requirement. Very often this results in irreparable damage to the customer being served. It is this type of problems that we analyze in chapters 5 and 6.

Chapter 5 discusses a queueing model with a single server offering many services to which arrival is according to a MAP forming a single line. The time taken for completing service is phase type distributed. A service could be appropriate or inappropriate for each customer. If the service starts in an inappropriate state with a positive probability, we assume a clock to start ticking simultaneously. In case the service time exceeds the realization of the clock, then that customer is compelled to leave the system forever without being eligible for the service that he actually requires. Otherwise the customer gets the required service and then leaves the system. Several system performance measures including the rate of loss of customers, rate of customers leaving with correct service, even if started in incorrect service, are computed. Numerical illustrations of the system behavior are also provided. Then this is compared with that of Madan [46] and Medhi [48]. Also we employ arbitrarily distributed service time in certain special cases of the model discussed here and analyze the system using supplementary variable technique [19].

In Chapter 6, we extend the discussion in previous chapter to a single server system offering n distinct services. Arrival of customers is according to MAP and service time has phase type distribution. For a customer any one among the n services is required and the remaining $n-1$ are damaging (undesirable/ inessential) for him. For different customers the exact service requirement may differ. When the service starts, a timer starts simultaneously whose realization determines the success of service. Several performance measures, including the expected service time of a customer, are evaluated. Effects of various parameters on the performance measures are numerically investigated.

Finally a section of “concluding remarks and suggestions for future study” is included.

Chapter 2

Priority Queues Generated through Customer Induced Service Interruption

The purpose of this chapter is to introduce a priority queue through 'self interruption' of service by customers. Such self interrupted customers are asked to join a lower priority queue. Earlier reported works considered server induced interruptions only. This can be in the form of breakdown or going on vacation or to attend higher priority customers and so on. Varghese Jacob [63] in his doctoral thesis describes several queueing models where customers in service interrupt their own service. However, he gives priority to such interrupted customers. Further he assumes availability of only a finite waiting space for such self interrupted customers. In contrast here we give lower priority for self interrupted customers and the limitation in waiting for such customers as Varghese Jacob [63] is taken

---

Some results of this chapter are appeared in Neural, Parallel and Scientific Computations: *A. Krishnamoorthy and Manjunath A. S.* : **On Priority Queues Generated through Customer Induced Service Interruption**, Neural, Parallel and Scientific Computations. 23 (2015), 459-486

away.

The priority queue considered by Miller [49] has two waiting lines, each of infinite capacity which are served by a single server. The arrival process to the two queues form two independent Poisson streams with parameters λ_1 and λ_2. The low as well as high priority customers, whether in service or in queue, is counted as the number of such customers in the system. The service time duration for high(low) priority customer has exponential distribution with parameter $\mu_1(\mu_2)$. Both preemptive and non-preemptive service disciplines are considered. The system is analyzed as a three dimensional continuous time Markov chain. The system is stable when $\frac{\lambda_1}{\mu 1} + \frac{\lambda_2}{\mu_2} < 1$. As an extension of the above, Sapna and Stanford [58] studied a single server queue with arrivals from N classes of customers on a non-preemptive priority basis. These arrivals follow independent Poisson processes with rates $\lambda_i, i = 1, 2, ..., N$ with class dependent phase type service. The capacity of each waiting line is assumed to be infinite. They analyze the queue length and waiting time processes by deriving a matrix geometric solution for the stationary distribution of the underlying Markov chain.

The above models consider distinct streams of independent Poisson arrivals to the system. In contrast, in this chapter we consider a 2 priority queueing system where input streams are dependent. The high priority($\mathcal{P}_1$) line has input from outside the system (external arrival) according to a Poisson process of rate λ, whereas the low priority($\mathcal{P}_2$) line has input from the high priority waiting line. Thus, low priority queue is generated from within the system. Hence the system that we consider is a highly dependent one as far as the formation of the low priority waiting line is concerned unlike the priority queues with infinite waiting lines that are so far considered in the literature. The same server serves different customers one at a time according to their priority. As an example consider the queue of patients($\mathcal{P}_1$) waiting to consult a physician. A patient while being examined may have to be referred to a specialist. After consulting the specialist the patient returns to the first physician and waits in the second queue($\mathcal{P}_2$). Unlike in [20, 26, 37] here we do not associate any specific distribution for the duration of interruption of a customer; rather we assume that, once an interrupted customer comes to $\mathcal{P}_2$, he is ready to receive service.

We do an extensive analysis in the two priority case: high priority of external (primary or $\mathcal{P}_1$) customers and a second queue (low priority or $\mathcal{P}_2$) of customers who interrupted their service while being served in the high priority queue. With a maximum of a single interruption permitted, we analyze the system as a three dimensional continuous time Markov Chain. Customers from each waiting class is taken for service according to the head of the queue discipline. When no high priority customer is available at a service completion epoch the server starts service of the head of the low priority queue. By a suitable arrangement and adjustment, we produce an upper triangular (infinite dimensional) rate matrix R. Once this is achieved, we will be in a position to compute the steady-state probability vector. Then this is utilized in the computation of performance of the system. The performance measures here, unlike in other set up, will be of a bit of curiosity as well. This is due to the dependence of the second queue on the first for its generation. Having done these, we proceed to the case of 3 queues (one primary and the other two generated from previous higher priority). Finally we briefly extend our results to the case of $N + 1$ queues, $N \geq 3$. In all these the systems are studied under steady-state. Therefore first we establish the condition for stability of the system and then proceed to the analysis. A special feature of the present model, unlike in classical priority models, is that when the server is in P_i queue all P_j queues except $\mathcal{P}_1$ queue turn out to be empty for $i > j$.

This chapter is arranged as follows: In section 1, the case of two priorities is extensively analyzed for the preemptive case. Section 2 is devoted to the study of two priority, non-preemptive service discipline. The discussion in section 2 is extended to three priority set up in section 3 and finally section 4 provides a brief description of $N + 1$ priority system with $N \geq 3$.

2.1 Two priority queues -Preemptive priority

Consider a single server infinite capacity queuing system in which customers from outside arrive according to a Poisson process with rate λ. Service time of the external customers ($\mathcal{P}_1$) are exponentially distributed with parameter μ_1. Customers in primary queue interrupt their service according to an exponentially distributed time with parameter θ_1, in which case they have to go to the lower priority($\mathcal{P}_2$) queue; else, complete service and leave the system forever. Suppose at the time when a $\mathcal{P}_1$ customer leaves the server by self interruption, and hence joins $\mathcal{P}_2$, finds that none is ahead of him and there was none left behind him in $\mathcal{P}_1$. In this case he is immediately taken for service in $\mathcal{P}_2$. Lower priority customers are taken for service one at a time from the head of the line whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a preemptive service discipline following an exponential distribution with parameter μ_2. That is, the arrival of a $\mathcal{P}_1$ customer interrupts the ongoing service of a $\mathcal{P}_2$ customer and hence he joins back as the head of the $\mathcal{P}_2$ queue. Consider the case where not more than one interruption is permitted, that is $N = 1$. Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers including the one in service if any and $N_2(t)$ the number of $\mathcal{P}_2$ customers waiting to get service. Whenever $\mathcal{P}_1$ is nonempty, the head of that line will be under service.

Then $\Omega = \{(N_1(t), N_2(t))\,/t \geq 0\}$ is a continuous time Markov chain with state space $\{0\} \cup \{(i,j)/i \geq 0, j \geq 0\}$. Here 0 represents an idle server and $(0,0)$ is the state where a $\mathcal{P}_2$ customer is in service with no $\mathcal{P}_1$ customer in the system.

The infinitesimal generator Q has as entries block matrices of infinite dimension since the phases (capacity of waiting line for interrupted customers) is infinite. It is given by

$$Q = \begin{pmatrix} B_{00} & B_{01} & & & \\ B_{10} & B_1 & B_0 & & \\ & B_2 & B_1 & B_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

with

$$B_{00} = \begin{pmatrix} -\lambda & & & \\ \mu_2 & -(\lambda+\mu_2) & & \\ & \mu_2 & -(\lambda+\mu_2) & \\ & & \ddots & \ddots \end{pmatrix}, B_{10} = B_2 = \begin{pmatrix} \mu_1 & \theta_1 & & \\ & \mu_1 & \theta_1 & \\ & & \ddots & \ddots \\ & & & \end{pmatrix}$$

$B_{01} = B_0 = \lambda I_\infty$ and $B_1 = -(\lambda + \mu_1 + \theta_1) I_\infty$.

We now establish the system stability requirement.

Theorem 2.1.1. *The condition for stability of the system is*

$$\rho = \frac{\lambda}{(\mu_1 + \theta_1)} + \frac{\lambda\,\theta_1}{(\mu_1 + \theta_1)\,\mu_2} \;<\; 1.$$

Proof. By interchanging the level and phase in the model, the matrices B_0, B_1 and B_2 are $B_0 = \begin{cases} \theta_1,\ i = 1,2,3,...; j = i-1 \\ 0,\ \text{elsewhere} \end{cases}$,

$$B_1 = \begin{cases} -(\lambda+\mu_2),\ i = j = 0 \\ -(\lambda+\mu_1+\theta_1),\ i = j = 1,2,... \\ \lambda,\ i = 0,1,2,...; j = i+1 \\ \mu_1,\ i = 1,2,3,...; j = i-1 \\ 0,\ \text{elsewhere} \end{cases} \quad \text{and } B_2 = \begin{cases} \mu_2,\ i = j = 0 \\ 0,\ \text{elsewhere} \end{cases}.$$

Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, ...)$ be the steady-state probability vector of the matrix $B(= B_0 + B_1 + B_2)$. Solving the relations $\boldsymbol{\pi} B = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$, we get

$\pi_j = \left(\frac{\lambda}{\mu_1+\theta_1}\right)^j \pi_0, j \geq 1$. As we have a level independent QBD model, the system is stable if $\boldsymbol{\pi} B_0 \mathbf{e} < \boldsymbol{\pi} B_2 \mathbf{e}$, which simplifies to $\rho < 1$. □

The infinitesimal generator Q constitutes a quasi birth and death(QBD) process with exceptional boundary behavior and an infinite number of sub-levels. The matrix geometric form of the steady-state distributions for both preemptive and non-preemptive priority single server queues were investigated by Neuts [52] in the case when number of phases in each level is finite. This is extended to blocks of infinite size in Miller [49] and is contained in the following theorem.

Theorem 2.1.2. *Let $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots)$ denote the invariant probability vector for the QBD process Q, where $\boldsymbol{y}_i$ is the probability vector of infinite dimension corresponding to level i . Then the solution for $\boldsymbol{y}$ possesses a matrix geometric structure*

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i R, \ i \geq 1. \tag{2.1}$$

where the rate matrix R is the minimal non negative solution to

$$R^2 B_2 + R B_1 + B_0 = O. \tag{2.2}$$

The matrix geometric structure in equation (2.1) extended to level '0' is

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 \left(\frac{1}{\lambda} B_{01}\right) R. \tag{2.3}$$

Proof. The relations (2.1) and (2.2) are proved in [49].
From $\mathbf{y}Q = 0$, the two boundary equations involving $\mathbf{y}_0$ are

$$\mathbf{y}_0 B_{00} + \mathbf{y}_1 B_{10} = \mathbf{0}, \tag{2.4}$$

$$\mathbf{y}_0 B_{01} + \mathbf{y}_1 [B_1 + R B_2] = \mathbf{0}. \tag{2.5}$$

From (2.2) it follows that

$$R[R B_2 + B_1] = -B_0.$$

Since $B_0 = \lambda I_\infty$, the matrix R is invertible and (2.5) now simplifies to (2.3). □

Theorem 2.1.3. *The infinite matrix R possesses the Toeplitz structure*

$$R = \begin{pmatrix} r_0 & r_1 & r_2 & r_3 & \cdots \\ 0 & r_0 & r_1 & r_2 & \cdots \\ 0 & 0 & r_0 & r_1 & \cdots \\ 0 & 0 & 0 & r_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where r_k are computed as

$$r_0 = \frac{(\lambda + \mu_1 + \theta_1) - \sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}}{2\mu_1},$$

$$r_1 = \frac{r_0^2 \theta_1}{\sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}},$$

$$r_k = \frac{\theta_1 \left[\sum_{i=0}^{k-1} r_i r_{k-1-i}\right] + \mu_1 \left[\sum_{i=1}^{k-1} r_i r_{k-i}\right]}{\sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}}, k > 1.$$

Proof. The structure of the process revealed by matrices in Q and the interpretation of rate matrix imply the special structure of R. On expanding (2.2), the following relations are obtained;

$$r_0^2 \mu_1 - (\lambda + \mu_1 + \theta_1) r_0 + \lambda = 0.$$

$$\left(\sum_{i=0}^{k-1} r_i r_{k-1-i}\right) \theta_1 + \left(\sum_{i=0}^{k} r_i r_{k-i}\right) \mu_1 - (\lambda + \mu_1 + \theta_1) r_k = 0, \quad k \geq 1.$$

Solving these, the expressions for $r_k, k = 0, 1, 2, ...$ are established. □

2.1.1 The joint and marginal probabilities

The Joint Probabilities

The steady-state probability vector $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, \ldots)$ of Q is computed first. The probability of idle state $\mathbf{y}_0 = (\mathbf{y}_0, \mathbf{y}_{00}, \mathbf{y}_{01}, \mathbf{y}_{02}, ...)$ with $\mathbf{y}_{00}$ denoting the probability that service is providing to a $\mathcal{P}_2$ customer when none is waiting in either queue. $\mathbf{y}_i = (\mathbf{y}_{i0}, \mathbf{y}_{i1}, \mathbf{y}_{i2}, ...)$ with $\mathbf{y}_{ij}$ representing the probability that the number of $\mathcal{P}_1$ customers in the system is i and that in $\mathcal{P}_2$ queue is j for $i > 1$. Equations (2.3) and (2.4) give

$$\begin{aligned}
\mathbf{y}_0 &= 1 - \rho; \quad \rho = \tfrac{\lambda}{(\mu_1+\theta_1)} + \tfrac{\lambda\,\theta_1}{(\mu_1+\theta_1)\mu_2}, \\
\mathbf{y}_{00} &= \tfrac{1}{\mu_2}\left(\lambda - r_0\mu_1\right)\mathbf{y}_0, \\
\mathbf{y}_{01} &= \tfrac{1}{\mu_2}\left\{\left(\lambda + \mu_2 - r_0\mu_1\right)\mathbf{y}_{00} - \left(r_0\theta_1 + r_1\mu_1\right)\mathbf{y}_0\right\},
\end{aligned}$$

$$\mathbf{y}_{0j} = \frac{1}{\mu_2}\Bigg\{\left(\lambda + \mu_2 - r_0\mu_1\right)\mathbf{y}_{0,j-1} - \theta_1\sum_{k=0}^{j-2} r_k\mathbf{y}_{0,j-2-k} - \mu_1\sum_{k=1}^{j-1} r_k\mathbf{y}_{0,j-1-k} - (r_{j-1}\theta_1 + r_j\mu_1)\mathbf{y}_0\Bigg\}, \quad j > 1.$$

Thus $\mathbf{y}_{0j}$ is recursively computed up to the desired range of values.
Substituting for $\mathbf{y}_0$ in equation (2.3) and expanding, $\mathbf{y}_{1j}, j = 0, 1, 2, ...$ are computed as

$$\begin{aligned}
\mathbf{y}_{10} &= (1-\rho)\,r_0, \\
\mathbf{y}_{11} &= (1-\rho)\,r_1 + \mathbf{y}_{00}\,r_0, \\
\mathbf{y}_{1j} &= (1-\rho)\,r_j + \sum_{k=0}^{j-1}\mathbf{y}_{0k}\,r_{j-1-k}, \; j = 2, 3, ...
\end{aligned}$$

Finally expression (2.1) on expansion results in

$$\mathbf{y}_{ij} = \sum_{k=0}^{j}\mathbf{y}_{i-1,k}\,r_{j-k}, i > 1. \tag{2.6}$$

After obtaining $\mathbf{y}_{0j}$ and $\mathbf{y}_{1j}$ for $j = 0, 1, 2, ...,$ the probabilities $\mathbf{y}_{ij}, i > 1$ are recursively computed using (2.6).

The Marginal Probabilities

Let the marginal probabilities of the number of $\mathcal{P}_1$ customers in the system be denoted by $\mathbf{y}_{i\bullet} = \sum_{j=0}^{\infty} \mathbf{y}_{ij}, \quad i \geq 0$. Then in recursive form

$$\mathbf{y}_{i\bullet} = \sum_{j=0}^{\infty} \sum_{k=0}^{j} \mathbf{y}_{i-1,k}\, r_{j-k} = \left(\sum_{j=0}^{\infty} \mathbf{y}_{i-1,j} \right) \left(\sum_{i=0}^{\infty} r_i \right) = \mathbf{y}_{(i-1)\bullet}\; \rho_1.$$

Remark: As an arrival of a $\mathcal{P}_1$ customer preempts a $\mathcal{P}_2$ customer in service, the system behaves as an M/M/1 queue as far as marginal probabilities of $\mathcal{P}_1$ customers are concerned. Hence

$$\mathbf{y}_{i\bullet} = \rho_1^i (1 - \rho_1), i \geq 0; \quad \rho_1 = \frac{\lambda}{\mu_1 + \theta_1}$$

The marginal distribution of $\mathcal{P}_2$ customers is computed numerically from

$$\mathbf{y}_{\bullet j} = \sum_{i=0}^{\infty} \mathbf{y}_{ij}, \quad j \geq 0.$$

2.1.2 Waiting time analysis

Waiting time of high priority customers

As an arriving $\mathcal{P}_1$ customer preempts a $\mathcal{P}_2$ customer under service if there is any, the distribution of waiting time in $\mathcal{P}_1$ line is same as in the case of an M/M/1 queue. Hence expected waiting time of a $\mathcal{P}_1$ customer in the system is

$$E(W_{\mathcal{P}_1}) = \frac{\rho_1}{\lambda\,(1 - \rho_1)} = \frac{1}{\mu_1 + \theta_1 - \lambda}$$

Waiting time of Low priority customers

Expected waiting time $E(WT_{\mathcal{P}_2})$ of an interrupted customer, provided he is the head of the $\mathcal{P}_2$ line, is the sum of the following: expected busy cycle generated by the primary customers left behind by this customer when he interrupted own service while in $\mathcal{P}_1$, the sum of the expected busy cycles generated at each preemption while chosen for service from $\mathcal{P}_2$ line and expected time taken to complete service without a preemption. We get

$$E(WT_{\mathcal{P}_2}) = \frac{1}{(\mu_1+\theta_1)}\frac{\rho_1}{(1-\rho_1)^2} + \frac{1}{\mu_2(1-\rho_1)}$$

Expected waiting time of a $\mathcal{P}_2$ customer if he is anywhere in the $\mathcal{P}_2$ line is

$$E(W_{\mathcal{P}_2}) = \frac{1}{(\mu_1+\theta_1)}\frac{\rho_1}{(1-\rho_1)^2} + \frac{1}{\mu_2}\frac{1}{(1-\rho_1)}E(\mathcal{P}_2); \quad E(\mathcal{P}_2) = \sum_{r=1}^{\infty} r\mathbf{y}_{\bullet r}$$

2.2 Two priority queues - Non-preemptive priority

We analyze a two-priority queueing model similar to that in the previous section, except that service to $\mathcal{P}_2$ customers is according to a non-preemptive service discipline. That is the arrival of a $\mathcal{P}_1$ customer does not interrupt the ongoing service of a $\mathcal{P}_2$ customer. We follow the same notations. Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system including the one in service if any, $N_2(t)$ be the number in the $\mathcal{P}_2$ waiting line and $S(t)$ the status of the server which is 1 or 2 according as the server is busy with $\mathcal{P}_1$ customers or $\mathcal{P}_2$ customers . Thus we get a continuous time Markov chain $\Omega = \{X(t), t \geq 0\} = \{(N_1(t), N_2(t), S(t))/t \geq 0\}$. Its state space is given as $\{(0,0)\} \cup \{(0,j,2)/j \geq 0\} \cup \{(i,j,k)/i > 0, j \geq 0, k = 1,2\}$. Clearly, the system is stable if $\frac{\lambda}{(\mu_1+\theta_1)} + \frac{\lambda\,\theta_1}{(\mu_1+\theta_1)\mu_2} < 1,$

The infinitesimal generator of this continuous time Markov chain consists of block entries of infinite dimension and is obtained as

$$Q^* = \begin{pmatrix} A_{00} & A_{01} & & & \\ A_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$$A_{00} = \begin{pmatrix} -\lambda & & & \\ \mu_2 & -(\lambda+\mu_2) & & \\ & \mu_2 & -(\lambda+\mu_2) & \\ & & \ddots & \ddots \end{pmatrix}, A_{01} = \begin{pmatrix} \lambda & & & & & & & \\ & \lambda & 0 & \dots & & & & \\ & & & \lambda & 0 & \dots & & \\ & & & & & \lambda & 0 & \\ & & & & & & \ddots & \end{pmatrix},$$

$$A_{10} = \begin{pmatrix} \begin{pmatrix} \mu_1 \\ 0 \end{pmatrix} & \begin{pmatrix} \theta_1 & 0 \\ 0 & 0 \end{pmatrix} & & \\ & \begin{pmatrix} \mu_1 & \theta_1 \\ 0 & 0 \end{pmatrix} & & \\ & & \begin{pmatrix} \mu_1 & \theta_1 \\ 0 & 0 \end{pmatrix} & \\ & & & \ddots \end{pmatrix},$$

$$A_2 = \begin{pmatrix} M_2 & M_3 & & \\ & M_2 & M_3 & \\ & & \ddots & \ddots \end{pmatrix}, \quad A_1 = \begin{pmatrix} M_1 & & \\ & M_1 & \\ & & \ddots \end{pmatrix}, \quad A_0 = \lambda I_\infty;$$

$$M_1 = \begin{pmatrix} -(\lambda+\mu_1+\theta_1) & 0 \\ \mu_2 & -(\lambda+\mu_2) \end{pmatrix}, \; M_2 = \begin{pmatrix} \mu_1 & 0 \\ 0 & 0 \end{pmatrix}, \; M_3 = \begin{pmatrix} \theta_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The infinitesimal generator Q^* constitutes a quasi birth and death(QBD) process with infinite number of sub-levels. As Q^* is irreducible and recurrent, following a similar argument to theorem 3 of Miller [49] we have,

Theorem 2.2.1. *Let $\mathbf{x} = [\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots]$ denote the invariant probability vector for the QBD process Q^* with infinite number of sub levels(phases), where $\mathbf{x}_i$ is the probability vector corresponding to level i of infinite dimension. Then* *the solution for* $\mathbf{x}$ *possesses a matrix geometric structure*

$$\mathbf{x}_i = \mathbf{x}_{i-1}\boldsymbol{R},\ i > 1. \tag{2.7}$$

where the rate matrix $\boldsymbol{R}$ *is the minimal non negative solution to*

$$\boldsymbol{R}^2 A_2 + \boldsymbol{R}A_1 + A_0 = O. \tag{2.8}$$

Theorem 2.2.2. *The* $\boldsymbol{R}$ *matrix, which is the minimal non negative solution to equation (2.8) possesses a Toeplitz structure* $(\boldsymbol{R}_0, \boldsymbol{R}_1, \boldsymbol{R}_2, ...)$*. That is* $\boldsymbol{R}$ *has the form*

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_0 & \boldsymbol{R}_1 & \boldsymbol{R}_2 & \boldsymbol{R}_3 & \cdots \\ 0 & \boldsymbol{R}_0 & \boldsymbol{R}_1 & \boldsymbol{R}_2 & \cdots \\ 0 & 0 & \boldsymbol{R}_0 & \boldsymbol{R}_1 & \cdots \\ 0 & 0 & 0 & \boldsymbol{R}_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where each of the matrices $\boldsymbol{R}_k$ *is of order 2 represented as* $\boldsymbol{R}_k = \begin{bmatrix} a_k & 0 \\ b_k & c_k \end{bmatrix}$.

Proof. The interpretation of $\boldsymbol{R}$ in Neuts [52] and the structure of the matrices in the generator matrix Q proves the theorem. □

Theorem 2.2.3. *The elements* $\boldsymbol{R}_k (k > 0)$ *in theorem 2.2.2 are computed as,*

$$\begin{aligned} a_k &= \frac{(\sum_{i=o}^{k-1} a_i a_{k-1-i})\theta + (\sum_{i=1}^{k-1} a_i a_{k-i})\mu_1}{(\lambda + \mu_1 + \theta) - 2a_0\mu_1}, \\ b_k &= \frac{(\sum_{i=1}^{k-1} a_i b_{k-1-i} + b_{k-1}(a_0 + c_0))\theta + (\sum_{i=1}^{k} a_i b_{k-i})\mu_1}{(\lambda + \mu_1 + \theta) - (a_0 + c_0)\mu_1}, \\ c_k &= 0, \qquad k = 1, 2, 3, \ldots \end{aligned}$$

and entries of r_0 are

$$a_0 = \frac{(\lambda + \mu_1 + \theta) - \sqrt{(\lambda + \mu_1 + \theta)^2 - 4\mu_1\lambda}}{2\mu_1},$$
$$b_0 = \frac{\mu_2 c_0}{(\lambda + \mu_1 + \theta) - (a_0 + c_0)\mu_1},$$
$$c_0 = \frac{\lambda}{\lambda + \mu_2}.$$

Proof. Expanding (2.8), we obtain the following relations:

$$\boldsymbol{R}_0^2 M_2 + \boldsymbol{R}_0 M_1 + \lambda I = \mathrm{O},$$
$$\boldsymbol{R}_0^2 M_3 + \left(\sum_{k=0}^{l} \boldsymbol{R}_k \boldsymbol{R}_{l-k}\right) M_2 + \boldsymbol{R}_l M_1 = \mathrm{O}, l \geq 1.$$

The result is established when these equations are expanded with respect to the phases. □

2.2.1 The joint and marginal probabilities

In this section recursive formulae for the joint distribution of i $\mathcal{P}_1$ customers in the system and j $\mathcal{P}_2$ customers in the queue and marginal distributions of each are derived. First we establish the following.

Theorem 2.2.4. *The matrix geometric structure*

$$\mathbf{x}_i = \mathbf{x}_{i-1}\boldsymbol{R}, \quad i > 1$$

given in theorem 2.2.1 extended to level 0 is

$$\mathbf{x}_i = \mathbf{x}_0 \left(\frac{1}{\lambda} A_{01}\right) \boldsymbol{R}^i, \quad i \geq 1.$$

Proof. From $\mathbf{x}Q^* = \mathbf{0}$, the two boundary equations involving $\mathbf{x}_0$ are

$$\mathbf{x}_0 A_{00} + \mathbf{x}_1 A_{10} = \mathbf{0}, \tag{2.9}$$

$$\mathbf{x}_0 A_{01} + \mathbf{x}_1[A_1 + \boldsymbol{R}A_2] = \mathbf{0}. \tag{2.10}$$

From (2.8) it follows that

$$\boldsymbol{R}[\boldsymbol{R}A_2 + A_1] = -A_0. \tag{2.11}$$

Since $A_0 = \lambda I_\infty$, $\boldsymbol{R}$ is invertible.
From (2.10) and (2.11) we get

$$\mathbf{x}_1 = \mathbf{x}_0 \left(\frac{1}{\lambda} A_{01} \boldsymbol{R} \right). \tag{2.12}$$

Combining relations (2.7) and (2.12) we obtain

$$\mathbf{x}_i = \mathbf{x}_0 \left(\frac{1}{\lambda} A_{01} \right) \boldsymbol{R}^i, \quad i \geq 1. \tag{2.13}$$

□

The Joint probability distribution

Let $\mathbf{x}_{ij}$ be the probability that there are i $\mathcal{P}_1$ customers in the system and j $\mathcal{P}_2$ customers waiting in queue. Further let the marginal distribution of the number of $\mathcal{P}_1$ customers in the system be denoted by $\mathbf{x}_{i\bullet}$. Then

$$\mathbf{x}_{i\bullet} = \sum_{j=0}^{\infty} \mathbf{x}_{ij} \ , \quad i \geq 0.$$

To know the type of customer in service we partition $\mathbf{x}_{ij}$ as

$$\mathbf{x}_{ij} = (\mathbf{x}_{ij}(1), \mathbf{x}_{ij}(2)).$$

We proceed to determine the joint probability vectors $\mathbf{x}_{ij}$. Considering the interrupted customers and the type of customer under service, equation (2.7) gives

$$\mathbf{x}_{ij} = \mathbf{x}_{i-1,j} \boldsymbol{R} \ , \quad i > 1, j \geq 0. \tag{2.14}$$

where

$$\mathbf{x}_{ij} = (\mathbf{x}_{ij}(1), \mathbf{x}_{ij}(2)).$$

Expanding (2.14) w.r.t. j

$$(\mathbf{x}_{i0}, \mathbf{x}_{i1}, \cdots) = (\mathbf{x}_{i-1,0}, \mathbf{x}_{i-1,1}, \cdots) \times \begin{pmatrix} \boldsymbol{R}_0 & \boldsymbol{R}_1 & \boldsymbol{R}_2 & \cdots \\ 0 & \boldsymbol{R}_0 & \boldsymbol{R}_1 & \cdots \\ 0 & 0 & \boldsymbol{R}_0 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

In general,

$$\mathbf{x}_{ij} = \sum_{k=0}^{j} \mathbf{x}_{i-1,k} \ \boldsymbol{R}_{j-k} \ , \quad i > 1, j \geq 0. \tag{2.15}$$

Expanding these equations once more to reveal the dependence on the type of service, we obtain

$$\mathbf{x}_{ij}(1) = \sum_{k=0}^{j} [a_{j-k} \ \mathbf{x}_{i-1,k}(1) + b_{j-k} \ \mathbf{x}_{i-1,k}(2)] \tag{2.16}$$

$$\mathbf{x}_{ij}(2) = c_0 \ \mathbf{x}_{i-1,j}(2) \ ; \qquad i > 1, j \geq 0. \tag{2.17}$$

Equation (2.12) on expansion gives

$$\mathbf{x}_{1j}(1) = a_j(1-\rho) + \sum_{k=0}^{j} b_{j-k} \ \mathbf{x}_{0k}(2) \tag{2.18}$$

$$\mathbf{x}_{1j}(2) = c_0 \ \mathbf{x}_{0j}(2) \ ; \qquad i = 1, j \geq 0. \tag{2.19}$$

Hence the joint probabilities depend on $\mathbf{x}_{0k}(2)$ for $k = 0, 1, 2, \ldots j$. We compute $\mathbf{x}_{0k}(2)$ in the desired range in the next section.

Marginal distribution of high priority customers

Adding equation(2.15) over j which is the low priority queue length, the marginal distribution $\mathbf{x}_{i\bullet}$ for the number of $\mathcal{P}_1$ customers in the system is

$$\begin{aligned} \mathbf{x}_{i\bullet} &= \sum_{j=0}^{\infty} \mathbf{x}_{ij} = \sum_{j=0}^{\infty}\sum_{k=0}^{j} \mathbf{x}_{i-1,k}\ \boldsymbol{R}_{j-k} = \sum_{k=0}^{\infty} \mathbf{x}_{i-1,k} \left(\sum_{j=0}^{\infty} \boldsymbol{R}_j \right) \\ &= \mathbf{x}_{(i-1)\bullet} \mathcal{R}_+ \qquad (2.20) \\ &= \mathbf{x}_{1\bullet} \mathcal{R}_+^{i-1}, \qquad i \geq 2; \qquad (2.21) \end{aligned}$$

where

$$\mathcal{R}_+ = \sum_{j=0}^{\infty} \boldsymbol{R}_j = \begin{bmatrix} \sum_{r=0}^{\infty} a_r & 0 \\ \sum_{r=0}^{\infty} b_r & c_0 \end{bmatrix}.$$

Now, expanding (2.20) based on the type of service, we have

$$\begin{aligned} \begin{pmatrix} \mathbf{x}_i(1), & \mathbf{x}_i(2) \end{pmatrix} &= \begin{pmatrix} \mathbf{x}_{i-1}(1), & \mathbf{x}_{i-1}(2) \end{pmatrix} \begin{bmatrix} \sum a_r & 0 \\ \sum b_r & c_0 \end{bmatrix}, i \geq 1 \\ &= \begin{pmatrix} \mathbf{x}_{i-1}(1)\,(\sum a_r) + \mathbf{x}_{i-1}(2)\,(\sum b_r), & \mathbf{x}_{i-1}(2)\ c_0 \end{pmatrix} \end{aligned}$$

So we obtain

$$\begin{aligned} \mathbf{x}_{i\bullet}(1) &= \mathbf{x}_{i-1,\bullet}(1)\,(\textstyle\sum a_r) + \mathbf{x}_{i-1,\bullet}(2)\,(\textstyle\sum b_r) \\ \mathbf{x}_{i\bullet}(2) &= \mathbf{x}_{i-1,\bullet}(2)\ c_0 \end{aligned}$$

Adding equations (2.18) and (2.19) over j

$$\begin{aligned} \mathbf{x}_{1\bullet}(1) &= (1-\rho)\left(\sum a_r\right) + \mathbf{x}_{0\bullet}(2)\left(\sum b_r\right) \\ \mathbf{x}_{1\bullet}(2) &= c_0\ \mathbf{x}_{0\bullet}(2) \end{aligned}$$

which in turn gives

$$\mathbf{x}_{1\bullet} = ((1-\rho), \mathbf{x}_{0\bullet}(2))\,\mathcal{R}_+ \qquad (2.22)$$

Combining equations (2.21) and (2.22) we get

$$\mathbf{x}_{i\bullet} = ((1-\rho), \mathbf{x}_{0\bullet}(2))\,\mathcal{R}_+^i \quad ; \quad i \geq 1.$$

Write $\mathbf{x}_{0\bullet} = ((1-\rho), \mathbf{x}_{0\bullet}(2))$ then

$$\mathbf{x}_{i\bullet} = \mathbf{x}_{0\bullet}\mathcal{R}_+^i \quad ; \quad i \geq 1.$$

Expanding this we get the high priority marginals as

$$\mathbf{x}_{i\bullet}(1) = (1-\rho)\left(\sum a_r\right)^i + \mathbf{x}_{0\bullet}(2)\sum_{k=0}^{i-1}\left(\sum a_r\right)^k\left(\sum b_r\right)c_0^{i-1-k}$$
$$\mathbf{x}_{i\bullet}(2) = \mathbf{x}_{0\bullet}(2)c_0^i.$$

From the above relations it is clear that the marginal probabilities depend on the probability that no $\mathcal{P}_1$ customer in the system and a $\mathcal{P}_2$ customer in service, which is given by

$$\mathbf{x}_{0\bullet}(2) = \sum_{j=0}^{\infty}\mathbf{x}_{0j}(2). \tag{2.23}$$

To compute $\boldsymbol{x}_{0\bullet}(2)$:

Substituting equation (2.12) in (2.9) we have

$$\mathbf{x}_0\left[A_{00} + \frac{1}{\lambda}(A_{01}\boldsymbol{R}A_{10})\right] = \mathbf{0} \tag{2.24}$$

where $\mathbf{x}_0 = ((1-\rho), \mathbf{x}_{00}(2), \mathbf{x}_{01}(2), \mathbf{x}_{02}(2), ..., ...)$.
Expanding(2.24), the following relations are obtained:

$$\begin{aligned}
\mathbf{x}_{00}(2)\left[b_0\mu_1+\mu_2\right] + (1-\rho)\left[a_0\mu_1-\lambda\right] = {} & \quad 0 \\
\mathbf{x}_{01}(2)\left[b_0\mu_1+\mu_2\right] + \mathbf{x}_{00}(2)\left[b_0\theta_1+b_1\mu_1-(\lambda+\mu_2)\right] + {} & \\
(1-\rho)\left[a_0\theta_1+a_1\mu_1\right] = {} & \quad 0 \\
(1-\rho)\left[a_{j-1}\theta_1+a_j\mu_1\right] + \sum_{k=0}^{j-2}\mathbf{x}_{0k}(2)\left[b_{j-k-1}\theta_1+b_{j-k}\mu_1\right] + {} & \\
\mathbf{x}_{0(j-1)}(2)\left[b_0\theta_1+b_1\mu_1-(\lambda+\mu_2)\right] + \mathbf{x}_{0j}(2)\left[b_0\mu_1+\mu_2\right] = {} & \quad 0, j \geq 2.
\end{aligned}$$

On solving these, we obtain

$$\mathbf{x}_{00}(2) = \frac{(\lambda - a_0\mu_1)(1-\rho)}{b_0\mu_1 + \mu_2} \tag{2.25}$$

$$\mathbf{x}_{01}(2) = \frac{1}{b_0\mu_1 + \mu_2}\Big\{ [(\lambda + \mu_2) - (b_0\theta_1 + b_1\mu_1)]\,\mathbf{x}_{00}(2) - (1-\rho)\,[a_0\theta_1 + a_1\mu_1] \Big\} \tag{2.26}$$

$$\mathbf{x}_{0j}(2) = \frac{1}{b_0\mu_1 + \mu_2}\Big\{ [(\lambda + \mu_2) - (b_0\theta_1 + b_1\mu_1)]\,\mathbf{x}_{0(j-1)}(2) - (1-\rho)\,[a_{j-1}\theta_1 + a_j\mu_1] - \sum_{k=0}^{j-2} \mathbf{x}_{0k}(2)\,[b_{j-k-1}\theta_1 + b_{j-k}\mu_1] \Big\},\ j \geq 2. \tag{2.27}$$

Hence $\mathbf{x}_{0\bullet}(2)$ in equation(2.23) is computed. Also the joint probabilities given by relations (2.16) to (2.19) are evaluated.

Marginal distribution of low priority customers

Define $\mathbf{x}_{\bullet j}(1) = \sum_{i=1}^{\infty} \mathbf{x}_{ij}(1)$ and $\mathbf{x}_{\bullet j}(2) = \sum_{i=0}^{\infty} \mathbf{x}_{ij}(2)$ for $j \geq 0$. Summing equations (2.16) from $i = 2$ to ∞ and adding this to (2.18) we obtain

$$\mathbf{x}_{\bullet j}(1) = a_j(1-\rho) + \sum_{k=0}^{j} [a_{j-k}\ \mathbf{x}_{\bullet k}(1) + b_{j-k}\ \mathbf{x}_{\bullet k}(2)] \tag{2.28}$$

Similarly adding equations (2.17) from $i = 2$ to ∞ and adding this to (2.19),

$$\mathbf{x}_{\bullet j}(2) = \mathbf{x}_{0j}(2) + c_0\ \mathbf{x}_{\bullet j}(2)$$

which gives

$$\mathbf{x}_{\bullet j}(2) = \frac{1}{1-c_0}\ \mathbf{x}_{0j}(2).$$

Hence the marginal probabilities of low priority customers while a $\mathcal{P}_2$ customer is under service, is determined once we evaluate $x_{0j}(2)$ for the desired range of values of j, which is done through equations (2.25) and (2.27). The marginal probabilities of low priority customers, while a $\mathcal{P}_1$ customer is under service, is determined as follows. Substituting for $\mathbf{x}_{0j}(2)$ and putting $k = 0, 1, 2, ..., j$ in (2.28) we get

$$\begin{aligned}
\mathbf{x}_{\bullet 0}(1) &= \frac{a_0\,(1-\rho) + b_0\ \mathbf{x}_{\bullet 0}(2)}{1-a_0}, \\
\mathbf{x}_{\bullet j}(1) &= \frac{a_j\,(1-\rho) + \sum\limits_{k=0}^{j-1} a_{j-k}\ \mathbf{x}_{\bullet k}(1) + \sum\limits_{k=0}^{j} b_{j-k}\ \mathbf{x}_{\bullet k}(2)}{1-a_0}, \quad j \geq 1.
\end{aligned}$$

2.2.2 Waiting time distribution

High priority waiting time distribution

First we compute the expected waiting time of a $\mathcal{P}_1$ customer who joins as the n^{th} customer $n(>0)$, in the queue at the time when he joins. We construct a Markov chain $\{N(t), t \geq 0\}$, where $N(t)$ is the rank of the customer at time t. The rank of a customer is r if he is the r^{th} customer in the queue at time t. His rank improves by 1 as the customers ahead of him leave the system after completing/ self interrupting service. Two cases are to be considered according to whether a $\mathcal{P}_1$ or a $\mathcal{P}_2$ customer is under service when the tagged customer joins.

State space of the Markov chain when a $\mathcal{P}_1$ customer is in service is $\{0\} \cup \{(r,1)\} \cup \{n : 1 \leq n < r\}$ and that when a $\mathcal{P}_2$ customer is in service is $\{0\} \cup \{(r,2)\} \cup \{n : 1 \leq n < r\}$, where $\{0\}$ is the absorbing state indicating that the tagged customer is selected for service. The corresponding infinitesimal generator matrices of dimension $r+1$ are denoted by $\mathcal{W}_1$ and $\mathcal{W}_2$ respectively, and are

$$\mathcal{W}_1 = \begin{bmatrix} T_r & T_r^0 \\ \mathbf{0} & 0 \end{bmatrix}, \mathcal{W}_2 = \begin{bmatrix} S_r & S_r^0 \\ \mathbf{0} & 0 \end{bmatrix} \text{ where,}$$

$$T_r = \begin{cases} -\mu_1\ ,\ i=j=1,2,...,r. \\ \mu_1\ ,\ j=i+1, i=1,2,...,r-1 \\ 0\ ,\ elsewhere \end{cases} ; S_r = \begin{cases} -\mu_2\ , i=j=1 \\ \mu_2\ , i=1, j=2 \\ -\mu_1\ ,\ i=j=1,2,...,r. \\ \mu_1\ ,\ j=i+1, i=1,2,...,r-1 \\ 0\ ,\ elsewhere \end{cases}$$

and $T_r^0 = S_r^0 = [0,\ ...,\ 0,\ \mu_1]^T$.

The expected waiting time of the r^{th} tagged customer is $-(T_r^{-1} + S_r^{-1})\boldsymbol{e}$.

Hence the expected waiting time of a $\mathcal{P}_1$ customer in the queue, with $\boldsymbol{\alpha} = (1,\ 0,\ ...,\ 0)$ a row vector of dimension r is,

$$W_{\mathcal{P}_1} = \sum_{r=1}^{\infty} \left[\left(-\alpha T_r^{-1}\mathbf{e}\right)\ \mathbf{x}_{(r+1)\bullet}(1) + \left(-\alpha S_r^{-1}\mathbf{e}\right) \mathbf{x}_{r\bullet}(2) \right]$$

.

Low priority waiting time distribution

We compute the bounds on the distribution of waiting time of an interrupted (tagged) customer in the system. Suppose the tagged customer joins as $r^{th}(r \geq 1)$ in the system. Upon arrival a tagged customer observes either a free server or the server is busy with a $\mathcal{P}_1$ customer or a $\mathcal{P}_2$ customer. The probability of these events are respectively $1-\rho$, $\boldsymbol{x}_{(r-1)\bullet}(1)$ and $\boldsymbol{x}_{(r-1)\bullet}(2)$. In the first case waiting time is merely his service time. The distribution of waiting time in the second case is Erlang of order r with rate parameter μ_1. For the third case waiting time distribution is the convolution of $exp(\mu_2)$ with service time of r $\mathcal{P}_1$ customers. Hence the distribution of waiting time in the system until the customer feedback is $\mathbf{F}_0(\cdot) = (1-\rho)exp(\mu_1) + \sum_{r;r\geq 2} \mathbf{E}(r,\mu_1)\ \boldsymbol{x}_{(r-1)\bullet}(1) + exp(\mu_2) * \sum_{r;r\geq 1} \mathbf{E}(r,\mu_1)\ \boldsymbol{x}_{(r-1)\bullet}(2)$. $\mathbf{E}(i,\alpha)$ stands for Erlang distribution of order i and parameter α.

