Stochastic Modelling: Analysis and Applications

QUEUES WITH POSTPONED WORK UNDER N-POLICY

Thesis submitted to the
Cochin University of Science and Technology
for the award of the degree of
DOCTOR OF PHILOSOPHY
under the Faculty of Science

By

AJAYAKUMAR C.B.

Department of Mathematics
Cochin University of Science and Technology
Cochin - 682022

August 2011

Certificate

This is to certify that the thesis entitled '**Queues with postponed work under N-policy**' submitted to the Cochin University of Science and Technology by Mr. Ajayakumar C.B. for the award of the degree of Doctor of Philosophy under the Faculty of Science is a bonafide record of studies carried out by him under my supervision in the Department of Mathematics, Cochin University of Science and Technology. This report has not been submitted previously for considering the award of any degree, fellowship or similar titles elsewhere.

Dr. A. Krishnamoorthy (Supervisor)
Professor (Retd.)
Department of Mathematics
Cochin University of Science and Technology
Cochin- 682022, Kerala.

Cochin-22

31-08-2011.

Declaration

I, Ajayakumar C.B. hereby declare that this thesis entitled '**Queues with postponed work under N-policy**' contains no material which had been accepted for any other Degree, Diploma or similar titles in any University or institution and that to the best of my knowledge and belief, it contains no material previously published by any person except where due references are made in the text of the thesis.

Ajayakumar C.B.
Research Scholar (Reg. No.2950)
Department of Mathematics
Cochin University of Science and Technology
Cochin-682022, Kerala.

Cochin-22
31-08-2011.

To

My Parents

Acknowledgement

This thesis is the fulfillment of my desire of research that formed in early studies of undergraduate mathematics. In the longest path of the desire, I am indebted to a lot of individuals for the realization of the thesis.

Words fail me to express gratitude to my supervisor and guide Dr. A. Krishnamoorthy, Professor(Retd), Department of Mathematics, Cochin University of science and Technology, for his support throughout my work with his patience and knowledge. The thesis would not have been possible without the inspiring discussions that I had with him.

I owe a special debt of gratitude to Dr. T. Thrivikraman, former Head, Department of Mathematics, who had introduced me to my guide to start my Doctoral work and blessed with his valuable suggestions. Also I wish to express sincere thanks to my Doctoral committee member Dr. M.N. Narayanan Namboodiri, for his support and advices.

I am also grateful to Dr. A. Vijayakumar, Head of the department of Mathematics, for his valuable suggestions and support for my research. I thank other faculty members of the department Dr. B. Lakshmy and Dr.P.G. Romeo for their inspiring words. I deeply obliged to Dr. M. Jathavedan, Dr.R.S. Chakravarti and Dr.M.K. Ganapathi for their support during my research work.

I thank the office staff Ms. Indukumari.S, Ms. Vidhya, Ms. Sheeba, Ms. Omana, Librarian Shajtha C. and the former Librarian V.K. Sailaja in the Department of Mathematics and the administrative authorities of Cochin University of Science and Technology for the facilities they pro-

vided.

I am indebted to Dr. S.R. Chakravarthi, Department of Industrial and Manufacturing Engineering, Kettering University, Flint, USA for some helpful discussions and his motivating suggestions. I thank Dr. T.G. Deepak, Department of Mathematics, Indian Institute of Space Science and Technology, Trivandrum, for his technical help in my research work.

I extend my heartfelt gratefulness to my friend Dr.Pramod P.K. for his valuable help in my research work. I acknowledge with gratitude to Dr. Viswanath C. Narayan, Dr. S. Babu, Dr. K.P.Jose and Dr. Lalitha K. for their motivations. I am indebted to Mr. Varghese Jacob for his valuable help and discussions. Also I thank Mr. Manikantan for his help in many situations. I also express my gratitude to research scholars Mr. C. Sreenivasan, Mr. Sajeev S. Nair, Mr. Gopakumar G., Ms. Deepthi C.P., and Mr. Sathyan M.K. in the same area for their willingness to share their bright ideas with me. I thank for the great suggestions and help from the research scholars Ms. Seema Varghese, Mr. Tonny K.B., Mr. Pravas K. and Mr. Kiran. I want to express my thanks to fellow research scholars Dr. Aparna Lakshman, Mr. Jayaprasad, Ms. Anu, Ms. Chitra M.R., Mr. Shinoj K.M., Mr. Didimos M.K., Mr.Tijo , Mr. Ratheesh for their interest in my work.

I express sincere thanks to Prof. P.V. Sugathan, The Principal, College of Engineering, Thalassery for his considerations and encouraging words. Also I thank Dr. Praseetha Lakshmi, The principal, College of Engineering, Kidangoor for motivating suggestions. Also my colleagues Mr. Jais Kurian, Mr. Sajith C. Subramanian, Mr. Varun, Mr. Tiju Baby, Mr. Siby, Mr. Suneesh Kurian, Ms. Mary James, Mr. Suresh deserve special

mention for their encouraging words. I express deepest appreciation to Mr. Krishnamohan for his help in my work.

I remember all of my teachers who had inspired their students with knowledge, magical voices and advices. I cannot forget the inspirations and loving support from my friends Mr. Savio James, Mr. Vishnu S., Mr. Robins Mathew, Mr. Tijo Thomas, Mr. Sreejith Nair, Mr. Jithin Abraham and Mr. Santhosh Thomas.

Last but not the least, is the love of my father, mother, two brothers and sister. I fail to find a word to express my gratitude to them.

Ajayakumar C.B.

QUEUES WITH POSTPONED WORK UNDER N-POLICY

Contents

List of Acronyme

CTMC	-	Continuous Time Markov Chain
DTMC	-	Discrete Time Markov Chain
FCFS	-	First Come-First Served
FIFO	-	First In-First Out
LCFS	-	Last Come-First Served
LIFO	-	Last In-First Out
LIQBD	-	Level Independent Quasi Birth-Death Process
MC	-	Markov Chain
MP	-	Markov Process
PH	-	Phase type distribution
QBD	-	Quasi Birth-Death Process
SIP	-	Service In Priority
SIRO	-	Service In Random Order
TPM	-	Transition Probability Matrix
UB	-	Upper Bound

List of symbols

Geo	-	Geometric distribution
$exp(t)$	-	Exponential function with parameter t
E_d	-	Discrete Erlang distribution
PH_d	-	Discrete phase type distribution
$[x]$	-	Greatest integer less than or equal to x
$\otimes$	-	Kronecker product
P_{FIFO}	-	Probability for FIFO violation
$m \times n$	-	m by n
$exp(t)$	-	Exponential distribution with parameter t
μ_{POOL}	-	Mean number of pool customers
μ_{BUFFER}	-	Mean number of buffer customers
θ_{LOST}	-	Loss rate
θ_{TR}	-	Transfer rate
$diag[A_1, A_2]$	-	Diagonal Matrix with elements A_1 and A_2
R	-	Rate matrix
$sp(R)$	-	Spectral radius of R
I	-	Identity matrix
I_r	-	Identity matrix of dimension r
e	-	Column vector of ones of appropriate order

List of Figures

Chapter 1

Introduction

Process is a phenomenon that takes place in time. In many practical situations, the result of a process at any time may not be certain. Such a process is called a stochastic process. As uncertainties lead to random variables, stochastic process requires a probabilistic setting. So many real world phenomena can be modelled and analysed by using the theory of stochastic processes where deterministic laws fail. This is called stochastic modelling. One of the most important part in stochastic modelling is the field of Queueing Thoery.

1.1 Stochastic process

A stochastic process is a family of random variables and each random variable is a function of a parameter say time t. It is denoted by $\{X(t), t \in T\}$

where T is an index set. The set of possible values of the random variable $X(t)$ is called its *state space* and a value of the random variable $X(t)$ is called a *state*. According to the nature of the state space and index set, a stochastic process is classified in to four categories. (i) discrete time - discrete state space (ii) discrete time - continuous state space. (iii) continuous time - discrete state space (iv) continuous time - continuous state space. If the state space is discrete, then the Stochastic process may be called as a *chain* , otherwise or in general we use the word process.

1.1.1 Poisson process

A continuous time stochastic process $\{X(t) : t \in T, T = [0, \infty)\}$ is called a Poisson process with parameter λ if and only if it satisfies the following conditions. (i) $X(0) = 0$ (ii) the increments $X(s_i + t_i) - X(s_i)$, over an arbitrary finite set of disjoint intervals $(s_i, s_i + t_i)$ are independent random variables. (iii)for each $s \geqslant 0$, $t \geqslant 0$, $X(t) = X(s + t) - X(s) = n$ has the Poisson distribution $\frac{e^{-\lambda t}(\lambda t)^n}{n!}$ with mean λt.

1.1.2 Markov process

A stochastic process is said to be *Markov* if $P(a < X_t \leqslant b | X_{t_1} = x_1, ..., X_{t_n} = x_n) = P(a < X_t \leqslant b | X_{t_n} = x_n)$ whenever $t_1 < t_2 < ..... < t_n < t$. That is, it is a process with the property that given the value of X_t, the values of X_s, $s > t$ do not depend on the values of X_u, $u < t$. That is the probability of any particular future behaviour of the process, when its present state is known exactly, is not altered by additional knowledge concerning

its past behaviour. A Markov process having a finite or denumerable state space is called a *Markov chain*. If the time is discrete, the MC is called discrete time Markov chain(DTMC) and if the time is continuous, it is called continuous time Markov chain(CTMC).

1.2 Queueing Theory

Queue is a waiting line. The imperfect matching between the customers and service facilities creates queues. Queueing theory originated as a very practical subject. It has many applications in telecommunications. The earliest works studied the telephone traffic congestion. The first work related to this is "The Theory of Probabilities and Telephone Conversations" which was published by A.K.Erlang in 1909. The thoery of queues is applied in many other practical situations of traffic, internet, facility designs like banks, amusement parks, fastfood restaurents, hospitals and post offices.

1.2.1 Queueing system

A system having arrivals, service facilities, and departures is called a queueing system. A diagrammatic representation of the queueing system is given in Figure 1.1 . For a complete description of a queueing system, we consider the following characteristics.

1. **Arrival pattern of customers**

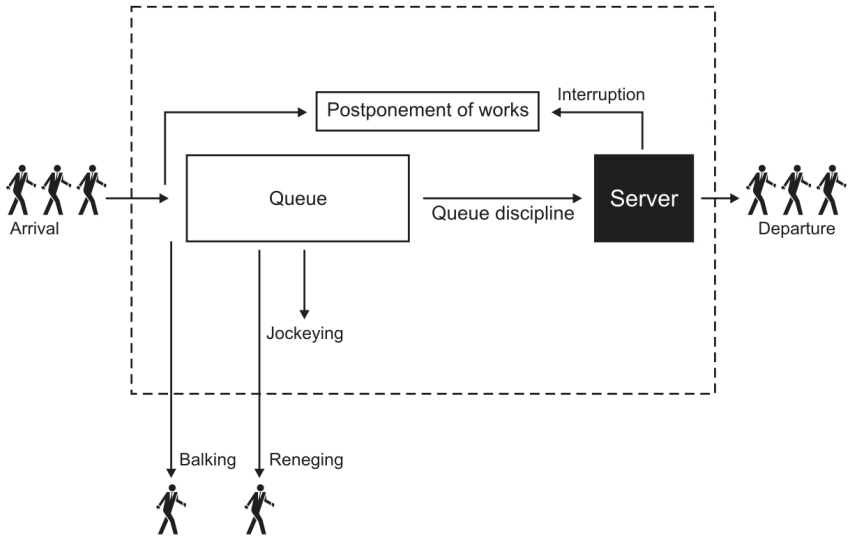

Fig 1.1: Queueing system

In a queueing system, time between two successive customer arrival is called *inter arrival time*. Practically it is stochastic. So it is necessary to know its probability distribution. Also customers may arrive in batch or bulk. So the size of a batch is another random variable and it is necessary to identify its probability distribution. *Customer behaviour* is another important one. A customer may decide not to enter the queue by seeing the queue too long. Then the customer is said to have *balked*. Some customers after joining the queue, wait for some time, and leave the service system due to intolerable delay. In this case the customer is said to have *reneged*. In case of parallel waiting lines, customers may move from one queue to another hoping to receive service more quickly. This is called *jockeying*. An arrival pattern that does not change with time is called a *stationary arrival pattern*. One that is not time independent is called

nonstationary.

2. Service pattern

Just like Arrival pattern, a probability distribution function is needed to describe service times. Service may be in single or in batch. Service process which depend on the number of customers waiting is called *state dependent service*. Like arrivals, service can be stationary or nonstationary with respect to time. If the system has no customers, then the server becomes *idle*. Then the server may leave the system for *vacation* with a random period of time. These vacations may be used by the server for doing other jobs. Server vacation period may be limited by some control policies like N, D, and T. An idle server starts service only when N are present in the queue and once he starts serving, goes on serving till the system size become zero. This is called N*-policy*. Under T*-policy*, the service facility is turned off for a fixed period of time T, from the instant of each service completion leaving the system empty. According to D*-policy*, the service facility re-opens as soon as the total workload exceeds a critical level D.

Also *interruption* may take place while a service is going on, for reasons like server breakdown or the arrival of a high priority customer. On completion of interruption, the interrupted work may be resumed or repeated. Also working vacations and vacation interruptions are recent concepts in service pattern. In *working vacation*, a customer is served at a lower rate rather than completely stoping the service during a vacation. But in *vacation interruption* policy, the server will come back from the vacation without completing it.

3. Queue discipline

Queue discipline refers to the manner in which customers are selected for service when a queue has formed. The most common Queue discipline is *first in-first out* (FIFO) [or first come - first served(FCFS)]. Some others in common usage are *last in-first out* (LIFO) [or last come-first served(LCFS)], service in random order(SIRO), Service in priority(SIP). There are two general situations in priority disciplines: *preemptive priority* and *non-preemptive priority*. When the high priority customer enters the system, the low priority customer in service is preempted. This is called preemptive priority. But in non-preemptive priority, the high priority customer goes to the head of the queue but cannot get in to the service until the customer presently in service is served completely, even though this customer has low priority.

4. System capacity

Physical space of the waiting room is called system capacity. It may be finite or infinite. In the case of finite system capacity, customers may be forced to balk if the capacity is full. In such situations, work may be *postponed*. In this thesis, it is discussed in great detail.