Now assume that the tagged customer interrupts his service. Probability to interrupt service is θ_1. We may assume, without loss of generality, that the tagged customer leaves behind i $\mathcal{P}_1$ customers at his service interruption and join as j^{th} in the $\mathcal{P}_2$ line. Each of these i $\mathcal{P}_1$ customers generate a busy cycle exponentially distributed with parameter $(\mu_1 + \theta_1 - \lambda)$. So the service time of all these customers is the i-fold convolution of $exp(\mu_1 + \theta_1 - \lambda)$ with itself. The probability to see i customers behind the tagged customer in $\mathcal{P}_1$ line is $\boldsymbol{x}_{i\bullet}(1)$ and thus the distribution of service time of these i customers is

$$\mathbf{F}_1(\cdot) = \sum_i \mathbf{E}(i, \mu_1 + \theta_1 - \lambda)\, \boldsymbol{x}_{i\bullet}(1)$$

. **The lower bound**.

The waiting time of the tagged customer is minimum if all the $(j-1)$ customers ahead of him in $\mathcal{P}_2$ line complete service in a row once service started in $\mathcal{P}_2$ queue. Assume that service started in $\mathcal{P}_2$ line and no $\mathcal{P}_1$ customer arrives until the tagged(interrupted) customer is taken for service.

The probability that there are $(j-1)$ $\mathcal{P}_2$ ahead of tagged $\mathcal{P}_2$,

$$q_j^{'} = \boldsymbol{x}_{0(j-2)}(2) + \boldsymbol{x}_{\bullet(j-1)}(1).$$

The probability that no $\mathcal{P}_1$ arrived during the service time of a $\mathcal{P}_2$ customer,

$$p_0 = \int_0^\infty e^{-\lambda t} \mu_2 e^{-\mu_2 t}\, dt.$$

Therefore the probability that no $\mathcal{P}_1$ customer arrived during the service time of $((j-1)\ \mathcal{P}_2)$,

$$q_{j-1} = p_0^{j-1}$$

. Hence the distribution of service time until tagged customer completes service is j-fold convolution of $exp(\mu_2)$ with itself multiplied by the probabilities $q_j^{'}$ and q_{j-1}. Therefore the service time distribution of j $\mathcal{P}_2$ customers is

$$\mathbf{F}_2(\cdot) = \sum_j \mathbf{E}(j, \mu_2)\, q_j^{'}\, q_{j-1}.$$

So we get lower bound for the waiting time distribution in the system as

$$\mathbf{F}_{\min wait}(\cdot) = \mathbf{F}_0 * \theta_1 \mathbf{F}_1 * \mathbf{F}_2.$$

Here $*$ stands for the convolution of distributions.

The upper bound.

The waiting time of the tagged customer is maximum if $\mathcal{P}_1$ customers arrive during the service of each of $(j-1)$ customers ahead of the tagged customer in $\mathcal{P}_2$ line. Hence immediately after the service of each $\mathcal{P}_2$ the server goes to $\mathcal{P}_1$ line and returns when no one in the $\mathcal{P}_1$ line. We suppose k $\mathcal{P}_1$ customers lined up during the service of a $\mathcal{P}_2$ customer. The probability of occurrence of this event is

$$p_k = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^k}{k!} \mu_2 e^{-\mu_2 t} \, dt.$$

Service time distribution of these k $\mathcal{P}_1$ customers is k-fold convolution of $exp(\mu_1 - \lambda)$ with itself as each $\mathcal{P}_1$ arrival generates a busy cycle. Therefore service time distribution for the $\mathcal{P}_1$ arrivals during the service of a $\mathcal{P}_2$ customer from among the $(j-1)$ $\mathcal{P}_2$ customers is

$$\sum_k E(k, \mu_1 + \theta_1 - \lambda) \, p_k.$$

Waiting time distribution generated by the service of all $(j-1)$ $\mathcal{P}_2$ ahead of the tagged customer is

$$\mathbf{F}_3(\cdot) = \sum_j \left[\exp(\mu_2) * \sum_k \mathbf{E}(k, \mu_1 + \theta_1 - \lambda) \, p_k \right]^{*(j-1)} q_j^{'}.$$

Here $*r$ denotes the r-fold convolution of a function with itself. Hence the distribution of the maximum waiting time of a feedback customer in the system is

$$\mathbf{F}_{\max wait}(\cdot) = \mathbf{F}_0 * \theta \ \mathbf{F}_1 * \mathbf{F}_3 * \ \exp(\mu_2).$$

2.2.3 Additional performance measures

In what to follow, by a cycle we shall mean, the time duration starting with the arrival of a $\mathcal{P}_1$ customer to an idle server, until all subsequent arrivals are also served out from the $\mathcal{P}_1$ line, resulting in no $\mathcal{P}_1$ customer in the system.

1. The probability that all the $\mathcal{P}_1$ customers served in a given cycle complete service without any interruption is

$$P_{AC} = \frac{\mu_1(\mu_1 + \theta_1 - \lambda)}{(\mu_1 + \theta_1)^2 - \lambda\mu_1}.$$

   This is equivalent to seeking the probability that there is no inflow to $\mathcal{P}_2$ from $\mathcal{P}_1$ during that cycle.

2. The probability that all the $\mathcal{P}_1$ customers served in a given cycle interrupt before completing service and hence join $\mathcal{P}_2$ line is

$$P_{AI} = \frac{\theta_1(\mu_1 + \theta_1 - \lambda)}{(\mu_1 + \theta_1)^2 - \lambda\theta_1}$$

   This is the probability for the other extreme case of 1.

We demonstrate below the impact of fixed values of $\lambda, \mu_1, \text{and} \mu_2$ on P_{AC} and P_{AI} with variations of θ_1. In tables 1 and 2, P_{AC} and P_{AI} have identical values corresponding to $\theta_1 = 6$. However, this seems to be more input specific.

θ_1	0	1	2	3	4	5	6	7
P_{AC}	1	.6316	.5294	.4706	.4286	.3954	**.3684**	.3453
P_{AI}	0	.0455	.1111	.1818	.2500	.3125	**.3684**	.4179

Table 2.1: $\lambda = 5, \mu_1 = 6, \mu_2 = 5$

The tables clearly shows that as the value of θ_1 increases P_{AC} decreases and P_{AI} increases.

θ_1	0	1	2	3	4	5	6	7
P_{AC}	1	.7200	.6000	.5263	.4737	.4330	**.4000**	.3724
P_{AI}	0	.0667	.1429	.2174	.2857	.3465	**.4000**	.4468

Table 2.2: $\lambda = 4, \mu_1 = 6, \mu_2 = 5$

2.3 Case of three priorities, non-preemptive:

As in the previous models here also we consider a single server infinite capacity queuing system to which customers arrival ($\mathcal{P}_1$) is according to a Poisson process with rate λ and form a queue if server is busy. Service time are exponentially distributed with parameter μ_1. Customers in $\mathcal{P}_1$ queue interrupt own service according to a Poisson process of rate θ_1, in which case he has to go to the lower priority queue($\mathcal{P}_2$). Else, he completes service and leaves the system forever. $\mathcal{P}_2$ customers are taken for service according to head of the line priority whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a non-preemptive service discipline and the service time are independent and identically distributed exponential random variables with parameter μ_2. A customer from $\mathcal{P}_2$ queue may also interrupt his service and if so it is according to a Poisson process of rate θ_2, up on which he has to go to a third waiting line $\mathcal{P}_3$ (of infinite capacity) and wait for his turn of service. The service time of customers in the third queue are independent and identically distributed exponential random variables with parameter μ_3. Their service is also according to non-preemptive service discipline and customers leave the system after completing service without further interruption. When the server is in $\mathcal{P}_3$ line, $\mathcal{P}_2$ line will be empty whereas in $\mathcal{P}_1$ there may be none, one or more customers.

Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system, $N_j(t)$ that of $\mathcal{P}_j$ customers in the queue for $j = 2, 3$; $S(t)$ the status of the server which is 1, 2 or 3 according as the server is busy with a $\mathcal{P}_1$, $\mathcal{P}_2$ or $\mathcal{P}_3$ customer respectively.

Then $\Omega = \{(N_1(t), N_2(t), N_3(t), S(t))\, / t \geq 0\}$ is a CTMC with state space $\{0\} \cup$ $\{(0, n_2, n_3, k)/n_2 \geq 0, n_3 \geq 0, k = 2, 3\} \cup$ $\{(n_1, n_2, n_3, k)/n_1 > 0, n_2 \geq 0, n_3 \geq 0, k = 1, 2, 3\}$. The condition for stability of the system is

$$\frac{\lambda}{(\mu_1 + \theta_1)} + \frac{\lambda\theta_1}{(\mu_1 + \theta_1)(\mu_2 + \theta_2)} + \frac{\lambda\theta_1\theta_2}{(\mu_1 + \theta_1)(\mu_2 + \theta_2)\mu_3} < 1.$$

The infinitesimal generator is obtained as

$$\boldsymbol{Q} = \begin{pmatrix} A_{00}^{(3)} & A_{01}^{(3)} & & & \\ A_{10}^{(3)} & A_1^{(3)} & A_0^{(3)} & & \\ & A_2^{(3)} & A_1^{(3)} & A_0^{(3)} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$A_0^{(3)} = \lambda I_\infty$.

$$A_1^{(3)} = I_\infty \otimes H_3, H_3 = \begin{pmatrix} L_3 & U_3^{(2)} & & & & \\ & L_3 & U_3^{(2)} & & & \\ & & L_3 & U_3^{(2)} & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \end{pmatrix}.$$

$\dim(L_3) = 3, \dim(U_3^{(2)}) = 3.$

$$(L_3)_{ij} = \begin{cases} -(\lambda + \mu_i + \theta_i) & ; i = j = 1, 2. \\ -(\lambda + \mu_3) & ; i = j = 3. \\ \mu_i & ; j = 1, i = 2, 3. \\ 0 & ; \text{otherwise} \end{cases}, (U_3^{(2)})_{ij} = \begin{cases} \theta_2 & ; i = 2, j = 1 \\ 0 & ; \text{otherwise} \end{cases}.$$

$$A_2^{(3)} = \begin{pmatrix} I_\infty \otimes M_3 & I_\infty \otimes N_3 & & & \\ & I_\infty \otimes M_3 & I_\infty \otimes N_3 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}$$

$\dim(M_3) = 3, \dim(N_3) = 3.$

$$(M_3)_{ij} = \begin{cases} \mu_1 & ; i = j = 1 \\ 0 & ; \text{otherwise} \end{cases}, (N_3)_{ij} = \begin{cases} \theta_1 & ; i = j = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$A_{01}^{(3)} = \begin{pmatrix} K_3^{(0)} & & & \\ I_\infty \otimes K_3 & & & \\ & I_\infty \otimes K_3 & & \\ & & I_\infty \otimes K_3 & \\ & & & \ddots \end{pmatrix},$$

$$K_3^{(0)} = \begin{pmatrix} \lambda & 0 & 0 & ... \end{pmatrix}, dim(K_3) = 2 \times 3, (K_3)_{ij} = \begin{cases} \lambda\ ; & j = i + 1, i = 1, 2. \\ 0\ ; & \text{elsewhere.} \end{cases}$$

$$A_{10}^{(3)} = \begin{pmatrix} C_3^* & C_3^{(0)} + I_\infty \otimes C_3^{(1)} & & & \\ & I_\infty \otimes C_3^{(2)} & I_\infty \otimes C_3^{(1)} & & \\ & & I_\infty \otimes C_3^{(2)} & I_\infty \otimes C_3^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \ddots \\ & & & & \end{pmatrix}, C_3^* = \begin{pmatrix} \mu_1 \\ 0 \\ 0 \\ \vdots \\ \end{pmatrix}$$

$$C_3^{(0)} = \begin{pmatrix} 0 & & & \\ B_1 & & & \\ & B_1 & & \\ & & B_1 & \\ & & & \ddots \end{pmatrix}, \ \dim(B_1) = 3 \times 2, (B_1)_{ij} = \begin{cases} \mu_1\ ; & i = 1, j = 2 \\ 0\ \ ; & \text{elsewhere} \end{cases}$$

$\dim(C_3^{(1)}) = \dim(C_3^{(2)}) = 3 \times 2.$

$$(C_3^{(1)})_{ij} = \begin{cases} \theta_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}, (C_3^{(2)})_{ij} = \begin{cases} \mu_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$A_{00}^{(3)} = \begin{pmatrix} -\lambda & 0 & & & \\ M & E_3^{(0)} & & & \\ & E_3^{(2)} & E_3^{(1)} & & \\ & & E_3^{(2)} & E_3^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}, M = \begin{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

$$E_3^{(1)} = I_\infty \otimes D_{31},\ \dim(D_{31}) = 2,\ (D_{31})_{ij} = \begin{cases} -(\lambda+\mu_2+\theta_2)\ ; & i=j=1 \\ -(\lambda+\mu_3)\ ; & i=j=2 \end{cases}$$

$$E_3^{(2)} = E_3^{(21)} + E_3^{(22)}$$

$$E_3^{(21)} = I_\infty \otimes D_3^{(2)}, \dim(D_3^{(2)}) = 2,\ \left(D_3^{(2)}\right)_{ij} = \begin{cases} \mu_{i+1}\,; & j=1, i=1,2 \\ 0 \quad ; & \text{elsewhere} \end{cases}$$

$$E_3^{(22)} = \begin{pmatrix} 0 & g_3^{(2)} & & \\ & & g_3^{(2)} & \\ & & & \ddots \\ & & & \end{pmatrix}, \quad \begin{array}{l} \dim(g_3^{(2)}) = 2 \\ \left(g_3^{(2)}\right)_{ij} = \begin{cases} \theta_2\,; & j=i=1 \\ 0\ ; & \text{elsewhere} \end{cases} \end{array}$$

$$E_3^{(0)} = \begin{pmatrix} J_3^{(1)} & & & \\ J_3^{(11)} & J_3^{(1)} & & \\ & J_3^{(11)} & J_3^{(1)} & \\ & & \ddots & \ddots \end{pmatrix}$$

$\dim(J_3^{(1)}) = \dim(J_3^{(11)}) = 2.$

$$\left(J_3^{(1)}\right)_{ij} = \begin{cases} -(\lambda+\mu_2+\theta_2) & ; \quad i=j=1 \\ -(\lambda+\mu_3) & ; \quad i=j=2 \\ \theta_2 & ; \quad i=1, j=2 \\ 0 & ; \quad \text{elsewhere} \end{cases}, \left(J_3^{(11)}\right)_{ij} = \begin{cases} \mu_{i+1} ; & i=1,2\,;j=2 \\ 0 \quad ; & \text{elsewhere} \end{cases}$$

Theorem 2.3.1. *Let* $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots]$ *denote the invariant probability vector for the QBD process with infinite number of sub levels, where* $\boldsymbol{x}_i$ *is the probability vector of infinite dimension corresponding to level* i. *Then the solution for* $\boldsymbol{x}$ *possesses a matrix geometric structure*

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1}R, \ i > 1.$$

where the rate matrix R *is the minimal non negative solution to*

$$R^2A_2 + RA_1 + A_0 = O. \tag{2.29}$$

Theorem 2.3.2. *The rate matrix* R *in the above theorem possesses a block upper triangular structure given by*

$$R = \begin{pmatrix} R_0 & R_1 & R_2 & \cdots \\ & R_0 & R_1 & \cdots \\ & & R_0 & \cdots \\ & & & \ddots \end{pmatrix}, \text{where } R_j = \begin{pmatrix} R_{j0} & R_{j1} & R_{j2} & \cdots \\ & R_{j0} & R_{j1} & \cdots \\ & & R_{j0} & \cdots \\ & & & \ddots \end{pmatrix}; j \geq 0,$$

in which R_{ji} *are matrices of the form* $R_{ji} = \begin{pmatrix} a_{ji} & 0 & 0 \\ b_{ji} & c_{ji} & 0 \\ d_{ji} & 0 & f_{ji} \end{pmatrix}; i \geq 0,$ *whose entries are as follows.*

Let $x = \lambda + \mu_1 + \theta_1, \ y = \lambda + \mu_2 + \theta_2,$ *and* $z = \lambda + \mu_3$. *Then,*

$$a_{00} = \frac{x - \sqrt{x^2 - 4\mu_1\lambda}}{2\mu_1}, \qquad b_{00} = \frac{2\mu_2\lambda}{xy + y\sqrt{x^2 - 4\mu_1\lambda} - 2\mu_1\lambda}, \qquad c_{00} = \frac{\lambda}{y},$$

$$d_{00} = \frac{2\mu_3\lambda}{xz + z\sqrt{x^2 - 4\mu_1\lambda} - 2\mu_1\lambda}, \qquad f_{00} = \frac{\lambda}{z}, \qquad a_{0k} = 0; \quad k \geq 1,$$

$$b_{01} = \frac{\theta_2 c_{00}}{x - (a_{00} + c_{00})\mu_1}, \qquad b_{0k} = 0; k \geq 2, \qquad c_{0k} = d_{0k} = f_{0k} = 0; \quad k \geq 1,$$

$$a_{l0} = \frac{\theta_1 \sum_{k=0}^{l-1} a_{k0}a_{(l-1-k)0} + \mu_1 \sum_{k=1}^{l-1} a_{k0}a_{(l-k)0}}{x - 2a_{00}\mu_1}; \qquad l \geq 1,$$

$$b_{l0} = \frac{\mu_1 \sum_{i=1}^{l} a_{i0}b_{(l-i)0} + \theta_1 \left[\sum_{j=0}^{l-1} a_{j0}b_{(l-1-j)0} + b_{(l-1)0}c_{00}\right]}{x - (a_{00} + c_{00})\mu_1}; \qquad l \geq 1,$$

$$d_{l0} = \frac{\mu_1 \sum_{i=1}^{l} a_{i0}d_{(l-i)0} + \theta_1 \left[\sum_{j=0}^{l-1} a_{j0}d_{(l-1-j)0} + d_{(l-1)0}f_{00}\right]}{x - (a_{00} + f_{00})\mu_1}; \qquad l \geq 1,$$

$$c_{l0} = f_{l0} = 0; \quad l \geq 1,$$

$$a_{mn} = c_{mn} = d_{mn} = f_{mn} = 0; \quad m, n \geq 1,$$

$$b_{m1} = \frac{\mu_1 \sum_{i=1}^{l} a_{i0}b_{(m-i)1} + \theta_1 \left[\sum_{j=0}^{m-1} a_{j0}b_{(m-1-j)1} + b_{(m-1)1}c_{00}\right]}{x - (a_{00} + c_{00})\mu_1}; \qquad m \geq 1,$$

$$b_{mn} = 0; \quad m \geq 1, n \geq 2 \qquad .$$

Proof: Expansion of equation (2.29) gives the following system of equations:

$$
\begin{aligned}
R_{00}^2 M_3 + R_{00} L_3 + \lambda I_3 &= \mathrm{O} \\
\sum_{j=0}^{m} R_{0j} R_{0,m-j} M_3 + R_{0,m-1} U_3^{(2)} + R_{0m} L_3 &= \mathrm{O} \\
\sum_{k=0}^{l-1} R_{k0} R_{l-1-k,0} N_3 + \sum_{k=0}^{l} R_{k0} R_{l-k,0} M_3 + R_{l0} M_1 &= \mathrm{O} \\
\sum_{k=0}^{l-1} \sum_{j=0}^{m} R_{kj} R_{l-1-k,m-j} N_3 + \sum_{k=0}^{l} \sum_{j=0}^{m} R_{kj} R_{l-k,m-j} M_3 + & \\
R_{l,m-1} U_3^{(2)} + R_{lm} L_3 &= \mathrm{O}.
\end{aligned}
$$

Solving this system of equations the required result is obtained.

2.3.1 Joint and Marginal Probabilities

Let $\boldsymbol{x}_{ijk}$ be the probability of i high priority customers in the system, j customers waiting in the $\mathcal{P}_2$ queue and k customers waiting in the $\mathcal{P}_3$ queue.

Then the marginal probability of i number of $\mathcal{P}_1$ customers is

$$\boldsymbol{x}_{i\bullet\bullet} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \boldsymbol{x}_{ijk}.$$

We have, from theorem (4.1) $\boldsymbol{x}_i = \boldsymbol{x}_{i-1} R$ and proceeding as in the section 3,

$$\boldsymbol{x}_{ijk} = \sum_{l=0}^{j} \sum_{m=0}^{k} \boldsymbol{x}_{i-1,lm} R_{j-l,k-m} \text{ for} j,k \geq 0.$$

To know the type of customer under service, we expand the above equation to get the recursive formulas,

$$
\begin{aligned}
&\boldsymbol{x}_{ijk}(1) = \sum_{l=0}^{j} a_{j-l,0} \boldsymbol{x}_{i-1,lk}(1) + \sum_{l=0}^{j} \sum_{m=0}^{l} b_{j-l,l-m} \boldsymbol{x}_{i-1,lm}(2) + \sum_{l=0}^{j} d_{j-l,0} \boldsymbol{x}_{i-1,lk}(3), \\
&\boldsymbol{x}_{ijk}(2) = c_{00} \boldsymbol{x}_{i-1,jk}(2), \\
&\boldsymbol{x}_{ijk}(3) = f_{00} \boldsymbol{x}_{i-1,jk}(3); \qquad j,k \geq 0, i \geq 1.
\end{aligned}
$$

2.3.2 High Priority Marginal Distribution

Marginal distribution of high priority customers in the system is

$$\begin{aligned}\boldsymbol{x}_{i\bullet\bullet} &= \textstyle\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\boldsymbol{x}_{ijk}\\ &= \boldsymbol{x}_{i-1,\bullet\bullet}R_{+}\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{R}_{+} &= \textstyle\sum_{j=0}^{\infty}R_j\\ &= \textstyle\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}R_{jk}.\end{aligned}$$

Expanding the above equation in lowest phases gives,

$$\boldsymbol{x}_{i\bullet\bullet} = [\boldsymbol{x}_{i\bullet\bullet}(1),\ \boldsymbol{x}_{i\bullet\bullet}(2),\ \boldsymbol{x}_{i\bullet\bullet}(3)] = [(1-\rho),\ \boldsymbol{x}_0(2),\ \boldsymbol{x}_0(3)]\,\boldsymbol{R}_{+}^{i}, \quad i \geq 1.$$

2.4 Case of $N+1$ priorities, non-preemptive:

Model Description: Now we extend the number of priorities to $N+1$. Thus a customer can interrupt at most N times, exactly once while in a particular priority, except the last. Consider a single server infinite capacity queuing system in which customers from outside arrive according to a Poisson process with rate λ and form a queue if server is busy. Service time are exponentially distributed with parameter μ_1. Customers in primary queue interrupt service according to a Poisson process of rate θ_1, in which case he has to go to a lower priority queue. Else, he completes service and leaves the system forever. Lower priority customers are taken for service according to head of the line priority whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a non-preemptive service discipline. A customer from this low priority queue may interrupt his service according to a Poisson process of rate θ_2 up on which he has to go to a third waiting line (of infinite capacity) and wait for his turn for service. The service time of customers in the i_{th} queue are independent and identically distributed exponential random variables with parameter μ_i. Customers in the i^{th} priority queue also interrupt their service according to a Poisson process with rate θ_i or else completes service with service time exponentially distributed with parameter μ_i. A maximum of N service interruptions is allowed for any customer so that $i = 2, 3, \ldots, N$. Thus there are $N + 1$ queues, the first one constituted solely by external (primary) customers and the remaining queues are generated by customers from the just preceding higher priority queue. Thus N dependent queues and one independent stream of customers served by a single server, form our system. At the service completion epoch of a low priority customer, the server checks whether there is any higher priority customer in the system. If there is one in the highest priority, he takes the head in that queue; else takes the one, if any, from the second queue and so on. From the $(N + 1)^{th}$ queue, a customer in service leaves on completion of service (following an exponential distribution with parameter μ_{N+1}) or interrupts his service according to a Poisson process of rate θ_{N+1}. In the latter case the customer leaves the system paying a heavy penalty.

The infinitesimal generator is

$$\widehat{Q} = \begin{pmatrix} A_{00}^{(n)} & A_{01}^{(n)} & & & \\ A_{10}^{(n)} & A_{1}^{(n)} & A_{0}^{(n)} & & \\ & A_{2}^{(n)} & A_{1}^{(n)} & A_{0}^{(n)} & \\ & & \ddots & \ddots & \end{pmatrix}$$

where,

$A_0^{(n)} = \lambda I_\infty, A_1^{(n)} = I_\infty \otimes H_n$

$$H_n = \begin{pmatrix} L_n & U_n^{(n-1)} & 0 & \ldots & U_n^{(n-2)} & 0 & \ldots & U_n^{(3)} & 0 & \ldots & U_n^{(2)} & 0 & \ldots \\ 0 & L_n & U_n^{(n-1)} & 0 & \ldots & U_n^{(n-2)} & 0 & \ldots & U_n^{(3)} & 0 & \ldots & U_n^{(2)} & 0 \\ & & & \ddots & \ddots & & & & \ddots & & & & \\ & & & & \ddots & \ddots & & & & \ddots & & & \end{pmatrix}$$

$\dim(Ln) = n,\ n \geq 3,\ \dim(Un^{(k)}) = n,\ k = 2, 3, ..., n-1$

$$(L_n)_{ij} = \begin{cases} -(\lambda+\theta_i+\mu_i) & j=i=1,2,...,n-1. \\ -(\lambda+\mu_n) & j=i=n \\ \mu_i & j=1, i=2,3,...,n \\ 0 & \text{otherwise} \end{cases}, (Un^{(k)})_{ij} = \begin{cases} \theta_k & j=1, i=k \\ 0 & \text{otherwise} \end{cases}$$

$$A_2^{(n)} = \begin{pmatrix} I_\infty \otimes M_n & I_\infty \otimes N_n & & \\ & I_\infty \otimes M_n & I_\infty \otimes N_n & \\ & & \ddots & \ddots \\ & & & \ddots \\ & & & \end{pmatrix},$$

$$\dim(Mn) = \dim(Nn) = n$$

$$(M_n)\,ij = \begin{cases} \mu_1 & ; j=i=1 \\ 0 & ;\text{otherwise} \end{cases}$$

$$(N_n)\,ij = \begin{cases} \theta_1 & ; j=i=1 \\ 0 & ;\text{otherwise} \end{cases}$$

$$A_{10}^{(n)} = \begin{pmatrix} C_n^* & C_n^{(0)} + I_\infty \otimes C_n^{(1)} & & & \\ & I_\infty \otimes C_n^{(2)} & I_\infty \otimes C_n^{(1)} & & \\ & & I_\infty \otimes C_n^{(2)} & I_\infty \otimes C_n^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \ddots \\ & & & & \end{pmatrix}$$

$$C_n^* = \begin{pmatrix} \mu_1 \\ 0 \\ 0 \\ \vdots \\ \end{pmatrix},\ C_n^{(0)} = \begin{pmatrix} \mathbf{0} \\ I_\infty \otimes B_1 \\ I_\infty \otimes B_2 \\ \cdots \\ \cdots \\ I_\infty \otimes B_{n-2} \end{pmatrix},\quad \dim(B_k) = n \times (n-1), 1 \le k \le (n-k), \quad (B_k)_{ij} = \begin{cases} \mu_1\ ; & i=1, j=n-k \\ 0\ ; & \text{elsewhere} \end{cases}$$

with $\mathbf{0} = [\ 0 \quad 0 \quad \cdots \quad\ ]$

$\dim(C_n^{(1)}) = \dim(C_n^{(2)}) = n \times (n-1)$

$$(C_n^{(1)})_{ij} = \begin{cases} \theta_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases} \qquad (C_n^{(2)})_{ij} = \begin{cases} \mu_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$A_{01}^{(n)} = \begin{pmatrix} K_n^{(0)} & & & & \\ I_\infty \otimes K_n & & & & \\ & I_\infty \otimes K_n & & & \\ & & I_\infty \otimes K_n & & \\ & & & \ddots & \ddots \end{pmatrix},$$

$$K_n^{(0)} = [\ \lambda \quad 0 \quad 0 \quad ... \ ], dim(K_n) = (n-1)\times n,$$

$$(K_n)_{ij} = \begin{cases} \lambda\,; & j=i+1, i=1,2,...,(n-1). \\ 0\,; & elsewhere. \end{cases}$$

$$A_{00}^{(n)} = \begin{pmatrix} -\lambda & & & & \\ M & E_n^{(0)} & & & \\ & E_n^{(2)} & E_n^{(1)} & & \\ & & E_n^{(2)} & E_n^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}, M = \begin{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{pmatrix} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

$$E_n^{(1)} = I_\infty \otimes D_{n1}, \dim(D_{n1}) = n-1, (D_{n1})_{ij} = \begin{cases} -(\lambda+\mu_{i+1}+\theta_{i+1}) & 1\le i,j \le n-2 \\ -(\lambda+\mu_{i+1}) & i=j=n-1 \end{cases}$$

$$E_n^{(2)} = E_n^{(21)} + E_n^{(22)}$$

$$E_n^{(21)} = I_\infty \otimes D_n^{(n-1)}, \dim(D_n^{(n-1)}) = n-1,\ \left(D_n^{(n-1)}\right)_{ij} = \begin{cases} \mu_{i+1} & j=1, 1\le i\le n-1 \\ 0 & \text{elsewhere} \end{cases}$$

$$
E_n^{(22)} = \begin{pmatrix}
0 & G_n^{(n-1)} & ... & & G_n^{(n-2)} & ... & & G_n^{(2)} & ... & \\
 & & G_n^{(n-1)} & ... & & G_n^{(n-2)} & ... & & G_n^{(2)} & ... \\
 & & & \ddots & & & \ddots & & & \ddots \\
 & & & & & & & & & \\
 & & & & & & & & &
\end{pmatrix}
$$

$$
\dim(G_n^{(k)}) = (n-1); (G_n^{(k)})_{ij} = \begin{cases} \theta_k \ ; & i = k-1, j = 1 \\ 0 \ ; & elsewhere, k = 2, 3, ..., n-1 \end{cases}
$$

$$
E_n^{(0)} = \begin{pmatrix}
J_n^{(1)} & & & & & & & & & & \\
J_n^{(11)} & J_n^{(1)} & & & & & & & & & \\
 & J_n^{(11)} & J_n^{(1)} & & & & & & & & \\
\vdots & & \ddots & \ddots & & & & & & & \\
J_n^{(21)} & J_n^{(22)} & \cdots & & J_n^{(2)} & & & & & & \\
 & J_n^{(21)} & J_n^{(22)} & \cdots & & J_n^{(2)} & & & & & \\
\vdots & & \ddots & \ddots & & & \ddots & & & & \\
J_n^{(31)} & J_n^{(32)} & \cdots & & J_n^{(33)} & \cdots & & J_n^{(3)} & & & \\
 & & \ddots & & & \ddots & & & \ddots & & \\
J_n^{((n-2)1)} & J_n^{((n-2)2)} & \cdots & & J_n^{((n-2)3)} & \cdots & & & & J_n^{((n-2))} & \\
\vdots & & \ddots & \ddots & & & & & & & \ddots
\end{pmatrix}
$$

$$
\dim(J_n^{(k)}) = \dim(J_n^{(km)}) = n-1 \ ; \ \begin{cases} k = 1, 2, ..., , (n-2) \\ m = 1, 2, ..., , k \end{cases}
$$

$$
\left(J_n^{(k)}\right)_{ij} = \begin{cases} -(\lambda + \mu_{i+1} + \theta_{i+1}) \ ; & i = j = 1, 2, ..., (n-2) \\ -(\lambda + \mu_{i+1}) \ ; & i = j = (n-1) \\ \theta_{i+1} \ ; & j = i+1, i = 1, 2, ..., (n-k+1) \\ 0 \ ; & elsewhere \end{cases}
$$

$$
\left(J_n^{(k1)}\right)_{ij} = \begin{cases} \mu_{i+1} \ ; & i = 1, 2, ..., (n-1) \ ; j = n-k \\ 0 \ ; & elsewhere \end{cases}
$$

$$\left(J_n^{(km)}\right)_{ij} = \begin{cases} \theta_{n-m+1} \; ; & i = n-m, \; ; j = n-k \\ 0 \quad ; & elsewhere, \end{cases} \qquad : m = 2, 3, ..., k,$$

The stability of the system is given to be

$$\frac{\lambda}{\mu_1 + \theta_1}\left[1 + \sum_{i=1}^{N-1}\prod_{j=1}^{i}\frac{\theta_j}{\mu_{j+1} + \theta_{j+1}} + \frac{\theta_N}{\mu_{N+1}}\prod_{j=1}^{N-1}\frac{\theta_j}{\mu_{j+1} + \theta_{j+1}}\right] < 1.$$

The performance measures are not computed for this case. One can proceed on the same lines as indicated in cases $N = 1$ and 2 respectively.

Chapter 3

Queues with Priority and Feedback

The feedback queues discussed in literature fall in either of the following two categories: Upon completion of service, the customer may decide to get his service repeated with a positive probability and joins either the tail end of the queue or occupies the server immediately after the first service is completed. In either case there is no separate queue for feedback customers. Also identifying a feedback customer in the system is not an easy task. Further, how many feedback, if any, customer has taken is not considered.

In this chapter we consider a priority queueing system with feedback distinct from what is being discussed in literature. Low priority ($\mathcal{P}_2$) as well as high priority ($\mathcal{P}_1$) customers arriving according to two independent Poisson processes, queue up separately for service at a busy service station and are served on priority basis. We assume infinite waiting space for both priorities. If a $\mathcal{P}_1$ customer decides to feedback immediately after completing service, then he has to join the

Some results of this chapter are included in the following paper.
A. Krishnamoorthy, Manjunath A. S.: **On queues with priority determined by feedback** (communicated).

$\mathcal{P}_2$ line as the last in that queue. Customers from $\mathcal{P}_2$ line are served only if none is left in the $\mathcal{P}_1$ queue. The case of exactly one feedback is analyzed in detail. That is no further feedback beyond $\mathcal{P}_2$. Both preemptive and non-preemptive cases are studied. Thus the low priority queue has both feedback customers and fresh customers.The case of more than one feedback can be analyzed in almost the same way as the one with no more than one feedback, though the complexities grow manifold. In this model customers complete their service and feedback for another service whereas in the previous chapter customers interrupt their service while service is in progress. Such customers form the immediately next lower priority queue and the two models differ.

Section 1 analyzes feed back queue with non-preemptive priority. Here service times are phase type distributed. Algorithms for computing the joint and the marginal probabilities are presented in this section. Waiting time distributions are also discussed.In section 2 we discuss the problem with exponential service time. Feed back queue with preemptive priorities and exponential service time are presented in section 3.

3.1 M/PH/1 Feedback queue with non-preemptive priority

We consider a single server queueing system with two distinct queues to which customers of two different priorities arrive according to Poisson processes of rates $\lambda_i, i = 1, 2$. Service time of both type of customers are phase type distributed with representation (α, T) and (β, S) of orders m and n, respectively. High priority ($\mathcal{P}_1$) customers who are not satisfied with the service already provided to them, join(feedback) the low priority ($\mathcal{P}_2$) line immediately after completing service. The probability of a $\mathcal{P}_1$ customer to feedback is θ on completion of his service. Low priority ($\mathcal{P}_2$) customers are taken for service one at a time from the head of the line whenever the $\mathcal{P}_1$ queue is found to be empty at a service completion epoch. The service of $\mathcal{P}_2$ customers is according to a non-preemptive service discipline; that is the arrival of a $\mathcal{P}_1$ customer doesn't interrupt the ongoing service of a $\mathcal{P}_2$ customer. No feedback is permitted to customers in $\mathcal{P}_2$ line. This means that there are only two types of customers $\mathcal{P}_1$ and $\mathcal{P}_2$.

Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system including the one in service if any, $N_2(t)$ be the number in the $\mathcal{P}_2$ line and $S(t)$ the status of the server which is 1 or 2 according as the server is busy with $\mathcal{P}_1$ customers or $\mathcal{P}_2$ customers and $M(t)$ the phase of service. Thus we get a continuous time Markov chain $\Omega = \{X(t), t \geq 0\} = \{(N_1(t), N_2(t), S(t), M(t)) / t \geq 0\}$ with state space $\{\mathbf{0}\} \cup \{(0, n_2, 2, j) / \ n_2 \geq 0, 1 \leq j \leq n\} \cup \{(n_1, n_2, k, j) / \ n_1 \geq 1, n_2 \geq 0, 1 \leq j \leq m, k = 1, 2\}$. The state $\mathbf{0}$ represents an idle server.

The infinitesimal generator of this continuous time Markov chain consists of block entries of infinite dimension and is obtained as

$$\widetilde{Q} = \begin{pmatrix} A_{00} & A_{01} & & & \\ A_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Let $\lambda = \lambda_1 + \lambda_2$, then we have $A_0 = \lambda_1 I_\infty$,

$$A_1 = \begin{pmatrix} M_1 & M_2 & & \\ & M_1 & M_2 & \\ & & \ddots & \ddots \\ & & & \end{pmatrix}; \qquad M_1 = \begin{bmatrix} T - \lambda I_m & 0 \\ S^0\alpha & S - \lambda I_n \end{bmatrix}$$

$$M_2 = \lambda_2 I_{m+n}$$

$$A_2 = \begin{pmatrix} M_3 & M_4 & & & \\ & M_3 & M_4 & & \\ & & \ddots & \ddots & \\ & & & & \\ & & & & \end{pmatrix}; \quad \begin{array}{l} M_3 = \begin{bmatrix} (1-\theta)T^0\alpha & 0 \\ 0 & 0 \end{bmatrix} \\ \\ M_4 = \begin{bmatrix} \theta T^0\alpha & 0 \\ 0 & 0 \end{bmatrix} \end{array}$$

$$A_{01} = \begin{pmatrix} M_5^* & & \\ M_5 & & \\ & M_5 & \\ & & \ddots \\ & & \end{pmatrix}; \quad \begin{array}{l} M_5^* = \begin{bmatrix} \lambda_1\alpha & 0 \end{bmatrix} \\ \\ M_5 = \begin{bmatrix} 0 & \lambda_1 I_n \end{bmatrix} \end{array}$$

$$A_{10} = \begin{pmatrix} M_6^* & M_7 & & & \\ & M_6 & M_7 & & \\ & & M_6 & M_7 & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}; \quad \begin{array}{l} M_6^* = \begin{bmatrix} (1-\theta)T^0 \\ 0 \end{bmatrix} \\ M_6 = \begin{bmatrix} (1-\theta)T^0\beta \\ 0 \end{bmatrix} \\ M_7 = \begin{bmatrix} \theta T^0\beta \\ 0 \end{bmatrix} \end{array}$$

$$A_{00} = \begin{pmatrix} -\lambda & \lambda_2\beta & & & \\ S^0 & M_8 & M_9 & & \\ & M_{10} & M_8 & M_9 & \\ & & M_{10} & M_8 & M_9 \\ & & & \ddots & \ddots \end{pmatrix}; \quad \begin{array}{ll} M_8 & = S - \lambda I_n \\ M_9 & = \lambda_2 I_n \\ M_{10} & = S^0\beta \end{array}$$

System stability

If μ_1 and μ_2 denote the mean service rate of $\mathcal{P}_1$ and $\mathcal{P}_2$ customers respectively, then $\frac{1}{\mu_1} = -\alpha T^{-1}\mathbf{e}$ and $\frac{1}{\mu_2} = -\beta S^{-1}\mathbf{e}$. The fraction of time the server is busy with $\mathcal{P}_i$ customers is $\tilde{\rho}_i = \frac{\lambda_i}{\mu_i}$ for $i = 1, 2$. We assume that $\tilde{\rho}_1 + \tilde{\rho}_2 < 1$, under this condition the system is stable.

The infinitesimal generator $\widetilde{Q}$ constitutes a QBD process with infinite number of sub-levels. As $\widetilde{Q}$ is irreducible and recurrent, following an argument similar in theorem 3 of Miller [49] we have,

Theorem 3.1.1. *Let $\tilde{\boldsymbol{x}} = [\tilde{\boldsymbol{x}}_0, \tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \ldots]$ denote the invariant probability vector for the QBD process $\widetilde{Q}$ with infinite number of sub levels(phases), where $\tilde{\boldsymbol{x}}_i$ is the probability vector corresponding to level i of infinite dimension. Then the solution for $\tilde{\boldsymbol{x}}$ possesses a matrix geometric structure*

$$\tilde{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_{i-1}\boldsymbol{R},\ i > 1. \tag{3.1}$$

where the rate matrix $\boldsymbol{R}$ is the minimal non negative solution to

$$\boldsymbol{R}^2 A_2 + \boldsymbol{R} A_1 + A_0 = O. \tag{3.2}$$

We now compute the elements of the infinite dimensional matrix $\boldsymbol{R}$.

Theorem 3.1.2. *The $\boldsymbol{R}$ matrix, which is the minimal non negative solution to equation (3.2) possesses a Toeplitz structure $(R_0, R_1, R_2, ...)$. That is, $\boldsymbol{R}$ has the form*

$$\boldsymbol{R} = \begin{pmatrix} R_0 & R_1 & R_2 & R_3 & \cdots \\ 0 & R_0 & R_1 & R_2 & \cdots \\ 0 & 0 & R_0 & R_1 & \cdots \\ 0 & 0 & 0 & R_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where each of the matrices R_k is of the form $R_k = \begin{pmatrix} R_{kA} & 0 \\ R_{kB} & R_{kC} \end{pmatrix}$ where R_{kA} and R_{kC} are matrices of order m and n respectively and R_{kB} is a matrix of order $m \times n$.

Proof. Follows from the interpretation of $\boldsymbol{R}$ in Neuts [52] and the structure of the matrices in the generator matrix $\widetilde{Q}$. □

Theorem 3.1.3. *The entries of the matrix $R_k (k > 0)$ in theorem 3.1.2 are computed from the following relations.*

$$(1-\theta)R_{0A}^2 T^0\alpha + R_{0A}(T-\lambda I_m) + \lambda_1 I_m = O \tag{3.3}$$

$$(1-\theta)(R_{0B}R_{0A} + R_{0C}R_{0B})T^0\alpha + R_{0B}(T-\lambda I_m) + R_{0C}S^0\alpha = O \tag{3.4}$$

$$R_{0C}(S-\lambda I_n) + \lambda_1 I_n = O \tag{3.5}$$

$$\theta\left(\sum_{i=0}^{k-1} R_{iA}R_{(k-1-i)A}\right)T^0\alpha + (1-\theta)\left(\sum_{i=0}^{k} R_{iA}R_{(k-i)A}\right)T^0\alpha + \lambda_2 R_{(k-1)A} + R_{kA}(T-\lambda I_m) = O \tag{3.6}$$

$$\left(\sum_{i=0}^{k-1}\left(R_{iB}R_{(k-1-i)A} + R_{iC}R_{(k-1-i)B}\right)\right)\theta T^0\alpha + \left(\sum_{i=0}^{k} R_{iB}R_{(k-i)A} + R_{iC}R_{(k-i)B}\right)(1-\theta)T^0\alpha + \lambda_2 R_{(k-1)B} + R_{kB}(T-\lambda I_m) + R_{kC}S^0\alpha = O \tag{3.7}$$

$$\lambda_2 R_{(k-1)C} + R_{kC}(S-\lambda I_n) = O \tag{3.8}$$

Proof. We obtain the following relations after expanding (3.2).

$$R_0^2 M_3 + R_0 M_1 + \lambda_1 I_{m+n} = \mathrm{O},$$

$$\left(\sum_{i=0}^{k-1} R_i R_{k-1-i}\right) M_4 + \left(\sum_{i=0}^{k} R_i R_{k-i}\right) M_3 + R_{k-1}M_2 + R_k M_1 = \mathrm{O}, k \geq 1.$$

The result is established when these relations are expanded with respect to the phases (the phases of the system are the server status(idle/ busy with $\mathcal{P}_1$ customers/ busy with $\mathcal{P}_2$ customers) and the number of $\mathcal{P}_2$ customers in the queue; the level of the system is the number of $\mathcal{P}_1$ customers in the system). □

3.1.1 The joint and marginal probabilities

The recursive formulas for joint distribution of the number of $\mathcal{P}_1$ customers in the system and $\mathcal{P}_2$ customers in the queue and marginal distributions of each are derived below. First we establish the following.

Theorem 3.1.4. *The matrix geometric structure*

$$\tilde{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_{i-1}\boldsymbol{R}, \quad i > 1$$

given in Theorem 3.1.1 extended to level 0 is

$$\tilde{\boldsymbol{x}}_i = \tilde{\boldsymbol{x}}_0 \left(\frac{1}{\lambda_1} A_{01} \right) \boldsymbol{R}^i, \quad i \geq 1. \tag{3.9}$$

Proof. From $\tilde{\boldsymbol{x}}\widetilde{Q} = 0$, the two boundary equations involving $\tilde{\boldsymbol{x}}_0$ are

$$\tilde{\boldsymbol{x}}_0 A_{00} + \tilde{\boldsymbol{x}}_1 A_{10} = \boldsymbol{0}, \tag{3.10}$$

$$\tilde{\boldsymbol{x}}_0 A_{01} + \tilde{\boldsymbol{x}}_1 [A_1 + \boldsymbol{R} A_2] = \boldsymbol{0}. \tag{3.11}$$

From (3.2) it follows that

$$\boldsymbol{R}[\boldsymbol{R}A_2 + A_1] = -A_0. \tag{3.12}$$

Since $A_0 = \lambda_1 I_\infty$, $\boldsymbol{R}$ is invertible.
From (3.11) and (3.12) we get

$$\tilde{\boldsymbol{x}}_1 = \tilde{\boldsymbol{x}}_0 \left(\frac{1}{\lambda_1} A_{01} \boldsymbol{R} \right). \tag{3.13}$$

Combining relations (3.1) and (3.13) we obtain (3.9). □

The joint probability distribution

Let $\tilde{\boldsymbol{x}}_{ij}$ be the probability that there are i $\mathcal{P}_1$ customers in the system and j $\mathcal{P}_2$ customers waiting in the queue. Further let the marginal distribution of the number of $\mathcal{P}_1$ customers in the system be denoted by

$$\tilde{\boldsymbol{x}}_{i\bullet} = \sum_{j=0}^{\infty} \tilde{\boldsymbol{x}}_{ij} \ , \quad i \geq 0.$$

To identify the type of customer in service, partition $\tilde{\boldsymbol{x}}_{ij}$ as

$$\tilde{\boldsymbol{x}}_{ij} = (\tilde{\boldsymbol{x}}_{ij}(1), \tilde{\boldsymbol{x}}_{ij}(2)), \qquad i \geq 1, j \geq 0,$$

where $\tilde{\boldsymbol{x}}_{ij}(k)$ represents a $\mathcal{P}_k$ customer in service for $k = 1, 2$. We proceed to determine the joint probability vectors $\tilde{\boldsymbol{x}}_{ij}$.

Considering the $\mathcal{P}_2$ line, equation (3.1) gives

$$\tilde{\boldsymbol{x}}_{ij} = \tilde{\boldsymbol{x}}_{i-1,j} \ \mathbf{R} \ , \quad i > 1, j \geq 0.$$

Expanding this w.r.t. j

$$\begin{pmatrix} \tilde{\boldsymbol{x}}_{i0}, & \tilde{\boldsymbol{x}}_{i1}, & \cdots \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{x}}_{i-1,0}, & \tilde{\boldsymbol{x}}_{i-1,1}, & \cdots \end{pmatrix} \times \begin{pmatrix} R_0 & R_1 & R_2 & \cdots \\ 0 & R_0 & R_1 & \cdots \\ 0 & 0 & R_0 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

That is, for $j \geq 0$,

$$\tilde{\boldsymbol{x}}_{ij} = \sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{i-1,k} \ R_{j-k} \ , \quad i > 1. \tag{3.14}$$

On expanding this, incorporating the type of service, we obtain

$$\tilde{\boldsymbol{x}}_{ij}(1) = \sum_{k=0}^{j} \left[\tilde{\boldsymbol{x}}_{i-1,k}(1) \ R_{(j-k)A} + \tilde{\boldsymbol{x}}_{i-1,k}(2) \ R_{(j-k)B}\right] \tag{3.15}$$

$$\tilde{\boldsymbol{x}}_{ij}(2) = \sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{i-1,k}(2)\ R_{(j-k)C}; \qquad i > 1, j \geq 0. \tag{3.16}$$

Denoting the probability of an idle server by $1-\tilde{\rho}$, the relation (3.13) on expansion gives

$$\tilde{\boldsymbol{x}}_{1j}(1) = (1-\tilde{\rho})\ \alpha R_{jA} + \sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{0k}(2)\ R_{(j-k)B} \tag{3.17}$$

$$\tilde{\boldsymbol{x}}_{1j}(2) = \sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{0k}(2)\ R_{(j-k)C} \qquad ; j \geq 0. \tag{3.18}$$

We observe that the joint probabilities depend on $\tilde{\boldsymbol{x}}_{0k}(2)$ for $k = 0, 1, 2, \ldots j$ and are computed in the desired range in the next section.

Marginal distribution of $\mathcal{P}_1$ customers

Adding equation(3.14) over j, the marginal probability $\tilde{\boldsymbol{x}}_{i\bullet}$, when there are i $\mathcal{P}_1$ customers in the system is given by

$$\begin{aligned} \tilde{\boldsymbol{x}}_{i\bullet} &= \sum_{j=0}^{\infty}\sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{i-1,k}\ R_{j-k} = \sum_{k=0}^{\infty} \tilde{\boldsymbol{x}}_{i-1,k} \left(\sum_{j=0}^{\infty} R_j \right) \\ &= \tilde{\boldsymbol{x}}_{(i-1)\bullet}\ \mathcal{R} \qquad (3.19) \\ &= \tilde{\boldsymbol{x}}_{1\bullet}\ \mathcal{R}^{i-1}, \qquad i > 1; \qquad (3.20) \end{aligned}$$

where

$$\mathcal{R} = \sum_{k=0}^{\infty} R_k = \begin{bmatrix} \sum_{k=0}^{\infty} R_{kA} & 0 \\ \sum_{k=0}^{\infty} R_{kB} & \sum_{k=0}^{\infty} R_{kC} \end{bmatrix} = \begin{bmatrix} R_A & 0 \\ R_B & R_C \end{bmatrix}$$

Now, expanding (3.19) on the type of customer in service, we have

$$\begin{aligned} \left(\tilde{\boldsymbol{x}}_{i\bullet}(1), \ \ \tilde{\boldsymbol{x}}_{i\bullet}(2) \right) &= \left(\tilde{\boldsymbol{x}}_{(i-1)\bullet}(1), \ \ \tilde{\boldsymbol{x}}_{(i-1)\bullet}(2) \right) \begin{bmatrix} R_A & 0 \\ R_B & R_C \end{bmatrix} \\ &= \left(\tilde{\boldsymbol{x}}_{(i-1)\bullet}(1) R_A + \tilde{\boldsymbol{x}}_{(i-1)\bullet}(2) R_B, \ \ \tilde{\boldsymbol{x}}_{(i-1)\bullet}(2) R_C \right) \end{aligned}$$

which implies for $i > 1$

$$\begin{aligned}\tilde{\boldsymbol{x}}_{i\bullet}(1) &= \tilde{\boldsymbol{x}}_{(i-1)\bullet}(1)\ R_A + \tilde{\boldsymbol{x}}_{(i-1)\bullet}(2)\ R_B \\ \tilde{\boldsymbol{x}}_{i\bullet}(2) &= \tilde{\boldsymbol{x}}_{(i-1)\bullet}(2)\ R_C\end{aligned}$$

Also by adding equations (3.17) and (3.18) over j, we obtain

$$\begin{aligned}\tilde{\boldsymbol{x}}_{1\bullet}(1) &= (1-\tilde{\rho})\ \alpha\ R_A + \tilde{\boldsymbol{x}}_{0\bullet}(2)\ R_B \\ \tilde{\boldsymbol{x}}_{1\bullet}(2) &= \tilde{\boldsymbol{x}}_{0\bullet}(2)\ R_C\end{aligned}$$

This in turn gives

$$\tilde{\boldsymbol{x}}_{1\bullet} = \Big(\ (1-\tilde{\rho})\alpha, \quad \tilde{\boldsymbol{x}}_{0\bullet}(2)\ \Big)\mathcal{R} \tag{3.21}$$

From relations (3.20) and (3.21) we get

$$\tilde{\boldsymbol{x}}_{i\bullet} = \Big(\ (1-\tilde{\rho})\alpha, \quad \tilde{\boldsymbol{x}}_{0\bullet}(2)\ \Big)\mathcal{R}^i\ ; \qquad i \geq 1.$$

Writing $\tilde{\boldsymbol{x}}_0 = \Big(\ (1-\tilde{\rho})\alpha, \quad \tilde{\boldsymbol{x}}_{0\bullet}(2)\ \Big)$,

$$\tilde{\boldsymbol{x}}_{i\bullet} = \tilde{\boldsymbol{x}}_0\mathcal{R}^i\ ; \qquad i \geq 1. \tag{3.22}$$

On expanding the relation (3.22), the marginal probabilities of $\mathcal{P}_1$ are given as

$$\tilde{\boldsymbol{x}}_{i\bullet}(1) = (1-\tilde{\rho})\,\alpha R_A^i + \tilde{\boldsymbol{x}}_{0\bullet}(2)\sum_{k=0}^{i-1} R_C^{i-1-k} R_B R_A^k \tag{3.23}$$

$$\tilde{\boldsymbol{x}}_{i\bullet}(2) = \tilde{\boldsymbol{x}}_{0\bullet}(2) R_C^i. \tag{3.24}$$

Clearly the marginal probabilities depend on $\tilde{\boldsymbol{x}}_{0\bullet}(2) = \sum_{j=0}^{\infty} \tilde{\boldsymbol{x}}_{0j}(2)$, the probability that a $\mathcal{P}_2$ customer is in service with no $\mathcal{P}_1$ customer in the system.

Theorem 3.1.5. *R_A, R_B and R_C are explicitly given by*

$$R_A = \lambda_1 \left(\lambda_1 I_m - \lambda_1 \boldsymbol{e}\alpha - T\right)^{-1}. \tag{3.25}$$

$$R_B = \boldsymbol{e}\alpha R_A. \tag{3.26}$$

$$R_C = \lambda_1 \left(\lambda_1 I_n - S\right)^{-1}. \tag{3.27}$$

Proof. Adding relation(3.6) for all values of k from 1 to ∞ and this added to relation (3.3) yields

$$R_A^2 T^0 \alpha + R_A (T - \lambda_1 I_m) + \lambda_1 I_m = O \tag{3.28}$$

Performing the same operation on relation(3.7) and (3.4) and then on relation(3.8) and (3.5) yield

$$(R_B R_A + R_C R_B) T^0 \alpha + R_B (T - \lambda_1 I_m) + R_C S^0 \alpha = O \tag{3.29}$$

$$\lambda_2 R_C + R_C (S - \lambda I_n) + \lambda_1 I_n = O \tag{3.30}$$

One obtains (3.27) directly from the relation (3.30). As customers in $\mathcal{P}_2$ line play no role on those waiting in $\mathcal{P}_1$ line, R_A is same as the rate matrix R of the $M/PH/1$ queue in Neuts[52][p.84]. So using theorem 3.2.1 of Neuts we get (3.25).

Now R_B is obtained as follows. Multiply equation (3.28) by inverse of R_A and rearranging one gets the relation

$$R_A T^0 \alpha = -(T - \lambda_1 I_m) - \lambda_1 R_A^{-1} \tag{3.31}$$

Pre-multiplying equation (3.28) by inverse of the matrix R_C and substituting for $R_A T^0 \alpha$ yields

$$R_B T^0 \alpha + S^0 \alpha - R_C^{-1} R_B \left(\lambda_1 R_A^{-1}\right) = 0$$

Substituting for $\lambda_1 R_A^{-1}$ from relation (3.25) and for R_C^{-1} from (3.27) we get

$$R_B T^0 \alpha + S^0 \alpha + \lambda_1^{-1} (\lambda_1 I_n - S) R_B (-\lambda_1 I_m + \lambda_1 \mathbf{e}\alpha + T) = 0 \tag{3.32}$$

Multiplying the above relation throughout by $\mathbf{e}$ on right side yields

$$R_B T^0 = \lambda_1 \mathbf{e} \tag{3.33}$$

Now proper substitution of (3.25) in relation (3.32) gives

$$R_B T^0 \alpha = (\lambda_1 I_n - S) R_B R_A^{-1} - S^0 \alpha \tag{3.34}$$

Post multiplying relation (3.33) by $\mathbf{e}$ and equating the right hand side of the resulting expression with right side of (3.34) yields the relation (3.26). □

Computation of $\tilde{\boldsymbol{x}}_{0\bullet}(2)$

Substituting equation (3.13) in (3.10) gives

$$\tilde{\boldsymbol{x}}_0 \left(A_{00} + \frac{1}{\lambda_1}(A_{01}\mathbf{R}A_{10}) \right) = \tilde{\boldsymbol{x}}_0 B = \mathbf{0}, \tag{3.35}$$

where $\tilde{\boldsymbol{x}}_0 = (1-\tilde{\rho},\ \tilde{\boldsymbol{x}}_{00}(2),\ \tilde{\boldsymbol{x}}_{01}(2),\ \tilde{\boldsymbol{x}}_{02}(2),\ \ldots,\ \ldots,)$ and the matrix B is given as follows.

$$B_{ij} = \begin{cases} (1-\theta)\alpha R_{0A}T^0 - \lambda & ;\ i=j=1 \\ \alpha\left((1-\theta)R_{1A} + \theta R_{0A}\right)T^0\beta + \lambda_2\beta & ;\ i=1, j=2 \\ \alpha\left((1-\theta)R_{(j-1)A} + \theta R_{(j-2)A}\right)T^0\beta & ;\ i=1, j\geq 3 \\ (1-\theta)R_{0B}T^0 + S^0 & ;\ i=2, j=1 \\ \left((1-\theta)R_{1B} + \theta R_{0B}\right)T^0\beta + S - \lambda I_n & ;\ i=j=2,3,... \\ \left((1-\theta)R_{2B} + \theta R_{1B}\right)T^0\beta + \lambda_2 I_n & ;\ i\geq 2, j=i+1 \\ \left((1-\theta)R_{0B}T^0 + S^0\right)\beta & ;\ i\geq 3, j=i-1 \\ \left((1-\theta)R_{(j-i+1)B} + \theta R_{(j-i)B}\right)T^0\beta & ;\ j\geq 4, 2\leq i\leq j-2 \end{cases}.$$

Expanding (3.35), the following relations are obtained:

$$(1-\tilde{\rho})\left((1-\theta)\alpha R_{0A}T^0 - \lambda\right) + \tilde{\boldsymbol{x}}_{00}(2)\left((1-\theta)R_{0B}T^0 + S^0\right) = \mathbf{0}. \tag{3.36}$$

$$\begin{aligned} &(1-\tilde{\rho})\left(\alpha\left((1-\theta)R_{1A} + \theta R_{0A}\right)T^0\beta + \lambda_2\beta\right) + \\ &\qquad \tilde{\boldsymbol{x}}_{00}(2)\left(\left((1-\theta)R_{1B} + \theta R_{0B}\right)T^0\beta + S - \lambda I_n\right) + \\ &\qquad\qquad \tilde{\boldsymbol{x}}_{01}(2)\left((1-\theta)R_{0B}T^0 + S^0\right)\beta = 0. \end{aligned} \tag{3.37}$$

$$\begin{aligned} &\theta\sum_{k=0}^{j-1}\tilde{\boldsymbol{x}}_{0k}(2)R_{(j-1-k)B}T^0\beta + (1-\theta)\sum_{k=0}^{j}\tilde{\boldsymbol{x}}_{0k}(2)R_{(j-k)B}T^0\beta + \\ &\qquad (1-\tilde{\rho})\alpha\left((1-\theta)R_{jA} + \theta R_{(j-1)A}\right)T^0\beta + \lambda_2\tilde{\boldsymbol{x}}_{0(j-2)}(2) + \\ &\qquad\qquad \tilde{\boldsymbol{x}}_{0(j-1)}(2)\left(S - \lambda I_n\right) + \tilde{\boldsymbol{x}}_{0j}(2)S^0\beta = 0, j\geq 2. \end{aligned} \tag{3.38}$$

Post multiplying the relation (3.36) by β and then adding to relations (3.37) and (3.38) yields the following equation on simplification.

$$\tilde{\boldsymbol{x}}_{0\bullet}(2)\left(R_B T^0 \beta + S^0 \beta + S - \lambda_1 I_n\right) = (1-\tilde{\rho})\left(\lambda_1 - \alpha R_A T^0\right)\beta \tag{3.39}$$

Post multiplying $\left(\lambda_1 - \alpha R_A T^0\right)$ by α and using relations (3.25) and (3.31) we have

$$\left(\lambda_1 - \alpha R_A T^0\right)\alpha = \lambda_1 \alpha \left(I_m - \mathbf{e}\alpha\right)$$

Multiplying both sides of the above relation by $\mathbf{e}$ yields $\lambda_1 - \alpha R_A T^0 = 0$. Then the expression on the right hand side of (3.39) becomes zero and we get

$$\tilde{\boldsymbol{x}}_{0\bullet}(2)\left(R_B T^0 \beta + S^0 \beta + S - \lambda_1 I_n\right) = \mathbf{0}$$

Now substituting for $R_B T^0$ from (3.33) in the above equation yields

$$\tilde{\boldsymbol{x}}_{0\bullet}(2)\left(S - \lambda_1 I_n\right)\left(I_n - e\beta\right) = \mathbf{0} \tag{3.40}$$

In order to get the unique solution we need to replace any one of the n linear equations in (3.40) by a normalizing condition. For this consider the probability $\tilde{\rho}_2$ that the server is busy with a $\mathcal{P}_2$ customer. That is $\tilde{\rho}_2 = \tilde{\boldsymbol{x}}_{0\bullet}(2) + \sum\limits_{i=1}^{\infty} \tilde{\boldsymbol{x}}_{i\bullet}(2)$. But relation(3.24) yields

$$\tilde{\rho}_2 = \tilde{\boldsymbol{x}}_{0\bullet}(2)\left(I - R_C\right)^{-1}\mathbf{e} \tag{3.41}$$

Replacing R_C from (3.27), equation(3.41) takes the form

$$\tilde{\rho}_2 = \tilde{\boldsymbol{x}}_{0\bullet.}(2)\left(I - \lambda_1 S^{-1}\right)\mathbf{e} \tag{3.42}$$

Equation(3.40) together with (3.41) or (3.42) provides the solution for $\tilde{\boldsymbol{x}}_{0\bullet}(2)$.

Marginal distribution of low priority($\mathcal{P}_2$) customers

Define $\tilde{\boldsymbol{x}}_{\bullet j}(1) = \sum_{i=1}^{\infty} \tilde{\boldsymbol{x}}_{ij}(1)$ and $\tilde{\boldsymbol{x}}_{\bullet j}(2) = \sum_{i=0}^{\infty} \tilde{\boldsymbol{x}}_{ij}(2)$ for $j \geq 0$, the marginal probability of j $\mathcal{P}_2$ customers in the queue with a $\mathcal{P}_1$ or a $\mathcal{P}_2$ customer in service.

Adding equations (3.15) from $i = 2$ to ∞ and adding this to (3.17) we obtain

$$\tilde{\boldsymbol{x}}_{\bullet j}(1) \quad = \quad (1-\tilde{\rho})\alpha R_{jA} + \textstyle\sum_{k=0}^{j} \left(\tilde{\boldsymbol{x}}_{\bullet k}(1) R_{(j-k)A} + \tilde{\boldsymbol{x}}_{\bullet k}(2) R_{(j-k)B}\right) \tag{3.43}$$

Similarly, adding equations (3.16) from $i = 2$ to ∞ and adding this to (3.18) we have

$$\tilde{\boldsymbol{x}}_{\bullet j}(2) = \tilde{\boldsymbol{x}}_{0j}(2) + \sum_{k=0}^{j} \tilde{\boldsymbol{x}}_{\bullet j}(2) R_{(j-k)C} \tag{3.44}$$

Hence it is required to compute $\tilde{\boldsymbol{x}}_{0j}(2)$ for the desired range of values of j in determining the marginal probabilities of low priority customers with a $\mathcal{P}_2$ customer under service. This is done by using the generating function method.

Let $U\left(\tilde{\boldsymbol{x}}_{0j}(2)\,;z\right) = \sum\limits_{j=0}^{\infty} \tilde{\boldsymbol{x}}_{0j}(2) z^j$ be a generating function for $\tilde{\boldsymbol{x}}_{0j}(2)$ for $j \geq 0$ and $R_\Lambda(z) = \sum\limits_{j=0}^{\infty} R_{j\Lambda} z^j$ for $\Lambda = A, B$ and C. Multiplying relations (3.36) by β, (3.37) by z and (3.38) by z^j and adding over all j we obtain

$$\begin{aligned} U\left(\tilde{\boldsymbol{x}}_{0j}(2)\,;z\right) = (1-\tilde{\rho}) &\left[(\lambda - \lambda_2 z)\beta - (1+\theta z)\alpha R_A(z) T^0 \beta\right] \\ &\left[(1-\theta+\theta z) R_B(z) T^0 \beta + S^0 \beta + (S - \lambda I_n) z + \lambda_2 z^2 I_n\right]^{-1} \end{aligned} \tag{3.45}$$

It remains to derive $R_\Lambda(z)$ for $\Lambda = A, B$ and C.
Multiplying equation(3.8) by z^j, adding over all values of j and adding the resultant to (3.5) gives

$$R_C(z)(S - \lambda I_n - \lambda_2 z I_n) + \lambda_1 I_n = \mathrm{O} \tag{3.46}$$

Performing the same operation on equations (3.6) and (3.3) and then on (3.7) and (3.4) yield

$$\left(1+\theta(z-1)\right) R_A^2(z) T^0 \alpha + R_A(z)\left(\lambda_2 z I_m + T - \lambda I_m\right) + \lambda_1 I_m = \mathrm{O} \tag{3.47}$$

$$(1+\theta(z-1))\left(R_B(z)R_A(z)+R_C(z)R_B(z)\right)T^0\alpha+$$
$$R_B(z)\left(\lambda_2 z I_m + T - \lambda I_m\right) + R_C(z)S^0\alpha = \mathrm{O} \quad (3.48)$$

The last three equations give $R_\Lambda(z)$ for $\Lambda = A, B$ and C

3.1.2 Waiting time analysis

Expected waiting time of $\mathcal{P}_1$ customers

The expected waiting time of an $n^{th}(n > 0)$ $\mathcal{P}_1$ customer in the queue is computed first. We construct a Markov chain $\{(N(t), S(t), M(t)), t \geq 0\}$, where $N(t)$ is the rank of the customer, $S(t)$ the status of server and $M(t)$ is the phase of service at time t. The rank of a customer is r if he is the r^{th} customer in the queue at time t. The rank decreases to 1 as the customers ahead of him leave the system after completing service. State space of the Markov chain is $\{(k, 1, j)/\ 1 \leq k \leq r, 1 \leq j \leq m\} \cup \{(r, 2, j)/1 \leq j \leq n\} \cup \{\Delta\}$ where Δ is the absorbing state indicating that the tagged customer is selected for service.

The infinitesimal generator of dimension $r+1$ is $W_r = \begin{bmatrix} G_r & G_r^0 \\ \mathrm{O} & \mathbf{0} \end{bmatrix}$ where,

$$G_r = \begin{pmatrix} N_1 & N_2 & & & \\ & T & T^0\alpha & & \\ & & T & T^0\alpha & \\ & & & \ddots & \\ & & & & T \end{pmatrix}, G_r^0 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ T^0\alpha \end{pmatrix}; \quad N_1 = \begin{pmatrix} T & 0 \\ S^0\alpha & S \end{pmatrix} \quad N_2 = \begin{pmatrix} T^0\alpha \\ 0 \end{pmatrix}$$

If $\alpha_r = \begin{pmatrix} \frac{\boldsymbol{x}_{r\bullet}}{\boldsymbol{x}_{r\bullet}\mathbf{e}}, & 0, & 0, & \cdots & ,0 \end{pmatrix}$ is a row vector of dimension $rm+n$ and $\mathbf{e}$ is a column vector of ones then expected waiting time of the r^{th} tagged customer according to the position at the time of his arrival is $-\alpha_r G_r^{-1}\mathbf{e}$. Hence the expected waiting time of a $\mathcal{P}_1$ customer in the queue, if he does not feedback is

$$W_{\mathcal{P}_1} = \sum_{r=1}^{\infty} -\alpha_r G_r^{-1} e$$

.

Distribution of waiting time

We use the Uniformization method to determine the waiting time distribution of feedback customers as well as customers who do not feedback.

Waiting time if a customer does not feedback

The later arriving $\mathcal{P}_1$ customers as well as those waiting in $\mathcal{P}_2$ line has no role on the waiting time of a $\mathcal{P}_1$ customer already in the system. So further arrivals to the system are not considered or we set $\lambda = 0$. This produces a Markov chain with an absorbing state, where 'absorbed' means system has reached idle state. The infinitesimal generator of this CTMC is

$$\widetilde{Q}_1 = \begin{pmatrix} 0 & & & & & \\ S^0 & S & & & & \\ t & O & U_0 & & & \\ & & U_{-1} & U_0 & & \\ & & & U_{-1} & U_0 & \\ & & & & \ddots & \ddots \end{pmatrix}$$

where

$$t = \begin{pmatrix} T^0 \\ \mathbf{0} \end{pmatrix} \quad U_0 = \begin{pmatrix} T & O \\ S^0\alpha & S \end{pmatrix} \quad U_{-1} = \begin{pmatrix} T^0\alpha & O \\ O & O \end{pmatrix}.$$

Define $\upsilon = \max |T_{ii}, S_{jj}|$. Then $\widetilde{T}_1 = \upsilon^{-1}\widetilde{Q}_1 + I$ is the transition probability matrix of a discrete time Markov chain of the embedded process for which the time between transitions is exponentially distributed with rate υ. Writing

$V_0^* = \upsilon^{-1}S + I$, $V_0 = \upsilon^{-1}U_0 + I$ and $V_{-1} = \upsilon^{-1}U_{-1}$,

$$\widetilde{T}_1 = \begin{pmatrix} 1 & O & & & & \\ \upsilon^{-1}S^0 & V_0^* & O & & & \\ \upsilon^{-1}t & O & V_0 & & & \\ & & V_{-1} & V_0 & & \\ & & & V_{-1} & V_0 & \\ & & & & \ddots & \ddots \end{pmatrix}.$$

The matrix V_0 contain the probabilities that the embedded process drops no level in one step and V_1 for dropping 1 level in one step. Let the matrix $D_r^{(n)}$ contain the probabilities that the embedded process drops precisely 'r' non-boundary levels in 'n' steps. $D_r^{(n)}$ is recursively given by $D_r^{(n)} = D_{r-1}^{(n-1)}V_{-1} + D_r^{(n-1)}V_0$, $r = 0, 1, 2, ..., n$ with $D_i^{(j)} = 0$ for $i < 0$ and $i > j$.

The number of transitions by time t of the embedded process follows a Poisson process with parameter υt. Therefore the probability that n transitions of the Poisson process have occurred by time t is $p_n(t) = e^{-\upsilon t}\frac{(\upsilon t)^n}{n!}$. Suppose n transitions of the embedded process have occurred in time t and q_n denote the probability that the idle state has not reached in these n transitions. Then the probability that waiting time does not exceed t is $1 - \sum\limits_{n=1}^{\infty} p_n(t)q_n(t)$. Two possibilities are to be considered in obtaining q_n.

First consider the case where a customer entering to the system finds a $\mathcal{P}_2$ in service and no one in $\mathcal{P}_1$ line. Then the probability that the customer is still waiting after n transitions is $q_n(1) = \boldsymbol{x}_{0\bullet}(2)(V_0)^n\mathbf{e}$.

On the other hand an arriving customer joins the $\mathcal{P}_1$ line in the i^{th} position. This time the customer is not absorbed in n transitions if the maximum number of level drops in the embedded process is $i - 1$ and the corresponding probability is $q_n(2) = \sum\limits_{i=1}^{\infty} \boldsymbol{x}_{i\bullet} \sum\limits_{r=0}^{i-1} D_r^{(n)}\mathbf{e}$.

Substitution for $\boldsymbol{x}_{i\bullet}$ from eq.(3.22) yields the relation $q_n(2) = \tilde{\boldsymbol{x}}_0\mathcal{R}(I-\mathcal{R})^{-1}\mathcal{F}^{(n)}\mathbf{e}$.

Here $\mathcal{F}^{(0)} = I$ and $F^{(n)} = \sum_{r=0}^{n} R^r D_r^{(n)}$ and, are recursively computed from the relation $\mathcal{F}^{(n)} = \mathcal{R}\mathcal{F}^{(n-1)}V_{-1} + \mathcal{F}^{(n-1)}V_1$. Finally $q_n = q_n(1) + q_n(2)$.

Waiting time of a customer in the feedback queue

We proceed along the same lines as above after setting $\lambda_2 = 0$, $\theta = 0$ and determine the first passage time until the idle state is reached since there are other recurrent states. The generator matrix of newly constructed Markov chain is

$$\widetilde{Q}_2 = \begin{pmatrix} A_{00}^* & A_{01} & & & \\ A_{10}^* & A_1^* & A_0 & & \\ & A_2^* & A_1^* & A_0 & \\ & & A_2^* & A_1^* & A_0 \\ & & & \ddots & \ddots & \ddots \\ & & & & & \end{pmatrix}$$

, where the elements of this generator matrix are

$A_2^* = \text{diag}(U_{-1}, U_{-1}, ...)$.

$A_1^* = \text{diag}(U_0 - \lambda_1 I, U_0 - \lambda_1 I, ...)$.

$A_{10}^* = \text{diag}((1-\theta)T^0, T^0\beta, T^0\beta, ...)$.

$$A_{00}^* = \begin{pmatrix} -\lambda_1 & & & \\ S^0 & S - \lambda_1 I_n & & \\ & S^0\beta & S - \lambda_1 I_n & \\ & & \ddots & \ddots \end{pmatrix}$$

Letting the uniformization parameter $\tilde{\upsilon} = \upsilon + \lambda_1$, the transition matrix of the embedded Markov chain is

$$\widetilde{P}_2 = \tilde{\upsilon}^{-1}Q + I = \begin{pmatrix} V_{00} & V_{01} & & & \\ V_{10} & V_1 & \tilde{V}_0 & & \\ & \tilde{V}_{-1} & V_1 & \tilde{V}_0 & \\ & & \tilde{V}_{-1} & V_1 & \tilde{V}_0 \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}.$$

$$V_{00} = \tilde{\upsilon}^{-1}A^*_{00} + I,\ V_{01} = \tilde{\upsilon}^{-1}A_{01},\ V_{10} = \tilde{\upsilon}^{-1}A^*_{10},\ V_1 = \tilde{\upsilon}^{-1}A^*_1 + I,$$

$$\tilde{V}_0 = \tilde{\upsilon}^{-1}A_0,\ \tilde{V}_{-1} = \tilde{\upsilon}^{-1}A^*_2.$$

Next we consider a special case of the problem discussed above.

3.2 M/M/1 Feedback queue with non-preemptive priority

We consider a queueing model similar to that in the previous section, except that the service time are exponentially distributed with respective parameters μ_1 and μ_2 for $\mathcal{P}_1$ and $\mathcal{P}_2$ customers. Moreover, we use the same notations to represent the arrival rates and feedback probability. Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system including the one in service if any, $N_2(t)$ be the number of $\mathcal{P}_2$ waiting to get service and $S(t)$ the status of the server which is 1 or 2 according as the server is busy with $\mathcal{P}_1$ or $\mathcal{P}_2$ customers. Thus we get a continuous time Markov chain $\Omega = \{X(t), t \geq 0\} = \{(N_1(t), N_2(t), S(t))/t \geq 0\}$. Its state space is given as $\{(0,0)\} \cup \{(0,j,2)/\ j \geq 0\} \cup \{(i,j,k)/\ i > 0, j \geq 0, k = 1,2\}$.

It is not hard to derive the condition for system stability as

$$\rho = \frac{\lambda_1}{\mu_1} + \frac{\lambda_1\,\theta + \lambda_2}{\mu_2} \ < \ 1.$$

The infinitesimal generator of this continuous time Markov chain consists of block entries of infinite dimension and is obtained as

$$Q^* = \begin{pmatrix} E_{00} & E_{01} & & & \\ E_{10} & E_1 & E_0 & & \\ & E_2 & E_1 & E_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$$E_{00} = \begin{pmatrix} -\lambda & \lambda_2 & & \\ \mu_2 & -(\lambda+\mu_2) & \lambda_2 & \\ & \mu_2 & -(\lambda+\mu_2) & \lambda_2 \\ & \ddots & \ddots & \ddots \end{pmatrix},$$

$$E_{01} = \begin{pmatrix} \lambda & & & & & & \\ & \lambda_1 & 0 & 0 & & & \\ & & & \lambda_1 & 0 & 0 & \\ & & & & & \lambda_1 & 0 \\ & & & & & & \ddots \end{pmatrix},$$

$$E_{10} = \begin{pmatrix} M_{00} & M_{01} & & \\ & M_0 & & \\ & & M_0 & \\ & & & \ddots \end{pmatrix}, \quad E_2 = \begin{pmatrix} M_1 & M_2 & & \\ & M_1 & M_2 & \\ & & \ddots & \ddots \end{pmatrix},$$

$$E_1 = \begin{pmatrix} M_3 & M_4 & \\ & M_3 & M_4 \\ & & \ddots \end{pmatrix}, \text{ and } E_0 = \lambda_1 I_\infty.$$

Here,

$$M_{00} = \begin{bmatrix} \mu_1(1-\theta) \\ 0 \end{bmatrix}, \quad M_{01} = \begin{bmatrix} \mu_1\theta & 0 \\ 0 & 0 \end{bmatrix}, \quad M_0 = \begin{bmatrix} \mu_1(1-\theta) & \mu_1\theta \\ 0 & 0 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} \mu_1(1-\theta) & 0 \\ 0 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} \mu_1\theta & 0 \\ 0 & 0 \end{bmatrix},$$
$$M_3 = \begin{bmatrix} -\lambda-\mu_1 & 0 \\ \mu_2 & -\lambda-\mu_2 \end{bmatrix}, M_4 = \begin{bmatrix} \lambda_2 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

The infinitesimal generator Q^* constitutes a QBD process with infinite number of sub-levels. If $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots]$ denotes the stationary probability vector of Q^* where $\boldsymbol{x}_i$ is the probability vector corresponding to level i of infinite dimension, then along the same lines of Theorems 3.1.1, 3.1.2 and 3.1.4 we have the following.

The solution for $\boldsymbol{x}$ possesses a matrix geometric structure

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1}R, \; i > 1. \tag{3.49}$$

which is extended to level 0 as

$$\boldsymbol{x}_i = \boldsymbol{x}_0 \left(\frac{1}{\lambda_1}A_{01}\right) R^i, \quad i \geq 1. \tag{3.50}$$

where the rate matrix R is the minimal non negative solution to

$$R^2A_2 + RA_1 + A_0 = \mathrm{O}. \tag{3.51}$$

The R matrix has the form

$$R = \begin{pmatrix} R_0 & R_1 & R_2 & R_3 & \ldots \\ 0 & R_0 & R_1 & R_2 & \ldots \\ 0 & 0 & R_0 & R_1 & \ldots \\ 0 & 0 & 0 & R_0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

where each of the matrices R_k is of order 2 represented as $R_k = \begin{bmatrix} a_k & 0 \\ b_k & c_k \end{bmatrix}$.

The entries of R are explicitly computed in the following Theorem.

Theorem 3.2.1. *The elements $R_k (k > 0)$ of the R matrix are computed as,*

$$a_k = \frac{\mu_1\theta\sum_{i=o}^{k-1} a_i a_{k-1-i} + \mu_1(1-\theta)\sum_{i=1}^{k-1} a_i a_{k-i} + \lambda_2 a_{k-1}}{(\lambda+\mu_1) - 2a_0\mu_1(1-\theta)},$$

$$b_k = \frac{1}{(\lambda+\mu_1) - (a_0+c_0)\mu_1(1-\theta)}\left\{\mu_1\theta\sum_{i=0}^{k-1} b_i(a_{k-1-i}+c_{k-1-i}) + \mu_1(1-\theta)\sum_{i=0}^{k-1} b_i(a_{k-i}+c_{k-i}) + \lambda_2 b_{k-1} + \mu_2 c_k\right\},$$

$$c_k = \frac{\lambda_1\lambda_2^k}{(\lambda+\mu_2)^{k+1}}, \qquad k = 1,2,3,\ldots$$

and entries of R_0 are

$$a_0 = \frac{(\lambda+\mu_1) - \sqrt{(\lambda+\mu_1)^2 - 4\mu_1\lambda_1(1-\theta)}}{2\mu_1(1-\theta)},$$

$$b_0 = \frac{\mu_2 c_0}{(\lambda+\mu_1) - (a_0+c_0)\mu_1(1-\theta)},$$

$$c_0 = \frac{\lambda_1}{\lambda+\mu_2}.$$

Proof. Expanding (3.51), we obtain the following relations:

$$R_0^2 M_1 + R_0 M_3 + \lambda_1 I_2 = \mathrm{O},$$

$$\left(\sum_{k=0}^{l-1} R_k R_{l-1-k}\right) M_2 + \left(\sum_{k=0}^{l} R_k R_{l-k}\right) M_1 + R_k M_3 + R_{k-1} M_4 = \mathrm{O}, l \geq 1.$$

The result is established when these equations are expanded with respect to the phases (the phases of the system are the server status(idle/ busy with $\mathcal{P}_1$ customers/ busy with $\mathcal{P}_2$ customers) and the number of $\mathcal{P}_2$ customers in the queue; the level of the system is the number of $\mathcal{P}_1$ customers in the system). □

3.2.1 The joint and marginal probabilities

The recursive formulas for the joint distribution of i $\mathcal{P}_1$ customers in the system and j $\mathcal{P}_2$ customers in the queue and marginal distributions of each are derived below.

The Joint Probability Distribution

Denote by $\boldsymbol{x}_{ij}$ the probability that there are i $\mathcal{P}_1$ customers in the system and j $\mathcal{P}_2$ customers waiting in queue. Let the marginal distribution of the number of $\mathcal{P}_1$ customers in system be denoted by

$$\boldsymbol{x}_{i\bullet} = \sum_{j=0}^{\infty} \boldsymbol{x}_{ij} \ , \quad i \geq 0.$$

Partition $\boldsymbol{x}_{ij}$ to distinguish the type of customer in service as

$$\boldsymbol{x}_{ij} = (x_{ij}(1), x_{ij}(2)), \qquad i \geq 1, j \geq 0.$$

Counting the feedback customers also, relation (3.49) gives

$$\boldsymbol{x}_{ij} = \boldsymbol{x}_{i-1,j} R \ , \quad i > 1, j \geq 0.$$

Expanding this over j,

$$\begin{pmatrix} \boldsymbol{x}_{i0}, & \boldsymbol{x}_{i1}, & \cdots \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{i-1,0}, & \boldsymbol{x}_{i-1,1}, & \cdots \end{pmatrix} \times \begin{pmatrix} R_0 & R_1 & R_2 & \cdots \\ 0 & R_0 & R_1 & \cdots \\ 0 & 0 & R_0 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}.$$

In general,

$$\boldsymbol{x}_{ij} = \sum_{k=0}^{j} \boldsymbol{x}_{i-1,k} \ R_{j-k} \ , \quad i > 1, j \geq 0.$$

Expanding this fixing the type of customer in service, we obtain

$$\boldsymbol{x}_{ij}(1) = \sum_{k=0}^{j} [a_{j-k} \ \boldsymbol{x}_{i-1,k}(1) + b_{j-k} \ \boldsymbol{x}_{i-1,k}(2)] \tag{3.52}$$

$$\boldsymbol{x}_{ij}(2) = \sum_{k=0}^{j} c_{j-k} \ \boldsymbol{x}_{i-1,k}(2); \qquad i > 1, j \geq 0. \tag{3.53}$$

Putting $i = 1$ in (3.50) and expanding we get

$$\boldsymbol{x}_{1j}(1) = a_j(1-\rho) + \sum_{k=0}^{j} b_{j-k}\ \boldsymbol{x}_{0k}(2), \tag{3.54}$$

$$\boldsymbol{x}_{1j}(2) = \sum_{k=0}^{j} c_{j-k}\ \boldsymbol{x}_{0,k}(2); \qquad j \geq 0. \tag{3.55}$$

Marginal distribution of $\mathcal{P}_1$ customers

The marginal distribution $\boldsymbol{x}_{i\bullet}$ for the number of $\mathcal{P}_1$ customers in the system is

$$\boldsymbol{x}_{i\bullet} = \boldsymbol{x}_{(i-1)\bullet}\mathcal{R}_\epsilon \tag{3.56}$$

Or,

$$\boldsymbol{x}_{i\bullet} = \boldsymbol{x}_{1\bullet}\mathcal{R}_\epsilon^{i-1}, \quad i \geq 2; \tag{3.57}$$

where

$$\mathcal{R}_\epsilon = \sum_{j=0}^{\infty} R_j = \begin{bmatrix} \sum_{r=0}^{\infty} a_r & 0 \\ \sum_{r=0}^{\infty} b_r & \sum_{r=0}^{\infty} c_r \end{bmatrix}.$$

Indicating the type of customer in service, relation (3.56) gives

$$(\boldsymbol{x}_{i\bullet}(1), \boldsymbol{x}_{i\bullet}(2)) = \big(\boldsymbol{x}_{(i-1)\bullet}(1), \boldsymbol{x}_{(i-1)\bullet}(2)\big) \begin{bmatrix} \sum a_r & 0 \\ \sum b_r & \sum c_r \end{bmatrix}, \qquad i > 1.$$

so that

$$\begin{aligned} \boldsymbol{x}_{i\bullet}(1) &= \boldsymbol{x}_{(i-1)\bullet}(1)\,(\textstyle\sum a_r) + \boldsymbol{x}_{(i-1)\bullet}(2)\,(\textstyle\sum b_r) \\ \boldsymbol{x}_{i\bullet}(2) &= \boldsymbol{x}_{(i-1)\bullet}(2)(\textstyle\sum c_r), \qquad i > 1. \end{aligned}$$

Adding equations (3.54) and (3.55) over j

$$\begin{aligned} \boldsymbol{x}_{1\bullet}(1) &= (1-\rho)\left(\sum a_r\right) + \boldsymbol{x}_{0\bullet}(2)\left(\sum b_r\right) \\ \boldsymbol{x}_{1\bullet}(2) &= \boldsymbol{x}_{0\bullet}(2)(\textstyle\sum c_r) \end{aligned}$$

which in turn gives

$$\boldsymbol{x}_{1\bullet} = \Big(\ (1-\rho), \quad \boldsymbol{x}_{0\bullet}(2) \ \Big) \mathcal{R}_{\epsilon} \tag{3.58}$$

Combining relations (3.57) and (3.58) we get

$$\boldsymbol{x}_{i\bullet} = \Big(\ (1-\rho), \quad \boldsymbol{x}_{0\bullet}(2) \ \Big) \mathcal{R}_{\epsilon}^{i} \ ; \qquad i \geq 1.$$

Write $\boldsymbol{x}_{0\bullet} = \Big(\ (1-\rho), \quad \boldsymbol{x}_{0\bullet}(2) \ \Big)$, then

$$\boldsymbol{x}_{i\bullet} = \boldsymbol{x}_{0\bullet} \mathcal{R}_{\epsilon}^{i} \ ; \qquad i \geq 1.$$

Expanding on both sides the marginal probabilities of $\mathcal{P}_1$ customers are obtained as

$$\begin{aligned} \boldsymbol{x}_{i\bullet}(1) &= (1-\rho)\left(\sum a_r\right)^{i} + \boldsymbol{x}_{0\bullet}(2) \sum_{k=0}^{i-1} \left(\sum a_r\right)^{k} \left(\sum b_r\right) \left(\sum c_r\right)^{i-1-k}. \\ \boldsymbol{x}_{i\bullet}(2) &= \boldsymbol{x}_{0\bullet}(2) \left(\sum c_r\right)^{i}. \end{aligned}$$

Clearly the marginal probabilities depend on the probability that a $\mathcal{P}_2$ customer is under service and there is no $\mathcal{P}_1$ customer in the system which is given by $\boldsymbol{x}_{0\bullet}(2) = \sum_{j=0}^{\infty} \boldsymbol{x}_{0j}(2)$.