5. Number of servers

There are queueing systems with number of parallel service stations, which can serve customers simultaneously. So the number of servers is essential to describe a queueing system.

6. Stages of service

There may be only one stage of service or may have several stages. In the case of several stages, a customer may not pass through all stages.

1.2.2 Notation of a queueing system

In the development of a queueing system, a notation has evolved to describe its essential characteristics, called Kendall-Lee notation. Here we notate a queueing system by $a/b/c/d/e$ where a denotes the arrival pattern, b the service pattern, c the number of servers, d the system capacity, e the queue discipline. If the queueing system is represented by $a/b/c$, then it is understood that the system capacity is infinite and the queue discipline is FCFS. However this notation is not sufficient to describe the whole characteristics of modern queueing systems.

1.2.3 Analysis of queueing models

Queueing models can be classified in to *Markovian* and *non-Markovian* models. If the inter arrival time of customers and service times are exponentially distributed, then the queueing model is called Markovian queueing model. Queueing models with inter arrival times and/or service times which are not exponential distributions are called non-Markovian queueing models. Matrix geometric method developed by Neuts is useful for analysing complicated queueing models in steady state.

$M/M/1$ queue in continuous time is a simple Markovian birth-death queueing model. Let $1/\lambda$ be the mean inter arrival time and $1/\mu$ be the mean service time. Then to analyse its steady state behaviour, we first

form the steady state probability distribution P_n for the system to have n units, by using difference-differential equation method. Then $P_n = \rho^n(1-\rho)$ if $\rho < 1$ where $\rho = \lambda/\mu$ called *traffic intensity*. It is a geometric distribution. If the system capacity is finite, we have the $M/M/1/K$ model, where $P_n = \frac{(1-\rho)\rho^n}{1-\rho^{K+1}}$ if $\rho \neq 1$ and $P_n = \frac{1}{K+1}$ if $\rho = 1$ for $n > 0$ and $P_0 = \frac{1-\rho}{1-\rho^{K+1}}$ if $\rho \neq 1$ and $P_0 = \frac{1}{K+1}$ if $\rho = 1$. This geometric nature of the solutions are the main motivating fact to the introduction of matrix geometric solutions for extended models.

1.3 Matrix analytic methods

Most of the modern queueing problems are difficult to analyse by making difference-differential equations and solving by the method of generating functions and Laplace transforms. This difficulty can be overcome by using Matrix analytic methods introduced by Neuts (see [45]). Here usual birth-death process can be extended to *quasi-birth-death(QBD) process*. If the tridiagonal elements in the intensity matrix of a birth-death process are matrices, then such a process is called a QBD process. Here the state space consists of states of the form $(i, j, ...)$. The first dimension is called the *level* of the process, while the other dimensions are called *phases*. The transitions are restricted to the same level or to the two adjacent levels. Thus it is possible only to move from (n, j) to (m, k) in one step if $m = n + 1$, n or $n - 1$ for $n \geqslant 1$ and $m = 0, 1$ for $n = 0$. If the transitions rates are level independent, the resulting QBD process is called *level independent quasi-birth-death process*(LIQBD).

Let the infinitesimal generator of a LIQBD process be

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

where the matrices B_0, B_2, A_0, A_2 are non negative and the matrices B_1 and A_1 have non negative off diagonal elements but strictly negative diagonal elements. The row sums of Q are necessarily equal to zero.

Let x be a stationary vector. Then $xQ = 0$ and $xe = 1$ where e is a column vector of ones of appropriate order. Let x be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form $x_i = x_1 R^{i-1}$ for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation $A_0 + RA_1 + R^2 A_2 = 0$ and the vectors x_0, x_1 are obtained by solving the equtions $x_0 B_1 + x_1 B_2 = 0$ and $x_0 B_0 + x_1(A_1 + RA_2) = 0$ subject to the normalising condition $x_0 e + x_1 (I - R)^{-1} e = 1$.

For the existence of stationary solution, spectral radius of R; $sp(R) < 1$, which is analogues to the condition $\rho < 1$ in familiar $M/M/1$ queueing model. R is called *rate matrix*. Once R is determined, the geometric nature of the solution is established. If the matrix $A = A_0 + A_1 + A_2$ is irreducible, then $sp(R) < 1$ iff $\pi A_0 e < \pi A_2 e$ where π is the stationary probability vector of the generator matrix A. That is π is the solution of $\pi A = 0$ and $\pi e = 1$. One can use the iterative formula $R = -A_0(A_1 + R_{n-1}A_2)^{-1}$ for $n \geqslant 1$ with an initial value R_0 which converges to R if $sp(R) < 1$.

The following are some distributions frequently used in queueing theory.

1. Exponential and Geometric distribution

Consider a Poisson process $\{N(t), t \geqslant 0\}$ with parameter λ where $N(t)$ represents total number of arrivals in an interval of duration t. Then the time between two successive arrivals will follow exponential distribution with probability density function $f(x) = \lambda e^{-\lambda x}$, $x \geqslant 0$. The distribution function is $F(x) = 1 - e^{-\lambda x}$. The exponential distribution is the only one continuous distribution which exhibits Markovian property. This property states that the probability that a customer currently in service has t units of remaining service is independent of how long it has already been in service. That is, $P[T \leqslant t_1 | T \geqslant t_0] = P[0 \leqslant T \leqslant t_1 - t_0]$. The only other distribution to exhibit this property is the geometric distribution which is the discrete analogue of the exponential distribution. Probability density function of geometric distribution is $f(x) = pq^x$ where $0 < p < 1$ and $q = 1 - p$.

2. Continuous time phase type distribution

Let $\{X_t, t \geqslant 0\}$ be a finite MC with statespace $\{1, 2, ..., m+1\}$ and generator $Q = \begin{bmatrix} T & T^0 \\ \overline{0} & 0 \end{bmatrix}$. Here $1, 2, ..., m$ are transient states and $m+1$ or 0 is absobing state. T is a square matrix of order m satisfying $T_{ii} < 0$ for $1 \leqslant i \leqslant m$ and $T_{ij} > 0$ for $i \neq j$. Also $Te + T^0 = 0$ where e is a column vector of ones of order m. Let the initial probability vector be (α, α_{m+1}) with α a row vector of dimension m, so that $\alpha e + \alpha_{m+1} = 1$. Let $Z = inf[t \geqslant 0, X_t = m+1]$ be the random variable of the time until absorption in state $m+1$. The distribution of Z is called phase

type distribution. We denote it by $PH(\alpha, T)$. The dimension m of T is called the *order* of the phase type distribution. The states $1, 2, ...m$ are called *phases*. If Z follows $PH(\alpha, T)$, the distribution function of Z is given by $F(t) = P(Z \leqslant t) = 1 - \alpha.exp(Tt).e$, $\forall t \geqslant 0$ and the density function is $f(t) = \alpha.exp(Tt).T^0$, $\forall t \geqslant 0$. It is possible to approximate any distribution on the non negative real numbers by a PH-distribution. Moments of Z are given by $E(Z^n) = (-1)^n n! \alpha T^{-n} e$, $\forall n \in N$. So $E(Z) = -\alpha T^{-1} e$ is the mean time to absorption.

3. Discrete time phase type distribution

Let $Z = min[n \in N_0, X_n = m + 1]$ denote the time until absorption in the state $m + 1$. The transition probability matrix(TPM) has the form $P = \begin{bmatrix} T & T^0 \\ \bar{0} & 1 \end{bmatrix}$ where T is a square matrix of order m such that $I - T$ is non-singular and $Te + T^0 = e$. The distribution of Z is called a discrete PH-distribution. $P(Z = n) = \alpha T^{n-1} T^0$ and $P(Z \leqslant n) = 1 - \alpha T^n e$, $\forall n \in N$. The mean time to absorption is given by $E(Z) = \alpha(I - T)^{-1} e$.

4. Erlang distribution

An Erlang distribution with n degrees of freedom (or stages) and parameter λ is the distribution of the sum of n exponential random variables with parameter λ. It has the density function $f(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}$, $\forall t \geqslant 0$. It can be represented as the holding time in the transient state set $\{1, 2, ..., n\}$ of a MC with absorbing state $n + 1$ where the only possible transitions occur from a state k to the next state $k + 1$ $(k = 1, 2, ..., n)$ with rate λ each. This can be approximated to PH distribution with

$\alpha = (1, 0, 0, ..., 0)$,

$$T = \begin{bmatrix} -\lambda & \lambda & & & & \\ & -\lambda & \lambda & & & \\ & & \ddots & \ddots & & \\ & & & -\lambda & \lambda & \\ & & & & -\lambda & \end{bmatrix}, T^0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \lambda \end{bmatrix}.$$

In the discrete case, general Erlang distribution with m stages is approximated to PH distribution with $\alpha = (1, 0, 0, ..., 0)$,

$$T = \begin{bmatrix} s_{11} & s_{12} & & & & \\ & s_{22} & s_{23} & & & \\ & & \ddots & \ddots & & \\ & & & & & \\ & & & & & s_{mm} \end{bmatrix}, T^0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ s_{mo} \end{bmatrix}.$$

where $Te + T^0 = e$.

1.4 Summary of the thesis

In many real life situations a work may be postponed for several reasons. It may be to attened a more important job or to go on vacation. To bring the postponed work in the usual service track, we introduce N-policy. Considering the physical limitation of a system, finite capacity queues are more realistic than infinite capacity queues. But this will result in overflow of jobs and make cosiderable loss to the system. Models with

postponement are an alternative to finite capacity queues to minimise such a loss.

A paper to deal with postponed work was introduced by Deepak et.al. (see [16]). They analysed such a system in the stationary case and provided a number of system performance measures. No further development in this is reported so far. Nevertheless this notion of postponement of work has been introduced into inventory by a few researchers (see Krishnamoorthy and Islam [36], Arivarignan et.al. [2], Paul Manuel et.al. [46], Sivakumar and Arivarignan [48]).

The thesis entitled "Queues with postponed work under N-policy" is divided in to 6 chapters.

Chapter 1 is an introductory chapter containing basic definitions and terminologies of stochastic process, queueing theory and matrix analytic methods. We provide a brief of the work done so far in queues with postponed work; some associated work is reviewed in this chapter.

Chapter 2 describes an $M/PH/1$ queue with postponed work under N-policy. Here we extend the model described in [16] by introducing N-policy for transfer of customers from the pool. When a buffer having capacity K is full, newly arriving jobs are not necessarily lost. They can accept the offer of joining a pool of postponed work having infinite capacity with probability γ. With probability $1-\gamma$, such customers do not join the system. When at the end of a service, if there are postponed customers, the system operates as follows. If the buffer is empty, the one ahead of all waiting in the pool gets trasferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$ at a

service completion epoch, then again the job at the head of the buffer starts service and with probability p, the head of the queue in pool is transferred to the finite buffer and positioned as the last among the waiting customers in the buffer. With probability $q = 1 - p$, no such transfer takes place.

N-policy ensures an early service for pooled customers. If the pool contains atleast one postponed work, continuously served customers from the buffer since the last transfer under N-policy, is counted at each service completion epoch. When it reaches a pre-assigned number N, then the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. The N-policy introduced here differs from the classical N-policy as explained below. In the classical case, N customers are to queue up to start the new service cycle once the system becomes empty. However in the present case N-policy is applied to determine a priority service to be given to a customer from the pool.

A stability condition based on first passage time probability and stationary distribution has been obtained. We derived the expected waiting time of a tagged customer (i) in the buffer and (ii) in the pool, the expected duration (i) between two consecutive transfers under N-policy (ii) for the first N-policy transfer in a busy cycle and the expectation of FIFO violation. Several system performance measures and an optimization problem involving N are discussed. Numerical illustrations are also provided.

In Chapter 3, we modify the model discussed in chapter 2. The entry to the buffer is restricted by the system with the increasing number of work in the buffer. If the buffer is empty, an arriving customer can enter in to it and his service starts immediately. Otherwise there is a probability depending on the number of work in the buffer. But at every time, when a

customer is not allowed to enter the buffer, he may join a pool of postponed work having infinite capacity with probability δ. If the buffer is full, a customer may select the pool with probability γ_1. But at that time, system may reject him with probability γ_2. Usual transfer from the buffer to the pool with some probability p and N-policy is considered for the service of postponed work. We studied its long run behaviour. Several system performance measures, and numerical illustrations are provided. By treating server and customer as players, we give a game theoretic approach to the model and found the mixed strategies of the players and the value of the game.

Chapter 4 discusses an $M/M/1$ Queue with Postponed work and service interruption under N-policy. At a service completion epoch, if the buffer size drops to a pre-assigned level or below, a postponed work is trasferred to the buffer for immediate service with some probability. During the service of such a pooled customer, if the buffer size rises to a pre-assigned higher level, then the postponed work at server will be interrupted, again postponed and wait at the head of the queue in pool. Just after the interruption, we start to count the number of continuously served customers from the buffer. When it reaches a pre-assigned number N at a service completion epoch, the interrupted pooled customer gets transferred to the buffer for immediate service, and further interruption is not allowed for such a work. We studied its long run behaviour and obtained several system performance measures. Several numerical illustrations are also provided.

Chapter 5 analyses a discrete time $Geo/PH_d/1$ queue with postponed work under N-policy. It is the discrete time counter part of the continuous time model discussed in chapter 2. Continuous time models describe the

event in a very short interval of time. But in this discrete time queueing system, time axis is divided in to intervals of equal length called slots, and where all queueing activities takeplace at the slot boundaries. Both arrival and departure may happen in a slot. We consider late arrival system. That is departures occur at the moment immediately before the slot boundaries and arrivals occur at the moment immediately after the slot boundaries. The time between two successive arrivals is is governed by a geometrical law with parameter α and service time of each customer by a discrete phase-type distribution. The model is studied as a quasi birth-death(QBD) process and a solution of the classical matrix geometric type is obtained.