To compute $\boldsymbol{x}_{0\bullet}(2)$

From $\boldsymbol{x}Q^{*} = 0$, the two boundary equations involving $\boldsymbol{x}_0$ are

$$\boldsymbol{x}_0 A_{00} + \boldsymbol{x}_1 A_{10} = \mathbf{0}, \tag{3.59}$$

$$\boldsymbol{x}_0 A_{01} + \boldsymbol{x}_1 [A_1 + R A_2] = \mathbf{0}. \tag{3.60}$$

Substitute for $\boldsymbol{x}_1$ in (3.59) from (3.50) yields

$$\boldsymbol{x}_0 \left(A_{00} + \frac{1}{\lambda_1} (A_{01} R A_{10}) \right) = 0 \tag{3.61}$$

where $\boldsymbol{x}_0 = ((1-\rho),\ \boldsymbol{x}_{00}(2), \boldsymbol{x}_{01}(2),\ \boldsymbol{x}_{02}(2),\ \ldots)$.
Expanding(3.61), the following relations are obtained.

$$(1-\rho)\left[a_0\mu_1(1-\theta)-\lambda\right]+\boldsymbol{x}_{00}(2)\left[b_0\mu_1(1-\theta)+\mu_2\right]=0, \tag{3.62}$$

$$\begin{aligned}(1-\rho)\left[a_0\mu_1\theta+a_1\mu_1(1-\theta)+\lambda_2\right]+\\ \boldsymbol{x}_{00}(2)\left[b_0\mu_1\theta+b_1\mu_1(1-\theta)-(\lambda+\mu_2)\right]+\\ \boldsymbol{x}_{01}(2)\left[b_0\mu_1(1-\theta)+\mu_2\right]=0,\end{aligned} \tag{3.63}$$

$$\begin{aligned}\boldsymbol{x}_{0(j-1)}(2)\left[b_0\theta+b_1\mu_1-(\lambda+\mu_2)\right]+\boldsymbol{x}_{0j}(2)\left[b_0\mu_1(1-\theta)+\mu_2\right]+\\ \sum_{k=0}^{j-2}\boldsymbol{x}_{0k}(2)\left[b_{j-k-1}\theta+b_{j-k}\mu_1\right]+\lambda_2 x_{0,j-2}(2)+\\ (1-\rho)\left[a_{j-1}\mu_1\theta+a_j\mu_1(1-\theta)\right]=0; j\geq 2\end{aligned} \tag{3.64}$$

Solving these we get

$$\boldsymbol{x}_{00}(2)=\frac{(\lambda-a_0\mu_1(1-\theta))\,(1-\rho)}{b_0\mu_1(1-\theta)+\mu_2}.$$

$$\begin{aligned}\boldsymbol{x}_{0j}(2)=\frac{1}{b_0\mu_1(1-\theta)+\mu_2}\Bigg\{&\left[(\lambda+\mu_2)-(b_0\mu_1\theta+b_1\mu_1(1-\theta))\right]\boldsymbol{x}_{0(j-1)}(2)-\\ &(1-\rho)\left[a_{j-1}\mu_1\theta+a_j\mu_1(1-\theta)\right]-\lambda_2\boldsymbol{x}_{0,j-2}(2)-\\ &\sum_{k=0}^{j-2}\boldsymbol{x}_{0k}(2)\left[b_{j-k-1}\mu_1\theta+b_{j-k}\mu_1(1-\theta)\right]\Bigg\}, j\geq 1.\end{aligned}$$

Hence $\boldsymbol{x}_{0\bullet}(2) = \sum_{j=0}^{\infty}\boldsymbol{x}_{0j}(2)$ is computed. Also the joint probabilities given by relations (3.52) to (3.55) are evaluated.

Marginal distribution of $\mathcal{P}_2$ customers

Define $\boldsymbol{x}_{\bullet j}(1)=\sum_{i=1}^{\infty}\boldsymbol{x}_{ij}(1)$ and $\boldsymbol{x}_{\bullet j}(2)=\sum_{i=0}^{\infty}\boldsymbol{x}_{ij}(2)$ for $j\geq 0$.

Summing equations (3.52) for $i = 2$ to ∞ and adding it to (3.54) we obtain

$$\boldsymbol{x}_{\bullet j}(1) = a_j(1-\rho) + \sum_{k=0}^{j} \left[a_{j-k}\, \boldsymbol{x}_{\bullet k}(1) + b_{j-k}\, \boldsymbol{x}_{\bullet k}(2) \right]. \tag{3.65}$$

Similarly adding (3.53) for $i = 2$ to ∞ and adding the resulting to (3.55) we have

$$\boldsymbol{x}_{\bullet j}(2) = \boldsymbol{x}_{0j}(2) + c_0\, \boldsymbol{x}_{\bullet j}(2)$$

This implies,

$$\boldsymbol{x}_{\bullet j}(2) = \frac{1}{1-c_0} \left(\boldsymbol{x}_{0j}(2) + \sum_{k=0}^{j-1} c_{j-k}\, x_{\bullet k}(2) \right).$$

Hence the marginal probabilities of $\mathcal{P}_2$ customers, while a $\mathcal{P}_2$ customer is under service, is determined once we compute $x_{0j}(2)$ for the desired range of values of j, which is done through relations (3.62) and (3.64). The marginal probabilities of $\mathcal{P}_2$ customers, while a $\mathcal{P}_1$ customer is under service, is determined by substituting for $\boldsymbol{x}_{0j}(2)$ and putting $k = 0, 1, 2, ..., j$ in (3.65) to get

$$\begin{aligned} \boldsymbol{x}_{\bullet 0}(1) &= \frac{a_0\,(1-\rho) + b_0\, \boldsymbol{x}_{\bullet 0}(2)}{1-a_0} \\ \boldsymbol{x}_{\bullet j}(1) &= \frac{a_j\,(1-\rho) + \sum_{k=0}^{j-1} a_{j-k}\, \boldsymbol{x}_{\bullet k}(1) + \sum_{k=0}^{j} b_{j-k}\, \boldsymbol{x}_{\bullet k}(2)}{1-a_0}, \quad j \geq 1 \end{aligned}$$

3.2.2 Waiting time analysis

Expected waiting time in $\mathcal{P}_1$ queue

We construct a Markov chain $\{(N(t), S(t)), t \geq 0\}$, where $N(t)$ is the rank of the (tagged)customer at time t. The rank of a customer is r if he is the r^{th} customer in the queue at time t. His rank decreases by 1 as the customers ahead of him leave the system after completing service. Two cases are to be considered according to whether a $\mathcal{P}_1$ or a $\mathcal{P}_2$ customer is under service at the time when the tagged customer joins.

State space of the Markov chain is $\{(n,1):\ 1 \leq n \leq r\} \cup \{(r,2)\} \cup \{\Delta\}$ where $\{\Delta\}$ is the absorbing state indicating that the tagged customer is selected for service. The corresponding infinitesimal generator matrix of dimension $r+2$ is $\mathcal{W} = \begin{bmatrix} T_r & T_r^0 \\ O & 0 \end{bmatrix}$ where,

$$T_r = \begin{cases} -\mu_2,\ i=j=1 \\ \ \mu_2,\ i=1, j=2 \\ -\mu_1,\ \ i=j=2,3,...,r+1 \\ \ \mu_1,\ \ j=i+1, i=2,3,...,r \\ \ 0,\ \ \text{elsewhere} \end{cases} \qquad \text{and} \quad T_r^0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mu_1 \end{pmatrix}$$

Let $\boldsymbol{\alpha}_r = \frac{1}{\boldsymbol{x}_{(r-1)\bullet}} \big(\boldsymbol{x}_{(r-1)\bullet}(2),\ \boldsymbol{x}_{(r-1)\bullet}(1),\ 0,\ \ldots\ ,\ 0\big)$
$\boldsymbol{x}_{(r-1)\bullet} = \boldsymbol{x}_{(r-1)\bullet}(1) + \boldsymbol{x}_{(r-1)\bullet}(2)$ is a row vector of dimension $r+1$. Then expected waiting time of the r^{th} tagged customer is $-\boldsymbol{\alpha}_r\, T_r^{-1}\mathbf{e}$. Hence the expected waiting time of a $\mathcal{P}_1$ customer who does not feedback is

$$W_{\mathcal{P}_1} = \sum_{r=1}^{\infty} -\boldsymbol{\alpha}_r T_r^{-1}\mathbf{e}$$

Waiting time distribution of low priority/feedback customers

We compute the bounds on the distribution of waiting time if a customer feedback in the system. Suppose the tagged customer joins as $r^{th}(r \geq 1)$ in the system. Upon arrival a tagged customer observes either a free server or the server is busy with a $\mathcal{P}_1$ customer or a $\mathcal{P}_2$ customer. The probability of these events are respectively $1-\rho$, $\boldsymbol{x}_{(r-1)\bullet}(1)$ and $\boldsymbol{x}_{(r-1)\bullet}(2)$. Repeating the argument in section 2.2.2 of chapter 2 to compute the bounds on the distribution of waiting time of a $\mathcal{P}_2$ customer, we arrive at the following:
The distribution of waiting time in the system until the tagged customer feedback

is

$$\mathbf{F}_0 = (1-\rho)exp(\mu_1) + \sum_{r;r\geq 2} \mathbf{E}(r,\mu_1)\ \boldsymbol{x}_{(r-1)\bullet}(1) + exp(\mu_2) * \sum_{r;r\geq 1} \mathbf{E}(r,\mu_1)\ \boldsymbol{x}_{(r-1)\bullet}(2).$$

Now assume that the tagged customer feedback in the system. Probability for feedback is θ. We assume that the tagged customer leaves behind i $\mathcal{P}_1$ customers at his feedback instant and join as j^{th} in the $\mathcal{P}_2$ line. Then
The distribution of service time of these i $\mathcal{P}_1$ customers is

$$\mathbf{F}_1(\cdot) = \sum_i \mathbf{E}(i,\mu_1-\lambda)\ \boldsymbol{x}_{i\bullet}(1)$$

. The probability that there are $(j-1)$ $\mathcal{P}_2$ ahead of tagged customer is

$$q_j^{'} = \boldsymbol{x}_{0(j-2)}(2) + \boldsymbol{x}_{\bullet(j-1)}(1).$$

The probability that no $\mathcal{P}_1$ arrived during the service time of a $\mathcal{P}_2$ customer is

$$p_0 = \int_0^\infty e^{-\lambda t}\mu_2 e^{-\mu_2 t}\,dt.$$

Therefore the probability that no $\mathcal{P}_1$ customer arrived during the service time of $((j-1)\ \mathcal{P}_2)$ customers is

$$q_{j-1} = p_0^{j-1}$$

Hence the service time distribution of the j $\mathcal{P}_2$ customers is

$$\mathbf{F}_2(\cdot) = \sum_j \mathbf{E}(j,\mu_2)\ q_j^{'}\ q_{j-1}.$$

- The lower bound for the waiting time distribution in the system is

$$\mathbf{F}_{\min wait} = \mathbf{F}_0 * \theta\mathbf{F}_1 * \mathbf{F}_2.$$

Let k $\mathcal{P}_1$ customers lined up during the service of a $\mathcal{P}_2$ customer. The probability of this event is

$$p_k = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^k}{k!}\mu_2 e^{-\mu_2 t}\,dt.$$

Waiting time distribution generated by the service of all $(j-1)$ $\mathcal{P}_2$ ahead of the tagged customer is

$$\mathbf{F}_3(\cdot) = \sum_j \left[\exp(\mu_2) * \sum_k \mathbf{E}(k, \mu_1 - \lambda)\ p_k\right]^{*(j-1)} q_j^{'}.$$

Then

- The distribution of the maximum waiting time of a feedback customer in the system is

$$\mathbf{F}_{\max wait}(\cdot) = \mathbf{F}_0 * \theta\ \mathbf{F}_1 * \mathbf{F}_3 *\ \exp(\mu_2).$$

3.2.3 Additional performance measures

1. The probability that all the $\mathcal{P}_1$ customers served in a given cycle complete service without any feedback

$$P_{nfb} = \frac{(\mu_1 - \lambda_1)(1-\theta)}{\mu_1 - \lambda_1(1-\theta)}.$$

   This is equivalent to seeking the probability that there is no inflow to $\mathcal{P}_2$ from $\mathcal{P}_1$ during that cycle.

2. The probability that all the $\mathcal{P}_1$ customers served in a given cycle feedback and hence go to $\mathcal{P}_2$ line

$$P_{afb} = \frac{\theta(\mu_1 - \lambda_1)}{(\mu_1 - \lambda_1\theta)}$$

   This is the probability for the other extreme of the case of no feedback in a cycle.

We demonstrate below the impact of fixed values of λ, μ_1 and μ_2 on P_{nfb} and P_{afb} with variations of θ. In tables 1 and 2, P_{nfb} and P_{afb} have identical values corresponding to $\theta = 0.5$.

θ	0	0.1	0.2	0.3	0.4	0.5	0.6
P_{nfb}	1	0.8617	0.7347	0.6176	0.5094	**0.4091**	0.3158
P_{afb}	0	0.0714	0.1475	0.2288	0.3158	**0.4091**	0.5094

Table 3.1: $\lambda_1 = 4, \mu_1 = 13, \mu_2 = 5$

θ	0	0.1	0.2	0.3	0.4	0.5	0.6
P_{nfb}	1	0.8741	0.7111	0.5895	0.4800	**0.3810**	0.2909
P_{afb}	0	0.0640	0.1333	0.2087	0.2909	**0.3810**	0.4800

Table 3.2: $\lambda_1 = 5, \mu_1 = 13, \mu_2 = 5$

The table clearly shows that as the value of θ increases P_{nfb} decreases and P_{afb} increases, as are expected.

Remark: Putting $\lambda_2 = 0$, and hence replacing $\lambda_1 = \lambda$, the whole problem reduces to the case where there is no external arrival to P_2 line. That is the case where there is only one type of customers arriving to the system. The above analysis can be extended to the case of more than one feedback. Dimension of the Markov chain increases by one for unit increase in the number of feedback allowed to a customer. This would result in infinite matrices within each phase. That is to say with a specific number of customers with one feedback, we have to look at all possible customers with 2 feedback and so on.

3.3 M/M/1 Feedback queue with preemptive priority

Here we analyze the feedback queueing system discussed above for preemptive service discipline. Arrival of customers form a Poisson stream and service time are exponentially distributed. The arrival of a $\mathcal{P}_1$ customer interrupts the ongoing service of a $\mathcal{P}_2$ customer and hence the latter joins back as the head of the $\mathcal{P}_2$ queue. Feedback is permitted only to $\mathcal{P}_1$ customers . Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system and $N_2(t)$, the number of $\mathcal{P}_2$ customers in the queue at time t. Whenever $\mathcal{P}_1$ is nonempty, the head of that line will be under service.

Then $\Omega = \{(N_1(t), N_2(t))\,/t \geq 0\}$ is a continuous time Markov chain with state space $\{0^*\} \cup \{(i,j)/i \geq 0, j \geq 0\}$. Here 0^* represents the state where there is no customer in the system(neither $\mathcal{P}_1$ nor $\mathcal{P}_2$) and $(0,0)$ is the state where a $\mathcal{P}_2$ customer is in service with no $\mathcal{P}_2$ customer n wait.

The infinitesimal generator $\hat{Q}$ has as entries block matrices of infinite dimension since the number of phase (capacity of waiting line for feedback customers) is countably infinite. It is given by

$$\hat{Q} = \begin{pmatrix} B_{00} & B_{01} & & & \\ B_{10} & B_1 & B_0 & & \\ & B_2 & B_1 & B_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$$B_{00} = \begin{pmatrix} -\lambda & \lambda_2 & & & \\ \mu_2 & -(\lambda+\mu_2) & \lambda_2 & & \\ & \mu_2 & -(\lambda+\mu_2) & \lambda_2 & \\ & & \ddots & \ddots & \ddots \\ & & & & \end{pmatrix}, \quad \lambda = \lambda_1 + \lambda_2.$$

$$\begin{array}{cc} & \begin{array}{ccccc} 0^* & 0 & 1 & 2 & \dots \end{array} \\ B_{10} = \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} & \begin{pmatrix} \mu_1(1-\theta) & \mu_1\theta & & & \\ & \mu_1(1-\theta) & \mu_1\theta & & \\ & & \mu_1(1-\theta) & \mu_1\theta & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix} \end{array}$$

$$B_2 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \dots \end{array} \\ \begin{pmatrix} \mu_1(1-\theta) & \mu_1\theta & & & \\ & \mu_1(1-\theta) & \mu_1\theta & & \\ & & \mu_1(1-\theta) & \mu_1\theta & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix} \end{array}$$

$$B_{01} = \begin{array}{c} 0^* \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \dots \end{array} \\ \begin{pmatrix} \lambda_1 & & & \\ & \lambda_1 & & \\ & & \lambda_1 & \\ & & & \ddots \\ & & & \end{pmatrix} \end{array}, \quad B_0 = \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & \dots \end{array} \\ \begin{pmatrix} \lambda_1 & & & \\ & \lambda_1 & & \\ & & \lambda_1 & \\ & & & \ddots \\ & & & \end{pmatrix} \end{array}$$

and

$$B_1 = \begin{pmatrix} -(\lambda+\mu_1) & \lambda_2 & & & \\ & -(\lambda+\mu_1) & \lambda_2 & & \\ & & -(\lambda+\mu_1) & \lambda_2 & \\ & & \ddots & \ddots & \\ & & & & \end{pmatrix}$$

We now establish the system stability requirement.

Theorem 3.3.1. *The condition for stability of the system is* $\hat{\rho} = \frac{\lambda_1}{\mu_1} + \frac{\lambda_1\,\theta+\lambda_2}{\mu_2} < 1$. *This is necessary and sufficient for system stability.*

Proof. By interchanging the level and phase in the model, entries of the matrices B_0, B_1, and B_2 are

$$(B_0)_{ij} = \begin{cases} \lambda_2, & i = j = 0, 1, 2, ... \\ \mu_1\theta, & i = 1, 2, 3...; \qquad j = i - 1 \\ 0, & \text{elsewhere} \end{cases},$$

$$(B_1)_{ij} = \begin{cases} -(\lambda + \mu_2), & i = j = 0 \\ -(\lambda + \mu_1), & i = j = 1, 2, ... \\ \lambda_1, & i = 0, 1, 2, ...; \quad j = i + 1 \\ \mu_1(1 - \theta), & i = 1, 2, 3...; \quad j = i - 1 \\ 0, & \text{elsewhere} \end{cases}$$

$$(B_2)_{ij} = \begin{cases} \mu_2, & i = j = 0 \\ 0, & \text{elsewhere} \end{cases}.$$

Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, ...)$ be the steady-state probability vector of the matrix $B(= B_0 + B_1 + B_2)$. Solving the relations $\boldsymbol{\pi} B = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$, we get $\pi_j = \left(\frac{\lambda_1}{\mu_1}\right)^j \pi_0, j \geq 1$. As we have a level independent QBD model, the system is stable if $\boldsymbol{\pi} B_0\mathbf{e} < \boldsymbol{\pi} B_2\mathbf{e}$, which simplifies to $\hat{\rho} < 1$, $\hat{\rho}$ being $\dfrac{\lambda_1}{\mu_1} + \dfrac{\lambda_1\,\theta + \lambda_2}{\mu_2}$. □

The infinitesimal generator $\hat{Q}$ constitutes a QBD process with exceptional boundary behavior and an infinite number of sub-levels. The matrix geometric form of the steady-state distributions for single server queues with preemptive priority also investigated by Neuts [52]when number of phases in each level is finite. An extension of this is done to blocks of infinite size in Miller [49] and is contained in the following theorem.

Theorem 3.3.2. *Let* $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, . . .)$ *denote the invariant probability vector for the QBD process, where* $\boldsymbol{y}_i$ *is the probability vector of infinite dimension corresponding to level* i *. Then the solution for* $\boldsymbol{y}$ *possesses a matrix geometric structure*

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i\mathrm{R},\ i \geq 1. \tag{3.66}$$

where the rate matrix R *is the minimal non negative solution to*

$$\mathrm{R}^2 B_2 + \mathrm{R} B_1 + B_0 = O. \tag{3.67}$$

The matrix geometric structure in relation (3.66) extended to level '0' is

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 \left(\frac{1}{\lambda_1} B_{01} \right) \mathrm{R}. \tag{3.68}$$

Proof. The relations (3.66) and (3.67) are proved in [49] extending the method discussed in [52].
From $\boldsymbol{y}\hat{Q} = 0$, the two boundary equations involving $\boldsymbol{y}_0$ are

$$\boldsymbol{y}_0 B_{00} + \boldsymbol{y}_1 B_{10} = \mathbf{0}, \tag{3.69}$$

$$\boldsymbol{y}_0 B_{01} + \boldsymbol{y}_1 [B_1 + \mathrm{R} B_2] = \mathbf{0}. \tag{3.70}$$

From (3.67) it follows that

$$\mathrm{R}[\mathrm{R} B_2 + B_1] = -B_0.$$

Since $B_0 = \lambda_1 I_\infty$, the matrix R is invertible and the relation (3.70) now simplifies to (3.68). □

Theorem 3.3.3. *The infinite matrix* R *possesses the Toeplitz structure*

$$\mathrm{R} = \begin{pmatrix} r_0 & r_1 & r_2 & r_3 & \cdots \\ 0 & r_0 & r_1 & r_2 & \cdots \\ 0 & 0 & r_0 & r_1 & \cdots \\ 0 & 0 & 0 & r_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where r_k are computed as

$$r_0 = \frac{(\lambda+\mu_1) - \sqrt{(\lambda+\mu_1)^2 - 4\lambda_1\mu_1(1-\theta)}}{2\mu_1(1-\theta)},$$

$$r_1 = \frac{r_0^2\mu_1\theta + \lambda_2 r_0}{\sqrt{(\lambda+\mu_1)^2 - 4\lambda_1\mu_1(1-\theta)}},$$

$$r_{k+1} = \frac{\mu_1\theta\left[\sum_{i=0}^{k} r_i r_{k-i}\right] + \mu_1(1-\theta)\left[\sum_{i=1}^{k} r_i r_{k+1-i}\right] + \lambda_2 r_k}{\sqrt{(\lambda+\mu_1)^2 - 4\lambda_1\mu_1(1-\theta)}}, \quad k \geq 1.$$

Proof. The structure of the process revealed by matrices in $\hat{Q}$ and the interpretation of rate matrix imply the special structure of R. On expanding (3.67), the following relations are obtained;

$$\mu_1(1-\theta)r_0^2 - (\lambda+\mu_1)\, r_0 + \lambda_1 = 0.$$

$$\mu_1\theta\left(\sum_{i=0}^{k} r_i r_{k-i}\right) + \mu_1(1-\theta)\left(\sum_{i=0}^{k+1} r_i r_{k+1-i}\right) + \lambda_2 r_k - (\lambda+\mu_1)\, r_{k+1} = 0, \quad k \geq 1.$$

Solving these, the expressions for $r_k, k = 0, 1, 2...$ are established. □

The nice structure of R and the computability of its elements enable us to have simple expression for the system state probability.

3.3.1 The joint and marginal probabilities

The joint probabilities

The steady-state probability vector $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots)$ of the Markov chain is computed first. Here $\boldsymbol{y}_0 = (y_0, y_{00}, y_{01}, y_{02}, ...)$, y_0 being the probability of idle server, y_{00} the probability of providing service to a $\mathcal{P}_2$ customer when none is waiting in either queues and y_{0j} representing the probability that number in the $\mathcal{P}_2$ line is $j (j \geq 1)$ and no $\mathcal{P}_1$ in the system. $\boldsymbol{y}_i = (y_{i0}, y_{i1}, y_{i2}, ...)$ with y_{ij} representing the probability that the number of $\mathcal{P}_1$ customers in the system is i and number in the $\mathcal{P}_2$ line is j for $i \geq 0$.

Substituting for $\boldsymbol{y}_1$ in (3.69) from (3.68) the following relations are obtained.

$$\begin{aligned} y_0 &= 1 - \hat{\rho}; \quad \hat{\rho} = \frac{\lambda}{\mu_1} + \frac{(\lambda_1\, \theta + \lambda_2)}{\mu_2}, \\ y_{00} &= \frac{1}{\mu_2} \left[\lambda - \mu_1(1-\theta) r_0\right] y_0, \\ y_{01} &= \frac{1}{\mu_2} \left\{ \left[\lambda + \mu_2 - \mu_1(1-\theta) r_0\right] y_{00} - \left[\mu_1 \theta r_0 + \mu_1(1-\theta) r_1 + \lambda_2\right] y_0 \right\}, \end{aligned}$$

$$\begin{aligned} y_{0j} = \frac{1}{\mu_2} \Bigg\{ & \left[\lambda + \mu_2 - \mu_1(1-\theta) r_0\right] y_{0,j-1} - \left[\mu_1 \theta r_{j-1} + \mu_1(1-\theta) r_j\right] y_0 \\ & - \sum_{k=0}^{j-2} \left[\mu_1 \theta r_{j-k-2} + \mu_1(1-\theta) r_{j-1-k}\right] y_{0k} - \lambda_2 y_{0,j-2} \Bigg\}, \quad j \geq 2. \end{aligned}$$

Thus we can compute y_{0j} recursively up to the desired range of values.
Substituting for $\boldsymbol{y}_0$ in (3.68) and expanding, y_{1j} for $j = 0, 1, 2, ...$ are computed as

$$\begin{aligned} y_{10} &= (1 - \hat{\rho})\, r_0, \\ y_{11} &= (1 - \hat{\rho})\, r_1 + y_{00}\, r_0, \\ y_{1j} &= (1 - \hat{\rho})\, r_j + \sum_{k=0}^{j-1} y_{0k}\, r_{j-1-k},\ j = 2, 3, ... \end{aligned}$$

Now the relation (3.66) gives

$$y_{ij} = \sum_{k=0}^{j} y_{i-1,k}\; r_{j-k}, i > 1.$$

The marginal probabilities

Next we compute the marginal probabilities of the system state. These, in turn help us compute the waiting time distribution. Denote the marginal probabilities of the number of high priority ($\mathcal{P}_1$) customers in the system be $\boldsymbol{y}_{i\bullet} = \sum_{j=0}^{\infty} y_{ij}, \quad i \geq 0$. Then

$$\boldsymbol{y}_{i\bullet} = \sum_{j=0}^{\infty}\sum_{k=0}^{j} y_{i-1,k}\; r_{j-k} = \left(\sum_{j=0}^{\infty} y_{i-1,j}\right)\left(\sum_{i=0}^{\infty} r_i\right) = \boldsymbol{y}_{(i-1)\bullet}\; \hat{\rho}_1.$$

Remark: As an arrival of a $\mathcal{P}_1$ customer preempts a $\mathcal{P}_2$ customer in service, the system behaves as an M/M/1 queue as far as marginal probabilities of $\mathcal{P}_1$ customers are concerned. Hence

$$\boldsymbol{y}_{i\bullet} = \hat{\rho}_1^i(1-\hat{\rho}_1), i \geq 0; \quad \hat{\rho}_1 = \frac{\lambda_1}{\mu_1}$$

The marginal distribution of $\mathcal{P}_2$ customers is computed numerically from

$$\boldsymbol{y}_{\bullet j} = \sum_{i=0}^{\infty} y_{ij}, \quad j \geq 0.$$

3.3.2 Waiting time analysis

Waiting time of high priority customers

As an arriving $\mathcal{P}_1$ customer preempts the $\mathcal{P}_2$ customer, if any under service, his waiting time distribution is same as in the case of an M/M/1 queue. Hence expected waiting time of $\mathcal{P}_1$ customer in the system is

$$E(W_{\mathcal{P}_1}) = \frac{\hat{\rho}_1}{\lambda\,(1-\hat{\rho}_1)} = \frac{1}{\mu_1 - \lambda_1}$$

Waiting time of feedback customers

Expected waiting time of a feedback customer is the sum of the following: expected busy cycle generated by the high priority customers left behind by this customer when he completed own service while in $\mathcal{P}_1$ queue, the sum of the expected busy cycles generated at each preemption while being served $\mathcal{P}_2$ line and expected time taken to complete service without a preemption. We get

$$E(W_{\mathcal{P}_2}) = \frac{1}{\mu_1}\frac{\hat{\rho}_1}{(1-\hat{\rho}_1)^2} + \frac{1}{\mu_2}\frac{1}{(1-\hat{\rho}_1)}E(\mathcal{P}_2); \quad E(\mathcal{P}_2) = \sum_{r=1}^{\infty} r\ \boldsymbol{y}_{\bullet r}$$

Chapter 4

A Multi-server Priority Queue with Preemption in Crowdsourcing

This chapter does not appear to have any connection with contents of chapters 2 and 3. Nevertheless, the theme discussed here is also priority queues; with the provision that the two priority queues are externally generated. High priority customers have finite capacity waiting space whereas low priority customers have waiting room of infinite capacity. We discuss a phenomenon called 'crowdsourcing' which is a common feature in supermarkets and shopping malls. This notion was introduced in to queues for the first time by Chakravarthy and Dudin [11].

In this chapter we analyze the impact of preemptive priority in the context of crowdsourcing. Crowdsourcing coined from 'crowd' and 'outsourcing' according to Howe [25] is the act of a company or institution taking a function once

---

Some results in this chapter are included in the following paper.
A. Krishnamoorthy, Dhanya Shajin, Manjunath A. S.: **On a multi-server priority queue with preemption in crowdsourcing** (communicated).

performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. For instance a store may have two type deliveries-one direct and other over phone. Crowdsourcing happens when the store decides to serve indirect customers through direct customers who are willing to serve, the store being main server and willing customers being servers for the store. For a discussion on the crowdsourcing queueing system one may refer to Chakravarthy and Dudin [11]. This is the first reported work on crowdsourcing modelled in the queueing theory context. The content of this chapter differs from that of Chakravarthy and Dudin in the fact that the former is on preemptive priority discipline. Thus several of system performance measures in the two cases differ significantly. Even the stability condition differ significantly in the two cases.

The rest of this chapter is arranged as follows. In section 1 the model under study is described. Section 2 provides the steady-state analysis of the model, including key performance measures. Waiting time analysis of customers is discussed in section 3. Numerical illustrations are presented in section 4.

4.1 Mathematical formulation

We consider a multi-server priority model with two types of customers $\mathcal{P}_1$ and $\mathcal{P}_2$ to which customers arrive according to Poisson process of rates λ_1 and λ_2 respectively. $\mathcal{P}_1$ has priority over $\mathcal{P}_2$, which is of preemptive nature. $\mathcal{P}_1$ and $\mathcal{P}_2$ customers are to be served by one of c servers and the service time are assumed to be exponentially distributed with respective parameters μ_1 and μ_2. Services are offered in the order of the arrivals of the customers. $\mathcal{P}_2$ customers may be served by a $\mathcal{P}_1$ customer also who has been just served out, provided he is available to act as a server. At the time of opting to serve there should be at least one $\mathcal{P}_2$ customer waiting to get a service. We assume that a served $\mathcal{P}_1$ customer will be available to serve a waiting $\mathcal{P}_2$ customer with probability p, $0 \leq p \leq 1$. With probability

$q = 1 - p$, the served $\mathcal{P}_1$ customer will leave the system. If a $\mathcal{P}_1$ customer decides to serve a $\mathcal{P}_2$ customer, then that $\mathcal{P}_2$ customer is immediately removed from the system as the system no longer needs to track that customer. $\mathcal{P}_2$ customers are taken for service one at a time from the head of the queue whenever the queue of $\mathcal{P}_1$ customers is found to be empty at a service completion epoch. The service of such customers is according to a preemptive service discipline, that is the arrival of a $\mathcal{P}_1$ customer interrupts the ongoing service of any one of $\mathcal{P}_2$ customers if any in service, and hence this preempted customer joins back as the head of the $\mathcal{P}_2$ queue. $\mathcal{P}_1$ customers have a limited waiting space L, $1 \leq L < \infty$, while $\mathcal{P}_2$ customers have unlimited waiting space.

Let $N_1(t), S(t)$ and $N_2(t)$ be the number of $\mathcal{P}_1$ customers in the system, the number of servers busy with $\mathcal{P}_2$ customers and the number of $\mathcal{P}_2$ customers in the queue respectively. Then $\Omega = \{(N_2(t), S(t), N_1(t)), t \geq 0\}$ is a continuous time Markov chain (CTMC) with state space

$$\{(0,0,k)/0 \leq k \leq c+L\} \cup \{(i,0,k)/i \geq 1, c \leq k \leq c+L\} \cup \\ \{(0,j,k)/1 \leq j \leq c, 0 \leq k \leq c-j\} \cup \{(i,j,k)/i \geq 1, 1 \leq j \leq c, k = c-j\}.$$

For convenience we group the set of states as follows.

$$\begin{aligned} \hat{\mathbf{0}} &= \{(0,0,k)/0 \leq k \leq c+L\} \cup \{(0,j,k)/1 \leq j \leq c, 0 \leq k \leq c-j\} \\ \hat{i} &= \{(i,0,k)/i \geq 1, c \leq k \leq c+L\} \cup \{(i,j,k)/i \geq 1, 1 \leq j \leq c, k = c-j\}, \\ &\quad \text{for } i \geq 1. \end{aligned}$$

The level $\hat{\mathbf{0}}$ has $c + L + 1 + \frac{c(c+1)}{2}$ states while the level $\hat{i}, i \geq 1$ has $c + L + 1$

states. The infinitesimal generator of this CTMC is a LIQBD and is of the form

$$\mathcal{Q} = \begin{pmatrix} B_1 & B_0 & & & & \\ B_2 & A_1 & A_0 & & & \\ B_3 & A_2 & A_1 & A_0 & & \\ & A_3 & A_2 & A_1 & A_0 & \\ & & A_3 & A_2 & A_1 & A_0 \\ & & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4.1}$$

Note that the CTMC at hand is not a QBD. At a later stage we adopt a procedure to make it a QBD. Define $H_{(i_1,j_1)}^{(i_2,j_2)}$ as the transition rates from $(i_1, j_1) \to (i_2, j_2)$ where $i_1(i_2)$ represents the number of $\mathcal{P}_2$ customers in service and $j_1(j_2)$ represents the number of $\mathcal{P}_1$ customers in the system. Then the transition rates of matrices appearing in $\mathcal{Q}$ are as follows:

$$B_{0(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} \lambda_2 & i_2 = i_1 = 0, j_2 = j_1, c \le j_1 \le c + L \\ \lambda_2 & i_2 = i_1, 1 \le i_1 \le c, j_2 = j_1, j_1 = c - i_1 \\ \lambda_1 & i_2 = i_1 - 1, 1 \le i_1 \le c, j_2 = j_1 + 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{1(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} \lambda_1 & i_2 = i_1 = 0, j_2 = j_1 + 1, \\ & 0 \le j_1 \le c + L - 1 \\ \lambda_1 & i_2 = i_1, 1 \le i_1 \le c - 1, \\ & j_2 = j_1 + 1, 0 \le j_1 \le c - i_1 - 1 \\ min\{j_1, c\}\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, \\ & 1 \le j_1 \le c + L \\ j_1\mu_1 & i_2 = i_1, 1 \le i_1 \le c - 1, \\ & j_2 = j_1 - 1, 1 \le j_1 \le c - i_1 \\ i_1\mu_2 & i_2 = i_1 - 1, 1 \le i_1 \le c, \\ & j_2 = j_1, 0 \le j_1 \le c - i_1 \\ \lambda_2 & i_2 = i_1 + 1, 0 \le i_1 \le c - 1, \\ & j_2 = j_1, 0 \le j_1 \le c - i_1 - 1 \\ -(\lambda_1 + \lambda_2) & i_2 = i_1 = 0, j_2 = j_1 = 0 \\ -(\lambda_1 + \lambda_2 + min\{j_1, c\}\mu_1) & i_2 = i_1 = 0, j_2 = j_1, \\ & 1 \le j_1 \le c + L - 1 \\ -(\lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1 = c + L \\ -(\lambda_1 + \lambda_2 + j_1\mu_1 + i_1\mu_2) & i_2 = i_1, 1 \le i_1 \le c \\ & j_2 = j_1, 0 \le j_1 \le c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{2(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} cp\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, c \leq j_1 \leq c + L \\ j_1 q\mu_1 & i_2 = i_1 + 1, 0 \leq i_1 \leq c - 1, j_2 = j_1 - 1, j_1 = c - i_1 \\ j_1 p\mu_1 & i_2 = i_1, 1 \leq i_1 \leq c - 1, j_2 = j_1 - 1, j_1 = c - i_1 \\ i_1\mu_2 & i_2 = i_1, 1 \leq i_1 \leq c, j_2 = j_1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{3(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} j_1 p\mu_1 & i_2 = i_1 + 1, 0 \leq i_1 \leq c - 1, j_2 = j_1 - 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{0(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} \lambda_2 & i_2 = i_1 = 0, j_2 = j_1, c \leq j_1 \leq c + L \\ \lambda_2 & i_2 = i_1, 1 \leq i_1 \leq c, j_2 = j_1, j_1 = c - i_1 \\ \lambda_1 & i_2 = i_1 - 1, 1 \leq i_1 \leq c, j_2 = j_1 + 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{2(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} cp\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, c + 1 \leq j_1 \leq c + L \\ j_1 q\mu_1 & i_2 = i_1 + 1, 0 \leq i_1 \leq c - 1, j_2 = j_1 - 1, j_1 = c - i_1 \\ i_1\mu_2 & i_2 = i_1, 1 \leq i_1 \leq c, j_2 = j_1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{3(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} j_1 p\mu_1 & i_2 = i_1 + 1, 0 \leq i_1 \leq c - 1, j_2 = j_1 - 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{1(i_1,j_1)}^{(i_2,j_2)} = \begin{cases} \lambda_1 & i_2 = i_1 = 0, j_2 = j_1 + 1, \\ & c + 1 \leq j_1 \leq c + L - 1 \\ cq\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, \\ & c + 1 \leq j_1 \leq c + L \\ -(\lambda_1 + \lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1, \\ & c \leq j_1 \leq c + L - 1 \\ -(\lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1 = c + L \\ -(\lambda_1 + \lambda_2 + j_1\mu_1 + i_1\mu_2) & i_2 = i_1, 1 \leq i_1 \leq c, j_2 = j_1, \\ & j_1 = c - i_1 \\ 0 & \text{otherwise.} \end{cases}$$

Rearranging the generator $\mathcal{Q}$ given in (4.1) by combining the set of states as

$$\widetilde{i} = \{2\hat{i} - 1, 2\hat{i}\}, \quad i \geq 1,$$

this model can be studied as QBD process with the generator $\tilde{\mathcal{Q}}$

$$\tilde{\mathcal{Q}} = \begin{pmatrix} B_1 & \tilde{B}_0 & & & \\ \tilde{B}_2 & \tilde{A}_1 & \tilde{A}_0 & & \\ & \tilde{A}_2 & \tilde{A}_1 & \tilde{A}_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} \tag{4.2}$$

where

$$\tilde{B}_0 = \begin{pmatrix} B_0 & \mathrm{O} \end{pmatrix}, \tilde{B}_2 = \begin{pmatrix} B_2 \\ B_3 \end{pmatrix}, \tilde{A}_0 = \begin{pmatrix} \mathrm{O} & \mathrm{O} \\ A_0 & \mathrm{O} \end{pmatrix}, \tilde{A}_1 = \begin{pmatrix} A_1 & A_0 \\ A_2 & A_1 \end{pmatrix},$$

and $\tilde{A}_2 = \begin{pmatrix} A_3 & A_2 \\ \mathrm{O} & A_3 \end{pmatrix}$.

4.2 Steady-state analysis

We proceed with the steady-state analysis of the queueing system under study. The first step in this direction is to look for the condition for stability.

4.2.1 Stability condition

Now we examine the stability of the system. Define $\tilde{A} = \tilde{A}_0 + \tilde{A}_1 + \tilde{A}_2$. Then

$$\tilde{A} = \begin{pmatrix} A_1 + A_3 & A_0 + A_2 \\ A_0 + A_2 & A_1 + A_3 \end{pmatrix}$$

is the infinitesimal generator of the finite state continuous time Markov chain. Let $\boldsymbol{\eta}$ be the steady-state probability vector of $\tilde{A}$. Then

$$\boldsymbol{\eta}\tilde{A} = \mathbf{0}, \quad \boldsymbol{\eta}\mathbf{e} = 1. \tag{4.3}$$

$\tilde{A}$ is a circulant matrix and hence the vector $\boldsymbol{\eta}$ is of the form

$$\boldsymbol{\eta} = \begin{pmatrix} \frac{\boldsymbol{\pi}}{2} & \frac{\boldsymbol{\pi}}{2} \end{pmatrix} \tag{4.4}$$

where $\boldsymbol{\pi}$ satisfies

$$\boldsymbol{\pi} A = \mathbf{0}, \quad \boldsymbol{\pi}\mathbf{e} = 1 \tag{4.5}$$

with $A = A_0 + A_1 + A_2 + A_3$.

From (4.5) we get

$$\pi_i(j) = \begin{cases} \frac{1}{j!}\left(\frac{\lambda_1}{\mu_1}\right)^j \pi_c(0) & 0 \le i \le c-1, j = c-i \\ \left(\frac{1}{c}\right)^{j-c} \frac{1}{c!}\left(\frac{\lambda_1}{\mu_1}\right)^j \pi_c(0) & i = 0, c+1 \le j \le c+L \end{cases} \tag{4.6}$$

where

$$\pi_c(0) = \left[1 + \sum_{i=1}^{c-1} \frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i + \frac{1}{c!}\left(\frac{\lambda_1}{\mu_1}\right)^c \sum_{i=0}^{L}\left(\frac{\lambda_1}{c\mu_1}\right)^i\right]^{-1}. \tag{4.7}$$

The following theorem provides the stability condition of the queueing system under study.

Theorem 4.2.1. *The system under study is stable if and only if*

$$\lambda_2 - p\lambda_1 a < cp\mu_1 a_1 + \mu_2 a_2 \tag{4.8}$$

where

$$a = \sum_{i=0}^{c-1} \frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i \pi_c(0),$$
$$a_1 = \frac{1}{c!}\left(\frac{\lambda_1}{\mu_1}\right)^c \sum_{i=1}^{L}\left(\frac{\lambda_1}{c\mu_1}\right)^i \pi_c(0),$$
$$a_2 = \sum_{i=0}^{c-1} \frac{c-i}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i \pi_c(0)$$

with $\pi_c(0)$ as given in (4.7).

Proof. The queueing system under study with the QBD type generator given in (4.2) is stable if and only if (see Neuts [52]) the left drift rate exceeds the right

drift rate.
That is,

$$\boldsymbol{\eta}\tilde{A}_0\mathbf{e} < \boldsymbol{\eta}\tilde{A}_2\mathbf{e}. \tag{4.9}$$

From (4.4) and (4.5) we have

$$\boldsymbol{\eta}\tilde{A}_0\mathbf{e} = \frac{\boldsymbol{\pi}A_0\mathbf{e}}{2} \tag{4.10}$$

and

$$\boldsymbol{\eta}\tilde{A}_2\mathbf{e} = \frac{\boldsymbol{\pi}A_2\mathbf{e}}{2} + \boldsymbol{\pi}A_3\mathbf{e}. \tag{4.11}$$

Equation (4.6) yields the following

$$\boldsymbol{\pi}A_0\mathbf{e} = \lambda_2 + \lambda_1 \sum_{i=0}^{c-1} \frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i \pi_c(0) \tag{4.12}$$

$$\boldsymbol{\pi}A_3\mathbf{e} = p\lambda_1 \sum_{i=0}^{c-1} \frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i \pi_c(0) \tag{4.13}$$

$$\boldsymbol{\pi}A_2\mathbf{e} = \left[cp\mu_1\frac{1}{c!}\left(\frac{\lambda_1}{\mu_1}\right)^c \sum_{i=1}^{L}\left(\frac{\lambda_1}{c\mu_1}\right)^i + q\lambda_1\sum_{i=0}^{c-1}\frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i + \mu_2\sum_{i=0}^{c-1}\frac{c-i}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i\right]\pi_c(0) \tag{4.14}$$

From relations (4.13) and (4.14) we get

$$\boldsymbol{\pi}A_2\mathbf{e} + 2\boldsymbol{\pi}A_3\mathbf{e} = \left[(p+1)\lambda_1\sum_{i=0}^{c-1}\frac{1}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i + \mu_2\sum_{i=0}^{c-1}\frac{c-i}{i!}\left(\frac{\lambda_1}{\mu_1}\right)^i + cp\mu_1\frac{1}{c!}\left(\frac{\lambda_1}{\mu_1}\right)^c\sum_{i=1}^{L}\left(\frac{\lambda_1}{c\mu_1}\right)^i\right]\pi_c(0). \tag{4.15}$$

Using (4.9), (4.12) and (4.15) we obtain the stated result. □

How the stability condition looks like for a single server queue, is given in

Theorem 4.2.2. *In the case of a single server, the queueing system under study is stable if and only if the following condition is satisfied*

$$\lambda_2 < p\mu_1 + (\mu_2 - p\mu_1)\left(1 - \frac{\lambda_1}{\mu_1}\right)\left[1 - \left(\frac{\lambda_1}{\mu_1}\right)^{L+2}\right]^{-1}. \tag{4.16}$$

Proof. When $c = 1$, the steady-state equations in (4.5) reduce to

$$\begin{aligned}
-(\lambda_1 + \mu_1)\pi_0(1) + \mu_1\pi_0(2) + \lambda_1\pi_1(0) &= 0, \\
\lambda_1\pi_0(i-1) - (\lambda_1 + \mu_1)\pi_0(i) + \mu_1\pi_0(i+1) &= 0, \quad 2 \le i \le L, \\
\lambda_1\pi_0(L) - \mu_1\pi_0(L+1) &= 0, \\
-\lambda_1\pi_1(0) - \mu_1\pi_0(1) &= 0,
\end{aligned} \tag{4.17}$$

subject to the normalizing condition

$$\sum_{i=1}^{L+1} \pi_0(i) + \pi_1(0) = 1. \tag{4.18}$$

Solving the set of equations in (4.17) we get

$$\pi_0(i) = \left(\frac{\lambda_1}{\mu_1}\right)^i \pi_1(0), \quad 1 \le i \le L+1 \tag{4.19}$$

Use (4.19) and the normalizing condition (4.18) to obtain

$$\pi_1(0) = \left(1 - \frac{\lambda_1}{\mu_1}\right)\left[1 - \left(\frac{\lambda_1}{\mu_1}\right)^{L+2}\right]^{-1}. \tag{4.20}$$

Using the relations

$$\begin{aligned}
\pi A_0\mathbf{e} &= \lambda_2 + \lambda_1\pi_1(0) \\
\frac{\pi A_2\mathbf{e}}{2} + \pi A_3\mathbf{e} &= \frac{1}{2}\left[p\mu_1 + \lambda_1\pi_1(0) - p\mu_1\pi_1(0) + \mu_2\pi_1(0)\right]
\end{aligned}$$

and substituting the expression for $\pi_1(0)$ we get stated result. □

Remark:

Under the assumption that $\lambda_1 < \mu_1$, when L goes to ∞,

$$\pi_1(0) \to \left(1 - \frac{\lambda_1}{\mu_1}\right).$$

Then the stability condition reduces to

$$\lambda_2 - p\lambda_1 < \mu_2 \left(1 - \frac{\lambda_1}{\mu_1}\right).$$

4.2.2 Steady-state probability vector

Let $\mathbf{y} = (\mathbf{y}_0, \mathbf{y}_1, \mathbf{y}_2, ...)$ be the steady-state probability vector of the generator $\tilde{\mathcal{Q}}$. That is,

$$\mathbf{y}\tilde{\mathcal{Q}} = \mathbf{0}, \text{ and } \mathbf{ye} = 1. \tag{4.21}$$

Note that $\mathbf{y}_0 = \mathbf{x}_0$ and $\mathbf{y}_i = (\mathbf{x}_{2i-1}, \mathbf{x}_{2i})$ for $i \geq 1$ where $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, ...)$ being the steady-state probability vector of $\mathcal{Q}$.

The vectors are partitioned as

$$\mathbf{x}_0 = \{x_0(0,k)/0 \leq k \leq c+L\} \cup \{x_0(j,k)/1 \leq j \leq c, 0 \leq k \leq c-j\} \text{ and}$$

$$\mathbf{x}_i = \{x_i(0,k)/c \leq k \leq c+L\} \cup \{x_i(j,k)/1 \leq j \leq c, k = c-j\} \text{ for } i \geq 1.$$

Under the stability condition given in (4.8) the steady-state probability vector $\mathbf{y}$ is obtained as

$$\mathbf{y}_i = \mathbf{y}_1 R^{i-1}, \quad i \geq 2$$

where R is the minimal non-negative solution to the matrix quadratic equation

$$R^2\tilde{A}_2 + R\tilde{A}_1 + \tilde{A}_0 = \mathrm{O} \tag{4.22}$$

and the boundary equations are given by

$$\begin{pmatrix} \mathbf{y}_0 & \mathbf{y}_1 \end{pmatrix} \begin{pmatrix} \tilde{B}_1 & \tilde{B}_0 \\ \tilde{B}_2 & \tilde{A}_1 + R\tilde{A}_2 \end{pmatrix} = \mathbf{0}.$$

The normalizing condition of (4.21) results in

$$\mathbf{y}_0\mathbf{e} + \mathbf{y}_1(I - R)^{-1}\mathbf{e} = 1.$$

The matrix R of the equation (4.22) is given by

$$R = \begin{pmatrix} \mathrm{O} & \mathrm{O} \\ R_1 & R_2 \end{pmatrix}. \tag{4.23}$$

Computation of R matrix

Logarithmic reduction algorithm developed by Latouche and Ramaswami [43] 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.

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

Step 1:

$$\begin{aligned} \mathcal{U} &= \mathcal{HL} + \mathcal{LH} \\ \mathcal{M} &= \mathcal{H}^2 \\ \mathcal{H} &\leftarrow (I - \mathcal{U})^{-1}\mathcal{M} \\ \mathcal{M} &\leftarrow \mathcal{L}^2 \\ \mathcal{L} &\leftarrow (I - \mathcal{U})^{-1}\mathcal{M} \\ \mathcal{G} &\leftarrow \mathcal{G} + \mathcal{TL} \\ \mathcal{T} &\leftarrow \mathcal{TH} \end{aligned}$$

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

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

Define the $(c + L + 1)$-dimensional vector $\boldsymbol{\xi}$ as

$$\boldsymbol{\xi} = \sum_{i=1}^{\infty} \mathbf{y}_i\mathbf{e} = \mathbf{y}_1(I - R)^{-1}(\mathbf{e} \otimes I) = (\mathbf{x}_1 \;\; \mathbf{x}_2)(I - R)^{-1}(\mathbf{e} \otimes I). \tag{4.24}$$

Using the form of R given in (4.23), we have '

$$\begin{aligned}
\boldsymbol{\xi} &= (\mathbf{x}_1,\ \mathbf{x}_2)\,(I-R)^{-1}\begin{pmatrix} I \\ I \end{pmatrix} \\
&= (\mathbf{x}_1,\ \mathbf{x}_2)\begin{pmatrix} I & 0 \\ -R_1 & I-R_2 \end{pmatrix}^{-1}\begin{pmatrix} I \\ I \end{pmatrix} \\
&= (\mathbf{x}_1,\ \mathbf{x}_2)\begin{pmatrix} I & 0 \\ (I-R_2)^{-1}R_1 & (I-R_2)^{-1} \end{pmatrix}\begin{pmatrix} I \\ I \end{pmatrix} \\
&= (\mathbf{x}_1,\ \mathbf{x}_2)\begin{pmatrix} I \\ (I-R_2)^{-1}(I+R_1) \end{pmatrix} \\
&= \mathbf{x}_1 +\ \mathbf{x}_2(I-R_2)^{-1}(I+R_1).
\end{aligned}$$

Partition $\boldsymbol{\xi} = (\boldsymbol{\xi}_0, \xi_1, ..., \xi_c)$ as

$$\begin{aligned}
&\boldsymbol{\xi}_0 = (\xi(0,c), \xi(0,c+1), ..., \xi(0,c+L)) \text{ and} \\
&\xi_j = \xi(j, c-j), \quad 1 \le j \le c.
\end{aligned}$$

Note that $\xi(j,k)$ gives the steady-state probability that j servers are busy with $\mathcal{P}_2$ customers and there are k $\mathcal{P}_1$ customers in the system.

4.2.3 System performance measures

1. Probability that the system is idle is,

$$P_{idle} = x_0(0,0)$$

2. Probability that j servers are busy is,

$$b_j = \begin{cases} x_0(0,0) & j=0 \\ \sum\limits_{k=0}^{j} x_0(k, j-k) & 1 \le j \le c-1 \\ \sum\limits_{i=0}^{\infty}\left[\sum\limits_{k=c}^{c+L} x_i(0,k) + \sum\limits_{k=1}^{c} x_i(k, c-k)\right] & j=c \end{cases}$$

3. Probability that j servers are busy with $\mathcal{P}_1$ customers is,

$$b_j^{(1)} = \begin{cases} x_0(0,0) + \sum_{k=1}^{c} x_0(k,0) + \sum_{i=1}^{\infty} x_i(c,0) & j = 0 \\ x_0(0,j) + \sum_{k=1}^{c-j} x_0(k,j) + \sum_{i=1}^{\infty} x_i(c-j,j) & 1 \leq j \leq c-1 \\ \sum_{i=0}^{\infty} \sum_{k=c}^{c+L} x_i(0,k) & j = c \end{cases}$$

4. Probability that j servers are busy with $\mathcal{P}_2$ customers is,

$$b_j^{(2)} = \begin{cases} \sum_{k=0}^{c+L} x_0(0,k) + \sum_{i=1}^{\infty} \sum_{k=c}^{c+L} x_i(0,k) & j = 0 \\ \sum_{k=0}^{c-j} x_0(j,k) + \sum_{i=1}^{\infty} x_i(j,c-j) & 1 \leq j \leq c \end{cases}$$

5. Probability that an arriving customer is lost due to lack of space in buffer is,

$$P_{lost} = x_0(0,c+L) + \xi(0,c+L)$$

6. Mean number of $\mathcal{P}_1$ customers in the queue is,

$$\mu_{N_1} = \sum_{i=0}^{\infty} \sum_{k=c+1}^{c+L} (k-c)x_i(0,k)$$

7. Mean number of $\mathcal{P}_2$ customers in the queue is,

$$\mu_{N_2} = \sum_{i=1}^{\infty} i \left[\sum_{k=c}^{c+L} x_i(0,k) + \sum_{j=1}^{c} x_i(j,c-j) \right]$$

8. Rate of $\mathcal{P}_2$ customers leaving with $\mathcal{P}_1$ customers denoted by $R_{\mathcal{P}_2 \to \mathcal{P}_1}$ upon service completion of $\mathcal{P}_1$ customers is,

$$R_{\mathcal{P}_2 \to \mathcal{P}_1} = p\mu_1 \sum_{i=1}^{\infty} \left[\sum_{k=c}^{c+L} c x_i(0,k) + \sum_{j=1}^{c-1} (c-j) x_i(j,c-j) \right]$$

9. Rate of $\mathcal{P}_2$ customers leaving the system denoted by $R_{\mathcal{P}_2\to S}$ upon getting service by one of $c-$servers is,

$$R_{\mathcal{P}_2\to S} = \mu_2 \left[\sum_{j=1}^{c}\sum_{k=0}^{c-j} j x_0(j,k) + \sum_{i=1}^{\infty}\sum_{j=1}^{c} j x_i(j,c-j)\right]$$

10. Rate of $\mathcal{P}_2$ customers preempted by $\mathcal{P}_1$ customers is,

$$R_{\mathcal{P}_2\to P} = \lambda_1 \sum_{i=0}^{\infty}\sum_{j=1}^{c} x_i(j,c-j)$$

11. Probability of $\mathcal{P}_2$ customers leaving with $\mathcal{P}_1$ customers upon service completion of $\mathcal{P}_1$ customers is,

$$\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1} = \frac{p\mu_1}{\lambda_2}\left[\sum_{k=c}^{c+L} c\xi(0,k) + \sum_{j=1}^{c}(c-j)\xi(j,c-j)\right]$$

12. Probability that $\mathcal{P}_2$ customers leaving the system upon getting service by one of $c-$servers is,

$$\mathcal{P}_{\mathcal{P}_2\to S} = \frac{\mu_2}{\lambda_2}\sum_{j=1}^{c}\left[\sum_{k=0}^{c-j} j x_0(j,k) + j\xi(j,c-j)\right]$$

13. Probability that a $\mathcal{P}_2$ customer is preempted by $\mathcal{P}_1$ customer is,

$$\mathcal{P}_{\mathcal{P}_2\to P} = \frac{\lambda_1}{\lambda_2}\sum_{j=1}^{c}\left[x_0(j,c-j) + \xi(j,c-j)\right]$$

4.3 Waiting time analysis

4.3.1 Waiting time of an admitted $\mathcal{P}_1$ customer

We assume that all servers are busy with $\mathcal{P}_1$ customers and less than L customers are waiting in $\mathcal{P}_1$ queue. For computing expected waiting time of an admitted

$\mathcal{P}_1$ customer in the queue, we consider the Markov chain $\{M(t), t \geq 0\}$ where $M(t)$ is the rank of the admitted $\mathcal{P}_1$ customer in the queue. We arrange the state space as $\{1, 2, ..., L\} \cup \{\Delta\}$ where $\{\Delta\}$ is the absorbing state denoting that the admitted $\mathcal{P}_1$ customer is taken for service. The infinitesimal generator is of the form

$$W = \begin{pmatrix} T & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{pmatrix}$$

where

$$T = \begin{pmatrix} -c\mu_1 & & & \\ c\mu_1 & -c\mu_1 & & \\ & \ddots & \ddots & \\ & & c\mu_1 & -c\mu_1 \end{pmatrix}, \mathbf{T}^0 = \begin{pmatrix} c\mu_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Thus waiting time of an admitted $\mathcal{P}_1$ customer follows a Phase type distribution with representation $(\boldsymbol{\alpha}, T)$ of order L with the initial probability vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_L)$ where

$$\alpha_j = \frac{1}{1 - P_{lost}}(x_0(0, c + j - 1) + \xi(0, c + j - 1).$$

That is, $\alpha_j,\ 1 \leq j \leq L$ is the probability that an admitted $\mathcal{P}_1$ customer finds $(j-1)$ $\mathcal{P}_1$ customers waiting in the queue with c servers busy with $\mathcal{P}_1$ customers. Since $\mathcal{P}_1$ customers have preemptive priority over $\mathcal{P}_2$ customers, there is no need to keep track of the number of $\mathcal{P}_2$ customers in the queue and future arrivals of any type.

After some algebra we get the expected waiting time of an admitted $\mathcal{P}_1$ customer in the queue as

$$\mu_W^{(1)} = -\boldsymbol{\alpha} T^{(-1)}\mathbf{e} = \frac{1}{c\mu_1}(\alpha_1 + 2\alpha_2 + \ldots + L\alpha_L).$$

4.3.2 Waiting time of $\mathcal{P}_2$ customers

Now we consider the system with preemptive priority. Then the probability for an arbitrary $\mathcal{P}_2$ in service being preempted is $\frac{\lambda_1}{i}$ when there are i $\mathcal{P}_2$ customers in service. In this case we have the infinitesimal generator of the form

$$\mathcal{W} = \begin{pmatrix} B_1^* & B_0^* & & & & \\ B_2 & A_1^* & A_0^* & & & \\ B_3 & A_2 & A_1^* & A_0^* & & \\ & A_3 & A_2 & A_1^* & A_0^* & \\ & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

where

$$B_{1(i_1,j_1)}^{*(i_2,j_2)} = \begin{cases} \lambda_1 & i_2 = i_1 = 0, j_2 = j_1 + 1, \\ & 0 \le j_1 \le c + L - 1 \\ \lambda_1 & i_2 = i_1, 1 \le i_1 \le c - 1, \\ & j_2 = j_1 + 1, 0 \le j_1 \le c - i_1 - 1 \\ min\{j_1, c\}\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, \\ & 1 \le j_1 \le c + L \\ j_1\mu_1 & i_2 = i_1, 1 \le i_1 \le c - 1, j_2 = j_1 - 1, \\ & 1 \le j_1 \le c - i_1 \\ i_1\mu_2 & i_2 = i_1 - 1, 1 \le i_1 \le c, j_2 = j_1, \\ & 0 \le j_1 \le c - i_1 \\ \lambda_2 & i_2 = i_1 + 1, 0 \le i_1 \le c - 1, \\ & j_2 = j_1, 0 \le j_1 \le c - i_1 - 1 \\ -(\lambda_1 + \lambda_2) & i_2 = i_1 = 0, j_2 = j_1 = 0 \\ -(\lambda_1 + \lambda_2 + min\{j_1, c\}\mu_1) & i_2 = i_1 = 0, j_2 = j_1, \\ & 1 \le j_1 \le c + L - 1 \\ -(\lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1 = c + L \\ -(\lambda_1 + \lambda_2 + j_1\mu_1 + i_1\mu_2) & i_2 = i_1, 1 \le i_1 \le c - 1, j_2 = j_1, \\ & 0 \le j_1 \le c - i_1 - 1 \\ -(\frac{\lambda_1}{i_1} + \lambda_2 + j_1\mu_1 + i_1\mu_2) & i_2 = i_1, 1 \le i_1 \le c, j_2 = j_1, \\ & j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{0(i_1,j_1)}^{*(i_2,j_2)} = \begin{cases} \lambda_2 & i_2 = i_1 = 0, j_2 = j_1, c \le j_1 \le c + L \\ \lambda_2 & i_2 = i_1, 1 \le i_1 \le c, j_2 = j_1, j_1 = c - i_1 \\ \frac{\lambda_1}{i_1} & i_2 = i_1 - 1, 1 \le i_1 \le c, j_2 = j_1 + 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{0(i_1,j_1)}^{*(i_2,j_2)} = \begin{cases} \lambda_2 & i_2 = i_1 = 0, j_2 = j_1, c \le j_1 \le c + L \\ \lambda_2 & i_2 = i_1, 1 \le i_1 \le c, j_2 = j_1, j_1 = c - i_1 \\ \frac{\lambda_1}{i} & i_2 = i_1 - 1, 1 \le i_1 \le c, j_2 = j_1 + 1, j_1 = c - i_1 \\ 0 & \text{otherwise,} \end{cases}$$

$$A^{*(i_2,j_2)}_{1(i_1,j_1)} = \begin{cases} \lambda_1 & i_2 = i_1 = 0, j_2 = j_1 + 1, \\ & c+1 \leq j_1 \leq c+L-1 \\ cq\mu_1 & i_2 = i_1 = 0, j_2 = j_1 - 1, \\ & c+1 \leq j_1 \leq c+L \\ -(\lambda_1 + \lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1, \\ & c \leq j_1 \leq c+L-1 \\ -(\lambda_2 + c\mu_1) & i_2 = i_1 = 0, j_2 = j_1 = c+L \\ -(\frac{\lambda_1}{i} + \lambda_2 + j_1\mu_1 + i_1\mu_2) & i_2 = i_1, 1 \leq i_1 \leq c, \\ & j_2 = j_1, j_1 = c - i_1 \\ 0 & \text{otherwise} \end{cases}$$

The matrices B_2, B_3, A_2, A_3 are given in section 4.1.

Let $\tilde{\mathbf{y}}$ be the steady-state probability vector of the generator

$$\tilde{\mathcal{W}} = \begin{pmatrix} B_1^* & \tilde{B}_0^* & & \\ \tilde{B}_2 & \tilde{A}_1^* & \tilde{A}_0^* & \\ & \tilde{A}_2 & \tilde{A}_1^* & \tilde{A}_0^* \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where

$$B_0 = \begin{pmatrix} B_0^* & \mathrm{O} \end{pmatrix}, A_0 = \begin{pmatrix} \mathrm{O} & \mathrm{O} \\ A_0^* & \mathrm{O} \end{pmatrix}, A_1 = \begin{pmatrix} A_1^* & A_0^* \\ A_2 & A_1^* \end{pmatrix}$$

and $\tilde{B}_2, \tilde{A}_2$ are given in section 4.1. Then

$$\tilde{\mathbf{y}}\tilde{\mathcal{W}} = \mathbf{0} \text{ and } \tilde{\mathbf{y}}\mathbf{e} = 1. \tag{4.25}$$

Here the stability condition is obtained as

$$\lambda_2 + \lambda_1 d < (p+1)\mu_1 d_1 + cp\mu_1 d_2 + \mu_2 d_3 \tag{4.26}$$

where

$$d = \sum_{i=1}^{c} \frac{\varsigma_i(c-i)}{i} \qquad d_i = \sum_{i=c+1}^{c+L} \varsigma_0(i)$$

$$d_2 = \sum_{i=0}^{c-1} (c-i)\varsigma_i(c-i) \qquad d_3 = \sum_{i=1}^{c} i\varsigma_i(c-i)$$

with ($\varsigma_i(j)$ denote the probability that i $\mathcal{P}_2$ customers in service and j $\mathcal{P}_1$ customers in the system)

$$\varsigma_i(j) = \begin{cases} \dfrac{1}{j!}\left(\dfrac{\lambda_1}{\mu_1}\right)^j \displaystyle\prod_{k=i+1}^{c} \dfrac{1}{k}\varsigma_c(0) & 0 \leq i \leq c-1, j = c-i \\ \left(\dfrac{1}{c!}\right)^2 \left(\dfrac{\lambda_1}{\mu_1}\right)^j \left(\dfrac{1}{c}\right)^{j-c} \varsigma_c(0) & i = 0, c+1 \leq j \leq c+L \end{cases}$$

and

$$\varsigma_c(0) = \left[1 + \left(\frac{1}{c!}\right)^2 \left(\frac{\lambda_1}{\mu_1}\right)^c \sum_{i=1}^{L} \left(\frac{\lambda_1}{c\mu_1}\right)^i + \sum_{i=0}^{c-1} \left(\frac{1}{(c-i)!}\right) \left(\frac{\lambda_1}{\mu_1}\right)^{c-i} \prod_{k=i+1}^{c} \frac{1}{k}\right]^{-1}.$$

Under the stability condition given in (4.26) the steady-state probability vector $\tilde{\mathbf{y}}$ is obtained as

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{y}}_1 \Re^{i-1}, \quad i > 1 \tag{4.27}$$

where $\Re$ is the minimal non-negative solution to the matrix quadratic equation:

$$\Re^2 \tilde{A}_2 + \Re \tilde{A}_1^* + \tilde{A}_0^* = \mathrm{O}$$

with boundary equations

$$\begin{aligned} \tilde{\mathbf{y}}_0 B_1^* + \tilde{\mathbf{y}}_1 \tilde{B}_2 &= \mathbf{0}, \\ \tilde{\mathbf{y}}_0 \tilde{B}_0^* + \tilde{\mathbf{y}}_1 [\tilde{A}_1^* + \Re \tilde{A}_2] &= \mathbf{0} \end{aligned}$$

subject to the normalizing condition

$$\tilde{\mathbf{y}}_0 \mathbf{e} + \tilde{\mathbf{y}}_1 (I - \Re)^{-1} \mathbf{e} = 1. \tag{4.28}$$

Note that $\tilde{\mathbf{y}}_0 = \tilde{\mathbf{x}}_0$ and $\tilde{\mathbf{y}}_i = (\tilde{\mathbf{x}}_{2i-1}, \tilde{\mathbf{x}}_{2i})$ for $i \geq 1$ where $\tilde{\mathbf{x}}_i,\ i \geq 0$ denote steady-state probability vector of $\mathcal{W}$.

An arriving $\mathcal{P}_1$ customer interrupt the service with equal probability, of any of the $\mathcal{P}_2$ customers in service. Using this assumption we compute the following performance measures:

Waiting time of $\mathcal{P}_2$ customer in the queue

For computing the expected waiting time of a $\mathcal{P}_2$ customer who joins as the r^{th} customer, $r(>0)$, in the queue since the time he joins (joining time taken as time origin) until taken for service for the first time. We consider the Markov chain $\{(N(t), N_2(t), N_1(t)), t \geq 0\}$ where $N(t)$ is the rank of the tagged $\mathcal{P}_2$ customer, $N_2(t)$ is the number of servers busy with $\mathcal{P}_2$ customers and $N_1(t)$ is the number of $\mathcal{P}_1$ customers in the system at time t. Rank of the tagged customer decreases to 1 as the customers ahead of him leave the system after completing service and increases to $r+c$ due to preemption. Thus the state space of the process is $\{(i,j,k)/1 \leq i \leq r+c, 1 \leq j \leq c, k = c-j\} \cup \{(i,0,k)/1 \leq i \leq r+c, c \leq k \leq c+L\} \cup \{\Delta\} \cup \{\Delta^*\}$ where $\{\Delta\}$ is the absorbing state indicating that the tagged customer first selected for service and $\{\Delta^*\}$ represents that the $\mathcal{P}_2$ customer leave the system with a $\mathcal{P}_1$ customer. Thus the infinitesimal generator $W_q^{(2)}(r)$ is of the form

$$W_q^{(2)}(r) = \begin{pmatrix} T_q^{(2)} & \mathbf{t}^0 & \mathbf{t}^{0*} \\ \mathbf{0} & 0 & 0 \end{pmatrix}$$

where

$$T_q^{(2)} = \begin{pmatrix} F_1 & F_2 & F_3 & & & \\ F_0 & F_1 & F_2 & F_3 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & F_0 & F_1 & F_2 & F_3 \\ & & & F_0 & F_1 & F_2 \\ & & & & F_0 & F_1 \end{pmatrix}, \mathbf{t}^0 = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{f} \\ \mathbf{f}_\Delta \end{pmatrix}, \mathbf{t}^{0*} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{f}_{\Delta^*} \end{pmatrix}.$$

Define $D_{(i,j)}^{(k,m)}, D = F_0, F_1, F_2, F_3$ as the transition rates from $(i,j) \to (k,m)$ where $i(k)$ represents the number of $\mathcal{P}_2$ customers in service and $j(m)$ denotes the number of $\mathcal{P}_1$ customers in the system.

$$F_{0(i,j)}^{(k,m)} = \begin{cases} \frac{\lambda_1}{i} & k = i-1, m = j+1, 1 \leq i \leq c, j = c-i \\ 0 & \text{otherwise} \end{cases}$$

$$F_{1(i,j)}^{(k,m)} = \begin{cases} \lambda_1 & k=i=0, m=j+1, c \le j \le c+L-1 \\ cq\mu_1 & k=i=0, m=j-1, c+1 \le j \le c+L \\ 0 & \text{otherwise} \end{cases}$$

$$F_{2(i,j)}^{(k,m)} = \begin{cases} cp\mu_1 & k=i=0, m=j-1, c+1 \le j \le c+L \\ jq\mu_1 & k=i+1, m=j-1, 0 \le i \le c-1, j=c-i \\ i\mu_2 & k=i, m=j, 1 \le i \le c, j=c-i \\ 0 & \text{otherwise} \end{cases}$$

$$F_{3(i,j)}^{(k,m)} = \begin{cases} jp\mu_1 & k=i+1, m=j-1, 0 \le i \le c-1, j=c-i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{f}_{(i,j)} = \begin{cases} jp\mu_1 & 0 \le i \le c-1, j=c-i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{f}_{\Delta(i,j)} = \begin{cases} jq\mu_1 + i\mu_2 & 0 \le i \le c, j=c-i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{f}_{\Delta*(i,j)} = \begin{cases} cp\mu_1 & i=0, c \le j \le c+L \\ jp\mu_1 & 1 \le i \le c-1, j=c-i \\ 0 & \text{otherwise} \end{cases}.$$

Expected waiting time of the tagged $\mathcal{P}_2$ customer in the queue just before taken for service is given by

$$\mu_{W_q}^{(2)} = -\boldsymbol{\psi} \left(T_q^{(2)}\right)^{(-1)} \mathbf{e}$$

where $\boldsymbol{\psi} = (\mathbf{0}, ..., \mathbf{0}, \boldsymbol{\psi}_r, \mathbf{0}, ..., \mathbf{0})$ is the initial probability vector having $(r+c)(c+L+1)$ elements with

$$\boldsymbol{\psi}_r = \frac{\{(\tilde{\mathbf{x}}_{r-1}(0,k)), c \le k \le c+L\} \cup \{(\tilde{\mathbf{x}}_{r-1}(j,k)), 1 \le j \le c, k=c-j\}}{\tilde{\mathbf{x}}_{r-1}\mathbf{e}}.$$

Waiting time of a $\mathcal{P}_2$ customer in the system

We compute the expected waiting time of a tagged $\mathcal{P}_2$ customer in the system. Let r be the rank of the $\mathcal{P}_2$ customer in the queue. Preemption of $\mathcal{P}_2$ customer by $\mathcal{P}_1$ customer has to be considered which, in fact, can be arbitrarily large for any sample $\mathcal{P}_2$ customer. So first we give a bound on this and proceed. Here we consider the Markov chain $\{(N(t), N_2(t), N_1(t)), t \geq 0\}$ where $N(t)$ is the rank of the tagged $\mathcal{P}_2$ customer, $N_2(t)$ is the number of servers busy with $\mathcal{P}_2$ customers and $N_1(t)$ is the number of $\mathcal{P}_1$ customers in the system at time t. Thus the state space of the process is $\{(0, j, k)/1 \leq j \leq c, 0 \leq k \leq c-j\} \cup \{(0, 0, k)/0 \leq k \leq c+L\} \cup \{(i, j, k)/1 \leq i \leq r+c, 1 \leq j \leq c, k = c-j\} \cup \{(i, 0, k)/1 \leq i \leq r+c, c \leq k \leq c+L\} \cup \{\Delta\} \cup \{\Delta_0\} \cup \{\Delta_1\} \cup ... \cup \{\Delta_n\}, n \geq 0$ where $\{\Delta\}$ is the absorbing state which means the tagged $\mathcal{P}_2$ customer leave the system with a $\mathcal{P}_1$ customer and $\{\Delta_i\}$ indicates that the tagged customer complete service with exactly i preemption. Thus the infinitesimal generator $W_s^{(2)}(r)$ is of the form.

$$W_s^{(2)}(r) = \begin{pmatrix} T_s^{(2)} & \tilde{\mathbf{t}} & \tilde{\mathbf{t}}_0 & \tilde{\mathbf{t}}_1 & ... & \tilde{\mathbf{t}}_n \\ \mathbf{0} & 0 & 0 & 0 & ... & 0 \end{pmatrix}$$

where

$$T_q^{(2)} = \begin{pmatrix} F_1 & F_2 & F_3 & & & \\ F_0 & F_1 & F_2 & F_3 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & F_0 & F_1 & F_2 & F_3 \\ & & & F_0 & F_1 & H_2 \\ & & & & H_0 & H_1 \end{pmatrix}, \tilde{\mathbf{t}} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{h} \\ \mathbf{0} \end{pmatrix}, \tilde{\mathbf{t}}_i = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{h}_i \end{pmatrix}, i \geq 0.$$

Define $E_{(i,j)}^{(k,m)}, E = H_0, H_1, H_2$ as the transition rates from $(i, j) \to (k, m)$ where $i(k)$ represents the number of $\mathcal{P}_2$ customers in service and $j(m)$ denotes the number of $\mathcal{P}_1$ customers in the system.

$$H_{0(i,j)}^{(k,m)} = \begin{cases} \frac{\lambda_1}{i} & k = i-1, m = j+1, 1 \leq i \leq c, j = c-i \\ 0 & \text{otherwise} \end{cases}$$

$$H_{1(i,j)}^{(k,m)} = \begin{cases} \lambda_1 & k = i = 0, m = j+1, 0 \le j \le c+L-1 \\ \lambda_1 & k = i, m = j+1, 1 \le i \le c-1, 0 \le\le c-i-1 \\ \lambda_2 & k = i+1, m = j, 1 \le i \le c-1, 0 \le j \le c-i-1 \\ minj, c\mu_1 & k = i = 0, m = j-1, 1 \le j \le c+L \\ j\mu_1 & k = i, m = j-1, 1 \le i \le c-1, 1 \le j \le c-i \\ (i-1)\mu_2 & k = i-1, m = j, 2 \le i \le c, 0 \le j \le c-i \\ 0 & \text{otherwise} \end{cases}$$

$$H_{2(i,j)}^{(k,m)} = \begin{cases} jq\mu_1 & k = i+1, m = j-1, 0 \le i \le c-1, j = c-i \\ i\mu_2 & k = i, m = j, 1 \le i \le c, j = c-i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{h}_{(i,j)} = \begin{cases} cp\mu_1 & i = 0, c \le j \le c+L \\ jp\mu_1 & 1 \le i \le c-1, j = c-i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{h}_{k(i,j)} = \begin{cases} p_k\mu_2 & 1 \le i \le c, 0 \le j \le c-i \\ 0 & \text{otherwise} \end{cases}, 0 \le k \le n$$

where $p_k = Prob.(k \text{ preemption})$, with $\sum_{k=0}^{n} p_k = 1$.

Expected waiting time of an admitted $\mathcal{P}_2$ customer who finds $(r-1)$ $\mathcal{P}_2$ customers in the queue with j servers busy with $\mathcal{P}_2$ customer and k $\mathcal{P}_1$ customers in the system, is given by

$$\mu_{W_s}^{(2)} = -\boldsymbol{\psi'} \left(T_s^{(2)}\right)^{(-1)} \mathbf{e}$$

where $\boldsymbol{\psi'} = (\mathbf{0}, ..., \mathbf{0}, \boldsymbol{\psi'}_r, \mathbf{0}, ..., \mathbf{0})$with

$$\boldsymbol{\psi'}_r = \frac{\{(\tilde{\mathbf{x}}_{r-1}(0,k)), c \le k \le c+L\} \cup \{(\tilde{\mathbf{x}}_{r-1}(j,k)), 1 \le j \le c, k = c-j\}}{\tilde{\mathbf{x}}_{r-1}\mathbf{e}}.$$

- Probability that r^{th} $\mathcal{P}_2$ customer leaves the system with a $\mathcal{P}_1$ customer is $-\boldsymbol{\psi'} \left(T_s^{(2)}\right)^{(-1)} \tilde{\mathbf{t}}$.

- Probability that r^{th} $\mathcal{P}_2$ customer completes his service with exactly k preemptions is $-\boldsymbol{\psi}'\left(T_s^{(2)}\right)^{(-1)}\tilde{\mathbf{t}}_k, 0 \leq k \leq n$.

4.3.3 A different approach to the waiting time analysis of $\mathcal{P}_2$ customers

In this section we consider the system in a different angle. A queue of preempted customers is introduced. Thus we get information about number of preempted customers. They are taken again for service according to their rank in the waiting line. These customers are assumed to have priority over $\mathcal{P}_2$ customers who have not yet been selected for service.

Let $N_1(t), S(t), N(t)$ and $N_2(t)$ be the number of $\mathcal{P}_1$ customers in the system, the number of servers busy with $\mathcal{P}_2$ customers, the number of preempted customers in the queue and the number of $\mathcal{P}_2$ customers in the queue respectively. Then $\Omega' = \{(N_2(t), \mathcal{N}(t), S(t), N_1(t)), t \geq 0\}$ is a continuous time Markov chain with state space

$\{(0,0,0,\ell)/0 \leq \ell \leq c+L\} \cup \{(0,0,k,\ell)/1 \leq k \leq c, 0 \leq \ell \leq c-k\} \cup$
$\{(0,j,k,\ell)/1 \leq j \leq c-1, 1 \leq k \leq c-j; \ell = c-k\} \cup$
$\{(i,j,0,\ell)/0 \leq j \leq c, c \leq \ell \leq c+L, i \geq 1\} \cup$
$\{(0,j,0,\ell)/1 \leq j \leq c, c \leq \ell \leq c+L\} \cup$
$\{(i,j,k,\ell)/0 \leq j \leq c-1, 1 \leq k \leq c-j; \ell = c-k, i \geq 1\}$.

Define $H^{(b_1,b_2,b_3)}_{(a_1,a_2,a_3)}$ as the transition rates from $(a_1,a_2,a_3) \to (b_1,b_2,b_3)$. Using this definition first note that the matrices appearing in $\mathcal{Q}$ (see (4.1)) are as follows:

$$A^{(b_1,b_2,b_3)}_{0(a_1,a_2,a_3)} = \begin{cases} \lambda_2 & b_1 = a_1, b_2 = a_2, b_3 = a_3 \\ & a_1 = 0, a_2 = 0, 0 \leq a_3 \leq c+L \\ & a_1 = 0, 1 \leq a_2 \leq c, 0 \leq a_3 \leq c - a_2 \\ & 0 \leq a_1 \leq c, a_2 = 0, c \leq a_3 \leq c+L \\ & 0 \leq a_1 \leq c-1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$B_{1(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} \lambda_1 & b_1 = a_1 = 0, b_2 = a_2 = 0, b_3 = a_3 + 1 \\ & 0 \leq a_3 \leq c + L - 1 \\ \lambda_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 + 1 \\ & 1 \leq a_1 \leq c, c \leq a_3 \leq c + L - 1 \\ \lambda_1 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 + 1 \\ & 1 \leq a_2 \leq c - 1, 0 \leq a_3 \leq c - a_2 - 1 \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & a_1 = 0, 1 \leq a_2 \leq c, a_3 = c - a_2 \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ min\{c, a_3\}\mu_1 & b_1 = a_1 = 0, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \leq a_3 \leq c + L \\ a_3\mu_1 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 - 1 \\ & 1 \leq a_2 \leq c - 1, 1 \leq a_3 \leq c - a_2 \\ a_2\mu_2 & b_1 = a_1 = 0, b_2 = a_2 - 1, b_3 = a_3 \\ & 1 \leq a_2 \leq c, 0 \leq a_3 \leq c - a_2 \\ \lambda_2 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 \\ & 0 \leq a_2 \leq c - 1, 0 \leq a_3 \leq c - a_2 - 1 \\ cp\mu_1 & b_1 = a_1 - 1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ cq\mu_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ a_3p\mu_1 & b_1 = a_1 - min\{a_1, 2\}, b_2 = a_2, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, 0 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_3q\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, 0 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_2\mu_2 & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$A_{2(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} a_3q\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ cp\mu_1 & b_1 = a_1 = 0, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & c + 1 \leq a_3 \leq c + L \\ a_2\mu_2 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 \\ & 1 \leq a_2 \leq c, a_3 = c - a_2 \\ a_3p\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & a_1 = 1, 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$A^{(b_1,b_2,b_3)}_{1(a_1,a_2,a_3)} = \begin{cases} \lambda_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 + 1 \\ & 0 \leq a_1 \leq c, c \leq a_3 \leq c + L \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & 0 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ cq\mu_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 0 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ a_3 q\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, 0 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ cp\mu_1 & b_1 = a_1 - 1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ a_3 p\mu_1 & b_1 = a_1 - 2, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 2 \leq a_1 \leq c, 0 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_2\mu_2 & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$A^{(b_1,b_2,b_3)}_{3(a_1,a_2,a_3)} = \begin{cases} a_3 p\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$B^{(b_1,b_2,b_3)}_{0(a_1,a_2,a_3)} = \begin{cases} \lambda_2 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 \\ & 0 \leq a_1 \leq c, c \leq a_3 \leq c + L \\ \lambda_2 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 \\ & 0 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$B^{(b_1,b_2,b_3)}_{3(a_1,a_2,a_3)} = \begin{cases} a_3 p\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$B^{(b_1,b_2,b_3)}_{2(a_1,a_2,a_3)} = \begin{cases} a_3 q\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ cp\mu_1 & b_1 = a_1 = 0, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & c + 1 \leq a_3 \leq c + L \\ a_2\mu_2 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 \\ & 1 \leq a_2 \leq c, a_3 = c - a_2 \\ a_3 p\mu_1 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ a_3 p\mu_1 & b_1 = a_1 = 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

In addition diagonal entries in A_1 and B_1 are non-positive, having numerical value equal to the sum of the remaining elements of the same row, found in $B_0, B_1, B_2, B_3, A_0, A_1, A_2$ and A_3.

Let $\tilde{\mathbf{z}}$ be the steady-state probability vector of $\tilde{\mathcal{Q}}$ (see (4.2)). Then

$$\tilde{\mathbf{z}}\tilde{\mathcal{Q}} = \mathbf{0}, \text{ and } \tilde{\mathbf{z}}\mathbf{e} = 1 \tag{4.29}$$

where $\tilde{\mathbf{z}}_0 = \mathbf{z}_0$ and $\tilde{\mathbf{z}}_i = (\mathbf{z}_{2i-1}, \mathbf{z}_{2i})$ for $i \geq 1$ with $\mathbf{z} = (\mathbf{z}_0, \mathbf{z}_1, \mathbf{z}_2, ...)$ is the steady-state probability vector of $\mathcal{Q}$.

Under the stability condition given in (4.8) the steady-state probability vector $\tilde{\mathbf{z}}$ is obtained as

$$\tilde{\mathbf{z}}_i = \tilde{\mathbf{z}}_1 R^{i-1}, \quad i \geq 2$$

where R is the minimal non-negative solution to the matrix quadratic equation

$$R^2\tilde{A}_2 + R\tilde{A}_1 + \tilde{A}_0 = \mathrm{O} \tag{4.30}$$

and the boundary equations are given by

$$\begin{pmatrix} \tilde{\mathbf{z}}_0 & \tilde{\mathbf{z}}_1 \end{pmatrix} \begin{pmatrix} \tilde{B}_1 & \tilde{B}_0 \\ \tilde{B}_2 & \tilde{A}_1 + R\tilde{A}_2 \end{pmatrix} = \mathrm{O}.$$

The normalizing condition of (4.29) results in

$$\tilde{\mathbf{z}}_0\mathbf{e} + \tilde{\mathbf{z}}_1(I - R)^{-1}\mathbf{e} = 1.$$

Waiting time of a $\mathcal{P}_2$ customer in the system

In this section we focus on the waiting time of a $\mathcal{P}_2$ customer in the system without having a bound on the number of pre-emptions. Suppose W_2 denotes the waiting time of a $\mathcal{P}_2$ customer in the system in steady-state. Since $\mathcal{P}_1$ customers have preemptive priority over $\mathcal{P}_2$ customers, the distribution of W_2 may depend on the arrival of $\mathcal{P}_1$ customers. The tagged $\mathcal{P}_2$ customer joins as the r^{th} customer in the queue. For computing the expected waiting time of a $\mathcal{P}_2$ customer we consider the Markov chain $\{(\mathcal{R}(t), \mathcal{N}(t), S(t), N_1(t)), t \geq 0\}$ where $\mathcal{R}(t)$ is the rank of the tagged $\mathcal{P}_2$ customer, $\mathcal{N}(t)$ is the number of preempted customers, $S(t)$ is the number of servers busy with $\mathcal{P}_2$ customers and $N_1(t)$ is the number of $\mathcal{P}_1$

customers in the system. Then the state space of this Markov chain is
$\{(0,0,k,\ell)/1 \leq k \leq c, 0 \leq \ell \leq c-k\} \cup$
$\{(0,j,k,\ell)/1 \leq j \leq c-1, 1 \leq k \leq c-j; \ell = c-k\} \cup$
$\{(i,j,0,\ell)/0 \leq j \leq c, c \leq \ell \leq c+L, 1 \leq i \leq r\} \cup$
$\{(0,j,0,\ell)/1 \leq j \leq c, c \leq \ell \leq c+L\} \cup$
$\{(i,j,k,\ell)/0 \leq j \leq c-1, 1 \leq k \leq c-j, \ell = c-k; 1 \leq i \leq r\} \cup \{\Delta_p\} \cup \{\Delta_{\mu_2}\}$,

where $\{\Delta_p\}$ is the absorbing state for which the tagged customer leave the system with a $\mathcal{P}_1$ customer and $\{\Delta_{\mu_2}\}$ represents that the tagged customer leave the system after completing his service.