In Chapter 6 we consider two models of discrete time $Geo/E_d/1$ queues with postponed work and protected stages. If a buffer having finite capacity is not full, a higher priority customer can enter it and a lower priority customer is directed to a pool of postponed work having infinite capacity. When the buffer is full, new arrivals of higher priority customers cannot join the system and will leave the system permanently. At that time, a new arrival of lower priority customer will join the pool with probability γ or it is lost to the system for ever with probability $1-\gamma$. At a service completion epoch, if buffer size drops to a pre-assigned lower level or below a postponed work is transferred to the buffer for immediate service with some probability. During the service of such a pooled customer, if the buffer size raises to a pre-assigned higher level, the postponed work at server, serving in unprotected stages will be lost for ever in the model-1 and will be interrupted in the model-2. Interrupted work is again postponed and wait at the head of the queue in the pool. After the interruption, when the continuously served customers from the buffer

reaches a pre-assigned number N, at a service completion epoch, the service of interrrupted customer will suddently repeat. We study its long run behaviour and obtained certain system performance measures. Several numerical illustrations are also provided.

In chapter 7, we compare the performance of all the models discussed through chapters 2 to 6. Concluding remarks and some further possible investigations are also included.

Chapter 2

An $M/PH/1$ Queue with Postponed work under N-policy

In many practical situations a work may be postponed for several reasons. Queueing theory deals with a variety of postponement of work dealing with multi priority system. When a higher priority customer arrives, service of lower priority customers waiting in the line, may be postponed. In the case of bulk service, postponement due to lack of quorum can happen. On many occassions, postponement of a work may be to attened a more important job or to go on vacation. Postponement of work may depend

Some results of this chapter are included in the following paper.
1. A.Krishnamoorthy, C.B.Ajayakumar, P.K.Pramod, An $M/PH/1$ Queue with Postponed work under N-policy (Communicated)

on both input and service process.

Even though postponement of work is not desirable, it turns out to be unavoidable in many real life situations. So naturally a question arises: how to bring the postponed work in the usual service track? In this chapter we consider the situation of postponement due to the finiteness of the buffer. If a customer on arrival, finds the buffer not full, it joins the same. Otherwise it proceeds to a pool of postponed work having infinite capacity, with a specified probability, or else leaves the system permanently. So a customer can decide whether to join the pool or not, according to the service process of the system. Here we emphasize that customers arriving when there are fewer customers in the buffer than its capacity, are not subjected to postponement. They wait for service in the buffer in the usual manner. So naturally the pool is occupied by postponed work. Pooled customers are transferred to the buffer with a known probability at a service completion epoch, if the number in the buffer at that time is less than a pre-assigned level. This transferred customer is positioned as the last among the waiting units. If there is no customer left in the buffer at a service completion epoch, and at least one is in the pool, the one at the head of the pool is transferred to the buffer with probability one for immediate service.

Models with postponement are an alternative to finite capacity queues in which overflow jobs are irrevocably lost. With this in view, a paper to deal with postponed work was introduced by Deepak et.al. (see [16]). They analysed such a system in great detail in the stationary case and provided a number of system performance measures. No further development in this is reported so far. Nevertheless this notion of postponement of work has been introduced into inventory by a few researchers (see Krish-

namoorthy and Islam [36], Arivarignan et.al. [2], Paul Manuel et.al. [46], Sivakumar and Arivarignan [48]). Here we extend the model described in [16] by introducing N-policy for customers from the pool as follows: If the pool contains at least one postponed work, continuously served customers from buffer since the last transfer under N-policy is counted at each service completion epoch. When it reaches a pre-assigned number N, then the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. The model discussed can be used to design queueing systems to minimise loss due to customers not joining the system when the buffer is full. In finite capacity queues, blocked customers are lost to the system for ever. However in the present model, the introduction of the N-policy, reduces the waiting time of pooled customers and this is an incentive for customers to join the pool when the buffer is full. Immediate commencement of service to the transferred customer from the pool under N-policy adds to the attraction of joining the pool. This fraction can be increased by suitably designing the system. A diagramatic representation of the model is given in figure 2.1.

Remark 2.0.1. The N-policy mentioned here differs from the classical N-policy. In the classical case, N customers are to queue up to start the new service cycle once the system becomes empty; it is a control policy. However in the present case N-policy is applied to determine a priority service to be given to a customer from the pool. This again is a control policy; nevertheless it counts the number of continuously served customers from the buffer.

In real life situations, the model described in this chapter is working quite naturally. A customer on arriving at a service station, seeing the server is busy with a specified number of works, may go to another sevice

station if he has emergency to get service. But if the customer is not bothered about the waiting time and his importance lies in the service of that particular service station, he can register there in a pool of postponed work.

2.1 Mathematical description

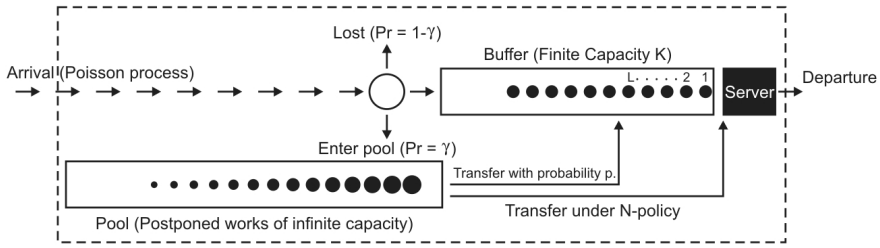

Fig 2.1: $M/PH/1$ queue with postponed work under N-policy

Consider an $M/PH/1$ queue with finite buffer of capacity K. If the buffer contains less than K customers including the one at server, newly arriving customers will join it. When the buffer is full with K customers, newly arriving jobs are not necessarily lost. They are offered the choice of leaving the system immediately or of being postponed until the system is less congested. That is a customer can accept the offer of postponement with probability γ $(0 \leqslant \gamma < 1)$. So he may join a pool of postponed work of infinite capacity. With probability $1-\gamma$, such customers do not join the system. In the absence of the pool the system behaves like the familiar $M/PH/1/K$ queue. When at the end of a service, if there are postponed

customers, the system operates as follows. If the buffer is empty, the one ahead of all waiting in the pool gets trasferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$ at a service completion epoch, then again the job at the head of the buffer starts service and with probability p, the head of the queue in pool is transferred (we call this a p-transfer) to the finite buffer and positioned as the last among the waiting customers in the buffer. With probability $q = 1 - p$, no such transfer takes place. No such transfer takes place at a service completion epoch if there is atleast L customers in the buffer. Also if the pool contain at least one postponed job, the continuously served customers from the buffer since the last transfer under N-policy is counted, at each service completion epoch. When it reaches N, $(N > 0)$ then the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. At this time, system does not consider the p-transfer. To be specific, if at a service completion epoch, if number of customers in the buffer is less than L and the number of continuously served customers from the buffer has reached N, then the transfer under N-policy is given preference.

Remark 2.1.1. It may be noted that the N-policy leads to violation of FIFO rule for customers in the pool. For example assume that there are two or more customers in the pool at a service completion epoch at which the number in the buffer droped to $L-1$ or below and the number of continuously served customers reached $N-1$. So the first in the pool may be selected under p-transfer and placed as the last in the buffer. When the next service is completed, the current head of the pool gets transferred to the buffer for immediate service there by violating the FIFO rule for pooled customers. Further it may be noted that this situation does not arise among the queued customers in the buffer. The probability of FIFO

violation among customers from the pool is calculated in section 2.3.5.

Customers arrive according to a homogeneous Poisson process of rate λ. The duration of the successive services, whether of regular or of postponed customers, are independent and identically distributed with the service time distribution following Phase Type(PH). Here the PH distribution has the irreducible representation $(\beta\ , S)$. There are m phases and the vector $S^0 = -Se$ containing elements S_{h0} denoting the absorption rate from the phase h, $h = 1, 2, ..., m$. Absorption (service completion) occurs with probability 1 from any phase i in $\{1, 2, ...m\}$ if and only if the matrix S is nonsingular (see [45]). Then the mean time until absorption is $-\beta S^{-1}e$. Also the equilibrium distribution of the excess life is $PH(\pi^*, S)$ where π^* is the stationary probability vector satisfying $\pi^* Q^* = 0$ and $\pi^* e = 1$ where $Q^* = S + S^0\beta$ (see [22]).

The model is studied as a Quasi Birth-Death(QBD) process and a solution of the classical matrix geometric type is obtained (see [45] and [38]). We define the state space of the QBD and exhibit the structure of its infinitesimal generator.

The state space consists of all tuples of the form (i, j, b, h) with $i \geqslant 1$; $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant\ m$, where i is the number of postponed work, j is the number of work in the finite buffer including the unit in service, b is the number of continuously served customers from the buffer at a service completion and h is the phase of the service in progress at a time t. For a given value of i, $K(N + 1)m$ states constitute the level i of the QBD. Now consider the boundary level $i = 0$. Then we denote the empty system $(0, 0, 0, 0)$ by 0. Also there are Km states of the form $(0, j, 0, h)$, $1 \leqslant j \leqslant K$; $1 \leqslant h \leqslant m$. This is due to the fact that when

the pool has no customers, N-policy is suspended. These have the same significance as before, except that in these states, no postponed jobs are present, but there are jobs in the finite buffer. These $Km+1$ states make up the boundary level 0 of the QBD.

The infinitesimal generator of the QBD describing the $M/PH/1/K$ queue with postponed customers under N-policy is of the form

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

where the matrix B_0 is of dimension $(Km+1) \times K(N+1)m$, B_1 is square matrix of order $Km+1$ and B_2 is of dimension $K(N+1)m \times (Km+1)$. A_0, A_1 and A_2 are square matrices of order $K(N+1)m$. Each of these matrices is itself highly structured.

The matrix B_1 corresponds to the transition from the level 0 to 0 is given below, where I is the identity matrix of order m and all non specified

entries are zeros:

$$B_1 = \begin{bmatrix} -\lambda & \lambda\beta & & & & & & & \\ S^0 & S-\lambda I & \lambda I & & & & & & \\ & S^0\beta & S-\lambda I & \lambda I & & & & & \\ & & S^0\beta & S-\lambda I & \lambda I & & & & \\ & & & \ddots & \ddots & \ddots & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & S^0\beta & S-\lambda I & \lambda I & \\ & & & & & & S^0\beta & S-\lambda I \end{bmatrix}.$$

Except for a single block $\lambda\gamma t_5 \otimes I_m$ at its south- east corner, the matrix B_0 is zero, where t_5 is a row vector of order $N+1$ with first element 1 and all other elements zero with I_m representing identity matrix of order m. The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \bar{0} & diag\Big(H_1, & H_2, & \cdots, & H_L, & H_{L+1}, & \cdots, & H_K \Big) \end{bmatrix}$$

where $\bar{0}$ is zero matrix of appropriate order and $diag(H_1, H_2, ..., H_L, H_{L+1}, ..., H_K)$ represents a diagonal block matrix of order K with diagonal block entries $H_1 = t_6 \otimes S^0\beta$, $H_2 = ... = H_L = t_7 \otimes S^0\beta$, $H_{L+1} = ... = H_K = t_8 \otimes S^0\beta$ and t_6 is a column vector of order $N+1$ with all entries are 1, t_7 is a column vector of order $N+1$ with all elements are p except a 1 at $(N,1)^{th}$ position, t_8 is a column vector of order $(N+1)$ with $(N,1)^{th}$ element is 1 and all other elements zero.

The matrix A_0 is zero except for a single block $\lambda\gamma I_{N+1} \otimes I_m$ at its south- east corner where I_{N+1} is the identity matrix of order $N+1$. The

matrix A_2 is given by

$$A_2 = diag\Big(\ \Lambda_1,\ \ \Lambda_2,\ \ \cdots,\ \ \Lambda_{_L},\ \ \Lambda_{L+1},\ \ \cdots,\ \ \Lambda_K\ \Big).$$

It denotes diagonal block matrix with block entries on main diagonal given by $\Lambda_1 = t_1 \otimes S^0\beta$, $\Lambda_2 = ... = \Lambda_L = t_2 \otimes S^0\beta$, $\Lambda_{L+1} = ... = \Lambda_K = t_3 \otimes S_0\beta$ where t_1 is a square matrix of order $N+1$, given by

$$t_1 = \begin{bmatrix} \bar{0} & I_N \\ 1 & \bar{0} \end{bmatrix}$$

where I_N is identity matrix of order N. t_2 is a square matrix of order $N+1$ given by

$$t_2 = \begin{bmatrix} 0 & p & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ p & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

and t_3 is a square matrix of order $N+1$ with $(N, N+1)^{th}$ entry 1 and all

other entries zero. The matrix A_1 is given by

$$A_1 = \begin{bmatrix} \zeta & \Omega & & & & & & \\ \Theta_1 & \zeta & \Omega & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \Theta_1 & \zeta & \Omega & & & \\ & & & \Theta_2 & \zeta & \Omega & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \Theta_2 & \zeta & \Omega \\ & & & & & & \Theta_2 & \eta \end{bmatrix}$$

where ζ corresponds to the transition of the buffer size from j to j for $j = 1, 2, ..., K-1$ and $\zeta = I_{N+1} \otimes (S - \lambda I_m)$; η corresponds to the transition of the buffer size from K to K where $\eta = I_{N+1} \otimes (S - \lambda\gamma I_m)$; Ω corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, ..., K-1$ and $\Omega = \lambda I_{N+1} \otimes I_m$; Θ_1 corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., L$ and $\Theta_1 = t_4 \otimes qS^0\beta$; Θ_2 corresponds to the transition of the buffer size from j to $j-1$ for $j = L+1, L+2, ..., K$ and $\Theta_2 = t_4 \otimes S^0\beta$. Also t_4 is a square matrix of order $N+1$ which is given below:

$$t_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

2.2 Analysis of the system

2.2.1 Stability criterion

Theorem 2.2.1. *The system is stable if and only if*

$$\lambda\gamma \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}.$$

Proof. Here we obtain the first passage time probability (fundamental period) from a level i to the level $i-1$ for $i \geqslant 1$ (see [45]).

Let $G_{ll'}(k,x)$ be the conditional probability that the QBD process starting in the state $l=(i,j,b,h)$ (for $i>1$) where $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant m$ at time $t=0$ reaches the state $l'=(i-1,j',b',h')$ where $1 \leqslant j' \leqslant K$; $0 \leqslant b' \leqslant N$; $1 \leqslant h' \leqslant m$, for the first time, involving exactly k transitions and completing before time x. That is

$$G_{ll'}(k,x) = P[\tau < \infty : \chi(\tau) = l' | \chi(0) = l]$$

where τ is the first passage time from the level i to the level $i-1$ and χ is the discussed QBD process. Because of the structure of Q, the probability $G_{ll'}(k,x)$ does not depend on i. The matrix with elements $G_{ll'}(k,x)$ is denoted by $G(k,x)$.