Let γ_1 be the probability that the tagged customer is chosen for service and γ_2 the probability that the tagged customer is chosen to leave the system with a $\mathcal{P}_1$ customer. Then the infinitesimal generator is of the form

$$\mathcal{W}^{(2)}(r) = \begin{pmatrix} \mathcal{T}^{(2)} & \mathbf{t}_p & \mathbf{t}_{\mu_2} \\ \mathbf{0} & 0 & 0 \end{pmatrix}$$

where

$$\mathcal{T}^{(2)} = \begin{pmatrix} \mathcal{Z} & A_2 & A_3 & & & & \\ & \mathcal{Z} & A_2 & A_3 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \mathcal{Z} & A_2 & A_3 & \\ & & & & \mathcal{Z} & A_2 & \mathcal{Z}_2 \\ & & & & & \mathcal{Z} & \mathcal{Z}_1 \\ & & & & & & \mathcal{Z}_0 \end{pmatrix}, \mathbf{t}_p = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathcal{Z}_p \\ \mathcal{Z}'_p \end{pmatrix}, \mathbf{t}^{0*} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ \mathcal{Z}_{\mu_2} \end{pmatrix}.$$

Define $D^{(b_1,b_2,b_3)}_{(a_1,a_2,a_3)}, D = \mathcal{Z}_0, \mathcal{Z}_1, \mathcal{Z}_2$ as the transition rates from
$(a_1,a_2,a_3) \to (b_1,b_2,b_3)$.

$$\mathcal{Z}_{0(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} \lambda_1 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 + 1 \\ & 1 \leq a_2 \leq c-1, 0 \leq a_3 \leq c - a_2 - 1 \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & a_1 = 0, 1 \leq a_2 \leq c, a_3 = c - a_2 \\ \lambda_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 + 1 \\ & 1 \leq a_1 \leq c, c \leq a_3 \leq c + L - 1 \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_3\mu_1 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 - 1 \\ & 1 \leq a_2 \leq c - 1, 1 \leq a_3 \leq c - a_2 \\ a_2\mu_2(1-\gamma_1) & b_1 = a_1 = 0, b_2 = a_2 - 1, b_3 = a_3 \\ & 2 \leq a_2 \leq c, 0 \leq a_3 \leq c - a_2 \\ cq\mu_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ cq\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c, a_2 = 0, a_3 = c \\ a_3p\mu_1(1-\gamma_2) & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 - 1 \\ & a_1 = 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_2\mu_2(1-\gamma_1) & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ a_3q\mu_1 & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 - 1 \\ & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ cp\mu_1(1-\gamma_2) & b_1 = a_1 - 1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 2 \leq a_1 \leq c, c + 1 \leq a_3 \leq c + L \\ cp\mu_1(1-\gamma_2) & b_1 = a_1 - 2, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 2 \leq a_1 \leq c, 0 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{p(a_1,a_2,a_3)}^{\Delta_p} = \begin{cases} cp\mu_1 & a_1 = 0, a_2 = 0, c \leq a_3 \leq c + L \\ a_3p\mu_1 & a_1 = 0, 1 \leq a_2 \leq c - 1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{p(a_1,a_2,a_3)}^{\prime\Delta_p} = \begin{cases} cp\mu_1 & a_1 = 1, a_2 = 0, c \leq a_3 \leq c + L \\ cp\mu_1\gamma_2 & 2 \leq a_1 \leq c, a_2 = 0, c \leq a_3 \leq c + L \\ a_3p\mu_1\gamma_2 & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{\mu_2(a_1,a_2,a_3)}^{\Delta_{\mu_2}} = \begin{cases} \mu_2 & a_1 = 0, a_2 = 1, 0 \leq a_3 \leq c - 1 \\ a_2\mu_2\gamma_1 & a_1 = 0, 2 \leq a_2 \leq c, 0 \leq a_3 \leq c - a_2 \\ a_2\mu_2\gamma_1 & 1 \leq a_1 \leq c - 1, 1 \leq a_2 \leq c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{1(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} a_3 q\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \le a_2 \le c-1, a_3 = c - a_2 \\ a_2\mu_2 & b_1 = a_1 = 0, b_2 = a_2, b_3 = a_3 \\ & 1 \le a_2 \le c, a_3 = c - a_2 \\ a_3 p\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & a_1 = 1, 0 \le a_2 \le c-1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{2(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} a_3 p\mu_1 & b_1 = a_1 = 0, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 0 \le a_2 \le c-1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{Z}_{(a_1,a_2,a_3)}^{(b_1,b_2,b_3)} = \begin{cases} \lambda_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 + 1 \\ & 0 \le a_1 \le c, c \le a_3 \le c + L \\ \lambda_1 & b_1 = a_1 + 1, b_2 = a_2 - 1, b_3 = a_3 + 1 \\ & 0 \le a_1 \le c-1, 1 \le a_2 \le c - a_1, a_3 = c - a_2 \\ cq\mu_1 & b_1 = a_1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 0 \le a_1 \le c, c+1 \le a_3 \le c + L \\ a_3 q\mu_1 & b_1 = a_1 - 1, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 1 \le a_1 \le c, 0 \le a_2 \le c - a_1, a_3 = c - a_2 \\ cp\mu_1 & b_1 = a_1 - 1, b_2 = a_2 = 0, b_3 = a_3 - 1 \\ & 1 \le a_1 \le c, c+1 \le a_3 \le c + L \\ a_3 p\mu_1 & b_1 = a_1 - 2, b_2 = a_2 + 1, b_3 = a_3 - 1 \\ & 2 \le a_1 \le c, 0 \le a_2 \le c - a_1, a_3 = c - a_2 \\ a_2\mu_2 & b_1 = a_1 - 1, b_2 = a_2, b_3 = a_3 \\ & 1 \le a_1 \le c-1, 1 \le a_2 \le c - a_1, a_3 = c - a_2 \\ 0 & \text{otherwise} \end{cases}$$

The diagonal entries in $\mathcal{Z}$ and $\mathcal{Z}_0$ are non-positive, having numerical value equal to the sum of other elements of the same row found in $\mathcal{Z}, \mathcal{Z}_0, \mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_p, \mathcal{Z}'_p, \mathcal{Z}_{\mu_2}$, A_2 and A_3.

Thus the expected waiting time of a $\mathcal{P}_2$ customer who finds $r-1$ $\mathcal{P}_2$ customers in the queue with i customers in the preempted queue, j servers busy with $\mathcal{P}_2$ customers and k $\mathcal{P}_1$ customers in the system, is given by

$$\mu_W^{(2)} = -\boldsymbol{\varphi} \left(\mathcal{T}^{(2)}\right)^{-1} \mathbf{e}$$

where $\boldsymbol{\varphi} = (\mathbf{0}, ..., \mathbf{0}, \boldsymbol{\varphi}_r, \mathbf{0}, ..., \mathbf{0})$ with

$$\boldsymbol{\varphi}_r = \frac{1}{\mathbf{z}_{r-1}\mathbf{e}}\Big\{(\mathbf{z}_{r-1}(i,0,k)), 0 \le i \le c, c \le k \le c+L\}\cup$$
$$\{(\mathbf{z}_{r-1}(i,j,k)), 0 \le i \le c-1, 1 \le j \le c-i, k = c-j\Big\}.$$

4.4 Numerical illustration

In this section we discuss a few numerical examples. In the following we define ρ as

$$\rho = \frac{\lambda_2}{p\lambda_1 a + cp\mu_1 a_1 + \mu_2 a_2}. \tag{4.31}$$

Whenever we need to fix a specific value for ρ, we can vary any of the system parameters $\lambda_1, \mu_1, \mu_2, L, c$ and p to arrive at that value. However, a, a_1, a_2 and the vector $\boldsymbol{\pi}$ are independent of λ_2. Thus, for a specific value of ρ from (4.31) we have λ_2.

Example:1

In this example we consider the behaviour of the measure $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$. We fix $\lambda_1 = 1, \mu_1 = \mu_2 = 1.1$, vary p to take values 0.5 and 1, c from 1 to 4 and ρ take values 0.1, 0.3, 0.5, 0.7, 0.9, 0.95 and 0.99 (see Table 4.1).

Table 4.1 gives a picture of the behaviour of $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ for $p = 0.5$ and 1 and with ρ varying from 0.1 to 0.99. We notice that the fraction $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ decreases with increasing value of ρ; in the case of single server and for fixed ρ, the fraction keeps increasing with increasing value of L. The latter behaviour is seen to be exhibited for the multi-server case also. However, when number of servers is 3 or more the fraction $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ increases with increasing value of ρ. This is so since more and more $\mathcal{P}_1$ customers get admitted to the system. However, for small values of λ_1, we notice that increase in value of c results in more and more $\mathcal{P}_2$

customers getting served in the absence of $\mathcal{P}_1$ customers. This explains the reason for small values for $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ for $c = 3$ and 4.

Example:2

Table 4.2 we investigate the behaviour of λ_2 at which the measure $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ attains its maximum. Fix $\lambda_1 = 1, \mu_1 = \mu_2 = 1.1$, vary p to 0.1, 0.2, 0.5, 0.8 and 1, c from 1 to 5, vary L to be 5, 10, 15 and 20. First we get the value of ρ at which $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$ attains its maximum then we obtain corresponding value of λ_2.

Example:3

In Table 4.3 we compute the optimum value of L, say L^* and value of λ_2 at L^*. The optimum L^* is such that the system measure P_{lost} is no larger than 10^{-4} when all other parameters are fixed. We fix $\lambda_1 = 1, \mu_1 = \mu_2 = 1.1$, vary p to 0, 0.5, 1, c from 1 to 5 and ρ take values 0.1, 0.3, 0.5, 0.8, 0.9 and 0.95.

Table 4.3 reveals certain interesting observation: for small values of ρ (hence small values of λ_2 the optimal value of L is relatively small compared to moderate to high values of L for larger values of ρ (hence large values of λ_2).

Example:4

In this example we discuss the benefits of the crowdsourcing queueing model as compared to the classical queueing model with two types of customers with one type having a finite buffer and higher preemptive priority over the other type. That is, we are comparing the model under study for the case when $p > 0$ with the model when $p = 0$. Here we consider the system where the probability of a $\mathcal{P}_1$ customer lost does not exceed 10^{-4}. According to the maximum arrival rate for $\mathcal{P}_2$ customers we define the following ratios:

$$\lambda_2^{Ratio} = \frac{\lambda_2^{p>0}}{\lambda_2^{p=0}}, \quad \mu_{N_1}^{Ratio} = \frac{\mu_{N_1}^{p>0}}{\mu_{N_1}^{p=0}}, \quad \mu_{N_2}^{Ratio} = \frac{\mu_{N_2}^{p>0}}{\mu_{N_2}^{p=0}}.$$

Fix $\lambda_1 = 1, \mu_1 = \mu_2 = 1.1$, vary p to 0.2, 0.4, 0.6, 0.8 and 1, c from 2 to 5 and ρ to take values 0.1, 0.3, 0.5, 0.8, 0.9, 0.95 and 0.99. At the optimum values (see Table 4.3) the ratios $\lambda_2^{Ratio}, \mu_{N_1}^{Ratio}, \mu_{N_2}^{Ratio}$ are given in Tables 4.4 and 4.5.

Tables 4.4 and 4.5 provides certain ratios for different values of p and ρ. For ρ values going up to 0.5 in Table 4.4 (0.99 in Table 4.5), $\mu_{N_1}^{Ratio}$ remains 1, the reason being that the number of $\mathcal{P}_2$ customers alone be affected by increasing value of p. However, for value of $\rho > 0.8$ (in Table 4.4), this ratio is seen to be increasing. Hence the $\mu_{N_2}^{Ratio}$ increases with p and ρ.

Revenue function

Define revenue function as

$$\mathcal{R}_f(\mu_1) = \mathcal{C}_1 \ R_{\mathcal{P}_2 \to \mathcal{P}_1} - \mathcal{C}_2 \ R_{\mathcal{P}_2 \to P} - \mathcal{C}_3 \ P_{lost} - \mathcal{C}_4 \ \mu_{N_2},$$

where $\mathcal{C}_1$: Revenue to the system on account of a waiting $\mathcal{P}_2$ customer, served by a departing $\mathcal{P}_1$ customer

$\mathcal{C}_2$: Preemption cost per unit $\mathcal{P}_2$ customer

$\mathcal{C}_3$: Cost of a $\mathcal{P}_1$ customer lost due to finite waiting space

$\mathcal{C}_4$: Holding cost per $\mathcal{P}_2$ customer

In order to study the variation in μ_1 on profit/ revenue function we fix the costs $\mathcal{C}_1 = \$50, \mathcal{C}_2 = \$10, \mathcal{C}_3 = \$15, \mathcal{C}_4 = \$5$.

For this profit function we get output as indicated in Table tab:6. There is an indication for this profit function to have a global optimum. In the present case the optimal service rate for $\mathcal{P}_1$ customers turn out to be $\mu_1 = 12$. Values of μ_1 above 12 result in very high preemption cost, whereas those below 12 result in large number of $\mathcal{P}_1$ customers loss.

c	p	ρ	$L=5$	$L=10$	$L=15$	$L=20$	$L=25$
		0.1	0.8352	0.8828	0.9005	0.909	0.9136
		0.3	0.8185	0.8729	0.8923	0.9016	0.9066
		0.5	0.7855	0.8492	0.8710	0.8813	0.8872
	0.5	0.7	0.7243	0.7920	0.8154	0.8276	0.8358
		0.9	0.6464	0.7129	0.7378	0.7523	0.7629
		0.95	0.6267	0.6927	0.7180	0.7329	0.7438
		0.99	0.6112	0.6768	0.7024	0.7176	0.7286
1							
		0.1	0.8995	0.9282	0.9390	0.9442	0.9470
		0.3	0.8927	0.9233	0.9348	0.9404	0.9434
		0.5	0.8773	0.9103	0.9224	0.9284	0.9320
	1	0.7	0.8336	0.8675	0.8805	0.8885	0.8944
		0.9	0.7605	0.7948	0.8101	0.8208	0.8290
		0.95	0.7402	0.7749	0.7907	0.8017	0.8101
		0.99	0.7239	0.7590	0.7750	0.7862	0.7946
		0.1	0.3648	0.3664	0.3664	0.3664	0.3664
		0.3	0.3589	0.3606	0.3607	0.3607	0.3607
		0.5	0.3499	0.3519	0.3522	0.3524	0.3527
	0.5	0.7	0.3379	0.3412	0.3431	0.3451	0.3473
		0.9	0.3236	0.3306	0.3369	0.3436	0.3505
		0.95	0.3198	0.3281	0.3360	0.3443	0.3527
		0.99	0.3167	0.3262	0.3354	0.3450	0.3549
2							
		0.1	0.4815	0.4828	0.4829	0.4829	0.4829
		0.3	0.4880	0.4893	0.4893	0.4893	0.4893
		0.5	0.4899	0.4914	0.4916	0.4917	0.4919
	1	0.7	0.4853	0.4883	0.4904	0.4924	0.4946
		0.9	0.4735	0.4825	0.4911	0.4997	0.5082
		0.95	0.4696	0.4810	0.4921	0.5031	0.5139
		0.99	0.4662	0.4798	0.4931	0.5062	0.5190
		0.1	0.1261	0.1262	0.1262	0.1262	0.1262
		0.3	0.1345	0.1346	0.1346	0.1346	0.1346
		0.5	0.1425	0.1426	0.1426	0.1426	0.1426
	0.5	0.7	0.1498	0.1499	0.1500	0.1501	0.1502
		0.9	0.1563	0.1567	0.1571	0.1576	0.1580
		0.95	0.1577	0.1583	0.1589	0.1595	0.1602
		0.99	0.1589	0.1596	0.1603	0.1611	0.1620
3							
		0.1	0.1897	0.1898	0.1898	0.1898	0.1898
		0.3	0.2111	0.2112	0.2112	0.2112	0.2112
		0.5	0.2310	0.2311	0.2311	0.2311	0.2311
	1	0.7	0.2487	0.2489	0.2490	0.2491	0.2493
		0.9	0.2636	0.2644	0.2653	0.2662	0.2671
		0.95	0.2669	0.2681	0.2693	0.2706	0.2719
		0.99	0.2694	0.2709	0.2725	0.2742	0.2759
		0.1	0.0344	0.0344	0.0344	0.0344	0.0344
		0.3	0.0398	0.0398	0.0398	0.0398	0.0398
		0.5	0.0457	0.0457	0.0457	0.0457	0.0457
	0.5	0.7	0.0520	0.0520	0.0520	0.0520	0.0520
		0.9	0.0585	0.0586	0.0586	0.0586	0.0586
		0.95	0.0602	0.0602	0.0602	0.0602	0.0602
		0.99	0.0615	0.0615	0.0615	0.0616	0.0616
4							
		0.1	0.0566	0.0566	0.0566	0.0566	0.0566
		0.3	0.0706	0.0706	0.0706	0.0706	0.0706
		0.5	0.0859	0.0859	0.0859	0.0859	0.0859
	1	0.7	0.1019	0.1019	0.1019	0.1019	0.1019
		0.9	0.1180	0.1180	0.1181	0.1181	0.1181
		0.95	0.1220	0.1220	0.1221	0.1221	0.1222
		0.99	0.1251	0.1252	0.1253	0.1253	0.1254

Table 4.1: Effect of c, p, ρ, L on $\mathcal{P}_{\mathcal{P}_2 \to \mathcal{P}_1}$

c	p	$L=5$		$L=10$		$L=15$		$L=20$	
		$\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1}$	λ_2	$\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1}$	λ_2	$\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1}$	λ_2	$\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1}$	λ_2
1	0.1	0.2715	0.2919	0.3543	0.2397	0.3971	0.2200	0.4218	0.2105
	0.2	0.4160	0.3805	0.5066	0.3340	0.5477	0.3165	0.5702	0.3081
	0.5	0.6112	0.6462	0.6768	0.6171	0.7024	0.6062	0.7176	0.6009
	0.8	0.6922	0.9119	0.7368	0.9003	0.7553	0.8959	0.7677	0.8938
	1	0.7239	1.0890	0.7590	1.0890	0.7750	1.0890	0.7862	1.0890
2	0.1	0.0905	0.7432	0.0931	0.7404	0.0952	0.7404	0.0973	0.7404
	0.2	0.1631	0.8416	0.1678	0.8394	0.1721	0.8394	0.1765	0.8394
	0.5	0.3167	1.1366	0.3262	1.1364	0.3354	1.1364	0.3450	1.1364
	0.8	0.4165	1.4317	0.4288	1.4333	0.4409	1.4334	0.4530	1.4334
	1	0.4662	1.6284	0.4798	1.6313	0.4931	1.6314	0.5062	1.6314
3	0.1	0.0382	0.8018	0.0383	0.8016	0.0385	0.8016	0.0386	0.8016
	0.2	0.0724	0.9007	0.0727	0.9006	0.0730	0.9006	0.0733	0.9006
	0.5	0.1589	1.1974	0.1596	1.1976	0.1603	1.1976	0.1611	1.1976
	0.8	0.2290	1.4940	0.2302	1.4946	0.2314	1.4946	0.2327	1.4946
	1	0.2694	1.6918	0.2709	1.6926	0.2725	1.6926	0.2742	1.6926
4	0.1	0.0122	0.6398	0.0122	0.6398	0.0122	0.6398	0.0122	0.6398
	0.2	0.0244	0.7387	0.0244	0.7388	0.0244	0.7388	0.0244	0.7388
	0.5	0.0615	1.0356	0.0615	1.0358	0.0615	1.0358	0.0616	1.0358
	0.8	0.0996	1.3325	0.0996	1.3328	0.0996	1.3328	0.0997	1.3328
	1	0.1251	1.5305	0.1252	1.5308	0.1253	1.5308	0.1253	1.5308
5	0.1	0.0028	0.4666	0.0028	0.4666	0.0028	0.4666	0.0028	0.4666
	0.2	0.0059	0.5656	0.0059	0.5656	0.0059	0.5656	0.0059	0.5656
	0.5	0.0176	0.8626	0.0176	0.8626	0.0176	0.8626	0.0176	0.8626
	0.8	0.03300	1.1596	0.0330	1.1596	0.0330	1.1596	0.0330	1.1596
	1	0.0452	1.3575	0.0452	1.3576	0.0452	1.3576	0.0452	1.3576

Table 4.2: Value of λ_2 at which $\mathcal{P}_{\mathcal{P}_2\to\mathcal{P}_1}$ attains its maximum and its maximum value

ρ	p	$c=1$		$c=2$		$c=3$		$c=4$		$c=5$	
		L	λ_2	L	λ_2	L	λ_2	L	λ_2	L	λ_2
0.1	0	47	0.0101	7	0.0649	4	0.0711	2	0.0551	1	0.0381
	0.5	47	0.0600	7	0.1148	4	0.1209	2	0.1036	1	0.0821
	1	47	0.1100	7	0.1647	4	0.1707	2	0.1522	1	0.1261
0.3	0	48	0.0303	7	0.1946	4	0.2133	2	0.1653	1	0.1144
	0.5	48	0.1801	7	0.3444	4	0.3627	2	0.3109	1	0.2463
	1	48	0.3300	7	0.4942	4	0.5121	2	0.4565	1	0.3783
0.5	0	51	0.0503	7	0.3243	4	0.3554	2	0.2754	1	0.1906
	0.5	60	0.3001	7	0.5740	4	0.6045	2	0.5182	1	0.4105
	1	74	0.5500	7	0.8236	4	0.8535	2	0.7609	1	0.6304
0.8	0	163	0.0800	8	0.5185	4	0.5687	2	0.4407	1	0.3050
	0.5	383	0.4800	8	0.9183	4	0.9672	2	0.8291	1	0.6569
	1	425	0.8800	9	1.3182	4	1.3657	2	1.2174	1	1.0087
0.9	0	169	0.0900	8	0.5834	4	0.6398	2	0.4958	1	0.3432
	0.5	483	0.5400	10	1.0331	4	1.0881	2	0.9327	1	0.7390
	1	537	0.9900	16	1.4831	4	1.5364	2	1.3696	1	1.1348
0.95	0	171	0.0950	8	0.6158	4	0.6753	2	0.5234	1	0.3622
	0.5	525	0.5700	13	1.0905	4	1.1485	2	0.9845	1	0.7800
	1	589	1.0450	22	1.5655	4	1.6217	2	1.4456	1	1.1979

Table 4.3: Optimum value of L and corresponding value of λ_2

ρ	p	$c = 2$			$c = 3$		
		λ_2^{Ratio}	$\mu_{N_1}^{Ratio}$	$\mu_{N_2}^{Ratio}$	λ_2^{Ratio}	$\mu_{N_1}^{Ratio}$	$\mu_{N_2}^{Ratio}$
0.1	0.2	1.3066	1	0.7966	1.2798	1	1.0243
	0.4	1.6147	1	0.7146	1.5597	1	1.0487
	0.6	1.9229	1	0.6694	1.8410	1	1.0975
	0.8	2.2311	1	0.6440	2.1209	1	1.1219
	1	2.5377	1	0.6299	2.4008	1	1.1707
0.3	0.2	1.3078	1	0.8624	1.2798	1	1.1086
	0.4	1.6156	0.9995	0.7969	1.5602	1	1.2173
	0.6	1.9234	0.9995	0.7641	1.8406	1	1.3260
	0.8	2.2317	0.9995	0.7489	2.1209	1	1.4456
	1	2.5395	0.9995	0.7430	2.4008	1	1.5652
0.5	0.2	1.3080	0.9995	0.9284	1.2805	1	1.1716
	0.4	1.6157	0.9987	0.8923	1.5607	1	1.3457
	0.6	1.9238	0.9982	0.8763	1.8410	1	1.5243
	0.8	2.2318	0.9974	0.9107	2.1212	1	1.7053
	1	2.5396	0.9970	0.8770	2.4015	1	1.8932
0.8	0.2	1.3083	1.0217	1.0262	1.2802	1	1.2549
	0.4	1.6169	1.0414	1.0530	1.5605	1.0032	1.5182
	0.6	1.9253	1.0592	1.0810	1.8408	1.0065	1.7899
	0.8	2.2337	1.0758	1.1108	2.1211	1.0098	2.0681
	1	2.5423	1.1430	1.1436	2.4014	1.0164	2.3529
0.9	0.2	1.3083	1.0420	1.0576	1.2802	1.0032	1.2805
	0.4	1.6165	1.1398	1.1144	1.5604	1.0098	1.5734
	0.6	1.9250	1.2637	1.1647	1.8407	1.0196	1.8764
	0.8	2.2336	1.5195	1.2144	2.1211	1.0295	2.1882
	1	2.5421	2.0109	1.2620	2.4013	1.0393	2.5072
0.95	0.2	1.3080	1.1083	1.0811	1.2803	1.0065	1.2932
	0.4	1.6165	1.3335	1.1584	1.5606	1.0163	1.6007
	0.6	1.9251	1.7884	1.2306	1.8409	1.0294	1.9198
	0.8	2.2336	2.4611	1.2890	2.1211	1.0424	2.2493
	1	2.5422	3.8968	1.3441	2.4014	1.0588	2.5865
0.99	0.2	1.3084	1.1431	1.0936	1.2801	1.0098	1.3028
	0.4	1.6171	1.5396	1.1925	1.5605	1.0228	1.6216
	0.6	1.9257	2.5222	1.2913	1.8407	1.0392	1.9534
	0.8	2.2344	3.8075	1.3540	2.1210	1.0588	2.2964
	1	2.5431	6.2835	1.4074	2.4012	1.0849	2.6476

Table 4.4: Ratios of λ_2, μ_{N_1} and μ_{N_2}

ρ	p	$c = 4$ λ_2^{Ratio}	$\mu_{N_1}^{Ratio}$	$\mu_{N_2}^{Ratio}$	$c = 5$ λ_2^{Ratio}	$\mu_{N_1}^{Ratio}$	$\mu_{N_2}^{Ratio}$
0.1	0.2	1.3520	1	1.25	1.4619	1	1
	0.4	1.7041	1	1.5	1.9238	1	1.1111
	0.6	2.0562	1	1.75	2.3858	1	1.1111
	0.8	2.4101	1	2	2.8477	1	1.1111
	1	2.7622	1	2.25	3.3097	1	1.1111
0.3	0.2	1.3520	1	1.4210	1.46153	1	1.5
	0.4	1.7047	1	1.8421	1.9222	1	2
	0.6	2.0568	1	2.3157	2.3837	1	3
	0.8	2.4095	1	2.8421	2.8452	1	4
	1	2.7616	1	3.3684	3.3068	1	5
0.5	0.2	1.3525	1	1.4791	1.4616	1	2
	0.4	1.7051	1	2.0416	1.9233	1	3
	0.6	2.0577	1	2.6667	2.3845	1	4.5
	0.8	2.4103	1	3.3750	2.8462	1	6.25
	1	2.7628	1	4.1250	3.3074	1	8.5
0.8	0.2	1.3526	1	1.6065	1.4616	1	2.1
	0.4	1.7050	1	2.3442	1.9229	1	3.7
	0.6	2.0574	1	3.2049	2.3842	1	5.8
	0.8	2.4100	1	3.6885	2.8459	1	8.5
	1	2.7624	1	5.3032	3.3072	1	11.8
0.9	0.2	1.3525	1	1.6433	1.4612	1	2.1538
	0.4	1.7049	1	2.4331	1.9224	1	3.8461
	0.6	2.0574	1	3.3757	2.3840	1	6.1538
	0.8	2.4098	1	4.4585	2.8452	1	9.0769
	1	2.7624	1	5.6815	3.3065	1	12.6923
0.95	0.2	1.3523	1	1.6610	1.4616	1	2.1333
	0.4	1.7048	1	2.4745	1.9229	1	3.8667
	0.6	2.0571	1	3.4519	2.3843	1	6.1333
	0.8	2.4096	1	4.5819	2.8456	1	9.1333
	1	2.7619	1	5.8587	3.3072	1	12.9333
0.99	0.2	1.3524	1	1.6701	1.4611	1	2.1176
	0.4	1.7048	1	2.5103	1.9226	1	3.7647
	0.6	2.0573	1	3.5154	2.3841	1	6.0588
	0.8	2.4097	1	4.6804	2.8452	1	9.1176
	1	2.7621	1	6	3.3067	1	12.8823

Table 4.5: Ratios of λ_2, μ_{N_1} and μ_{N_2}

μ_1	2	4	6	8	10	**12**	14	16	18
$\mathcal{R}_f(\mu_1)$	6.4723	12.6282	22.8559	27.8689	29.3834	**29.6287**	29.4716	29.2023	28.9182

Table 4.6: Effect of μ_1 on $\mathcal{R}_f(\mu_1)$ for $(c, L, \lambda_1, \mu_2, p, \rho) = (4, 10, 2, 2.5, 0.7, 0.9)$

Chapter 5

Single Server Queue with Several Services

In the previous chapters we considered interruption of service either by self-interruption (chapter 2), feedback (chapter 3) or through arrival of higher priority customers (chapter 4). In present and the chapter to follow, we analyze cases where permanent interruption (removal from service) takes place due to erroneous service offered or exactly needed service is offered after going through one or more undesired service. That is to say the previous chapters we followed the conventional assumption that the server is completely aware of the exact service requirement of a customer and customer is sure about the type of service he needs. The present and the next chapter discuss models where service requirement of a customer is exactly not known to him nor to the server(s) since a number of distinct services are offered by the service provider. For example patients approach a physician for medical help. The patient may not be aware of his exact health problem, nor the physician be able to diagnose it correctly. Quite often only one type of service is offered by the system and so conflict does not occur. In real life there are several service providing systems offering a multitude of service. The service may start inappropriate and will turn correct. There is also a chance of this service to continue in the incorrect mode and becomes an unsuccessful service. In the latter case the result could be disastrous, especially when life models are considered. The customer may even lose his life. We label such models as diagnostic problems and try to find out a solution to reduce the dilemma caused by this uncertainty.

Consider the example of a multi specialty hospital. A patient could be directed to a physician who has nothing to do with the patient's ailment. However, he still starts medication - as per his diagnosis; the patient and/ physician subsequently realizes that the nature of medication the patient needed was different and refers to some other physician of a different specialty. Here again the patient may end up in the same situation as in the first case. This process could go on until either the patient or physician arrive at the exact nature of medication or the patient reaches such a condition where no medication would work from that time point on. Even in a hospital/ clinic with a single physician the above described is a probable situation.

First we analyze the above described situation in a single server set up. A service system with a preliminary service and a main service is then examined which is found to be on similar lines. This model is then identified with that of Madan [46] and Medhi [48]. We employed arbitrarily distributed service time in certain special cases of the model discussed and analyze such system using supplementary variables [19] to produce a CTMC.

Rest of the chapter is organized as follows. The mathematical model is described in section 1. This section also provides the steady-state analysis and some performance measures. Various cases of the model are considered in Sections 2 and 3. An illustration of the problem is given in Section 4. Numerical example is described in Section 5. In Section 6 we extend the analysis in the case of arbitrarily distributed service time for the undesired and desired stages of service.

5.1 The MAP/PH/1 model

The assumptions leading to the formulation of the mathematical model are

- An infinite capacity queueing system where a single server is providing different kinds of service.

- Arrival of customers to the system is according to the MAP (Markovian arrival process). In a MAP, the customers arrival is directed by an irreducible $CTMC$ (continuous time Markov chain) $\{\boldsymbol{\phi}_t, t \geq 0\}$ with the state space $\{1, 2, ..., m\}$. The transition intensities of the Markov chain $\{\boldsymbol{\phi}_t, t \geq 0\}$ which are accompanied by arrival of $k(= 0, 1)$ customers are described by the matrices D_k. Vector $\boldsymbol{\eta}$ of the stationary distribution of the process $\{\boldsymbol{\phi}_t, t \geq 0\}$ is the unique solution to the system

$$\boldsymbol{\eta}(D_0 + D_1) = \boldsymbol{\eta}D = \mathbf{0} \text{ and } \boldsymbol{\eta}\mathbf{e} = 1. \tag{5.1}$$

  Fundamental rate λ of the MAP is given by $\lambda = \boldsymbol{\eta}D_1 e$.

- A customer is selected for desired (required) service with probability p or to the undesired (incorrect) service with probability $q = 1 - p$.

- PH-representation$(\boldsymbol{\beta}_1, S_1)$ of order n_1 gives the duration of the correct service time distribution when the service of a customer starts in correct service mode. Let S_1^0 be such that $S_1\boldsymbol{e} + S_1^0 = \mathbf{0}$. Let $\mu_1' = \boldsymbol{\beta}_1(-S_1)^{-1}\mathbf{e}$ be the mean of this PH-representation.

- PH-representation $(\boldsymbol{\beta}_2, S_2)$ of order n_2 gives the duration of the incorrect service time distribution when the service of a customer starts in incorrect service mode. The rate (vector) of loss of customers is then given by S_2^0 and the rate (vector) of getting into correct service mode is given by $\hat{S}_2^0$. Note that $S_2\boldsymbol{e} + S_2^0 + \hat{S}_2^0 = \mathbf{0}$. Let $\mu_2' = \boldsymbol{\beta}_2(-S_2)^{-1}\mathbf{e}$ be the mean of this PH-representation. A random threshold clock(timer) starts ticking from the beginning of service so that the customer is pushed out of the system if the clock expires before service completion in undesired service. This timer determines the vector S_2^0.

- PH-representation $(\boldsymbol{\beta}_3, S_3)$ of order n_3 gives the duration of the correct service time distribution when the customer has gone through incorrect service initially. Let S_3^0 be such that $S_3\boldsymbol{e} + S_3^0 = \mathbf{0}$. Let $\mu_3' = \boldsymbol{\beta}_3(-S_3)^{-1}\mathbf{e}$ be the mean of this PH-representation.

- Under the above assumptions the service time of a customer can be modeled as a PH-distribution with representation $(\boldsymbol{\beta}, S)$ of order $n = n_1 + n_2 + n_3$, where

$$\boldsymbol{\beta} = (p\boldsymbol{\beta}_1, q\boldsymbol{\beta}_2, \mathbf{0}) \tag{5.2}$$

$$S = \begin{pmatrix} S_1 & O & O \\ O & S_2 & \hat{S}_2^0\boldsymbol{\beta}_3 \\ O & O & S_3 \end{pmatrix}.$$

Let S^0 be such that $S\boldsymbol{e} + S^0 = \mathbf{0}$ and $S^0 = \begin{bmatrix} S_1^0 & S_2^0 & S_3^0 \end{bmatrix}^T$.

Let $N(t)$ be the number of customers in the system, $N^*(t)$ the nature of service going on− whether direct admission to required/ undesired or one that came from undesired service− designated by 1,2 and 3 respectively, $S(t)$ the phase of service and $A(t)$ the phase of arrival at time t. With these the process $\{(N(t), N^*(t), S(t), A(t)), t \geq 0\}$ is a continuous time Markov chain with state space $\Omega = \{\underline{0}, \underline{1}, \underline{2}, ...\}$, where

$$\underline{0} = \{(0, r)/1 \leq r \leq m\}$$

(in the level zero we need consider only the phase of arrival) and

$$\underline{i} = \{(i, j, k, r)/i \geq 1, 1 \leq j \leq 3, 1 \leq k \leq n_j, 1 \leq r \leq m\}, i \geq m.$$

Thus the infinitesimal generator of this $CTMC$ is a $LIQBD$ of the form

$$\mathbf{Q} = \begin{pmatrix} D_0 & A_{01} & & & \\ A_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}, \tag{5.3}$$

where $A_{01} = \boldsymbol{\beta} \otimes D_1, A_{10} = S^0 \otimes I_m, A_0 = I_n \otimes D_1, A_1 = S \oplus D_0, A_2 = S^0 \boldsymbol{\beta} \otimes I_m$.

5.1.1 Stability condition

Consider $\mathbf{A}(= A_0 + A_1 + A_2)$, the generator matrix of the Markov chain corresponding to the phase changes.

$$\begin{aligned} \mathbf{A} &= (S + S^0 \boldsymbol{\beta}) \oplus D \\ &= \begin{pmatrix} (pS_1^0 \boldsymbol{\beta}_1 + S_1) \oplus D & qS_1^0 \boldsymbol{\beta}_2 \otimes I_m & O \\ pS_2^0 \boldsymbol{\beta}_1 \otimes I_m & (qS_2^0 \boldsymbol{\beta}_2 + S_2) \oplus D & \hat{S}_2^0 \boldsymbol{\beta}_3 \otimes I_m \\ pS_3^0 \boldsymbol{\beta}_1 \otimes I_m & qS_3^0 \boldsymbol{\beta}_2 \otimes I_m & S_3 \oplus D \end{pmatrix}. \end{aligned}$$

Let $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3)$ be the steady-state probability vector of $(S + S^0 \boldsymbol{\beta})$. Then

$$\boldsymbol{\pi}(S + S^0 \boldsymbol{\beta}) = \mathbf{0} \text{ and } \boldsymbol{\pi}\mathbf{e} = 1. \tag{5.4}$$

From the relation $\boldsymbol{\pi}(S + S^0 \boldsymbol{\beta}) = \mathbf{0}$ we have

$$\boldsymbol{\pi}_1(pS_1^0 \boldsymbol{\beta}_1 + S_1) + \boldsymbol{\pi}_2 pS_2^0 \boldsymbol{\beta}_1 + \boldsymbol{\pi}_3 pS_3^0 \boldsymbol{\beta}_1 = \mathbf{0}, \tag{5.5}$$

$$\boldsymbol{\pi}_1 qS_1^0 \boldsymbol{\beta}_2 + \boldsymbol{\pi}_2(qS_2^0 \boldsymbol{\beta}_2 + S_2) + \boldsymbol{\pi}_3 qS_3^0 \boldsymbol{\beta}_2 = \mathbf{0}, \tag{5.6}$$

$$\boldsymbol{\pi}_2 \hat{S}_2^0 \boldsymbol{\beta}_3 + \boldsymbol{\pi}_3 S_3 = \mathbf{0}. \tag{5.7}$$

Multiplying equation (5.7) by $\mathbf{e}$ on right hand side we get

$$\boldsymbol{\pi}_3 S_3^0 = \boldsymbol{\pi}_2 \hat{S}_2^0. \tag{5.8}$$

Putting this in equation (5.5) yields

$$\boldsymbol{\pi}_1 S_1^0 = -\frac{p}{q}\boldsymbol{\pi}_2 S_2 \mathbf{e}. \tag{5.9}$$

Substitute relations (5.8) and (5.9) in equation (5.6) to get

$$\boldsymbol{\pi}_2(S_2^0\boldsymbol{\beta}_2 + \hat{S}_2^0\boldsymbol{\beta}_2 + S_2) = \mathbf{0}.$$

This implies, for an arbitrary constant c,

$$\boldsymbol{\pi}_2 = c\boldsymbol{\beta}_2(-S_2)^{-1}. \tag{5.10}$$

Then from (5.9) we get

$$\boldsymbol{\pi}_1 = \frac{cp}{q}\boldsymbol{\beta}_1(-S_1)^{-1}. \tag{5.11}$$

Let $\delta = \boldsymbol{\beta}_2(-S_2)^{-1}\hat{S}_2^0$ be the probability that a customer, starting with incorrect service, leaves the system after getting correct service. Then the relation (5.8) gives

$$\boldsymbol{\pi}_3 = c\delta\boldsymbol{\beta}_3(-S_3)^{-1}. \tag{5.12}$$

From the normalizing condition $\boldsymbol{\pi}\mathbf{e} = 1$, the value of c is computed as

$$c = \left[\frac{p}{q}\mu_1' + \mu_2' + \delta\mu_3'\right]^{-1}. \tag{5.13}$$

Now from (5.1) and (5.4) we get the steady-state probability vector of A as

$$\hat{\boldsymbol{\pi}} = \boldsymbol{\pi} \otimes \boldsymbol{\eta}.$$

Theorem 5.1.1. *The stability of the system is given by*

$$\lambda < (\boldsymbol{\pi} \otimes \boldsymbol{\eta})(S^0\boldsymbol{\beta} \otimes I_m)\mathbf{e}. \tag{5.14}$$

Proof. The queueing system under study with the *LIQBD* type generator given in (5.3) is stable if and only if rate of left drift is less than the rate of right drift, that is,

$$\hat{\boldsymbol{\pi}} A_0 \mathbf{e} < \hat{\boldsymbol{\pi}} A_2 \mathbf{e}. \tag{5.15}$$

The left drift rate is $\hat{\boldsymbol{\pi}}(\ I_n \otimes D_1)\mathbf{e}$ which when simplified reduces to λ. Now, the right drift rate is $(\boldsymbol{\pi} \otimes \boldsymbol{\eta})(S^0 \boldsymbol{\beta} \otimes\ I_m)\mathbf{e}$.

Let $\rho = \dfrac{\lambda}{(\boldsymbol{\pi} \otimes \boldsymbol{\eta})(S^0 \boldsymbol{\beta} \otimes\ I_m)\mathbf{e}}$. Then from (5.14), we have $\rho < 1$. □

5.1.2 Steady-state probability vector

A brief outline for the computation of the stationary probability vector of the system is as follows. Let $\boldsymbol{x}$ denote the steady-state probability vector of the generator $\mathbf{Q}$. Then

$$\boldsymbol{x}\mathbf{Q} = \ \mathbf{0} \text{ and } \boldsymbol{x}\mathbf{e} = 1. \tag{5.16}$$

Assuming that the stability condition (5.14) holds and partitioning $\boldsymbol{x}$ as $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, ...)$, we obtain

$$\boldsymbol{x}_n = \boldsymbol{x}_1 R^{n-1}, n \geq 1$$

where R is the minimal non negative solution to the matrix quadratic equation

$$R^2 A_2 + RA_1 + A_0 = \mathrm{O}.$$

The two boundary equations involving $\boldsymbol{x}_0$ are

$$\boldsymbol{x}_0 D_0 + \boldsymbol{x}_1 A_{10} = \mathbf{0},$$

$$\boldsymbol{x}_0 A_{01} + \boldsymbol{x}_1 [A_1 + RA_2] = \mathbf{0}.$$

These together with the normalizing condition in (5.16) gives

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 V \text{ where } V = -A_{01}[A_1 + RA_2]^{-1}$$

$$\boldsymbol{x}_0[I + V(I - R)^{-1}]\mathbf{e} = 1.$$

To see how the system performs, it is instructive to define $\boldsymbol{y} = \sum_{i=1}^{\infty} \boldsymbol{x}_i$. Then $\boldsymbol{y} = (\boldsymbol{y}_1 \ \ \boldsymbol{y}_2 \ \ \boldsymbol{y}_3)$ where the $\boldsymbol{y}_i$'s indicate status of the customer in service.

5.1.3 System performance measures

1. Probability that system is idle, $P_{idle} = \boldsymbol{x}_0\mathbf{e} = 1-\rho$.

2. Rate of loss of customers, $R_{loss} = \boldsymbol{y}_2 S_2^0 = \lambda q(1-\delta)$.

3. Probability that a customer is lost, $P_{loss} = q(1-\delta)$.

4. Mean number of customers in the system, $\mu_{NS} = \sum\limits_{i=1}^{\infty} i\boldsymbol{x}_i\boldsymbol{e}$.

5. Mean number of customers in the queue, $\mu_{NQ} = \sum\limits_{i=2}^{\infty}(i-1)x_i\boldsymbol{e}$.

6. Probability that the server is serving in required mode,
   $P_C = \boldsymbol{y}_1\boldsymbol{e} + \boldsymbol{y}_3\boldsymbol{e} = \rho - \lambda q\mu_2'$.