Now introduce the transform matrix,

$$\hat{G}(z,\theta) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-\theta x} dG(k,x)$$

for $|z| \leqslant 1$, $\theta > 0$. The matrix $\hat{G}(z,\theta)$ satisfies the matrix equation

$$\hat{G}(z,\theta) = z(\theta I - A_1)^{-1} A_2 + (\theta I - A_1)^{-1} A_0 \hat{G}^2(z,\theta).$$

Use the notations $C_0(\theta) = (\theta I - A_1)^{-1} A_2$ and $C_2(\theta) = (\theta I - A_1)^{-1} A_0$. Now the transform matrix $\hat{G}(z,\theta)$ is equal to the minimal non negative solution of the matrix quadratic equation

$$X(z,\theta) = zC_0(\theta) + C_2(\theta) X^2(z,\theta)$$

and it is obtained by successive substitutions starting with the zero matrix. Also we have

$$\lim_{z \to 1, \theta \to 0} \hat{G}(z,\theta) = G(k,x) = [G_{ll'}(k,x)].$$

Then G is obtained as the minimal non negative solution to the equation $G = C_0 + C_2 G^2$ where $C_0 = (-A_1)^{-1} A_2$ and $C_2 = (-A_1)^{-1} A_0$. That is, G is the minimal non negative solution of the matrix quadratic equation

$$A_2 + A_1 G + A_0 G^2 = 0.$$

The matrix G can be computed by using the logarithmic reduction algorithm.

Let $m_1 = [m_{1_l}]$ denotes the column vector of dimension $K(N+1)m$

where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started in the state l. Then,

$$m_1 = \left[-\frac{\partial}{\partial\theta}\hat{G}(z,\theta)e\right]_{\theta=0,z=1} = -(A_1 + A_0(I+G))^{-1}e.$$

Suppose the matrix $A = A_0+A_1+A_2$ is irreducible. Then the necessary and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0+A_1+A_2$. That is, it is the unique solution to $\pi A = 0$, $\pi e = 1$ and $A = A_0 + A_1 + A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. That is for the system stability, the rate of drift from level i to level $i-1$ should be greater than that to level $i+1$. The rate of drift from the level i to the level $i+1$ is given by $\lambda\gamma\sum_{b=0}^{N}\sum_{h=1}^{m}\pi_{Kbh}$ and the rate of drift from the level i to the level $i-1$ is given by $\frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}$.

It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to

$$\lambda\gamma\sum_{b=0}^{N}\sum_{h=1}^{m}\pi_{Kbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}.$$

□

So by an appropriate choice of γ, that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate

is greater than service rate.

2.2.2 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of Q be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{2.1}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0 \tag{2.2}$$

and the vectors x_0, x_1 are obtained by solving the equations

$$x_0 B_1 + x_1 B_2 = 0 \tag{2.3}$$

$$x_0 B_0 + x_1 (A_1 + RA_2) = 0 \tag{2.4}$$

subject to the normalising condition

$$x_0 e + x_1 (I - R)^{-1} e = 1. \tag{2.5}$$

From the above discussion it is clear that to determine x, a key step is the computation of the rate matrix R. Although there exist several algorithms for computing R, we use logarithmic reduction algorithm (see [38]), which is considered to be the most efficient one among the existing algorithms. The important steps of this algorithm is given below.

Assign $H := (-A_1)^{-1}A_0$; $L := (-A_1)^{-1}A_2$; $G := L$; and $T := H$;
and repeat
$U := HL + LH$; $M := H^2$; $H := (I - U)^{-1}M$; $M := L^2$;
$L := (I - U)^{-1}M$; $G := G + TL$; $T := TH$
until $||1 - G.e||_\infty \leqslant \epsilon$.
Then $R = -A_0(A_1 + A_0G)^{-1}$
Note that here, due to the special structure of the coefficient matrices A_0, A_1 and A_2 occuring in equation 2.2, the matrix R of order $K(N+1)m$ of the form

$$R = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ R_1 & R_2 & \cdots & R_k \end{bmatrix}$$

where each R_i, $1 \leqslant i \leqslant K$, is a square matrix of order $(N + 1)m$. We partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, x_{i3}, ..., x_{iK})$$

where $i \geqslant 1$ and x_{00} is a scalar; x_{0j}, $1 \leqslant j \leqslant K$, are vectors of order m and

$$x_{ij} = (x_{ij0}, x_{ij1}, ..., x_{ijN})$$

where $i \geqslant 1$; $1 \leqslant j \leqslant K$ and x_{ijb}, $0 \leqslant b \leqslant N$ are vectors of order m.

2.3 Computation of Expected values

In this section we derive the expected waiting time of a tagged customer (i) in the buffer and (ii) in the pool, the expected duration (i) between two consecutive transfers under N-policy (ii) for the first N-policy transfer in a busy cycle and the expectation of FIFO violation.

2.3.1 Expected waiting time in buffer

We denote the mean waiting time of customers who upon their arrival enter the buffer by $E(W_1)$.

Case 1. $N \geqslant K$

In this case the tagged customer is not affected by the new arrivals in buffer and in pool. So we can calculate the waiting time by considering the system state at which the tagged customer enters. Hence

$E(W_1) = \sum_i \sum_j \sum_b \sum_h E(\text{waiting time of the customer who finds the system in state } (i,j,b,h))\ Pr(\text{system is in state } (i,j,b,h))$

$$E(W_1) = \sum_{j=1}^{k-1} \sum_{h=1}^{m} -\beta S^{-1} e(j-1) x_{0j0h}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{b=0}^{N-1} \sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi) x_{ijbh}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{h=1}^{m}-\beta S^{-1}e(j-1)x_{ijNh}-\pi^{*}S^{-1}e$$

where $\pi^{*}Q^{*}=0$, $\pi^{*}e=1$ and $Q^{*}=S+S^{0}\beta$ and

$$\psi=1+\left[\frac{j-(N-b)}{N}\right],\quad 0\leqslant b<N$$

where [y] denotes the greatest integer value of y. $-\pi^{*}S^{-1}e$ is the additional time required to complete the service of the customer who is at the server when the tagged person enters the buffer.

Remark 2.3.1. In $M/M/1$ case with service rate μ,

$$E(W_1)=\sum_{j=1}^{K-1}\tfrac{1}{\mu}jx_{0j0}+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1}\tfrac{1}{\mu}(j+\psi)x_{ijb}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\frac{1}{\mu}jx_{ijN}$$

where

$$\psi=1+\left[\frac{j-(N-b)}{N}\right],\quad 0\leqslant b<N.$$

Case 2. $N<K$

In this case, the tagged customer in the buffer will be affected by the number of new arrivals in the pool and so the number of new arrivals in the buffer. So the waiting time of the tagged customer depends on the following susequent developments in the pool: one or more visits to zero level, and a finite number of customers joining the pool after the tagged customer. Because of the complexity of calculation, we may turn

to computing an upper bound on the waiting time, by keeping in mind, the fact that only a maximum finite number K of persons in the pool will affect the tagged person. In the worst case we have $N = 1$ which represents service alternating between buffer and pool. So an upper bound for the waiting time of a customer who upon his arrival enters the buffer in the state (i, j, b, h), is

$$UB(W_1) = \sum_{j=1}^{k-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+[\tfrac{j}{N}])x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi)x_{ijbh}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+[\frac{j-1}{N}])x_{ijNh} - \pi^* S^{-1} e$$

where $\pi^* Q^* = 0$, $\pi^* e = 1$ and $Q^* = S + S^0\beta$ and

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

$-\pi^* S^{-1} e$ is the additional time required to complete the service of the customer who is at the server when the tagged person enter buffer.

Remark 2.3.2. In $M/M/1$ case with service rate μ,

$$UB(W_1) = \sum_{j=1}^{K-1} \tfrac{1}{\mu}(j+[\tfrac{j}{N}])x_{0j0} + \sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1} \tfrac{1}{\mu}(j+\psi)x_{ijb}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1} \frac{1}{\mu}(j+[\frac{j-1}{N}])x_{ijN}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

2.3.2 Expected waiting time in pool

We denote the expected waiting time of a customer who upon his arrival enters the pool, by $E(W_2)$.

To find this, first we define the Markov process $\{X(t)\}$ as follows. $X(t) = (a, j, b, h)$ where a denotes the rank of the tagged customer entered pool, j denotes the number of customers in the buffer, b denotes the number of continuously served customers from buffer and h is the phase of the service process at time t. The rank a of the customer is assumed to be r if he joins as r^{th} customer in pool. His rank may decrease to 1 with the customers ahead of him transferred from the pool to the buffer. Since the customers who arrive after the tagged customer cannot change his rank, level changing transitions in $\{X(t)\}$ can takeplace only to one side of the diagonal. We arrange the statespace of $\{X(t)\}$ as

$$\{r, r-1, ..., 2, 1\} \times \{1, 2, ..., K\} \times \{0, 1, ..., N\} \times \{1, 2, ..., m\}$$

with absorbing state 0 in the sense that the tagged customer is either selected to be served under N-policy or placed in the buffer with probability p or to the server with probability 1 if the buffer size reduces to 0 at the end of a service. The infinitesimal generator of the process is

$$\tilde{Q} = \begin{bmatrix} T & T^0 \\ \bar{0} & 0 \end{bmatrix}$$

where

$$T = \begin{bmatrix} A_1 & A_2 & & & & \\ & A_1 & A_2 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & A_1 & A_2 \\ & & & & & A_1 \end{bmatrix}$$

of order $rK(N+1)m$ and

$$T^0 = \begin{bmatrix} \bar{0} \\ \vdots \\ \bar{0} \\ B_2 \end{bmatrix}.$$

Now the expected absorption time of a particular customer is given by the column vector

$$E_w^{(r)} = -\tilde{I}T^{-1}e$$

where $\tilde{I} = \begin{bmatrix} I_{K(N+1)m} & \bar{0} \end{bmatrix}$ having order $K(N+1)m \times rK(N+1)m$ and e is a column vector of ones of order $rK(N+1)m$. So the expected waiting time of the tagged customer is

$$W_L = \sum_{r=1}^{\infty} x_r E_w^{(r)}$$

where x_r is the steady state probability vector corresponding to $i = r$. W_L gives the waiting time of a customer in the pool up to the epoch of his transfer to the buffer.

Case 1. $N \geqslant K$

Expected waiting time in pool is

$$E(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L(x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+ \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh}(W_L + W^{(1)} p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$W^{(1)} = \sum_{j=1}^{L}\sum_{h=1}^{m} -\beta S^{-1} e(j-1) x_{0j0h}$$

$$+ \sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi) x_{ijbh}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

Case 2. $N < K$

In this case we get an upperbound $UB(W_2)$ for the waiting time in pool.

$$UB(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L(x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+ \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh}(W_L + UB(W^{(1)}) p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$UB(W^{(1)}) = \sum_{j=1}^{L}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\left[\frac{j-1}{N}\right])x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi)x_{ijbh}$$

and

$$\psi = 1+\left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

2.3.3 Expected duration between two consecutive transfers under N-policy

For computing expected duration between two consecutive transfers under N- policy, we consider the Markov process $\{X(t)\}$ described as follows. $X(t) = (b, i, j, h)$ where b is the number of continuously served customers from the buffer, if pool has at least one person, at time t, measured from the service completion of the last customer who was transferred under N-policy. So $b = 0, 1, 2, ..., N$. Here we regard $0, 1, 2, ..., N-1$ as transient states and N as absorbing state (that is the state at which a new N-policy transfer occurs). i denotes the number of postponed jobs at time t. Even if pool is of infinite capacity, we restrict here it to be a finite value say V for sufficiently large V. So $i = 0, 1, 2, ..., V$; $j(= 0, 1, 2, ..., K)$ denotes the number of customers in the buffer at time t. Also $h = 1, 2, ..., m$ denotes the phase of the service in progress at a time t. The process $\{X(t)\}$ has the state space

$$\{0, 1, 2, ..., N-1, N\} \times \{0, 1, 2, .., V\} \times \{0, 1, 2, ..., K\} \times \{1, 2, 3, ..., m\}$$

Generator of the process is

$$\hat{Q} = \begin{bmatrix} U & U^0 \\ \bar{0} & 0 \end{bmatrix}$$

where

$$U = \begin{bmatrix} C_0 & D_0 & & & & & \\ C_1 & C_3 & D_1 & & & & \\ C_1 & & C_3 & D_1 & & & \\ C_1 & & & C_3 & \ddots & & \\ \vdots & & & & \ddots & \ddots & \\ C_1 & & & & & \ddots & D_1 \\ C_2 & & & & & & C_3 \end{bmatrix}_{(NVKm+Km+1)\times(NVKm+Km+1)}$$

and

$$U^0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ D_2 \end{bmatrix}$$

$$C_0 = \begin{bmatrix} E_1 & E_4 & & & & \\ & E_2 & E_5 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & E_2 & E_5 \\ & & & & & E_3 \end{bmatrix}_{(VKm+Km+1)\times(VKm+Km+1)}$$

where E_1 is of order $(Km+1)\times(Km+1)$ and

$$E_1 = \begin{bmatrix} -\lambda & \lambda\beta & & & & & \\ S^0 & S-\lambda I_m & \lambda I_m & & & & \\ & S^0\beta & & & & & \\ & & & \cdots & & & \\ & & & & S-\lambda I_m & \lambda I_m & \\ & & & & S^0\beta & S-\lambda\gamma I_m & \end{bmatrix}$$

$$E_2 = \begin{bmatrix} S-\lambda I_m & \lambda I_m & & & & \\ & \cdots & \cdots & & & \\ & & \cdots & \cdots & & \\ & & & S-\lambda I_m & \lambda I_m & \\ & & & & S-\lambda\gamma I_m & \end{bmatrix}_{Km\times Km}$$

$$E_3 = \begin{bmatrix} S-\lambda I_m & \lambda I_m & & & & \\ & \cdots & \cdots & & & \\ & & \cdots & \cdots & & \\ & & & S-\lambda I_m & \lambda I_m & \\ & & & & S & \end{bmatrix}_{Km\times Km}$$

$$E_4 = \begin{bmatrix} \bar{0} & \bar{0} \\ \bar{0} & \lambda\gamma I_m \end{bmatrix}_{(Km+1)\times Km} \qquad E_5 = \begin{bmatrix} \bar{0} & \bar{0} \\ \bar{0} & \lambda\gamma I_m \end{bmatrix}_{Km\times Km}$$