7. Probability that the server is serving in undesired mode,
   $P_I = \boldsymbol{y}_2\boldsymbol{e} = \lambda q\mu_2'$.

8. Rate at which customers leave with required service starting in desired service mode, $R_C = \boldsymbol{y}_1 S_1^0 = \lambda p$.

9. Rate at which customers leave with correct service starting with undesired service, $R_I = \boldsymbol{y}_3 S_3^0 = \lambda q\delta$.

10. Expected waiting time in the system $W_S = \dfrac{\mu_{NS}}{\lambda}$

5.2 Poisson arrival and phase type service

In this section we analyze the system when arrival follows Poisson process. Service time is phase type distributed as in previous section. Then $\{(N(t), N^*(t), S(t)), t \geq 0\}$ (see section 5.1) is a continuous time Markov chain with state space $\{0, \underline{1}, \underline{2}, \ldots\}$ where $\underline{i} = \{(i,j,k)/1 \leq j \leq 3, 1 \leq k \leq n_j\}$ for $i \geq 1$.

Thus the infinitesimal generator is of the form

$$\mathbf{Q}' = \begin{pmatrix} -\lambda & \lambda\boldsymbol{\beta} & & & & \\ S^0 & S-\lambda I & \lambda I & & & \\ & S^0\boldsymbol{\beta} & S-\lambda I & \lambda I & & \\ & & S^0\boldsymbol{\beta} & S-\lambda I & \lambda I & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Theorem 5.2.1. *The system is stable if and only if* $\rho' < 1$ *where*

$$\rho' = \lambda\left[p\mu_1' + q(\mu_2' + \delta\mu_3')\right].$$

Proof. From the relation (5.15) we have $\lambda < \boldsymbol{\pi} S^0\boldsymbol{\beta}\mathbf{e}$ where $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3)$ (with $\boldsymbol{\pi}_i$'s as given in (5.10)-(5.12)) is the steady-state probability vector of $S + S^0\beta$. The right drift $\boldsymbol{\pi} S^0\beta\mathbf{e} = \sum_{i=1}^{3} \boldsymbol{\pi_i} S_i^0$.

Multiplying (5.5) by $\mathbf{e}$ on right hand side we get

$$\begin{aligned} \sum_{i=1}^{3} \boldsymbol{\pi_i} S_i^0 &= \frac{c}{q} - \frac{1}{p}\boldsymbol{\pi_i} S_1\mathbf{e} \\ &= \frac{1}{p}\frac{cp}{q}\boldsymbol{\beta_1}\mathbf{e} \text{ from (5.11)} \\ &= \frac{c}{q} \end{aligned}$$

where c is given in (5.13). Hence the condition for system stability is given by

$$\lambda < \frac{1}{p\mu_1' + q(\mu_2' + \delta\mu_3')}$$

□

The generator matrix corresponding to the phase changes is $S + S^0\beta$ and the stationary probability vector is $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3)$.

Theorem 5.2.2. *The steady-state probability vector* $\mathbf{x} = (x_0, \mathbf{x}_1, \mathbf{x}_2, \cdots)$ *of* $\boldsymbol{Q}'$ *is given by*

$$x_0 = 1 - \rho', \quad \mathbf{x}_i = (1-\rho')\boldsymbol{\beta} R^i, \ i \geq 1,$$

where R *is*

$$R = \lambda \begin{bmatrix} \lambda I - \lambda p \mathbf{e}\boldsymbol{\beta}_1 - S_1 & -\lambda q \mathbf{e}\boldsymbol{\beta}_2 & 0 \\ -\lambda p \mathbf{e}\boldsymbol{\beta}_1 & \lambda I - \lambda q \mathbf{e}\boldsymbol{\beta}_2 - S_2 & -\hat{S}_2^0 \boldsymbol{\beta}_3 \\ -\lambda p \mathbf{e}\boldsymbol{\beta}_1 & -\lambda q \mathbf{e}\boldsymbol{\beta}_2 & \lambda I - S_3 \end{bmatrix}^{-1}. \tag{5.17}$$

Proof. Let $\mathbf{x}$ be the steady-state probability vector of $\mathbf{Q}'$. Then $\mathbf{x}\mathbf{Q}' = \mathbf{0}$ and $\mathbf{x}\mathbf{e} = 1$.

The steady-state equations are given by

$$-\lambda x_0 + \mathbf{x}_1 \boldsymbol{S}^0 = 0, \tag{5.18}$$

$$\lambda x_0 \boldsymbol{\beta} + \mathbf{x}_1(S - \lambda I) + \mathbf{x}_2 \boldsymbol{S}^0 \boldsymbol{\beta} = \mathbf{0}, \tag{5.19}$$

$$\lambda \mathbf{x}_{i-1} + \mathbf{x}_i(S - \lambda I) + \mathbf{x}_{i+1} \boldsymbol{S}^0 \boldsymbol{\beta} = \mathbf{0}, i \geq 2. \tag{5.20}$$

From (5.18) we have

$$\mathbf{x}_1 \boldsymbol{S}^0 = \lambda x_0. \tag{5.21}$$

Multiplying equations (5.19) and (5.20) by the column vector $\mathbf{e}$ on the right hand side leads to

$$\mathbf{x}_{i+1} \boldsymbol{S}^0 = \lambda \mathbf{x}_i \mathbf{e} \text{ for } i \geq 1.$$

Writing $\mathcal{B} = \mathbf{e}.\boldsymbol{\beta}$ we get $\mathbf{x}_{i+1} \boldsymbol{S}^0 \boldsymbol{\beta} = \lambda \mathbf{x}_i \mathcal{B}$ for $i \geq 1$. Then from (5.19) and (5.20) we obtain

$$\mathbf{x}_1(\lambda I - \lambda \mathcal{B} - S) = \lambda x_0 \boldsymbol{\beta} \tag{5.22}$$

and

$$\mathbf{x}_i(\lambda I - \lambda \mathcal{B} - S) = \lambda \mathbf{x}_{i-1}, \text{ for } i \geq 2.$$

Denoting $(\lambda I - \lambda \mathcal{B} - S)$ by $\mathcal{K}$, relation (5.22) takes the form $\mathbf{x}_1 = \lambda x_0 \boldsymbol{\beta} \mathcal{K}^{-1}$, provided $\mathcal{K}$ is invertible. We now prove the non singularity of $\mathcal{K}$.

Let the vector $\mathbf{u}$ be in the left kernal of $\mathcal{K}$. Then

$$\lambda \mathbf{u} - \mathbf{u}S - \lambda(\mathbf{ue})\boldsymbol{\beta} = 0. \tag{5.23}$$

Suppose $\mathbf{ue} = 0$. Then (5.23) reduces to $\mathbf{u}(\lambda I - S) = 0$. But $(\lambda I - S)$ is nonsingular and hence $\mathbf{u} = 0$.

If $\mathbf{ue} \neq 0$, normalize $\mathbf{u}$ by setting $\mathbf{ue} = 1$. Post multiplying (5.23) by $\mathbf{e}$ gives

$$\mathbf{u}\boldsymbol{S}^0 = 0. \tag{5.24}$$

Substituting for $\mathbf{ue}$, (5.23) reduces to $\mathbf{u} = \lambda \boldsymbol{\beta}(\lambda I - S)^{-1}$.

From (5.24) we have

$$\lambda \boldsymbol{\beta}(\lambda I - S)^{-1}\boldsymbol{S}^0 = 0. \tag{5.25}$$

But $\boldsymbol{\beta}(\lambda I - S)^{-1}\boldsymbol{S}^0$ is the Laplace-Stieltjes transform at $s = \lambda > 0$, of the probability distribution $F(t) = 1 - \boldsymbol{\beta}\exp(St)\mathbf{e}$ for $t \geq 0$. Therefore (5.25) cannot hold and hence $\mathbf{u} = 0$. Thus $\mathcal{K}$ is nonsingular.

The irreducibility of the representation $(\boldsymbol{\beta}, S)$ leads to the irreducibility of the stable $\mathcal{K}$, so that the matrix R in (5.17) is positive.

We have $sp(R) < 1$, if $\rho' < 1$. Therefore the quantity x_0 is given by the normalizing equation

$$x_0 + x_0 \boldsymbol{\beta} R(I - R)^{-1}\mathbf{e} = 1.$$

Substitution for R leads to

$$x_0 - \lambda x_0 \boldsymbol{\beta}(\lambda \mathcal{B} + S)^{-1}\mathbf{e} = 1. \tag{5.26}$$

The inverse of $(\lambda\mathcal{B}+S)$ is calculated as

$$\begin{aligned}
(\lambda\mathcal{B}+S)^{-1} &= S^{-1}\left(I+\lambda\mathcal{B}S^{-1}\right)^{-1} \\
&= S^{-1}\sum_{n=0}^{\infty}(-1)^n\lambda^n\left(\mathcal{B}S^{-1}\right)^n \\
&= S^{-1}\left[I-\lambda\left[\sum_{n=0}^{\infty}(-1)^n\lambda^n\left(\mathcal{B}S^{-1}\right)^n\right]\mathcal{B}S^{-1}\right] \\
&= S^{-1}\left[I-\lambda\sum_{n=0}^{\infty}\rho'^n\mathcal{B}S^{-1}\right] \\
&= S^{-1}\left[I-\lambda\left(1-\rho'\right)^{-1}\mathcal{B}S^{-1}\right].
\end{aligned}$$

From (5.26) we have

$$\begin{aligned}
x_0-\lambda x_0\boldsymbol{\beta}(\lambda\mathcal{B}+S)^{-1}\mathbf{e} &= x_0-\lambda x_0\boldsymbol{\beta}\left[S^{-1}\left(I-\lambda\left(1-\rho'\right)^{-1}\mathcal{B}S^{-1}\right)\right]\mathbf{e} \\
&= x_0-\lambda x_0\boldsymbol{\beta}S^{-1}\mathbf{e}+\lambda^2x_0(1-\rho)^{-1}\boldsymbol{\beta}S^{-1}\mathcal{B}S^{-1}\mathbf{e} \\
&= x_0+\rho'x_0+\rho'^2(1-\rho')x_0 \\
&= (1-\rho')x_0=1,
\end{aligned}$$

so that $x_0=(1-\rho')$. □

Letting $\boldsymbol{y}=\sum_{i=1}^{\infty}\mathbf{x}_i$, it is obtained that $\boldsymbol{y}=\rho'\boldsymbol{\pi}$. In the sequel partition $\boldsymbol{y}=(\boldsymbol{y}_1,\boldsymbol{y}_2,\boldsymbol{y}_3)$, so that $\boldsymbol{y}_i=\rho'\boldsymbol{\pi}_i,\ 1\le i\le 3$.

5.3 Poisson arrival with exponentially distributed service time

In this section we consider customers to arrive according to a Poisson process with rate λ and desired (correct) service time follows exponential distribution with parameter μ ($\mu'_1=\mu'_3=\mu$) and the undesired (incorrect) part of service following phase type distribution with representation $(\boldsymbol{\beta}_2,S_2)$ of order n_2 (see section 5.1). Let $N(t)$ be the number of customers in the system, $N^*(t)$ the type of service and $S(t)$ the phase of service at time t. $S(t)$ assumes a value between 1 and n_2(including both) if server is in undesired phase of service, otherwise 0 or n_2+1 according as a desired service going on for a customer admitted directly or from undesired state. Then $\{(N(t), N^*(t), S(t)), t \geq 0\}$ is a continuous time Markov chain with state space $\{0, \underline{1}, \underline{2}, ...\}$ where

$$\underline{i} = \{(i,1,0),(i,3,n_2+1)\} \cup \{(i,2,j)/1 \leq j \leq n_2\} \text{ for } i \geq 1.$$

Thus the infinitesimal generator is of the form

$$\mathbf{Q} = \begin{array}{c} \\ 0 \\ \underline{1} \\ \underline{2} \\ \vdots \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & \underline{1} & \underline{2} & \underline{3} & .. & .. \end{array} \\ \begin{pmatrix} -\lambda & \mathbf{b}_0 & & & \\ \mathbf{c}_0 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \\ & & & & \end{pmatrix} \end{array}$$

where

$$\mathbf{b}_0 = \lambda(p, q\boldsymbol{\beta}_2, 0), \quad \mathbf{c}_0 = \begin{pmatrix} \mu \\ \hat{S}_2^0 \\ \mu \end{pmatrix}, A_0 = \lambda I$$

$$A_1 = \begin{pmatrix} -\lambda-\mu & \mathbf{0} & 0 \\ \mathbf{0} & S_2 - \lambda I & \hat{S}_2^0 \\ 0 & \mathbf{0} & -\lambda-\mu \end{pmatrix}, \quad A_2 = \begin{pmatrix} \mu p & \mu q \boldsymbol{\beta}_2 & 0 \\ p\ S_2^0 & q\ S_2^0 \boldsymbol{\beta}_2 & \mathbf{0} \\ \mu p & \mu q \boldsymbol{\beta}_2 & 0 \end{pmatrix}$$

with $S\mathbf{e} + \ S_2^0 + \ \hat{S}_2^0 = \mathbf{0}$.

5.3.1 Stability condition

Consider $A = A_0 + A_1 + A_2$

$$= \begin{pmatrix} -\mu q & \mu q \boldsymbol{\beta}_2 & 0 \\ p\ S_2^0 & S_2 + q\ S_2^0 \boldsymbol{\beta}_2 & \hat{S}_2^0 \\ \mu p & \mu q \boldsymbol{\beta}_2 & -\mu \end{pmatrix},$$

the generator matrix of the Markov chain corresponding to the phase changes. Let
$\Pi = (\pi_0, \hat{\boldsymbol{\pi}}, \pi_{r+1})$ be the steady-state probability matrix of A. Solving the relations

$$\Pi A = \mathbf{0}, \quad \Pi \mathbf{e} = 1 \tag{5.27}$$

we obtain

$$-\mu q \, \pi_0 + p\hat{\boldsymbol{\pi}} \;\; S_2^0 + \mu p \, \pi_{r+1} = 0 \tag{5.28}$$

$$\mu q \pi_0 \boldsymbol{\beta}_2 + \hat{\boldsymbol{\pi}}(\; S_2 + q \; S_2^0 \, \boldsymbol{\beta}_2) + \mu q \; \pi_{r+1}\boldsymbol{\beta}_2 = \mathbf{0} \tag{5.29}$$

$$\hat{\boldsymbol{\pi}} \;\; \hat{S}_2^0 - \mu \pi_{r+1} = 0. \tag{5.30}$$

From equations (5.28) and (5.30),

$$\mu q \; \pi_0 = p \left(\hat{\boldsymbol{\pi}} \;\; S_2^0 + \hat{\boldsymbol{\pi}} \;\; \hat{S}_2^0 \right). \tag{5.31}$$

This together with (5.29) gives

$$\hat{\boldsymbol{\pi}} \; \left(\; S_2 + \; S_2^0 \, \boldsymbol{\beta}_2 + \;\; \hat{S}_2^0 \, \boldsymbol{\beta}_2 \right) = 0$$

so that

$$\hat{\boldsymbol{\pi}} = c \, \boldsymbol{\beta}_2 (- \; S_2)^{-1}, \tag{5.32}$$

c being a constant and is computed from the normalizing condition. Let δ be the probability that a customer getting correct service following one or several incorrect services, and η the probability of staying back in incorrect services. Then

$$\delta = \boldsymbol{\beta}_2 \, (- \; S_2)^{-1} \;\; \hat{S}_2^0$$

and

$$\eta = \left(\boldsymbol{\beta}_2 \, (- \; S_2)^{-1} \, \mathbf{e} \right)^{-1}.$$

Then the probability that a customer leaves the system without getting required service is

$$1 - \delta = \boldsymbol{\beta}_2 \, (- \; S_2)^{-1} \;\; S_2^0$$

and the mean time a customer stay back in incorrect services is

$$\frac{1}{\eta} = \left(\boldsymbol{\beta}_2(-\ S_2)^{-1}\mathbf{e}\right).$$

The normalizing equation is

$$\pi_0 + \hat{\boldsymbol{\pi}}\ \mathbf{e} + \pi_{r+1} = 1.$$

Substituting for the components of Π which are now computed as

$$\pi_0 = \frac{pc}{\mu q},\ \hat{\boldsymbol{\pi}}\ \mathbf{e} = \frac{c}{\eta},\ \pi_{r+1} = \frac{c\delta}{\mu}$$

we get

$$\frac{pc}{\mu q} + \frac{c}{\eta} + \frac{c\delta}{\mu} = 1$$

which shows

$$c = \frac{\mu q \eta}{p\eta + \mu q + \delta q \eta}.$$

Theorem 5.3.1. *The stability of the system is given by* $\lambda < \frac{1}{q}\ c$.

Proof. The condition for the stability of the system is $\Pi A_0\mathbf{e} < \Pi A_2\mathbf{e}$. Simplification gives $\Pi A_0\mathbf{e} = \lambda$. Now $A_2\mathbf{e} = \left(\mu, S_2^0, \mu\right)^T$. Therefore $\Pi A_2\mathbf{e} = \mu\pi_0 + \hat{\boldsymbol{\pi}}\left(\ S_2^0 +\ \hat{S}_2^0\right)$. Substituting for $\mu\pi_0$, right hand side becomes $\frac{1}{q}\hat{\boldsymbol{\pi}}\left(\ S_2^0 +\ \hat{S}_2^0\right)$. Using equation(5.32) and the fact that $(\ S_2)^{-1}(\ S_2^0 +\ \hat{S}_2^0) = \mathbf{e}$, the result follows. Hence the system is stable if and only if $\rho < 1$, where

$$\rho = \lambda\ \frac{q}{c}. \tag{5.33}$$

□

5.3.2 Steady-state probability vector

Let the steady-state probability vector $\mathbf{x} = (x^*, \boldsymbol{x}(1), \boldsymbol{x}(2), ...)$ of $\mathbf{Q}$ be such that $\mathbf{xQ} = \mathbf{0}, \mathbf{xe} = 1$. Partitioning gives $\boldsymbol{x}(i) = \left(x_0(i), \widehat{x}(i), x_{r+1}(i)\right)$. The relation

$\mathbf{xQ} = 0$ gives the following system of equations:

$$-\lambda x^* + \boldsymbol{x}(1)\mathbf{c}_0 = 0, \tag{5.34}$$

$$x^*\mathbf{b}_0 + \boldsymbol{x}(1)A_1 + \boldsymbol{x}(2)A_2 = \mathbf{0}, \tag{5.35}$$

$$\text{For } i \geq 1, \quad \boldsymbol{x}(i-1)A_0 + \boldsymbol{x}(i)A_1 + \boldsymbol{x}(i+1)A_2 = \mathbf{0}. \tag{5.36}$$

From the matrix geometric structure we obtain

$$\boldsymbol{x}(i) = \boldsymbol{x}(1)R^{i-1}, \ i \geq 1$$

where R is the minimal non negative solution to the matrix quadratic equation

$$R^2A_2 + RA_1 + A_0 = \mathrm{O}. \tag{5.37}$$

Equation (5.34) shows

$$x^* = \frac{1}{\lambda}\boldsymbol{x}(1)\mathbf{c}_0.$$

Equation (5.35) together with normalizing condition gives

$$x^*\mathbf{b}_0 + \boldsymbol{x}(1)\left(A_1 + RA_2\right) = \mathbf{0}$$

$$\text{subject to} \qquad x^*\mathbf{e} + \boldsymbol{x}(1)\left(I - R\right)^{-1}\mathbf{e} = 1.$$

Substituting for x^* we get

$$\boldsymbol{x}(1)\left(A_1 + RA_2 + \frac{1}{\lambda}\mathbf{c}_0\mathbf{b}_0\right) = \mathbf{0}$$

$$\text{subject to} \qquad \boldsymbol{x}(1)\left(\frac{1}{\lambda}\mathbf{c}_0 + \left(I - R\right)^{-1}\mathbf{e}\right) = 1.$$

But $\mathbf{c}_0\mathbf{b}_0 = \lambda A_2$ which implies

$$\boldsymbol{x}(1)\left(A_1 + RA_2 + A_2\right) = \mathbf{0}$$

$$\text{subject to} \qquad \boldsymbol{x}(1)\left(\frac{1}{\lambda}\mathbf{c}_0 + \left(I - R\right)^{-1}\mathbf{e}\right) = 1.$$

Computation of R

R can computed explicitly along the following lines.
We have

$$A_2 = \begin{pmatrix} \mu p & \mu q \boldsymbol{\beta}_2 & 0 \\ p\ S_2^0 & q\ S_2^0 \boldsymbol{\beta}_2 & \mathbf{0} \\ \mu p & \mu q \boldsymbol{\beta}_2 & 0 \end{pmatrix} = \begin{bmatrix} \mu \\ S_2^0 \\ \mu \end{bmatrix} \begin{bmatrix} p & q\boldsymbol{\beta}_2 & 0 \end{bmatrix}$$

so that

$$A_2 \mathbf{e} = \begin{bmatrix} \mu \\ S_2^0 \\ \mu \end{bmatrix} = \mathbf{c}_0$$

Also from the relation $RA_2\mathbf{e} = A_0\mathbf{e}$, we obtain

$$RA_2\mathbf{e} = \lambda \mathbf{e} \tag{5.38}$$

Now,

$$R^2 A_2 = R^2 \begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} \begin{pmatrix} p & q\boldsymbol{\beta}_2 & 0 \end{pmatrix} = R^2 A_2 \mathbf{e} \begin{pmatrix} p & q\boldsymbol{\beta}_2 & 0 \end{pmatrix}$$

Substituting for RA_2 from (5.38), we get

$$R^2 A_2 = R\lambda \mathbf{e} \begin{pmatrix} p & q\boldsymbol{\beta}_2 & 0 \end{pmatrix}$$

Therefore equation (5.37) gives

$$\lambda R\mathbf{e} \begin{pmatrix} p & q\boldsymbol{\beta}_2 & 0 \end{pmatrix} + RA_1 + \lambda I = \mathrm{O}$$

This gives

$$R = \lambda \begin{pmatrix} \mu + \lambda q & -\lambda q \boldsymbol{\beta}_2 & 0 \\ -\lambda p \mathbf{e} & \lambda I - \lambda q \mathbf{e} \boldsymbol{\beta}_2 - \ S_2 & -\ \hat{S}_2^0 \\ -\lambda p & -\lambda q \boldsymbol{\beta}_2 & \lambda + \mu \end{pmatrix}^{-1}$$

Lemma 5.3.1. $x^* = 1 - \rho$ so that $\boldsymbol{x}(1)\,(I-R)^{-1}\,\mathbf{e} = \rho$

Proof. Multiplying by $\mathbf{e}$ on the right side of equation (5.35) and simplifying we get the relation

$$\lambda x^* + \boldsymbol{x}(1)\begin{pmatrix} -\lambda-\mu \\ S_2 - \lambda I + \hat{S}_2^0 \\ -\lambda-\mu \end{pmatrix} + \boldsymbol{x}(2)\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} = 0. \tag{5.39}$$

Equation (5.34) gives

$$\lambda x^* = \boldsymbol{x}(1)\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix}. \tag{5.40}$$

Putting this in (5.39) the following relation is obtained.

$$\boldsymbol{x}(2)\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} = \lambda \boldsymbol{x}(1)\mathbf{e}. \tag{5.41}$$

□

Multiplying equation(5.36) on right side by $\mathbf{e}$ and recursive use of the relation results in

$$\boldsymbol{x}(i)\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} = \lambda \boldsymbol{x}(i-1)\mathbf{e} \quad \text{for } i \geq 3. \tag{5.42}$$

Adding (5.40), (5.41) and (5.42)

$$\sum_{i=1}^{\infty} \boldsymbol{x}(i)\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} = \lambda. \tag{5.43}$$

Adding the system of equations (5.36) with equation (5.35) and using the fact that
$x^*\mathbf{b}_0 = \boldsymbol{x}(1)A_2$ we get

$$\sum_{i=1}^{\infty} \boldsymbol{x}(i)\ A = 0.$$

But the relation(5.27) says

$$\sum_{i=1}^{\infty} \boldsymbol{x}(i)\ = d\ \Pi \quad \text{for some constant } c$$

which in turn gives

$$\sum_{i=1}^{\infty} \boldsymbol{x}(i)\ = (1 - x^*)\ \Pi.$$

Multiplying on the right side by $\begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix}$ and using the relation in (5.33)

$$\sum_{i=1}^{\infty} \boldsymbol{x}(i) \begin{pmatrix} \mu \\ S_2^0 \\ \mu \end{pmatrix} = (1 - x^*)\ \frac{\lambda}{\rho}. \tag{5.44}$$

The result follows from(5.43) and (5.44).

5.3.3 System performance measures

1. Probability that the system is idle, $P_0 = x^*$.
2. Rate of loss, $R_{loss} = \sum_{i=1}^{\infty} \widehat{x}(i)\ S_2^0 = \lambda q(1 - \delta)$.
3. Probability of loss, $P_{loss} = q(1 - \delta)$.
4. Mean number of customers in the system,
   $\mu_{ns} = \sum_{i=1}^{\infty} i\boldsymbol{x}(i)\mathbf{e} = \boldsymbol{x}(1)\,(I - R)^{-2}\,\mathbf{e}$.

5. Mean number of customers in the queue,
$\mu_{nq} = \sum_{i=1}^{\infty} (i-1)\boldsymbol{x}(i)\mathbf{e} = \boldsymbol{x}(1)\,(I-R)^{-2}\,\mathbf{e} - \boldsymbol{x}(1)\,(I-R)^{-1}\,\mathbf{e}.$

6. Probability that the server is busy serving in correct mode,
$P_c = \sum_{i=1}^{\infty} \boldsymbol{x}(i) \begin{pmatrix} 1 \\ \mathbf{0} \\ 1 \end{pmatrix} = \rho\,(\pi_0 + \pi_{r+1}) = \rho - \frac{\lambda q}{\eta}.$

7. Probability that the server is busy serving in incorrect mode,
$P_i = \sum_{i=1}^{\infty} \boldsymbol{x}(i) \begin{pmatrix} 0 \\ \mathbf{e} \\ 0 \end{pmatrix} = \rho\widehat{\pi}\mathbf{e} = \frac{\lambda q}{\eta}.$

5.4 An illustration

In this section we consider a queueing model consisting of two service stations-preliminary service and main service. Customers arrive to this system according to a MAP (Markovian Arrival Process) with representation (D_0, D_1) of order m. A customer, who is taken for service is directly selected for main service with probability p or to the preliminary service with probability $q(= 1-p)$. A threshold clock starts ticking if a customer enters to preliminary service. When the duration of preliminary service exceeds the threshold clock, the customer moves out of the system, else he goes to main service. The threshold clock follows exponential distribution with parameter ζ. Service time of the customers at these stations follow phase type distributions with representation $(\boldsymbol{\alpha}, S_P)$, $(\boldsymbol{\gamma}, S_M)$ and of order a, b respectively. Write $S_P^0 + \zeta\mathbf{e} = -S_P\mathbf{e}$ and $S_M^0 = -S_M\mathbf{e}$ where $\mathbf{e}$ is a column vector of 1's of appropriate order. Hence service time of a customer can be modeled as a phase type distribution with representation $(\boldsymbol{\xi}, \mathbf{U})$ of order $a + 2b$ such that $\mathbf{U}\mathbf{e} + \mathbf{U}^0 = \mathbf{0}$ where

$$\boldsymbol{\xi} = \begin{pmatrix} p\boldsymbol{\gamma} & q\boldsymbol{\alpha} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} S_M & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & S_P & S_P^0\boldsymbol{\gamma} \\ \mathbf{0} & \mathbf{0} & S_M \end{pmatrix}, \mathbf{U}^0 = \begin{pmatrix} S_M^0 \\ \zeta\mathbf{e} \\ S_M^0 \end{pmatrix}.$$

Let $N(t), N^*(t), S(t), A(t)$ denote respectively the number of customers in the system, nature of service, phase of service and phase of arrival at time t with

$$N^*(t) = \begin{cases} 1 & \text{main service} \\ 2 & \text{preliminary service} \\ 3 & \text{one that come from preliminary service} \end{cases} .$$

The process $\Omega = \{(N(T), N^*(t), S(t), A(t)), t \geq 0\}$ is a continuous time Markov chain with state space $\{(n, i, j, k)/i = 1, 3; 1 \leq j \leq b, 1 \leq k \leq m\} \cup \{(n, 2, j, k)/1 \leq j \leq a, 1 \leq k \leq m\}$ for $n \geq 1$.
Note that when $N(t) = 0$, the only other component in the state vector is $A(t)$. Thus the infinitesimal generator of Ω is of the form

$$Q^* = \begin{pmatrix} D_0 & A_{01} & & & \\ A_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where $A_{01} = \boldsymbol{\xi} \otimes D_1, A_{10} = \mathbf{U}^0 \otimes I_m, A_0 = I_{a+2b} \otimes D_1, A_1 = \mathbf{U} \oplus D_0, A_2 = \mathbf{U}^0\boldsymbol{\xi} \otimes I_m$.

The infinitesimal generator Q^* is of the same form as $\mathbf{Q}$ of the model described initially. Thus the analysis of the Markov chain with infinitesimal generator Q^* can be done in the same way as for $\mathbf{Q}$.

The significance of this model is as follows: customer arriving to a single server belong to two categories, though they join the same waiting line. While taking for service the category will be decided. Call them category 1 and category 2, respectively. Category 1 are qualified for the main service without undergoing preliminary service. However, category 2 have to be given the preliminary service before admitted to main service. However, if such customers do not get service in preliminary before realization of the timer (random clock), they get disqualified and so leave the system forever. On the other hand those among category 2, completing service successfully before timer realization in preliminary, are immediately admitted to main service. On completion of that service such customers leave the system.

Remark 5.4.1. In telecommunication it is this type of situation that is often encountered. Packages have to identify the server in idle state; then wait for a while. But in the mean time another message may get through, making the server busy. Then the customer (packet) under consideration has to go through a series of contention windows. These passages could be regarded as unwanted service. In case the process of going through contention windows exceeds a threshold time limit (time out/ clock realization), the message will not get served.

Remark 5.4.2. The problem discussed in Madan [46] and Medhi [48] could be arrived at from our model as follows. Suppose that we reverse the order of preliminary and main service, that is, main service first and preliminary (hereafter we call the second as optional) service next. Then after completion of main service, the customer asks for an optional service with probability $1 - q$ (this optional service time has exponential distribution in Madan [46]). This could be regarded as an instantaneous feedback as head of waiting line and get served according to a different distribution. With probability q, the customer leaves the system immediately after main service completion.

5.5 Numerical illustration

The following numerical illustration is based on the description in Section 2. We fix parameters $n_1 = 2, n_2 = 3, n_3 = 4, \boldsymbol{\beta}_1 = (0.4 \;\; 0.6)$,

$\boldsymbol{\beta}_2 = (0.3 \ \ 0.5 \ \ 0.2), \boldsymbol{\beta}_3 = (0.2 \ \ 0.3 \ \ 0.3 \ \ 0.2)$,

$$S_1 = \begin{bmatrix} * & 6 \\ 8 & * \end{bmatrix}, \mathbf{S}_1^0 = \begin{bmatrix} 7 \\ 8 \end{bmatrix} \text{ with } S_1\mathbf{e} + \mathbf{S}_1^0 = \mathbf{0},$$

$$S_2 = \begin{bmatrix} * & 5 & 5 \\ 6 & * & 6 \\ 5 & 7 & * \end{bmatrix}, \mathbf{S}_2^0 = \begin{bmatrix} 3 \\ 3 \\ 2 \end{bmatrix}, \hat{\mathbf{S}}_2^0 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \text{ with } S_2\mathbf{e} + \mathbf{S}_2^0 + \hat{\mathbf{S}}_2^0 = \mathbf{0},$$

$$S_3 = \begin{bmatrix} * & 7 & 8 & 9 \\ 6 & * & 7 & 7 \\ 6 & 6 & * & 6 \\ 8 & 7 & 6 & * \end{bmatrix}, \mathbf{S}_3^0 = \begin{bmatrix} 6 \\ 7 \\ 8 \\ 9 \end{bmatrix} \text{ with } S_3\mathbf{e} + \mathbf{S}_3^0 = \mathbf{0}.$$

For the arrival process, we consider the following two sets of values for D_0 and D_1 as follows. The arrival processes labeled $MNCA$ and $MPCA$ respectively, have negative and positive correlation for two successive inter-arrival time with values -0.48891 and 0.48891. The standard deviation of the inter-arrival time of these two arrival processes are, respectively, 0.2819 and 0.2819.

1. MAP with negative correlation ($MNCA$):

$$D_0 = \begin{bmatrix} -5.0111 & 5.0111 & 0 \\ 0 & -5.0111 & 0 \\ 0 & 0 & -1128.75 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0.05011 & 0 & 4.96099 \\ 1117.4625 & 0 & 11.2875 \end{bmatrix}$$

2. MAP with positive correlation ($MPCA$):

$$D_0 = \begin{bmatrix} -5.0111 & 5.0111 & 0 \\ 0 & -5.0111 & 0 \\ 0 & 0 & -1128.75 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 4.96099 & 0 & 0.05011 \\ 11.2875 & 0 & 1117.4625 \end{bmatrix}$$

The output in Tables 5.1 and 5.2 are on expected lines. Note that P_{loss} decreases with increasing value of p. The value of $P_C(R_C)$ steadily increases with p and values of $P_I(R_I)$ and W_S decrease with increase in value of p, as expected.

p	P_{loss}	μ_{NS}	P_C	P_I	R_C	R_I	W_S
0.4	0.2136	7.5229	0.5242	0.3921	2	1.9320	1.5046
0.5	0.1780	4.9744	0.5483	0.3267	2.5	1.6100	0.9949
0.6	0.1424	3.6690	0.5724	0.2614	3	1.2880	0.7338
0.7	0.1068	2.8654	0.5965	0.1960	3.5	0.9660	0.5731
0.8	0.0712	2.3138	0.6206	0.1307	4	0.6440	0.4628
0.9	0.0356	1.9069	0.6447	0.0653	4.5	0.3220	0.3814

Table 5.1: Effect of p for $MNCA$

p	P_{loss}	μ_{NS}	P_C	P_I	R_C	R_I	W_S
0.4	0.2136	546.8179	0.5242	0.3921	2	1.9320	109.3646
0.5	0.1780	349.9587	0.5483	0.3267	2.5	1.6100	69.9924
0.6	0.1424	250.7699	0.5724	0.2614	3	1.2880	50.1545
0.7	0.1068	191.0008	0.5965	0.1960	3.5	0.9660	38.2005
0.8	0.0712	151.0402	0.6206	0.1307	4	0.6440	30.2083
0.9	0.0356	122.4351	0.6446	0.0653	4.5	0.3220	24.4873

Table 5.2: Effect of p for $MPCA$

The main comparison in Tables 5.1 and 5.2 is between values of μ_{NS} in $MNCA$ and $MPCA$. Both decrease with increase in value of p. However, $MNCA$ has much smaller values compared to their $MPCA$ counter parts. This indicates that positive correlation in the arrival process results in accumulation of large number of customers in the system.

5.6 $M/G/1$ Model

In this section we consider an $M/G/1$ system with two service stations – preliminary service and main service. Customers arrive to this system according to a Poisson process with rate λ. A customer, when taken for service, is directly selected for main service with probability p or to the preliminary service with probability q $(= 1 - p)$. A threshold clock starts ticking if a customer enters to preliminary service. When the duration of preliminary service exceeds the threshold clock, the customer moves out of the system, else he goes to main service. The threshold clock follows exponential distribution with parameter ζ. Here the service time, V_p, V_m of the preliminary and main services are independent having general distributions with distribution function $G_1(.), G_2(.)$, LST $G_1^*(.), G_2^*(.)$ respectively.

The (total) service time V of a unit is

$$V = \begin{cases} V_f & \text{with probability } q \bullet P(G_1(.) > \exp(\zeta)) \\ V_p & \text{with probability } q \bullet P(G_1(.) < \exp(\zeta)) \\ V_m & \text{with probability } p \end{cases}$$

where V_f is the duration of threshold clock realization. Thus

$$\begin{aligned} G(t) &= P(V \leq t) \\ &= q\left[\int_0^t \zeta e^{-\zeta u}(1 - G_1(u))du + \int_0^t e^{-\zeta u} G_1(u) dG_2(t-u)\right] + p\int_0^t dG_2(u) \end{aligned}$$

The LST $G^*(s)$ of V is given by

$$G^*(s) = \int_0^\infty e^{-st} dG(t).$$

Remark 5.6.1. This modelling closely resembles the protocol IEEE 802.11. This is so because of a message generated has to wait before checking for idle server; if server is busy it has to go through a series of contention windows and then look for idle server. In case this process takes longer duration than the life of message (before its significance is lost), then the message does not serve any purpose. In the opposite case it is transmitted before its expiry time.

Remark 5.6.2. Assume the random clock to be of infinite duration (ie., its rate of realization goes to zero). Now interchange the roles of preliminary and main services (in this case, we call the preliminary service, which is the second one now, as optional service). Invariably main service is given for all customers. Thus the main service is followed by an optional service to which customers, on completion of main service, proceed with probability q. Then our model reduces to Madan [46] with exponentially distributed optional service and to Medhi [48] in the case of arbitrarily distributed optional service time.

Transient solution

The supplementary variable technique (see Cox [19], Medhi [47]) could be used to get the transient solution. Suppose that the general distribution $G(x) = P(V \leq x)$ has the hazard function $h(x) = \dfrac{dG(x)}{1 - G(x)}$ and the probability density function of V is given by

$$g(x) = h(x)\exp\{-N(x)\}$$

where

$$N(x) = \int_0^x h(u)du; N(0) = 0 \text{ and } \frac{d}{dx}N(x) = h(x).$$

If V is the total service time, then $h(x)dx = P$(service will be completed in $(x, x + dx)$ given that service time exceeds x) and $E(V) = \int xg(x)dx = -G^{*(1)}(0)$.

The supplementary variable $X(t)$ considered is defined below. Let

$$
\begin{aligned}
N(t) &= \text{system size at time } t \\
X(t) &= \text{time already spent in service up to } t \text{ of a unit receiving service} \\
p_n(t) &= P(N(t) = n) \text{ with } p_0(0) = 1 \\
p_n(t,x)dx &= P(N(t) = n, x \le X(t) < x + dx),\ \ n \ge 1 \\
p_n(t) &= \int_0^\infty p_n(t,x)dx \\
Q(t,z) &= \sum_{n=0}^{\infty} p_n(t) z^n \\
Q(t,x,z) &= \sum_{n=1}^{\infty} p_n(t,x) z^n
\end{aligned}
$$

Now we have

$$p_0(t+\delta t) = [1 - \lambda\delta t + o(\delta t)]p_0(t) + \int_0^\infty p_1(t,x)h(x)dx\delta t.$$

As $\delta t \to 0$,

$$\frac{\partial}{\partial t} p_0(t) = -\lambda p_0(t) + \int_0^\infty p_1(t,x)h(x)dx. \tag{5.45}$$

For $\delta x > 0$,

$$p_1(t+\delta t, x+\delta x) = [1 - \lambda\delta t + o(\delta t)][1 - h(x)\delta x + o(\delta x)]p_1(t,x).$$

Subtracting and adding a term $p_1(t, x+\delta x)$ to the LHS, then dividing by $\delta t(\delta x)$ and taking as $\delta t \to 0(\delta x \to 0)$, we get

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) p_1(t,x) = -(\lambda + h(x))p_1(t,x). \tag{5.46}$$

For $n \geq 0$,

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) p_n(t,x) = -(\lambda + h(x))p_n(t,x) + \lambda p_{n-1}(t,x). \tag{5.47}$$

We have the following boundary conditions:

$$p_1(t,0) = \int_0^\infty p_2(t,x)h(x)dx + \lambda p_0(t) \tag{5.48}$$

and

$$p_n(t,0) = \int_0^\infty p_{n+1}(t,x)h(x)dx, \qquad n \geq 2. \tag{5.49}$$

Multiplying (5.47) by $z^n, n = 2,3,...$ and (5.46) by z, then adding all the terms we get

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) \sum_{n=1}^\infty p_n(t,x)z^n = -(\lambda + h(x)) \sum_{n=1}^\infty p_n(t,x) + \lambda \sum_{n=2}^\infty p_{n-1}(t,x)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right) Q(t,x,z) = -(\lambda - \lambda z + h(x))Q(t,x,z). \tag{5.50}$$

Now multiplying (5.49) by $z^n, n = 2,3,...$ and (5.48) by z, then adding the terms we have

$$Q(t,0,z) = \int_0^\infty \left(\sum_{n=1}^\infty p_{n+1}(t,x)z^n\right) h(x)dx + \lambda z p_0(t). \tag{5.51}$$

Now $\int_0^\infty \left(\sum_{n=1}^\infty p_{n+1}(t,x)z^n\right) h(x)dx$

$$= \int_0^\infty \left(\frac{1}{z}\right) \sum_{n=1}^{\infty} p_{n+1}(t,x) z^{n+1} h(x) dx$$

$$= \int_0^\infty \left(\frac{1}{z}\right) \left[\sum_{n=1}^{\infty} p_n(t,x) z^n - p_1(t,x) z\right] h(x) dx$$

$$= \left(\frac{1}{z}\right) \int_0^\infty \left[Q(t,x,z) - p_1(t,x) z\right] h(x) dx$$

$$= \left(\frac{1}{z}\right) \left[\int_0^\infty Q(t,x,z) h(x) dx - z(p_0'(t) + \lambda p_0(t))\right] \text{ by (5.45)}$$

Thus (5.51) reduces to

$$\begin{aligned} Q(t,0,z) &= \left(\frac{1}{z}\right) \left[\int_0^\infty Q(t,x,z) h(x) dx - z(p_0'(t) + \lambda p_0(t))\right] + \lambda z p_0(t) \\ &= \left(\frac{1}{z}\right) \left[\int_0^\infty Q(t,x,z) h(x) dx - z(p_0'(t) + \lambda p_0(t)) + \lambda z^2 p_0(t)\right] \end{aligned}$$

$$zQ(t,0,z) = \int_0^\infty Q(t,x,z) h(x) dx - z p_0'(t) + \lambda z(z-1) p_0(t). \tag{5.52}$$

The partial differential equation (5.50) can be solved using the boundary condition (5.52) and the normalizing condition $\sum_{n=0}^{\infty} p_n(t) = 1$.