Also

$$C_1 = \begin{bmatrix} E_6 & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}_{(KVm)\times(KVm+Km+1)}$$

$$C_2 = \begin{bmatrix} E_7 & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}_{(KVm)\times(KVm+Km+1)}$$

where

$$E_6 = \begin{bmatrix} 0 & S^0\beta & \bar{0} & \bar{0} \\ \bar{0} & \bar{0} & I_{L-1} \otimes pS^0\beta & \bar{0} \\ \bar{0} & \bar{0} & \bar{0} & \bar{0} \end{bmatrix}_{Km \times (Km+1)}$$

$$E_7 = \begin{bmatrix} \bar{0} & I_K \otimes S^0\beta \end{bmatrix}_{Km \times (Km+1)}$$

$$E_8 = \begin{bmatrix} \bar{0} & \bar{0} \\ I_{L-1} \otimes qS^0\beta & \bar{0} \\ \bar{0} & I_{K-L} \otimes S^0\beta \end{bmatrix}_{Km \times Km}$$

$$E_9 = \begin{bmatrix} S^0\beta & \bar{0} & \bar{0} \\ \bar{0} & I_{L-1} \otimes pS^0\beta & \bar{0} \\ \bar{0} & \bar{0} & \bar{0} \end{bmatrix}_{Km \times Km}$$

$$C_3 = \begin{bmatrix} E_2 & E_5 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & E_2 & E_5 \\ & & & & E_3 \end{bmatrix}_{KVm \times KVm}$$

$$D_1 = \begin{bmatrix} E_8 & & & \\ E_9 & \ddots & & \\ & \ddots & \ddots & \\ & & E_9 & E_8 \end{bmatrix}_{KVm \times KVm}$$

$$D_0 = \begin{bmatrix} \bar{0} \\ D_1 \end{bmatrix}_{KVm \times KVm} \qquad D_2 = \begin{bmatrix} \bar{0} & \bar{0} \\ I_{V-1} \otimes I_K \otimes S^0\beta & \bar{0} \end{bmatrix}_{KVm \times KVm}$$

where $\otimes$ denotes Kronecker product. The initial probability vector of $\hat{Q}$ is

$$\delta = \begin{bmatrix} \frac{1}{\sum_{r=0}^{KVm+Km} x_r} \begin{bmatrix} x_0 & x_1 & \cdots & x_{KVm+Km} \end{bmatrix} & \bar{0} \end{bmatrix}_{1 \times (NKVm+Km+1)}$$

where $x_0 = x_{0000}$. $x_r = x_{ij0h}, 0 \leqslant i \leqslant V; 1 \leqslant j \leqslant K; 1 \leqslant h \leqslant m$ and r varies from 1 to $KVm + Km$ according to its lexicographic order. Then we have the following lemma:

Lemma 2.3.1. *Expected duration between two consecutive transfers under N-policy follows PH distribution with representation (δ, U) and it is given by*

$$N_{ABSORB} = -\delta U^{-1} e.$$

Using this expected value, a cost function is defined in section 2.6 and the optimal value of N is determined.

2.3.4 Expected duration for the first N-policy transfer in a busy cycle

Here we compute the expected duration of the time elapsed from the epoch of the first arrival to an idle system until the first N-policy transfer is effected. This can be obtained as corollary to lemma 2.3.1.

Corollary 2.3.2. The time elapsed, starting with an arrival to an idle system, until the realization of the N-policy for the first time follows the PH-distribution with representation (α, U) where

$$\alpha = \frac{1}{\sum_{r=1}^{m} x_r} \begin{bmatrix} 0 & x_1 & x_2 & \cdots & x_m & \bar{0} \end{bmatrix}_{1\times(NKVm+Km+1)}$$

where $x_r = x_{010h}; 1 \leqslant h \leqslant m$, r varies from 1 to m and U is described in section 2.3.3.

Proof. At the epoch of the first arrival to an idle system, process starts with the service in one of the m phases with steady state probability $x_r = x_{010h}; 1 \leqslant h \leqslant m$, r varies from 1 to m. This justifies the form of the initial probability vector α as given above. □

Then the expected duration for the realization of the above random variable is

$$N_{FIRST} = -\alpha U^{-1} e.$$

2.3.5 Expected number of FIFO violation

Next we compute the expectation of the indicator random variable defined as FIFO violation in pool as explained in remark 2.1.1. Its expectation is the probability for FIFO violation in pool which is given by

$$P_{FIFO} = \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=N-j+1}^{N-1} x_{ijbh} p S_{h0}.$$

The FIFO may be violated by more than one customers who join the pool after the tagged customer joins the buffer when $N < L$. However this can be overcome by making N large than L. If $N \geqslant K$, a customer joining the pool will not overtake any of the customers in the buffer who had joined before his entering the pool. At this time, FIFO is violated by atmost one successor in pool. Even this can be overcome by a slight modification by redefining the N-policy by resetting b in (i, j, b, h) as zero at the time of p-transfer.

2.4 Performance measures

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{K} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{ijbh}$$

for $i \geqslant 1$ and

$$a_0 = x_{00} + \sum_{j=1}^{K} \sum_{h=1}^{m} x_{0j0h}.$$

2. The probability that there are j customers in the buffer (including the one in service)is

$$b_j = \sum_{h=1}^{m} x_{0j0h} + \sum_{i=1}^{\infty} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{ijbh}$$

for $1 \leqslant j \leqslant K$ and

$$b_0 = x_{00}.$$

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that a customer, on its arrival enters the pool is γb_K.

6. The probability that an arriving customer enters service immediately is b_0.

7. The rate at which the customer who finds the buffer full leave the system without service (mean number of customers not joining the system per unit time) is

$$\theta_{LOST} = \lambda(1 - \gamma) b_K.$$

That is

$$\theta_{LOST} = \lambda(1 - \gamma)(\sum_{h=1}^{m} x_{0K0h} + \sum_{i=1}^{\infty} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{iKbh}).$$

8. The rate at which pooled customers transfer in to the buffer is

$$\theta_{TR} = \sum_{i=1}^{\infty} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{i1bh} S_{h0} + \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=0}^{N-2} \sum_{h=1}^{m} x_{ijbh} p S_{h0}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0} + \sum_{i=1}^{\infty}\sum_{j=2}^{L}\sum_{h=1}^{m} x_{ijNh}pS_{ho}.$$

9. The rate at which pooled customers transfer under N-policy(mean number of transfers under N-policy per unit time) is

$$T_N = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0}.$$

10. Mean number of customers served out per unit time is

$$\mu_{SERVED} = (1 - b_o)\,\frac{1}{-\beta S^{-1}e} \quad .$$

2.5 Numerical results

We present some numerical results in order to illustrate the performance of the system. Take

$$\gamma = \frac{Lp}{K} + \frac{1}{N}$$

in order to bring out explicitly the dependence of γ on the system parameters.

This is justified as follows. Larger the L value, the customer encountering the buffer full, will be inclined to join the pool with higher probability. Also same is the relationship of γ with p. On the other hand, γ inversely varies with K. The additional term $\frac{1}{N}$ comes through N-policy. Here as N increases γ decreases so that γ and N vary inversely. But the relationship is feasible for those values of L, p, K and N such that $0 \leqslant \gamma \leqslant 1$. This is

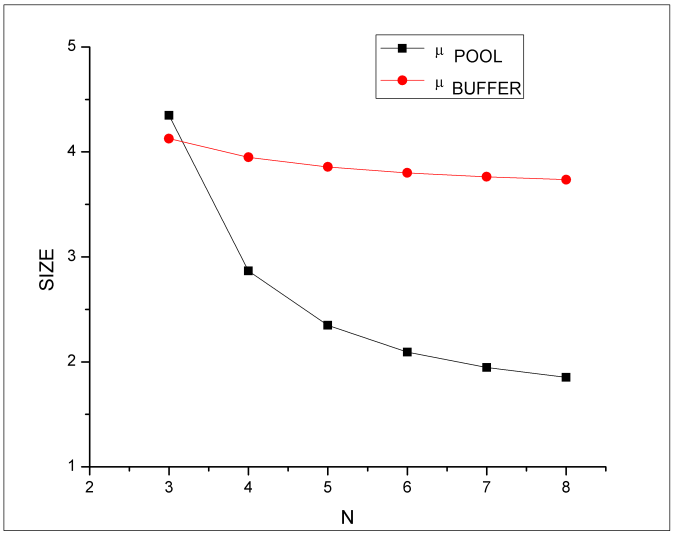

Fig 2.2: N versus μ_{POOL} and μ_{BUFFER}

possible if $N \geqslant K$ and such a selection is highly consistent. But N can be made less than K by suitably selecting other variables so that $0 \leqslant \gamma \leqslant 1$, and that can be considered as an incentive to customers joining the pool.

The impact of N on various measures of descriptors with $K = 6, L = 3, m = 2, \lambda = 7, p = 0.5, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \quad S = \begin{bmatrix} -12.5 & 6.0 \\ 6.0 & -12.5 \end{bmatrix} \quad S^0 = \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix}$$

is shown in figure 2.2 and figure 2.3. As N decreases μ_{POOL}, μ_{BUFFER}, θ_{TR} increase monotonically whereas θ_{LOST} decrease monotonically. This is due to the fact that by our assumption γ varies inversely as N and as a result,

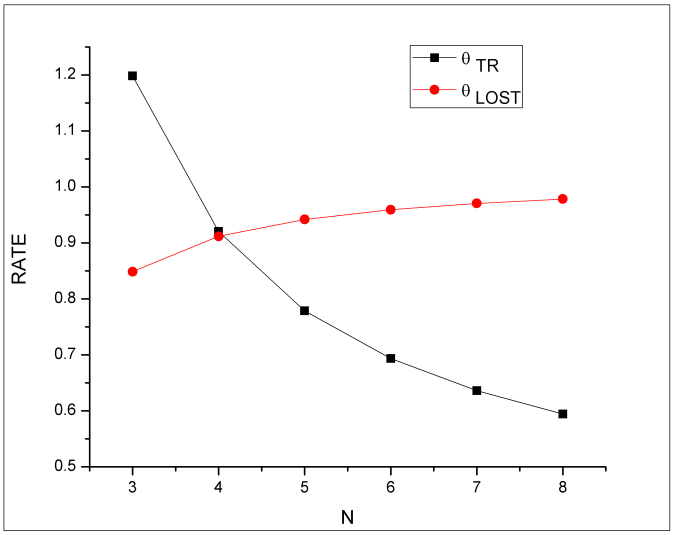

Fig 2.3: N versus θ_{TR} and θ_{LOST}

loss rate decreases and inflow rate to the pool increases as N decreases. As N decreases, transfer rate from pool to buffer increases, and thus mean buffer size increases. As a result, expected waiting time in buffer increases which is shown in figure 2.4 with $p = 0.5, \lambda = 7, K = 6, m = 2$.

By keeping $K = 6, L = 3, m = 2, \lambda = 7, N = 5, \gamma = \frac{Lp}{K} + \frac{1}{N}$ the effect of p on various measures is shown in figures 2.5 and 2.6. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing in p, as expected. The measures are numerically computed for various values of L and shown in table 2.1. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing as expected, in L. All the above are true due to the fact that by our assuption, γ varies directly as p and L. As a result, loss rate decreases

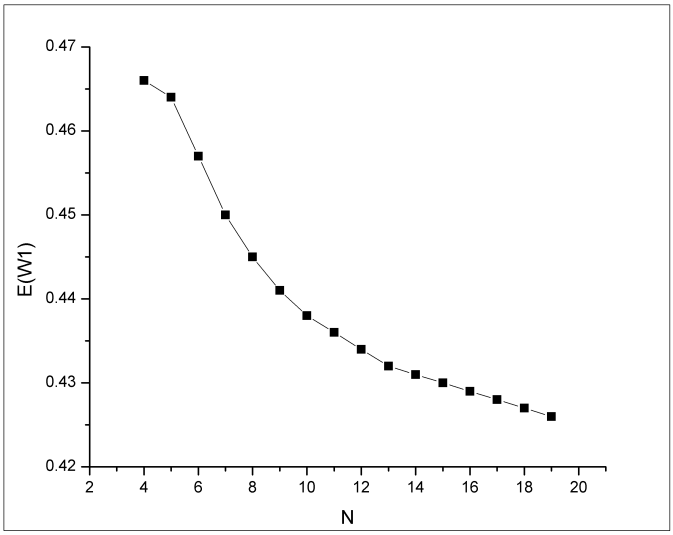

Fig 2.4: N versus expected waiting time in buffer

and inflow rate to the pool increases as p and L increases. This will make μ_{POOL} increasing. Also transfer rate from pool to buffer increases as p and L increases. So mean buffer size increases.

L	μ_{POOL}	μ_{BUFFER}	θ_{TR}	θ_{LOST}
2	1.4772367	3.6969533	0.5948937	1.0082242
3	2.3480656	3.8571582	0.7786083	0.9418701
4	3.2663956	4.0585852	1.0212065	0.8890664
5	3.8800666	4.2669754	1.3845636	0.8587388

Table 2.1: $K = 6, p = 0.5, m = 2, \lambda = 7, N = 5, \gamma = \frac{Lp}{K} + \frac{1}{N}$

Keeping service rate fixed and $p = 0.5, N = 3, L = 3, K = 6$ we can increase the arrival rate λ by reducing the value of γ (by assuming the

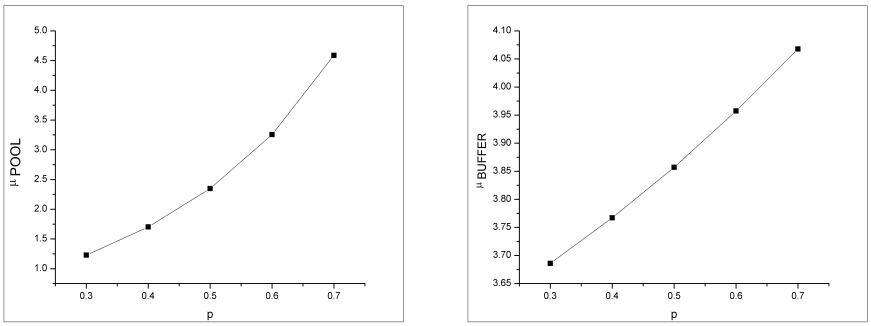

Fig 2.5: p versus μ_{POOL} and μ_{BUFFER}

independence of γ on L, p, K and N) so as to maintain the system stability. It can be seen that as γ tends to 0, arrival rate λ can approach ∞ for a fixed service rate as shown in figure 2.7, so as to satisfy the stability criterion given by the theorem 2.2.1 which is true, since this case results in the loss system $M/PH/1/K$ queue.