Steady-state distribution

Let

$$lim_{t\to\infty} p_n(t) = p_n, \ n \geq 0$$

and

$$\begin{aligned} \lim_{t\to\infty} p_n(t,x) &= p_n(x), \quad && x > 0, n \geq 1 \\ &= p_0(x) = 0, && x > 0. \end{aligned}$$

Then $\{p_n, n \geq 0\}$ gives the distribution of the general time system size.
Let

$$\begin{aligned} Q(x,z) &= \sum_{n=1}^{\infty} p_n(x) z^n \\ &= \sum_{n=1}^{\infty} \left[\lim_{t\to\infty} p_n(t,x) \right] z^n \\ &= \lim_{t\to\infty} \left[\sum_{n=1}^{\infty} p_n(t,x) z^n \right] \\ &= \lim_{t\to\infty} Q(t,x,z) \end{aligned}$$

and

$$Q(z) = \int_0^{\infty} Q(x,z)dx.$$

Then

$$(5.45) \Rightarrow \lambda p_0 = \int_0^{\infty} p_1(x)h(x)dx$$

$$(5.46) \text{ and } (5.47) \Rightarrow \frac{\partial}{\partial x} p_n(x) = -(\lambda + h(x))p_n(x) + \lambda p_{n-1}(x), \ n \geq 1$$

$$(5.48) \Rightarrow p_1(0) = \int_0^{\infty} p_2(x)h(x)dx + \lambda p_0$$

$$(5.49) \Rightarrow p_n(0) = \int_0^{\infty} p_{n+1}(x)h(x)dx, \ n \geq 2.$$

The partial differential equation (5.50) and the boundary condition (5.52) reduces to

$$\frac{d}{dx} Q(x,z) = -(\lambda - \lambda z + h(x))Q(x,z) \tag{5.53}$$

$$zQ(0,z) = \int_0^{\infty} Q(x,z)h(x)dx + \lambda z(z-1)p_0 \tag{5.54}$$

and

$$p_0 + Q(1) = 1. \tag{5.55}$$

From relation (5.53)

$$\begin{aligned}
\int \frac{dQ(x,z)}{Q(x,z)} &= \int -(\lambda - \lambda z + h(x))dx \\
log(Q(x,z)) &= logc\ (-\lambda(1-z)x - N(x)) \\
Q(x,z) &= c\exp(-\lambda(1-z)x - N(x)) \\
Q(0,z) &= c
\end{aligned}$$

$$Q(x,z) = Q(0,z)\exp(-\lambda(1-z)x - N(x)) \tag{5.56}$$

Substituting (5.56) in (5.54) we get

$$\begin{aligned}
zQ(0,z) &= \int_0^\infty Q(0,z)e^{(-\lambda(1-z)x - N(x))}h(x)dx + \lambda z(z-1)p_0 \\
&= Q(0,z)\int_0^\infty e^{-\lambda(1-z)x}\left[e^{-N(x)}h(x)\right]dx + \lambda z(z-1)p_0 \\
&= Q(0,z)G^*(\lambda(1-z)) + \lambda z(z-1)p_0.
\end{aligned}$$

Thus

$$Q(0,z) = \frac{\lambda z(z-1)p_0}{z - G^*(\lambda - \lambda z)}. \tag{5.57}$$

Now from (5.56) we have

$$\begin{aligned}
Q(z) &= \int_0^\infty Q(x,z)dx \\
&= \int_0^\infty Q(0,z)e^{(-\lambda(1-z)x - N(x))}dx \\
&= Q(0,z)\int_0^\infty e^{(-\lambda(1-z)x} \quad e^{-N(x)}dx \\
&= \frac{Q(0,z)}{\lambda(1-z)}\left[1 - \int_0^\infty e^{-\lambda(1-z)x}\left(e^{-N(x)}h(x)\right)dx\right]
\end{aligned}$$

$$Q(z) = \frac{Q(0,z)}{\lambda(1-z)}[1 - G^*(\lambda - \lambda z)]$$

From this and equation (5.57) we get

$$Q(z) = \frac{z[G^*(\lambda - \lambda z) - 1]p_0}{z - G^*(\lambda - \lambda z)}$$

Using L'Hospital rule, we get

$$\begin{aligned} Q(1) &= \lim_{z \to 1} Q(z) \\ &= p_0 \frac{[G^*(\lambda - \lambda z) - 1] + z\lambda G^{*(1)}(\lambda - \lambda z)}{1 + \lambda G^{*(1)}(\lambda - \lambda z)} \\ &= p_0 \frac{\lambda E(V)}{1 - \lambda E(V)} \end{aligned}$$

From (5.55) we obtain

$$p_0 = 1 - \lambda E(V).$$

Hence

$$Q(z) = \frac{z[G^*(\lambda - \lambda z) - 1][1 - \lambda E(V)]}{z - G^*(\lambda - \lambda z)}.$$

Busy period

Let T be the length of a busy period (starting with a customer arrival to an idle server, until the becomes idle again). Define $B(t) = P(T \leq t)$. Then $B(t)$ satisfies the relation

$$B(t) = \int_0^t \sum_{k=0}^{\infty} \frac{(\lambda u)^k}{k!} e^{-\lambda u} B^{*k}(t-u) dG(u) \tag{5.58}$$

The Laplace Stieltjes Transform (LST) of busy period $B(t)$ be denoted by $B^*(s)$. That is,

$$\begin{aligned}
B^*(s) &= \int_0^\infty e^{-st} dB(t) \qquad (\text{for } Re(s) > 0) \\
&= \int_0^\infty e^{-st} \int_0^t \sum_{k=0}^\infty \frac{(\lambda u)^k}{k!} e^{-\lambda u} B^{*k}(t-u) dG(u) dt \\
&= \int_0^\infty \sum_{k=0}^\infty \frac{(\lambda u)^k}{k!} e^{-\lambda u} e^{-su} \int_u^\infty e^{-s(t-u)} B^{*k}(t-u) dt dG(u) \\
&= \int_0^\infty \sum_{k=0}^\infty \frac{(\lambda u)^k}{k!} e^{-\lambda u} e^{-su} \left(B^*(s)\right)^k dG(u) \\
&= \int_0^\infty \sum_{k=0}^\infty \frac{(\lambda B^*(s) u)^k}{k!} e^{-(\lambda+s)u} dG(u) \\
&= \int_0^\infty e^{-(\lambda+s-\lambda B^*(s))u} dG(u)
\end{aligned}$$

Therefore

$$B^*(s) = G^*(\lambda + s - \lambda B^*(s)).$$

From this the mean and higher moments of the number of customers in the system can be computed.

Chapter 6

A $MAP/PH/1$ Queue with Uncertainty in Selection of Type of Service

In the previous chapter we assumed that the server offered n distinct services, of which only one was the needed/ desired service for each customer. However, due to certain complex situation neither the server nor the customer is aware of the exact needed service. The rest of the services may turn out to be harmful/ ineffective. A typical example is the Chikungunya, the symptoms of which varied from person to person. Accordingly physicians prescribed medicines to the patients; however, those who did not receive the right medication within a specified time were rendered physically/ mentally handicapped. In this chapter we extend the model described in chapter 5 to the case of $n(n > 1)$ distinct services offered by a server with distinct customers requiring any one among the n services which

Part of this chapter is included in the following paper.
A. Krishnamoorthy, A. S. Manjunath, and V. M. Vishnevsky: **An M/M/1 Queue with n Undesired Services and a Desired Service**, V. Vishnevsky and D. Kozyrev (Eds.): DCCN 2015, CCIS 601, pp. 102110, 2016. Springer International Publishing Switzerland 2016

we label as the desired service for that customer. The desired service may vary from customer to customer. For example, a customer dialing a customer care center for a specific service.

We analyze a single server system providing n distinct services; customer arrival follows a Markovian arrival process. At the time when taken for service the service requirement is correctly diagnosed with probability θ; with complementary probability $(1-\theta)$ the identification goes wrong. As a consequence of correct diagnosis, service in correct mode immediately starts, whose duration has exponential distribution with parameter μ_i if service required is in state i and the customer leaves the system after service. However, if initially the customer is admitted to one of the incorrect services (with probability p_i, it is diagnosed as requiring type i service, $i = 1, 2, ..., n$), it may stay in this class, moving from one incorrect to another incorrect, until finally all turn out to be failure and the customer turns out to be unfit for further service. It may also happen that at some stage of service in incorrect class, the service provider identifies that the customer is being served in the incorrect set of services and so immediately takes him to the actually required service stage. At this point, the customer starts required service and leaves the system on completion of service. But then how long is it possible to stay in service in the incorrect set of states? We assume that a timer with exponentially distributed duration starts ticking the moment a customer starts getting his service. If correct diagnosis is made of the desired service during its sojourn in the incorrect set of states before this random clock (timer) realizes, then the customer is immediately transferred for service to the correct state. On completion of service, assumed exponentially distributed with parameter μ_i, the customer leaves the system. On the other hand if the timer realizes before the customer's service need is correctly diagnosed, then no further service is provided to that customer since it is rendered useless as a consequence of service in the incorrect set of states.

In Section 1, the mathematical model is described. Section 2 provides the steady-state analysis and some performance measures including the expected service time of a customer. Effect of various parameters on performance measures of the system are numerically computed in Section 3.

6.1 Mathematical formulation

The assumptions leading to the formulation of the mathematical model are

- Arrival of customers to the system is according to the MAP. We use the same notations used in the previous chapter associated with MAP.

- The probability that a customer gets desired (correct) service from the very beginning is θ; denote $1-\theta=\hat{\theta}$.

- The probability that a customer requires the i^{th} type of service is p_i so that $p_1+p_2+\ldots+p_n=1$.

- If i is the required type of service for a customer and service starts correctly then corresponding service time is exponentially distributed with mean service rate $\mu_i, i=1,2,\ldots,n$.

- If i is the required type of service for a customer write $\boldsymbol{\beta}^{(i)}=\left(\boldsymbol{\beta}_k^{(i)}\right)$; $1\leq k\leq n, k\neq i$ where $\boldsymbol{\beta}_k^{(i)}$ is the probability that a customer in need of i^{th} service starts with $k^{th}(k\neq i)$. The rate of transition to j^{th} state of incorrect service after completing service in k^{th} state is μ_{kj} for, $k\neq i; j\in\{1,2,\ldots,n\}, j\neq i$.

- A random threshold clock(timer) which follows exponential distribution with mean rate γ starts ticking from the very beginning of service to the specific customer so that the customer is pushed out of the system if the clock expires before service completion in undesired service.

The last two assumptions indicate that only if the service requirement is correctly diagnosed right at beginning when taken for service, does the customer has an exponentially distributed service time. In the other case the service time turns out to be phase type distributed (initially in state(s) which are not the correct one and then get absorbed due to realization of timer or in the absence of realization of timer during service in the undesired states, thus escaping to the correct state of service, where there is additional exponentially distributed service requirement).

Let $N_1(t)$ be the number of customers in the system, $N_2(t)$ and $N_4(t)$ respectively the required service type and the type of service being provided and $N_3(t)$ the mode of service whether desired from the very beginning of service, unwanted or moved from undesired to desired, designated by 1, 2 and 3 respectively and $A(t)$ be the arrival phase at time t.

Then, $\{(N_1(t), N_2(t), N_3(t), N_4(t), A(t)), t \geq 0\}$ is a CTMC with state space $\Omega = \{(0, a)/1 \leq a \leq m\} \cup \{(i, j, k, \ell, a)/i \in Z^+, 1 \leq j \leq n, \ell = j; k = 1, 3; 1 \leq a \leq m\} \cup \{(i, j, 2, \ell, a)/i \in Z^+, 1 \leq j, \ell \leq n; \ell \neq j; 1 \leq a \leq m\}$. The infinitesimal generator $\boldsymbol{Q}$ of this CTMC is a LIQBD where

$$\boldsymbol{Q} = \begin{pmatrix} B_{00} & B_{01} & & & \\ B_{10} & B_1 & B_0 & & \\ & B_2 & B_1 & B_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

In the above matrix $B_{00} = D_0$, $B_{01} = \widetilde{\boldsymbol{\alpha}} \otimes D_1$, $B_{10} = C \otimes I_m$, $B_0 = I_{n(n+1)} \otimes D_1, B_1 = T \oplus D_0$, $B_2 = H \otimes I_m$. Here $\widetilde{\boldsymbol{\alpha}} = \begin{pmatrix} p_1\boldsymbol{\alpha}_1, & p_2\boldsymbol{\alpha}_2, & \cdots, & p_n\boldsymbol{\alpha}_n \end{pmatrix}$ with $\boldsymbol{\alpha}_i = \begin{pmatrix} \theta, & \hat{\theta}\boldsymbol{\beta}^{(i)}, & 0 \end{pmatrix}$, $1 \leq i \leq n$, and $C = \begin{pmatrix} c^{(1)}, & c^{(2)}, & \cdots, & c^{(n)} \end{pmatrix}^T$.

Let $\Delta_1^{(i)} = \begin{pmatrix} \mu_i \\ \mathbf{0} \\ \mu_i \end{pmatrix}$, and $\Delta_2 = \begin{pmatrix} 0 \\ \gamma\mathbf{e} \\ 0 \end{pmatrix}$. Then $c^{(i)} = \Delta_1^{(i)} + \Delta_2$.

$$H = [M_{ij}] \text{ where } M_{ij} = \begin{pmatrix} \mu_i p_j \theta & \mu_i p_j \hat{\theta}\boldsymbol{\beta}^{(j)} & 0 \\ \\ U^0 p_j \theta & U^0 p_j \hat{\theta}\boldsymbol{\beta}^{(j)} & \mathbf{0} \\ \\ \mu_i p_j \theta & \mu_i p_j \hat{\theta}\boldsymbol{\beta}^{(j)} & 0 \end{pmatrix}, \, 1 \leq i, j \leq n.$$

$$T = \text{diag}(T_1, T_2, ..., T_n), \, T_i = \begin{pmatrix} -\mu_i & \mathbf{0} & 0 \\ \\ \mathbf{0} & S^{(i)} & S_i^0 \\ \\ 0 & \mathbf{0} & -\mu_i \end{pmatrix}, \, 1 \leq i \leq n$$

with

$$S^{(i)} = \begin{array}{c} \\ 1 \\ 2 \\ \\ i-1 \\ i+1 \\ \\ n \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & & i-1 & i+1 & & n \end{array} \\ \begin{pmatrix} \mu_1^{(i)} & \mu_{12} & \cdots & \mu_{1(i-1)} & \mu_{1(i+1)} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_2^{(i)} & \cdots & \mu_{2(i-1)} & \mu_{2(i+1)} & \cdots & \mu_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mu_{(i-1)1} & \mu_{(i-1)2} & \cdots & \mu_{(i-1)}^{(i)} & \mu_{(i-1)(i+1)} & \cdots & \mu_{(i-1)n} \\ \mu_{(i+1)1} & \mu_{(i+1)2} & \cdots & \mu_{(i+1)(i-1)} & \mu_{(i+1)}^{(i)} & \cdots & \mu_{(i+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & & & \cdots & \mu_n^{(i)} \end{pmatrix} \end{array}.$$

Here, $\mu_k^{(i)} = -\left(\sum_{\substack{j=1 \\ j \neq k}}^{n} \mu_{kj} + \gamma \right)$, $k = 1, 2, ..., i-1, i+1, ..., n$ and

$$S_i^0 = \begin{pmatrix} \mu_{1i} & \mu_{2i} & \cdots & \mu_{(i-1)i} & \mu_{(i+1)i} & \cdots & \mu_{ni} \end{pmatrix}^T.$$

6.2 Steady-state analysis

We proceed with the steady-state analysis of the queueing system under study. Naturally we have to look for the condition for stability.

6.2.1 Stability condition

We consider the matrix $B(= B_0 + B_1 + B_2)$ representing the phase changes for determining the stability condition of the original system.
We have $B = (T + H) \oplus D$

where $T + H = (E_{ij})_{1 \leq i,j \leq n}$, $E_{ij} = M_{ij}$ for $i \neq j$

$$\text{and} \quad E_{ii} = \begin{pmatrix} -\mu_i(1 - p_i\theta) & \mu_i p_i \hat{\theta}\beta^{(i)} & 0 \\ U^0 p_i \theta & U^0 p_i \hat{\theta}\beta^{(i)} + S^{(i)} & S_i^0 \\ \mu_i p_i \theta & \mu_i p_i \hat{\theta}\beta^{(i)} & -\mu_i \end{pmatrix}.$$

Let $\widetilde{\boldsymbol{\pi}} = (\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2, ..., \hat{\boldsymbol{\pi}}_n)$ be the stationary probability vector of the Markov chain corresponding to the generator $T + H$ and $\boldsymbol{\eta}$ be that of D. Then

$$\widetilde{\boldsymbol{\pi}}(T + H) = \mathbf{0}, \ \widetilde{\boldsymbol{\pi}}\mathbf{e} = 1.$$
$$\boldsymbol{\eta} D = \mathbf{0}, \quad \boldsymbol{\eta}\ \mathbf{e} = 1.$$

Thus the stationary probability vector of B is $\mathbf{\Pi} = \widetilde{\boldsymbol{\pi}} \otimes \boldsymbol{\eta}$.
An algorithm for computing $\widetilde{\boldsymbol{\pi}}$ is given below:

$$\hat{\boldsymbol{\pi}}_{n-k} = \sum_{i=1}^{n-k-1} \hat{\boldsymbol{\pi}}_i F_{i(n-k)} \quad ; 0 \leq k \leq n-2$$

For $1 \leq i \leq n - 1$,

$$F_{i(n-k)} = \begin{cases} -B_{in}E_{nn}^{-1};\ k=0, \\ -\left(B_{i(n-k)} + \sum\limits_{m=1}^{k} U_{im}\right)\left(E_{(n-k)(n-k)+}\sum\limits_{m=1}^{k} U_{(n-k)m}\right)^{-1}; \\ \qquad 1 \le k \le n-2, \end{cases}$$

with

$$U_{im} = \sum_{\substack{J_{r-1}+1 \le j_r \le n-m+r \\ 1 \le r \le m,\ j_0 = n-k}} F_{ij_1}F_{j_1j_2}\ldots F_{j_{m-1}j_m}B_{j_m(n-k)}.$$

$\hat{\boldsymbol{\pi}}_1$ is obtained from

$$\hat{\boldsymbol{\pi}}_1\left(I + \sum_{m=1}^{n-1} V_m\right)e = 1,$$

where

$$V_m = \sum_{\substack{r_{j-1}+1 \le r_j \le n-(m-j) \\ 1 \le j \le m,\ r_0 = 1}} F_{1r_1}\prod_{j=1}^{m-1} F_{r_jr_{j+1}}.$$

The LIQBD description of the model indicates that the queueing system is stable if and only if the rate of left drift is larger than right drift rate (see Neuts [52]). That is

$$\mathbf{\Pi}B_0\mathbf{e} < \mathbf{\Pi}B_2\mathbf{e}.$$

This gives the stability condition as

Lemma 6.2.1. The system under study is stable if and only if

$$\lambda < (H \otimes I_m)\mathbf{e} \tag{6.1}$$

6.2.2 Steady-state probability vector

Assuming that equation (6.1) is satisfied, we briefly outline the computation of the steady-state probability of the system. Let $\boldsymbol{y}$ denote the steady-state probability vector of the generator $\boldsymbol{Q}$. Then

$$\boldsymbol{y}\boldsymbol{Q} = \mathbf{0}, \qquad \boldsymbol{y}\mathbf{e} = 1. \tag{6.2}$$

Assuming that the stability condition (6.1) holds and partitioning $\boldsymbol{y}$ as

$$\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, ...) \text{ with}$$

$$\boldsymbol{y}_i = \begin{cases} \boldsymbol{y}_0(a), & 1 \leq a \leq m, i = 0 \\ \boldsymbol{y}_i(j,1,j,a) \cup \boldsymbol{y}_i(j,2,\ell,a) \cup \boldsymbol{y}_i(j,3,j,a), & 1 \leq j, \ell \leq n; \ell \neq j, \\ & \quad 1 \leq a \leq m, i \geq 1 \end{cases}$$

we obtain

$$\boldsymbol{y}_n = \boldsymbol{y}_1 R^{n-1}, n \geq 2$$

where R is the minimal non negative solution to the matrix quadratic equation

$$R^2 B_2 + RB_1 + B_0 = \mathrm{O}.$$

The two boundary equations involving $\boldsymbol{y}_0$ are

$$\boldsymbol{y}_0 B_{00} + \boldsymbol{y}_1 B_{10} = \mathbf{0},$$

$$\boldsymbol{y}_0 B_{01} + \boldsymbol{y}_1 [B_1 + RB_2] = \mathbf{0}.$$

These together with the normalizing condition in (6.2) gives

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 V \text{ where } V = -B_{01}[B_1 + RB_2]^{-1}$$

$$\boldsymbol{y}_0[1 + V(I-R)^{-1}\mathbf{e}] = 1.$$

6.2.3 Expected service time of a customer

Let i be the required/correct service of tagged customer. Consider the Markov chain $\{(N_2(t), N_3(t), N_4(t))/t \geq 0\}$ with state space $\{(i, j, k)/j = 1, 3; k = i\} \cup \{(i, 2, k)/1 \leq k \neq i \leq n\} \cup \{\Delta_1\} \cup \{\Delta_2\}$, where $\{\Delta_1\}$ denotes the absorbing state which is completion of service from the in correct phases of service before the threshold clock is expired and $\{\Delta_2\}$ the absorbing state which represents the realization of the random threshold clock (that is, expulsion from service). The infinitesimal generator of this $CTMC$ is

$$\mathcal{W}_i = \begin{pmatrix} T_i & T^0_{\mu_i} & T^0_{\gamma} \\ \mathbf{0} & 0 & 0 \end{pmatrix}$$

where

$$T_i = \begin{pmatrix} -\mu_i & \mathbf{0} & 0 \\ \mathbf{0} & S^{(i)} & S^0_i \\ 0 & \mathbf{0} & -\mu_i \end{pmatrix}, \quad T^0_{\mu_i} = \begin{pmatrix} \mu_i \\ \mathbf{0} \\ \mu_i \end{pmatrix}, \quad T^0_{\gamma} = \begin{pmatrix} 0 \\ \gamma\mathbf{e} \\ 0 \end{pmatrix}.$$

The service time of a customer is the time until absorption of the Markov chain. The distribution of $\mathcal{W}_i$ is phase type with initial probability vector $\boldsymbol{\alpha}_i = \left(\theta, \hat{\theta}\boldsymbol{\beta}^{(i)}, 0\right)$ of order $n + 1$. The expected time a tagged customer spends in service is $E_{W_i} = -\boldsymbol{\alpha}_i T_i^{-1}\mathbf{e}$. Therefore the service time of an arbitrarily chosen customer is $E_{st} = \sum\limits_{i=1}^{n} p_i \, E_{W_i}$.

6.2.4 Performance measures

Now we look at a few of the system performance measures. Let a customer enters in to incorrect service with initial probability vector $(\boldsymbol{\psi}_1, \ \boldsymbol{\psi}_2, \ ..., \ \boldsymbol{\psi}_n)$, $\boldsymbol{\psi}_i$ being $p_i\boldsymbol{\beta}^{(i)}$ for $1 \leq i \leq n$. Let $l = (n+1)d$.

1. Probability that the system is idle, $P_{idle} = \boldsymbol{y}_0\mathbf{e}$.

2. Probability that the server is busy in direct correct mode, $\boldsymbol{Y}_1 = \sum_{i=1}^{\infty}\sum_{j=1}^{n}\sum_{a=1}^{m} \boldsymbol{y}_i(j,1,j,a)$.
3. Probability that the server is serving in the incorrect mode, $\boldsymbol{Y}_2 = \sum_{i=1}^{\infty}\sum_{j=1}^{n}\sum_{\ell=1,\ell\neq j}^{n}\sum_{a=1}^{m} \boldsymbol{y}_i(j,2,\ell,a)$.
4. Probability that the server is busy in correct, service of which started in incorrect mode, $\boldsymbol{Y}_3 = \sum_{i=1}^{\infty}\sum_{j=1}^{n}\sum_{a=1}^{m} \boldsymbol{y}_i(j,3,j,a)$.
5. Expected number of customers in the system, $\mu_{NS} = \sum_{i=1}^{\infty} i\, \boldsymbol{y}_i \mathbf{e}$.
6. Expected number of customers in the queue, $\mu_{NQ} = \sum_{i=2}^{\infty} (i-1)\, \boldsymbol{y}_i \mathbf{e}$.
7. Probability of customers leaving with correct service starting in incorrect service mode, $P_{cs} = \hat{\theta} \sum_{i=1}^{n} \boldsymbol{\psi}_i \, (-S_i)^{-1}\, S_i^0$.
8. Rate at which customers leave with correct service initially starting in incorrect service mode, $R_{cs} = \lambda P_{cs}$.
9. Probability that a customer is lost (leaving the system without getting correct service), $P_{loss} = \hat{\theta} \sum_{i=1}^{n} \boldsymbol{\psi}_i \, (-S_i)^{-1}\, \gamma \mathbf{e}$.
10. Rate of loss of customers due to incorrect service, $R_{loss} = \lambda P_{loss}$.
11. Rate of customers leaving successfully after being selected in correct service, $P_{lc} = \lambda\theta$.

6.3 Numerical illustration

In this section we provide numerical illustration of the system performance with variation in values of underlying parameters.

We fix parameters $n = 4, (p_1,\ p_2,\ p_3,\ p_4) = (0.1,\ 0.2,\ 0.3,\ 0.4), \mu_1 = 8, \mu_2 = 9, \mu_3 = 8, \mu_4 = 9$,

$$S^{(1)} = \begin{bmatrix} * & 6 & 8 \\ 7 & * & 6 \\ 5 & 9 & * \end{bmatrix}, S_1^0 = \begin{bmatrix} 5 \\ 8 \\ 5 \end{bmatrix}, \beta^{(1)} = \begin{bmatrix} 0.3 & 0.3 & 0.4 \end{bmatrix},$$

$$S^{(2)} = \begin{bmatrix} * & 7 & 6 \\ 8 & * & 6 \\ 5 & 9 & * \end{bmatrix}, S_2^0 = \begin{bmatrix} 6 \\ 7 \\ 5 \end{bmatrix}, \beta^{(2)} = \begin{bmatrix} 0.2 & 0.4 & 0.4 \end{bmatrix},$$

$$S^{(3)} = \begin{bmatrix} * & 6 & 6 \\ 5 & * & 8 \\ 5 & 5 & * \end{bmatrix}, S_3^0 = \begin{bmatrix} 7 \\ 6 \\ 9 \end{bmatrix}, \beta^{(3)} = \begin{bmatrix} 0.1 & 0.5 & 0.4 \end{bmatrix},$$

$$S^{(4)} = \begin{bmatrix} * & 6 & 7 \\ 5 & * & 6 \\ 8 & 7 & * \end{bmatrix}, S_4^0 = \begin{bmatrix} 6 \\ 8 \\ 6 \end{bmatrix}, \beta^{(4)} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \end{bmatrix},$$

$$D_0 = \begin{bmatrix} -5.0111 & 5.0111 & 0 \\ 0 & -5.0111 & 0 \\ 0 & 0 & -1128.75 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 4.96099 & 0 & 0.05011 \\ 11.2875 & 0 & 1117.4625 \end{bmatrix}.$$

Effect of θ

The entries in Table 6.1 are on expected lines: P_{idle} increases with increasing value of θ - this means that customers, when selected in incorrect mode of service, spent a long time in the system before departure; the value of $\boldsymbol{Y}_1$ steadily increases with θ since, for example all customers are selected for correct service at the beginning stage itself when $\theta = 1$; values of $\boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$ decrease with increase in value of θ, as expected and when $\theta = 1$, both turn out to be zero; in P_{cs} column all entries against corresponding values of θ decrease and reach zero when $\theta = 1$. Loss probability of customers, when admitted first to undesirable phases of service who leave without going to desired phase of service, decrease with increasing value of θ.

θ	P_{idle}	$\boldsymbol{Y}_1$	$\boldsymbol{Y}_2$	$\boldsymbol{Y}_3$	P_{cs}	P_{loss}	E_{st}
0.5	0.0647	0.2917	0.3625	0.2811	0.4819	0.0181	0.1871
0.6	0.1351	0.3500	0.2900	0.2249	0.3855	0.0145	0.1730
0.7	0.2055	0.4083	0.2175	0.1687	0.2891	0.0109	0.1589
0.8	0.2759	0.4667	0.1450	0.1124	0.1928	0.0072	0.1448
0.9	0.3463	0.5250	0.0725	0.0562	0.0964	0.0036	0.1307
1	0.4167	0.5833	0	0	0	0	0.1167

Table 6.1: Effect of θ for $\gamma = 0.25$

Effect of γ

γ	P_{idle}	$\boldsymbol{Y}_1$	$\boldsymbol{Y}_2$	$\boldsymbol{Y}_3$	P_{cs}	P_{loss}	R_{lc}	E_{st}
0.1	0.1969	0.4083	0.2223	0.1724	0.9852	0.0148	3.5	0.1606
0.2	0.2027	0.4083	0.2191	0.1699	0.9708	0.0292	3.5	0.1595
0.3	0.2083	0.4083	0.2159	0.1675	0.9568	0.0432	3.5	0.1583
0.4	0.2137	0.4083	0.2128	0.1651	0.9432	0.0568	3.5	0.1573
0.5	0.2190	0.4083	0.2098	0.1628	0.9301	0.0699	3.5	0.1562
0.6	0.2242	0.4083	0.2069	0.1605	0.9172	0.0828	3.5	0.1552

Table 6.2: Effect of γ for $\theta = 0.7$

The output in Table 6.2 also are on expected lines. Note that P_{loss} increases with increasing value of γ, since clock realized faster for higher value of γ. The column corresponding to $\boldsymbol{Y}_1$ has all entries with same value; this is so since the clock realization time does not affect the probability of getting into correct service. Columns corresponding to $\boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$ should have values decreasing with γ increasing since faster clock realization leads to moving out of incorrect service states faster. Further P_{cs} decrease with increase in value of γ. This is due again to the fact that clock realizes faster for larger values of γ, resulting in customers at undesirable phases of service leave the system (due to clock realization).

Effect of arrival process

For the arrival process, we consider the following five sets of values for D_0 and D_1 as follows.

1. Exponential ($EXPA$):

$$D_0 = \begin{bmatrix} -5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 5 \end{bmatrix}$$

2. Erlang ($ERLA$):

$$D_0 = \begin{bmatrix} -10 & 10 \\ 0 & -10 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 \\ 10 & 0 \end{bmatrix}$$

3. Hyper-exponential ($HEXA$):

$$D_0 = \begin{bmatrix} -9.5 & 0 \\ 0 & -0.95 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 8.55 & 0.95 \\ 0.855 & 0.095 \end{bmatrix}$$

4. MAP with negative correlation ($MNCA$):

$$D_0 = \begin{bmatrix} -5.0111 & 5.0111 & 0 \\ 0 & -5.0111 & 0 \\ 0 & 0 & -1128.75 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0.05011 & 0 & 4.96099 \\ 1117.4625 & 0 & 11.2875 \end{bmatrix}$$

5. MAP with positive correlation ($MPCA$):

$$D_0 = \begin{bmatrix} -5.0111 & 5.0111 & 0 \\ 0 & -5.0111 & 0 \\ 0 & 0 & -1128.75 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 0 & 0 & 0 \\ 4.96099 & 0 & 0.05011 \\ 11.2875 & 0 & 1117.4625 \end{bmatrix}$$

The above MAP processes will be normalized so as to have a specific arrival rate. However, these are qualitatively different in that they have different variance and correlation structure. The first three arrival processes, namely, $EXPA, ERLA$ and $HEXA$ have zero correlation for two successive inter-arrival times. The arrival processes labeled $MNCA$ and $MPCA$, respectively, have negative and positive correlation for two successive inter-arrival times with values -0.48891 and 0.48891. The standard deviation of the inter-arrival times of these five arrival processes are, respectively, 0.2, 0.14142, 0.44894, 0.2819 and 0.2819.

The main comparison in Tables 6.3 and 6.4 is between values of μ_{NS} in $MNCA$ and $MPCA$. Both decrease with increase in value of γ (θ) in both tables. However, $MNCA$ has much smaller values compared to their $MPCA$ counter parts. This indicates that positive correlation in the arrival process results in accumulation of large number of customers in the system.

γ	$EXPA$	$ERLA$	$HEXA$	$MNCA$	$MPCA$
0.1	4.0165	3.0730	10.5391	4.2115	203.9659
0.2	3.8699	2.9622	10.1056	4.0638	196.7258
0.3	3.7355	2.8609	9.7083	3.9285	190.0795
0.4	3.612	2.7677	9.3429	3.8041	184.2689
0.5	3.498	2.6819	9.0058	3.6892	178.2982

Table 6.3: Effect of γ for $\theta = 0.7$

θ	$EXPA$	$ERLA$	$HEXA$	$MNCA$	$MPCA$
0.5	13.3258	9.7995	40.6466	13.5708	721.7476
0.6	6.1198	4.5998	17.2625	6.3312	319.9786
0.7	3.8013	2.9105	9.9027	3.9947	193.3331
0.8	2.6298	2.0452	6.3133	2.8092	131.2621
0.9	1.9059	1.9059	4.2039	2.0729	94.3918

Table 6.4: Effect of θ for $\gamma = 0.25$

Concluding remarks and suggestions for future study:

In this thesis we discussed priority queueing models with self generation of lower priorities through interruption or feedback. A multi server priority model in the context of crowdsourcing was analyzed. Also discussed are queueing systems where uncertainty prevails in the selection of service.

Chapter 2 dealt with a highly dependent priority queueing system where low priority customers join the queue from immediately preceding waiting lines due to interruption of service by self. We assumed all underlying distributions to be exponential. Analytical expressions for system state probabilities were computed. The second chapter discussed an analogous situation but customers joined the low priority queue only after completing their service from high priority line. A multi server priority queueing model with two types of customers was discussed in chapter 4. The main advantage with the problem we analyzed in this chapter, in comparison with that of Chakravarthy and Dudin [11] is that the loss of high priority customers is reduced due to preemption. This results in a larger number of low priority customers being served by high priority customers. However, preemption of a low priority, sometimes even more than once, may lead to its longer waiting time in the system. Nevertheless if suitable incentive is provided to the high priority customer who serve a low priority customer on leaving the system, the probability to offer service may become close to 1, if not equal to 1. The thesis then focused some diagnostic problems where uncertainty in the selection of service type plays a prominent role. In chapter 5 we analyzed a situation where service starts without knowing whether it is going to be inappropriate for the customer, but service is compulsorily needed for customer arriving at the service point. We assumed the case of two types of services of which one is correct and services are offered in phases. In the last chapter we examined a queueing model offering n distinct services, but for any customer one among the n services was required and the remaining $n-1$ were damaging (undesirable/ inessential) and both of these cases were analyzed for a single server case.

In a future work we propose to extend the models in chapters 2 and 3 to the case of correlated arrivals. Crowdsourcing model is to be analyzed in the context of queueing-inventory scenario. In the diagnostic problems further analysis is needed when required service constitutes more than one correct service. The advantage in using multiple service channels to improve the performance of the system is to be explored. Also, analysis of the case of arbitrarily distributed service process is under progress.

Bibliography

[1] Alfa A. S. (2010) *Queueing theory for telecommunications: discrete time modelling of a single node system*, Springer, New York.

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

[3] Bini D., Latouche G. and Meini B (2005) *Numerical methods for structured Markov chains*, Oxford University Press, Oxford.

[4] 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.

[5] Bhat U.N. and Miller G.K. (2002) *Elements of Applied Stochastic Processes*, Wiley Science in Probability and Statistics, 3^{rd} Edition.

[6] Bhat U.N. (2008) *An Introduction to Queueing Theory: Modeling and Analysis in Applications*, Birkhauser Boston, Springer Science+Business Media, New York.

[7] O. J. Boxma and U. Yechiali (1997) *An M/G/1 queue with multiple types of feedback and gated vacations*. Journal of Applied Probability, 34(3): 773-784.

[8] Brodal G. S. (2013) *A Survey on Priority Queues: Volume 8066 of the series Lecture Notes in Computer Science*, 150-163 .

[9] Breuer L. and Baum D. (2005) *An introduction to queueing theory and matrix-analytic methods*, Springer, Dordrecht.

[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] S. R. Chakravarthy, A. N. Dudin(2016)(to appear) *A Queueing Model for Crowdsourcing: Journal of the Operational Research Society.*

[12] Chan W. C. (2014) *An elementary introduction to queueing systems*, World Scientific Publishing Co. Pre. Ltd.

[13] B. D. Choi and V. G. Kulkarni (1992) *Feedback retrial queueing system in Queueing and related models*, Oxford Univ. Press, New York, 93-105.

[14] B. D. Choi and B. Kim (2002) *M/G/1 queueing system with fixed feedback policy.* The ANZIAM Journal, 44(2): 283-297

[15] B. D. Choi, B. Kim and S. H. Choi (2003) *An M/G/1 queue with multiple types of feedback, gated vacations and FCFS policy.* Computer and Operation Research, 30(9): 1289-1309

[16] G. Choudhury and L. Tadj (2009) *An M/G/1 queue with two phases of service subject to the server breakdown and delayed repair.* Applied Mathematical Modelling 33 26992709

[17] A. Cobham (1954)*Priority Assignment in Waiting Line Problems.* Operations Research 2, 70-76.

[18] J. W. Cohen (1982) *The Single Server Queue*, 2nd ed., North-Holland, Amsterdam.

[19] D. R. Cox (1955) *The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables*, Proc. Cambridge Phil. Soc. 51, 433-441.

[20] A. N. Dudin, V. Jacob and A. Krishnamoorthy (2013) *A multi-server queueing system with service interruption, partial protection and repetition of service.* Annals of Operations Research, DOI 10.1007/s10479- 013-1318-3, Springer.

[21] A Gomez-Corral, A Krishnamoorthy, and V. C. Narayanan (2005) *The Impact of Self-generation of Priorities on Multi-server Queues with Finite Capacity.* Stochastic Models, 21: 427-447.

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

[23] Henk C. Tijms (1998) *Stochastic Models-An Algoritheoremic Approach* John Wiley & Sons.

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

[25] J. Howe (2006) *Crowdsourcing: A definition.* URL:http://www.crowdsourcing.com/ cs/2006/06/ crowdsourcing a.html.[Online] (2006).

[26] V. Jacob, S. R. Chakravarthy and A. Krishnamoorthy (2012) *On a Customer Induced Interruption in a service system.* Stochastic Analysis and Applications, 30 : 6, 949–962, DOI: 10.1080/07362994.2012.704845. Taylor & Francis

[27] Jain J. L., Mohanty S. G. and Bohm W (2007) *A course on queueing models*, Taylor & Francis Group, LLC.

[28] N. K. Jaiswal (1961) *Preemptive resume priority queue.* Operations Research 9, 732-742.

[29] N. K. Jaiswal (1962) *Time dependent solution of the head of the line priority queue.* Jr. Roy Statistic Soc. B. 24, 73-90.

[30] N. K. Jaiswal (1968) *Priority queues.* Academic Press, New York and London.

[31] J. Al Jararha and K. C. Madan (2003) *An M/G/1 Queue with Second Optional Service with General Service Time Distribution.* Information and Management Sciences Volume 14, Number 2, pp.47-56.

[32] E. M. Jewkes and J. A. Buzacott (1991) *Flow time distributions in a K class M/G/1 priority feedback queue.* Queueing Systems;8:183-202.

[33] Karlin S and Taylor H. E. (1975) *A first course in Stochastic Processes*, 2 nd ed., Elsevier.

[34] J. Keilson and A. Kooharian (1960) *On Time Dependent Queuing Processes*, The Annals of Mathematical Statistics, Vol. 31, No. 1, pp. 104-112

[35] L. Kleinrock (1975) *Queueing Systems Volume 1: Theory.* A Wiley-Interscience Publication John Wiley and Sons New York

[36] Kosten, L. (1973) *Stochastic theory of service systems.* Oxford: Pergamon Press

[37] A. Krishnamoorthy and V. Jacob (2012) *Analysis of Customer Induced Interruption in a multi server system.* Neural,Parallel and Scientific Computations, 20, 153–172. Dynamic publishers.

[38] A. Krishnamoorthy, S. Babu and Viswanath C. Narayanan (2009) *The MAP/(PH/PH)/1 queue with self-generation of priorities and non-preemptive service.* European Journal of Operational Research, 195, 174185

[39] A. Krishnamoorthy, P.K. Pramod, S.R. Chakravarthy(2014) *Queues with interruptions: a survey* Top, 22 (1), 290320 .

[40] B. Krishna Kumar, S. Pavai Madheswari and A. Vijayakumar (2002) *The M/G/1 retrial queue with feedback and starting failures*. Applied Mathematical Modelling 26: 1057-1075.

[41] B. Krishna Kumar, R. Rukmani and V. Thangaraj (2009) *On multi server feedback retrial queue with finite buffer*. Applied Mathematical Modelling 33: 2062-2083

[42] B. Krishna Kumar, G. Vijayalakshmi, A. Krishnamoorthy and S. Sadiq Basha (2010) *A single server feed back retrial queue with collisions*. Computers and Operations Research 37: 1247-1255

[43] G. Latouche, V. Ramaswami (1993) *A logarithmic reduction algorithm for quasi-birth-and-death processes*. Journal of Applied Probability, 30, 650-674 .

[44] Latouche G., and Ramaswami (1999) *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM., Philadelphia, PA.

[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] K. C. Madan (2000) *An M/G/1 queue with second optional service*, Queueing systems 34, 37-46.

[47] J. Medhi (1994) *Stochastic Processes*, 2^{nd} ed. Wiley, New York and Wiley Eastern, New Delhi.

[48] J. Medhi (2002) *A Single Server Poisson Input Queue with a Second Optional Channel*, Queueing systems 42, 239-242.

[49] D.R. Miller (1981) *Computation of Steady-State Probabilities for M/M/1 Priority queues*. Operations Research.

[50] M.F. Neuts (1979)*A versatile Markvian point process.* J. Appl. Prob. 16, 764-779.

[51] M.F. Neuts (1980)*The Probabilistic Significance of the Rate Matrix in Matrix- Geometric Invariant Vectors.* J. Appl. Prob. 17, 291-296.

[52] M.F. Neuts (1994) *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*, J2nd ed., Dover Publications, Inc., New York.

[53] Pattavina, A. and Parini, A. (2005) *Modelling voice call inter-arrival and holding time distributions in mobile networks*, Performance Challenges for Efficient Next Generation Networks - Proc. of 19th International Teletraffic Congress, pp. 729-738.

[54] 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.

[55] Qi-Ming He (2014) *Fundamentals of Matrix-Analytic Methods*, Springer Science and Business Media, New York.

[56] Richard J. Boucherie, Nico M. van Dijk (2010) *Queueing Networks: A Fundamental Approach* Springer Science and Business Media, New York.

[57] S. M. Ross (1996) *Stochastic Processes*, John Wiley and Sons.

[58] K. P. Sapna Isotupa and David A. Stanford (2002) *An Infinite-Phase Quasi-Birth-And-Death Model for the Non-preemptive Priority M/Ph/1 Queue.* Stochastic Models, 18:3, 387-424.

[59] B. Simon (1984) *Priority queues with feedback.* Journal of the Association for Computing Machinery, 31(1): 134-149

[60] L. Takacs (1963) *A single-server queue with feedback.* Bell System Technical Journal, 42(3): 509-519

[61] H. Takagi (1991)*Queueing Analysis. Volume 1: Vacations and Priority Systems*. North-Holland: Amsterdam.

[62] H. Takagi and Y. Kodera (1996) *Analysis of preemptive loss priority queues with preemption distance*. Queueing Systems 22, 367-381

[63] Varghese Jacob (2012) *Queues with Customer Induced Interruption*. Ph. D. Thesis submitted to the Cochin University of Science and Technology, June 2012.

[64] H. White and L. S. Christie (1958) *Queueing with Preemptive Priorities or with Breakdown*. Operations Research 6, 79-96.

Publications

1. *A. Krishnamoorthy and Manjunath A. S.*: **On Priority Queues Generated Through Customer Induced Service Interruption**, Neural, Parallel, and Scientific Computations 23 (2015) 459-486

2. *A. Krishnamoorthy, A. S. Manjunath, and V. M. Vishnevsky*: **An M/M/1 Queue with n Undesired Services and a Desired Service**, V. Vishnevsky and D. Kozyrev (Eds.): DCCN 2015, CCIS 601, pp. 102110, 2016. Springer International Publishing Switzerland 2016

3. *A. Krishnamoorthy, Manjunath A. S.*: **On queues with priority determined by feedback** (communicated).

4. *A. Krishnamoorthy, Dhanya Shajin, Manjunath A. S.*: **On a multi-server priority queue with preemption in crowdsourcing** (communicated).

CURRICULUM VITAE

Name :	Manjunath A S
Present Address :	Department of Mathematics, CUSAT Cochin, Kerala, India – 682 022.
Official Address :	Assistant Professor, Department of Mathematics, Govt. College Kottayam Kerala, India – 686 013.
Permanent Address :	Ashokanivas Nedumkunnam P.O. Kottayam, Kerala, India – 686 542. Mob: 94470 78934
Email :	manjunadem@gmail.com
Qualifications :	**M.Sc.**(Mathematics) CUSAT Cochin, Kerala, India – 682 022.
Research Interest :	Stochastic Modelling– Queueing Theory.