2.5.1 Comparison with model of Deepak et.al.[16]

Here we compare the present model with the model of Deepak et.al.[16] to emphasize the effect of N-policy. We call model of Deepak et.al.[16] as model 1 and the present one as model 2. The same numerical example for model 1 as given in Deepak et.al.[16] is taken here also. By keeping $K = 6, L = 3, m = 3, \lambda = 0.8, N = 3, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$\beta = \begin{bmatrix} 0 & 0.2 & 0.5 & 0.3 \end{bmatrix}, \quad S = \begin{bmatrix} -3 & 1 & 0.5 \\ 0.3 & -2 & 0.1 \\ 1 & 2 & -4 \end{bmatrix}, \quad S^0 = \begin{bmatrix} 1.5 \\ 1.6 \\ 1 \end{bmatrix},$$

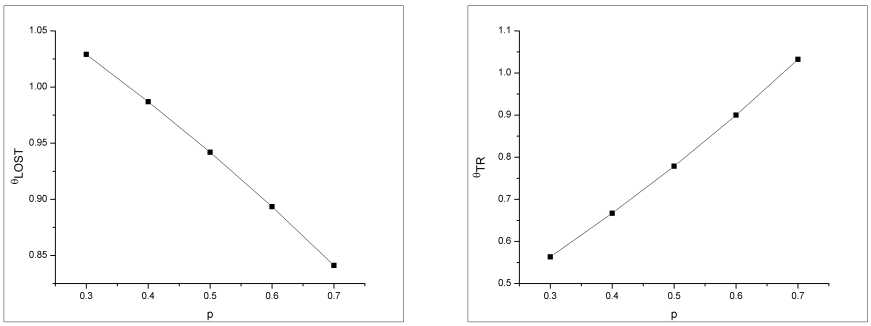

Fig 2.6: p versus θ_{LOST} and θ_{TR}

various measures in models 1 and 2 are plotted against p, and is shown in figures 2.8 and 2.9. From these figures it is clear that μ_{POOL}, μ_{BUFFER}, and θ_{TR} in model 2 are greater than that in model 1 and θ_{LOST} in model 2 is less than that in model 1 as expected which is a consequence of the N-policy. Tables 2.2 and 2.3 show the effect of L on various descriptors which are numerically computed for the models 1 and 2 by keeping $p = 0.5$. Here also μ_{POOL}, μ_{BUFFER}, and θ_{TR} in model 2 are greater than that in model 1 and θ_{LOST} in model 2 is less than that in model 1.

	μ_{POOL}		μ_{BUFFER}	
L	$Model1$	$Model2$	$Model1$	$Model2$
2	0.00753	0.01536	1.06938	1.08449
3	0.01154	0.01885	1.07122	1.08838
4	0.01572	0.02175	1.07298	1.09314
5	0.02007	0.02278	1.07466	1.09940

Table 2.2: Effect of L on μ_{POOL} and μ_{BUFFER} in models 1 and 2

Remark 2.5.1. The above $M/PH/1$ queue with postponed work under N policy can be approximated to other models by suitably fixing the

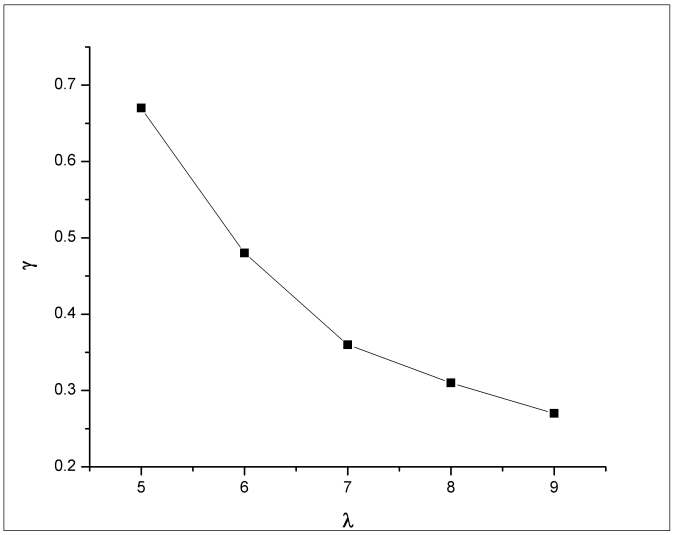

Fig 2.7: λ versus γ

variables.

If $\gamma \to 0$, it is $M/PH/1/K$ model.

If $N \to \infty$ then we get the model of Deepak et.al.[16]

If $\gamma \to 1$, $N \to \infty$, $p \to 1$, $L = K$ and $m = 1$ then it is $M/M/1/\infty$ model.

If $N = 1$ and $p \to 0$ then service alternates between buffer and pool.

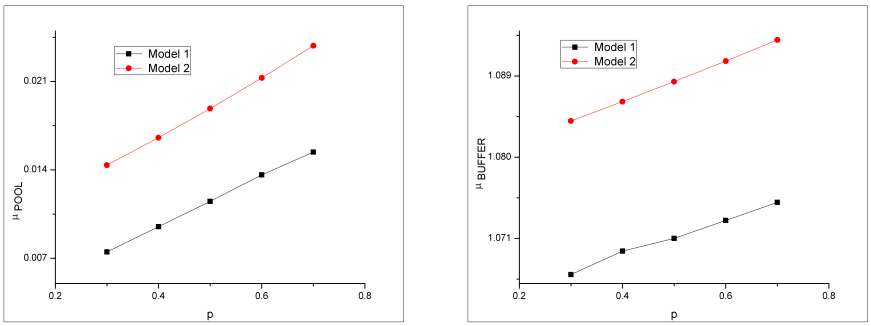

Fig 2.8: p versus μ_{POOL} and μ_{BUFFER} in models 1 and 2

	θ_{TR}		θ_{LOST}	
L	$Model1$	$Model2$	$Model1$	$Model2$
2	0.00145	0.00493	0.00723	0.00487
3	0.00217	0.00583	0.00652	0.00413
4	0.00290	0.00681	0.00580	0.00339
5	0.00363	0.00813	0.00508	0.00271

Table 2.3: Effect of L on θ_{TR} and θ_{LOST} in models 1 and 2

2.6 Cost function and determination of optimal N

Here we investigate the value of N which minimises a suitably defined cost function. The following important costs are included.

C_1 : Holding cost per customer per unit time in buffer

C_2 : Holding cost per customer per unit time in pool ($C_1 > C_2$)

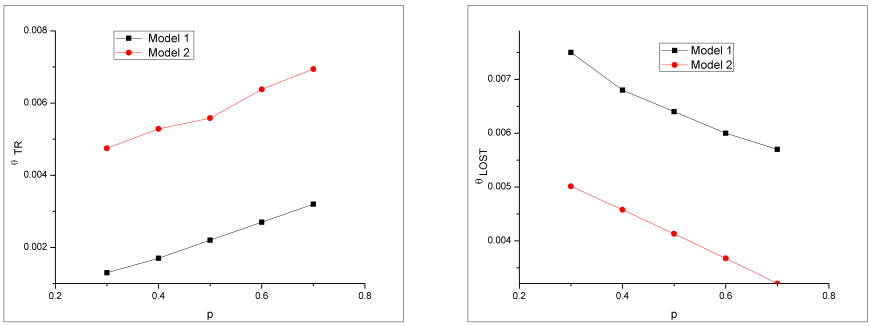

Fig 2.9: p versus θ_{TR} and θ_{LOST} in models 1 and 2

C_3 : Fixed cost of transfer of a customer from pool to buffer for immediate service by N-policy

Total expected cost $TC = C_1$ (mean buffer size) + C_2 (mean pool size) + C_3(1/ expected duration between two consecutive N-policy transfers)

That is

$$TC = C_1\ \mu_{BUFFER} + C_2\ \mu_{POOL} + C_3\ \ \frac{1}{N_{ABSORB}}$$

where

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j$$

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e$$

and

$$N_{ABSORB} = -\delta\ U^{-1}\ e.$$

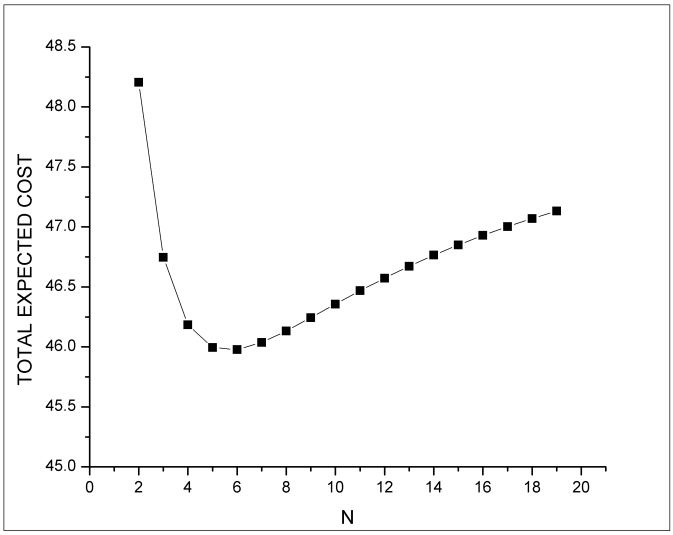

Fig 2.10: N versus total expected cost

In figure 2.10, total expected cost is plotted against N with $C_1 = 10, C_2 = 9, C_3 = 5, K = 6, L = 3, m = 2, \lambda = 7, p = 0.5, \gamma = Lp/K$,

$$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \quad S = \begin{bmatrix} -12.5 & 6.0 \\ 6.0 & -12.5 \end{bmatrix} \quad S^0 = \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix}$$

Figure 2.10 indicates that total expected cost of the system first decreases and then increases as N increases. For given parameters, it can be seen that beyond a certain value of N, this cost is asymptote to N axis. Thus we could anticipate a global minimum for the value of N.

Chapter 3

Modified $M/PH/1$ Queue with Postponed work under N-policy

In this chapter, we modify the model discussed in chapter 2. We consider a selection rule for an entry to the buffer if there is a vacancy in it. This can be analysed by giving a probability for each buffer entry. At the same situation, we may also take the interest of customers in to consideration. If a customer is not bothered about the waiting time and their importance lies in the service of that particular service station, the system will provide a pool of postponed work having infinite capacity. Such customers can

Some results of this chapter are included in the following paper.
1. A.Krishnamoorthy, C.B.Ajayakumar, Modified $M/PH/1$ Queue with Postponed work under N-policy (Communicated)

register there and wait as usual for a chance to get service from that particular service station. Here we emphasize that at each epoch, entry to pool is decided by the customer according to a specified probability law. When the buffer is full, entry to pool can also be restricted by the system with a probability, to retain system stability. So in this model, we consider the interest of both the server and the customers to reduce the total loss to the system.

Pooled customers are transferred to the buffer with a known probability at a service completion epoch, if the number in the buffer at that time is less than a pre-assigned level. This transferred customer is positioned as the last among the waiting units. If there is no customer left in the buffer at a service completion epoch, and at least one is in the pool, the one at the head of the pool is transferred to the buffer with probability one for immediate service. To work with N-policy if the pool contain at least one postponed work, continuously served customers from buffer is counted at each service completion epoch. When it reaches a pre-assigned number N, then the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. A diagramatic representation of the model is given in figure 3.1.

3.1 Mathematical Formulation

Consider an $M/PH/1$ queue with finite buffer of capacity K. If the system is empty, an arriving customer will join buffer and his service starts immediately. If the buffer has l persons where $1 \leqslant l \leqslant K-1$, then a newly arriving customer will be allowed to enter buffer with probability

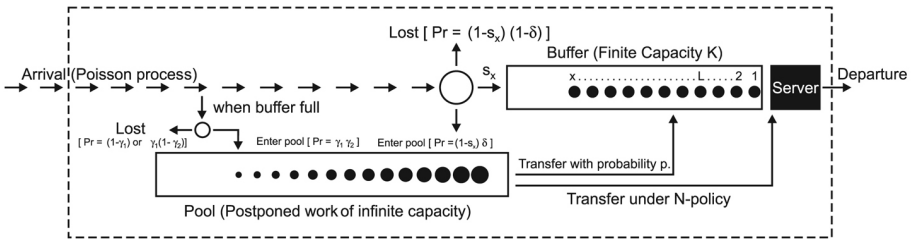

Fig 3.1: Modified $M/PH/1$ queue with postponed work under N-policy

s_l. We assume that as l increases, the probability s_l decreases. That is, $s_1 \geqslant s_2 \geqslant s_3 \geqslant .... \geqslant s_{K-1}$ with $s_0 = 1$ and $s_K = 0$. If a customer is not allowed to enter the buffer, he will be directed with probability $1 - s_l$ to a pool of postponed work having infinite capacity . In this case, however, the customer may decide to join the pool with probability δ or to leave the system forever with probability $1 - \delta$. So If the buffer has l persons where $1 \leqslant l \leqslant K - 1$, customers join buffer with rate λs_l, join pool with rate $\lambda(1-s_l)\delta$ and leave the system forever without getting service at rate $\lambda(1 - s_l)(1 - \delta)$. If the buffer is full, then the newly arriving customer decides to join pool, with probability γ_1 or to leave with probability $1-\gamma_1$. However in the former case, the server permits him to join pool with probability γ_2 or to decline admission with probability $1 - \gamma_2$. So in this case, a customer will join pool with rate $\lambda\gamma_1\gamma_2$ and will leave from the system with rate $\lambda(1 - \gamma_1) + \lambda\gamma_1(1 - \gamma_2)$.

When at the end of a service, if there are postponed customers, the system operates as described in chapter 2. That is if the buffer is empty,

the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$ at a service completion epoch, then again the job at the head of the buffer starts getting service and simultaneously with probability p the head of the queue in the pool is transferred to the buffer and positioned as the last among the waiting customers in the buffer. With probability $q = 1 - p$, no such transfer takes place. No such transfer takes place at a service completion epoch if there is atleast L customers in the buffer. Also if the pool contain at least one postponed job, the continuously served customers from the buffer since the last transfer under N-policy is counted, at each service completion epoch. When it reaches N $(N > 0)$, the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. At this time, system does not consider the p-transfer.

Customers arrive according to a homogeneous Poisson process of rate λ. The duration of the successive services whether of regular or of postponed customers are independent and identically distributed with the service time distribution following Phase Type(PH). Here the PH distribution has the irreducible representation $(\beta\ , S)$. There are m phases and the vector $S^0 = -Se$ containing elements S_{h0} denoting the absorption rate from the phase h, $h = 1, 2, ..., m$.

The model is studied as a Quasi Birth-Death(QBD) process and a solution of the classical matrix geometric type is obtained (see [45] and [38]). We define the statespace of the QBD and exhibit the structure of its infinitesimal generator.

The state space consists of all tuples of the form (i, j, b, h) with $i \geqslant 1$, $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant\ m$ where i is the number of postponed

work, j is the number of work in the finite buffer including the unit in service, b is the number of continuously served customers from the buffer at a service completion and h is the phase of the service in progress at a time t. For a given value of i, $K(N+1)m$ states constitute the level i of the QBD. Now consider the boundary level $i = 0$. Then we denote the empty system $(0,0,0,0)$ by 0. Also there are Km states of the form $(0,j,0,h)$, $1 \leqslant j \leqslant K$; $1 \leqslant h \leqslant m$. This is due to the fact that when the pool has no customers, N-policy is suspended. These have the same significance as before, except that in these states, no postponed jobs are present, but there are jobs in the finite buffer. These $Km+1$ states make up the boundary level 0 of the QBD.

The infinitesimal generator of the QBD describing the $M/PH/1/K$ queue with postponed customers under N-policy is of the form

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

where the matrix B_0 is of dimension $(Km+1) \times K(N+1)m$, B_1 is square matrix of order $Km+1$ and B_2 is of dimension $K(N+1)m \times (Km+1)$. A_0, A_1 and A_2 are square matrix of order $K(N+1)m$. Each of these matrices is itself highly structured.

The matrix B_1 corresponds to the transition from the level 0 to 0 is given below, where I is the identity matrix of order m and all non specified

entries are zeros.

$$B_1 = \begin{bmatrix} -\lambda & \lambda\beta & & & & & \\ S^0 & \Delta_1 & \Omega_1 & & & & \\ & S^0\beta & \Delta_2 & \Omega_2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & S^0\beta & \Delta_{K-1} & \Omega_{K-1} \\ & & & & & S^0\beta & \Delta \end{bmatrix}$$

where $\Delta = S - \lambda\gamma_1\gamma_2 I_m$, $\Delta_l = S - \epsilon_l I_m$, $\Omega_l = \lambda s_l I_m$, $\epsilon_l = \lambda(s_l + \delta - s_l\delta)$, $l = 1, 2, ..., K-1$.

$$B_0 = \begin{bmatrix} \bar{0} \\ diag(\Gamma_1, \Gamma_2, ..., \Gamma_{K-1}, \Gamma) \end{bmatrix}$$

where $\Gamma_l = \lambda(1 - s_l)\delta t_5 \otimes I_m$, $l = 1, 2, ..., K-1$; $\Gamma = \lambda\gamma_1\gamma_2 t_5 \otimes I_m$ and t_5 is a row vector of order $N+1$ with first element 1 and all other elements zero with I_m representing identity matrix of order m. Also $\bar{0}$ is zero matrix of appropriate order and $diag(\Gamma_1, \Gamma_2, ..., \Gamma_{K-1}, \Gamma)$ represents a diagonal block matrix of order K.

The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \bar{0} & diag\Big(& H_1, & H_2, & \ldots, & H_L, & H_{L+1}, & \ldots, & H_K, & \Big) \end{bmatrix}$$

where $diag(H_1, H_2, ..., H_L, H_{L+1}, ..., H_K)$ represents a diagonal block matrix of order K with diagonal block entries $H_1 = t_6 \otimes S^0\beta$, $H_2 = ... = H_L = t_7 \otimes S^0\beta$, $H_{L+1} = ... = H_K = t_8 \otimes S^0\beta$ and t_6 is a column vector of order $N+1$ with all entries are 1, t_7 is a column vector of order $N+1$ with

all elements are p except a 1 at $(N,1)^{th}$ position, t_8 is a column vector of order $(N+1)$ with $(N,1)^{th}$ element is 1 and all other elements zero.

The matrix A_0 is given by

$$A_0 = diag(\omega_1, \omega_2, ..., \omega_{K-1}, \omega)$$

where $\omega_l = \lambda(1-s_l)\delta I_{N+1} \otimes I_m$, $l = 1, 2, ..., K-1$; $\omega = \lambda\gamma_1\gamma_2 I_{N+1} \otimes I_m$ and I_{N+1} is the identity matrix of order $N+1$.

The matrix A_2 is given by

$$A_2 = diag(\Lambda_1, \Lambda_2, ..., \Lambda_L, \Lambda_{L+1}, ..., \Lambda_K)$$

It denotes diagonal block matrix with block entries on main diagonal given by $\Lambda_1 = t_1 \otimes S^0\beta$, $\Lambda_2 = ... = \Lambda_L = t_2 \otimes S^0\beta$, $\Lambda_{L+1} = ... = \Lambda_K = t_3 \otimes S_0\beta$ where t_1 is a square matrix of order $N+1$,given by

$$t_1 = \begin{bmatrix} [0] & I_N \\ 1 & [0] \end{bmatrix}$$

where I_N is identity matrix of order N and $[0]$ is zero matrix of appropriate

order. t_2 is a square matrix of order $N+1$ given by

$$t_2 = \begin{bmatrix} 0 & p & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ p & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

and t_3 is a square matrix of order $N+1$ with $(N, N+1)^{th}$ entry 1 and all other entries zero.

The matrix A_1 is given by

$$A_1 = \begin{bmatrix} \Theta_1 & \Phi_1 & & & & & & & \\ \zeta & \Theta_2 & \Phi_2 & & & & & & \\ & \zeta & \Theta_3 & & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & & & & & & \\ & & & & \zeta & \Theta_L & \Phi_L & & \\ & & & & & \eta & \Theta_{L+1} & & \\ & & & & & & \ddots & \ddots & \ddots \\ & & & & & & & \Theta_{K-1} & \Phi_{K-1} \\ & & & & & & & \eta & \Theta \end{bmatrix}$$

where ζ corresponds to the transition of the buffer size from j to $j-1$, $j = 2, 3, ..., L$ and $\zeta = t_4 \otimes qS^0\beta$; η corresponds to the transition of the buffer size from j to $j-1$, $j = L+1, ..., K$ and $\eta = t_4 \otimes S^0\beta$; Θ corresponds to the

transition of the buffer size from K to K and $\Theta = I_{N+1} \otimes (S - \lambda\gamma_1\gamma_2 I_m)$; Θ_j corresponds to the transition of the buffer size from j to j where $j = 1, 2, ..., L, L+1, ..., K-1$ and $\Theta_j = I_{N+1} \otimes (S - \epsilon_j I_m)$, $\epsilon_j = \lambda(s_j + \delta - s_j\delta)$, $j = 1, 2, ..., L, L+1, ..., K-1$; Φ_j corresponds to the transition of the buffer size from j to $j+1$ where $j = 1, 2, ..., L, L+1, ..., K-1$ and $\Phi_j = \lambda s_j I_{N+1} \otimes I_m$, $j = 1, 2, ..., L, L+1, ..., K-1$. Also t_4 is a square matrix of order $N+1$ which is given below:

$$t_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

3.2 Analysis of the system

3.2.1 Stability criterion

Theorem 3.2.1. *The system is stable if and only if*

$$\lambda\gamma_1\gamma_2 \sum_{b=0}^{N}\sum_{h=1}^{m} \pi_{Kbh} + \lambda\delta \sum_{j=1}^{K-1}\sum_{b=0}^{1}\sum_{h=1}^{n} (1 - s_j)\pi_{jbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}.$$

Proof. Let $G_{ll'}(k, x)$ be the conditional probability that the QBD pro-

cess starting in the state $l = (i, j, b, h)$ (for $i > 1$) where $1 \leqslant j \leqslant K$, $0 \leqslant b \leqslant N$, $1 \leqslant h \leqslant m$ at time $t = 0$ reaches the state $l' = (i-1, j', b', h')$ where $1 \leqslant j' \leqslant K$, $0 \leqslant b' \leqslant N$, $1 \leqslant h' \leqslant m$ for the first time, involving exactly k transitions (that is after exactly k service completions from the system) and completing before time x. Because of the structure of Q, the probability $G_{ll'}(k, x)$ does not depend on i. The matrix with elements $G_{ll'}(k, x)$ is denoted by $G(k, x)$.

Now introduce the transform matrix,

$$\hat{G}(z, \theta) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-\theta x} dG(k, x)$$

for $|z| \leqslant 1$, $\theta > 0$. The matrix $\hat{G}(z, \theta)$ satisfies the matrix equation

$$\hat{G}(z, \theta) = z(\theta I - A_1)^{-1} A_2 + (\theta I - A_1)^{-1} A_0 \hat{G}^2(z, \theta).$$

Use the notations $C_0(\theta) = (\theta I - A_1)^{-1} A_2$ and $C_2(\theta) = (\theta I - A_1)^{-1} A_0$. Now the transform matrix $\hat{G}(z, \theta)$ is equal to the minimal non negative solution of the matrix quadratic equation

$$X(z, \theta) = zC_0(\theta) + C_2(\theta) X^2(z, \theta)$$

and it is obtained by successive substitutions starting with the zero matrix. Also we have

$$\lim_{z \to 1, \theta \to 0} \hat{G}(z, \theta) = G(k, x) = [G_{ll'}(k, x)].$$

Suppose the matrix $A = A_0 + A_1 + A_2$ is irreducible. Then the necessary

and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0 + A_1 + A_2$. That is it is the unique solution to $\pi A = 0$, $\pi e = 1$ and $A = A_0 + A_1 + A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. G is obtained as the minimal non negative solution to the equation $G = C_0 + C_2 G^2$ where $C_0 = (-A_1)^{-1} A_2$ and $C_2 = (-A_1)^{-1} A_0$. That is, G is the minimal non negative solution of the matrix quadratic equation $A_2 + A_1 G + A_0 G^2 = 0$.

Let $m_1 = [m_{1_l}]$ denotes the column vector of dimension $K(N+1)m$ where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started in the state l. Then,

$$m_1 = \left[-\frac{\partial}{\partial \theta} \hat{G}(z, \theta) e \right]_{\theta=0, z=1} = -(A_1 + A_0(I + G))^{-1} e.$$

For the system stability, the rate of drift from level i to level $i-1$ should be greater than that to level $i+1$. This means that the Markov Chain(MC) is stable if and only if $\pi A_2 e > \pi A_0 e$. The rate of drift from level i to the level $i+1$ is given by $\lambda \gamma_1 \gamma_2 \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh} + \lambda \delta \sum_{j=1}^{K-1} \sum_{b=0}^{1} \sum_{h=1}^{n} (1 - s_j) \pi_{jbh}$. It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to

$$\lambda \gamma_1 \gamma_2 \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh} + \lambda \delta \sum_{j=1}^{K-1} \sum_{b=0}^{1} \sum_{h=1}^{n} (1 - s_j) \pi_{jbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}.$$

□

So by an appropriate choice of γ_1 and γ_2 , that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate is greater than service rate.

3.2.2 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of Q be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{3.1}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0 \tag{3.2}$$

and the vectors x_0, x_1 are obtained by solving the equtions

$$x_0 B_1 + x_1 B_2 = 0 \tag{3.3}$$

$$x_0 B_0 + x_1 (A_1 + RA_2) = 0 \tag{3.4}$$

subject to the normalising condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{3.5}$$

From the above discussion it is clear that to determine x, a key step is the computation of the rate matrix R. we use logarithmic reduction algorithm

as in section 2.2.2 in chapter 2. We can partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, x_{i3}, ..., x_{iK})$$

where $i \geqslant 1$ and x_{00} is a scalar and x_{0j} , $1 \leqslant j \leqslant K$ are vectors of order m and

$$x_{ij} = (x_{ij0}, x_{ij1}, ..., x_{ijN})$$

where $i \geqslant 1$, $1 \leqslant j \leqslant K$ and x_{ijb}, $0 \leqslant b \leqslant N$ are vectors of order m.

3.3 Computation of Expected values

In this section we derive the expected waiting time of a tagged customer (i) in the buffer and (ii) in the pool, the expected duration (i) between two consecutive transfers under N-policy (ii) for the first N-policy transfer in a busy cycle and the expectation of FIFO violation.

3.3.1 Expected waiting time in buffer

We denote the mean waiting time of customers who upon their arrivals enter the buffer by $E(W_1)$.

Case1: $N \geqslant K$.

In this case the tagged customer is not affected by the new arrivals in

buffer and in pool. So we can calculate the waiting time by considering the system state at which the tagged customer enters. Hence

$E(W_1) = \sum_i \sum_j \sum_b \sum_h E(\text{waiting time of the customer who finds the system in state } (i,j,b,h))\ Pr(\text{system is in state } (i,j,b,h))$

$$E(W_1) = \sum_{j=1}^{k-1} \sum_{h=1}^{m} -\beta S^{-1} e(j-1) x_{0j0h}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{b=0}^{N-1} \sum_{h-1}^{m} -\beta S^{-1} e(j-1+\psi) x_{ijbh}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{h=1}^{m} -\beta S^{-1} e(j-1) x_{ijNh} - \pi^* S^{-1} e$$

where $\pi^* Q^* = 0$, $\pi^* e = 1$ and $Q^* = S + S^0 \beta$ and

$$\psi = 1 + \left[\frac{j - (N-b)}{N} \right], 0 \leqslant b < N$$

where [y] denotes the greatest integer value of y. $-\pi^* S^{-1} e$ is the additional time required to complete the service of the customer who is at the server when the tagged person enters the buffer.

Remark 3.3.1. In $M/M/1$ case with service rate μ,

$$E(W_1) = \sum_{j=1}^{K-1} \frac{1}{\mu} j x_{0j0} + \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{b=0}^{N-1} \frac{1}{\mu} (j+\psi) x_{ijb}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\frac{1}{\mu}jx_{ijN}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

Case2: $N < K$.

In this case, the tagged customer in the buffer will be affected by the new arrivals in the pool and so the new arrivals in the buffer. So the waiting time of the tagged customer depends on the various susequent developments in the pool such as visits to zero level one or more, but a finite number in the pool joining after the tagged customer. Because of the complexity of calculation, we may turn to computing an upper bound on the waiting time, by keeping in mind, the fact that only a maximum finite number K of persons in the pool will affect the tagged person. In the worst case we have $N = 1$ which represents service alternating between buffer and pool. So an upper bound for the waiting time of a customer who upon his arrival enters the buffer in the state (i, j, b, h), is

$$UB(W_1) = \sum_{j=1}^{k-1}\sum_{h=1}^{m} -\beta S^{-1}e(j-1+[\frac{j}{N}])x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1}e(j-1+\psi)x_{ijbh}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{h=1}^{m} -\beta S^{-1}e(j-1+[\frac{j-1}{N}])x_{ijNh} - \pi^{*}S^{-1}e$$

where $\pi^* Q^* = 0$, $\pi^* e = 1$ and $Q^* = S + S^0\beta$ and

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

$-\pi^* S^{-1} e$ is the excess time required to complete the service of the customer who is at the server when the tagged person enter buffer.

Remark 3.3.2. In $M/M/1$ case with service rate μ,

$$UB(W_1) = \sum_{j=1}^{K-1} \frac{1}{\mu}(j + [\frac{j}{N}])x_{0j0} + \sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1} \frac{1}{\mu}(j+\psi)x_{ijb}$$

$$+ \sum_{i=1}^{\infty}\sum_{j=1}^{K-1} \frac{1}{\mu}(j + [\frac{j-1}{N}])x_{ijN}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

3.3.2 Expected waiting time in pool

We denote the expected waiting time of a customer who upon his arrival enters the pool, by $E(W_2)$.

To find this, first we define the Markov process $\{X(t)\}$ as follows. $X(t) = (a, j, b, h)$ where a denotes the rank of the tagged customer entered pool, j denotes the number of customers in the buffer, b denotes the number of continuously served customers from buffer and h is the phase

of the service process at time t. The rank a of the customer is assumed to be r if he joins as the r^{th} customer in pool. His rank may decrease to 1 with the customers ahead of him transferred from the pool to the buffer. Since the customers who arrive after the tagged customer cannot change his rank, level changing transitions in $\{X(t)\}$ can takeplace only to one side of the diagonal. We arrange the statespace of $\{X(t)\}$ as

$$\{r, r-1, ..., 2, 1\} \times \{1, 2, ..., K\} \times \{0, 1, ..., N\} \times \{1, 2, ..., m\}$$

with absorbing state 0 in the sense that the tagged customer is either selected to be served under N-policy or placed in the buffer with probability p or to the server with probability 1 if the buffer size reduces to 0 at the end of a service. The infinitesimal generator of the process is

$$\tilde{Q} = \begin{bmatrix} T & T^0 \\ \bar{0} & 0 \end{bmatrix}$$

where

$$T = \begin{bmatrix} A_1 & A_2 & & & & \\ & A_1 & A_2 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & A_1 & A_2 \\ & & & & & A_1 \end{bmatrix}$$

of order $rK(N+1)m$ and

$$T^0 = \begin{bmatrix} \bar{0} \\ \vdots \\ \bar{0} \\ B_2 \end{bmatrix}.$$

Now the expected absorption time of a particular customer is given by the column vector

$$E_w^{(r)} = -\tilde{I}T^{-1}e$$

where $\tilde{I} = \begin{bmatrix} I_{K(N+1)m} & \bar{0} \end{bmatrix}$ having order $K(N+1)m \times rK(N+1)m$. So the expected waiting time of the customer is

$$W_L = \sum_{r=1}^{\infty} x_r E_w^{(r)}$$

where x_r is the steady state probability vector corresponding to $i = r$. W_L gives the waiting time of a customer in the pool up to the epoch of his transfer to the buffer.

Case1: $N \geqslant K$

Expected waiting time in pool is

$$E(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L (x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+ \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} (W_L + W^{(1)} p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$W^{(1)} = \sum_{j=1}^{L}\sum_{h=1}^{m} -\beta S^{-1} e(j-1) x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi) x_{ijbh}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

Case2: $N < K$

In this case we get an upperbound $UB(W_2)$ for the waiting time in pool.

$$UB(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L (x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+\sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} (W_L + UB(W^{(1)}) p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$UB(W^{(1)}) = \sum_{j=1}^{L}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\left[\frac{j-1}{N}\right]) x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m} -\beta S^{-1} e(j-1+\psi) x_{ijbh}$$

where

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

3.3.3 Expected duration between two consecutive transfers under N-policy

For computing expected duration between two consecutive transfers under N- policy, we consider the Markov process $\{X(t)\}$ described as follows. $X(t) = (b, i, j, h)$ where b is the number of continuously served customers from the buffer, if pool has at least one person, at time t, measured from the service completion of the last customer who was transferred under N-policy. So $b = 0, 1, 2, ..., N$. Here we regard $0, 1, 2, ..., N-1$ as transient states and N as absorbing state (that is the state at which a new N-policy transfer occurs). i denotes the number of postponed jobs at time t. Even if pool is of infinite capacity, we restrict here it to be a finite value say V for sufficiently large V. So $i = 0, 1, 2, ..., V$; $j(= 0, 1, 2, ..., K)$ denotes the number of customers in the buffer at time t. Also $h = 1, 2, ..., m$ denotes the phase of the service in progress at a time t. The process $\{X(t)\}$ has the state space

$$\{0, 1, 2, ..., N-1, N\} \times \{0, 1, 2, .., V\} \times \{0, 1, 2, ..., K\} \times \{1, 2, 3, ..., m\}$$

Generator of the process is

$$\hat{Q} = \begin{bmatrix} U & U^0 \\ \bar{0} & 0 \end{bmatrix}$$

where

$$U = \begin{bmatrix} C_0 & D_0 & & & & & \\ C_1 & C_3 & D_1 & & & & \\ C_1 & & C_3 & D_1 & & & \\ C_1 & & & C_3 & \ddots & & \\ \vdots & & & & \ddots & \ddots & \\ C_1 & & & & & \ddots & D_1 \\ C_2 & & & & & & C_3 \end{bmatrix}_{(NVKm+Km+1)\times(NVKm+Km+1)}$$

and

$$U^0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ D_2 \end{bmatrix}$$

$$C_0 = \begin{bmatrix} E_1 & E_4 & & & & \\ & E_2 & E_5 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & E_2 & E_5 \\ & & & & & E_3 \end{bmatrix}_{(VKm+Km+1)\times(VKm+Km+1)}$$

where

$$E_1 = \begin{bmatrix} -\lambda & \lambda\beta & & & & \\ S^0 & S-\epsilon_1 I_m & \lambda s_1 I_m & & & \\ & S^0\beta & & & & \\ & & & \cdots & & \\ & & & & S-\epsilon_{K-1}I_m & \lambda s_{K-1}I_m \\ & & & & S^0\beta & S-\lambda\gamma_1\gamma_2 I_m \end{bmatrix}$$

which is of order $(Km+1)\times(Km+1)$ where $\epsilon_l = \lambda(s_l+\delta-s_l\delta)$.

$$E_2 = \begin{bmatrix} S-\epsilon_1 I_m & \lambda s_1 I_m & & & \\ & \cdots & \cdots & & \\ & & \cdots & \cdots & \\ & & & S-\epsilon_{K-1}I_m & \lambda s_{K-1}I_m \\ & & & & S-\lambda\gamma_1\gamma_2 I_m \end{bmatrix}_{Km\times Km}$$

$$E_3 = \begin{bmatrix} S-\epsilon_1 I_m & \lambda s_1 I_m & & & \\ & \cdots & \cdots & & \\ & & \cdots & \cdots & \\ & & & S-\epsilon_{K-1}I_m & \lambda s_{K-1}I_m \\ & & & & S \end{bmatrix}_{Km\times Km}$$

$$E_4 = \begin{bmatrix} \lambda(1-s_1)\delta I_m & & & \\ & \cdots & & \\ & & \cdots & \\ & & & \lambda(1-s_{K-1})\delta I_m \\ & & & \lambda\gamma_1\gamma_2 I_m \end{bmatrix}$$ having

order $(Km+1)\times Km$.

$$E_5 = \begin{bmatrix} \lambda(1-s_1)\delta I_m & & & & \\ & \cdots & & & \\ & & \cdots & & \\ & & & \lambda(1-s_{K-1})\delta I_m & \\ & & & & \lambda\gamma_1\gamma_2 I_m \end{bmatrix}$$ having order $Km \times Km$.

Also

$$C_1 = \begin{bmatrix} E_6 & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}_{(KVm)\times(KVm+Km+1)}$$

$$C_2 = \begin{bmatrix} E_7 & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}_{(KVm)\times(KVm+Km+1)}$$

where

$$E_6 = \begin{bmatrix} 0 & S^0\beta & \bar{0} & \bar{0} \\ \bar{0} & \bar{0} & I_{L-1}\otimes pS^0\beta & \bar{0} \\ \bar{0} & \bar{0} & \bar{0} & \bar{0} \end{bmatrix}_{Km\times(Km+1)}$$

$$E_7 = \begin{bmatrix} \bar{0} & I_K \otimes S^0\beta \end{bmatrix}_{Km\times(Km+1)}$$

$$E_8 = \begin{bmatrix} \bar{0} & \bar{0} \\ I_{L-1}\otimes qS^0\beta & \bar{0} \\ \bar{0} & I_{K-L}\otimes S^0\beta \end{bmatrix}_{Km\times Km}$$

$$E_9 = \begin{bmatrix} S^0\beta & \bar{0} & \bar{0} \\ \bar{0} & I_{L-1}\otimes pS^0\beta & \bar{0} \\ \bar{0} & \bar{0} & \bar{0} \end{bmatrix}_{Km\times Km}$$

$$C_3 = \begin{bmatrix} E_2 & E_5 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & E_2 & E_5 \\ & & & & E_3 \end{bmatrix}_{KVm \times KVm}$$

$$D_1 = \begin{bmatrix} E_8 & & & \\ E_9 & \ddots & & \\ & \ddots & \ddots & \\ & & E_9 & E_8 \end{bmatrix}_{KVm \times KVm}$$

$$D_0 = \begin{bmatrix} \bar{0} \\ D_1 \end{bmatrix}_{KVm \times KVm} \qquad D_2 = \begin{bmatrix} \bar{0} & \bar{0} \\ I_{V-1} \otimes I_K \otimes S^0\beta & \bar{0} \end{bmatrix}_{KVm \times KVm}$$

where $\otimes$ denotes Kronecker product. The initial probability vector of $\hat{Q}$ is

$$\delta = \begin{bmatrix} \frac{1}{\sum_{r=1}^{KVm+Km+1} x_{r-1}} \begin{bmatrix} x_0 & x_1 & \cdots & x_{KVm+Km+1} \end{bmatrix} & \bar{0} \end{bmatrix}_{1 \times (NKVm+Km+1)}$$

where $x_0 = x_{0000}$. $x_r = x_{ij0h}, 0 \leqslant i \leqslant V; 1 \leqslant j \leqslant K; 1 \leqslant h \leqslant m$ and r varies from 1 to $KVm + Km$ according to its lexicographic order. Then we have the following lemma:

Lemma 3.3.1. *Expected duration between two consecutive transfers under N-policy follows PH distribution with representation (δ, U) and it is given by*

$$N_{ABSORB} = -\delta U^{-1} e.$$

Using this expected value, a cost function is defined in section 7 and the optimal value of N is determined.

3.3.4 Expected duration for the first N-policy transfer in a busy cycle

Here we compute the expected duration of the time elapsed from the epoch of the first arrival to an idle system until the first N-policy transfer is effected. This can be obtained as corollary to lemma 4.1.

Corollary 3.3.2. The time elapsed, starting with an arrival to an idle system, until the realization of the N-policy for the first time follows the PH-distribution with representation (α, U) where

$$\alpha = \frac{1}{\sum_{r=1}^{m} x_r} \begin{bmatrix} 0 & x_1 & x_2 & \cdots & x_m & \bar{0} \end{bmatrix}_{1\times(NKVm+Km+1)}$$

where $x_r = x_{010h}; 1 \leqslant h \leqslant m$, r varies from 1 to m and U is described in section 4.3.

Proof. At the epoch of the first arrival to an idle system, process starts with the service in one of the m phases with steady state probability $x_r = x_{010h}; 1 \leqslant h \leqslant m$, r varies from 1 to m. This justifies the form of the initial probability vector α as given above. □

Then the expected duration for the realization of the above random

variable is

$$N_{FIRST} = -\alpha U^{-1} e.$$

3.3.5 Expected number of FIFO violation

It may be noted that the N-policy leads to violation of FIFO rule for customers in the pool. For example assume that there are two or more customers in the pool at a service completion epoch at which the number in the buffer droped to $L-1$ or below and the number of continuously served customers reached $N-1$. So the first in the pool may be selected under p-transfer and placed as the last in the buffer. When the next service is completed, the current head of the pool gets transferred to the buffer for immediate service there by violating the FIFO rule for pooled customers. Further it may be noted that this situation does not arise among the queued customers in the buffer. We compute the expectation of the indicator random variable defined as FIFO violation in pool. Its expectation is the probability for FIFO violation in pool which is given by

$$P_{FIFO} = \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=N-j+1}^{N-1} x_{ijbh} p S_{h0}.$$

The FIFO may be violated by more than one successors if $N < L$. However this can be overcome by making N sufficiently large than L. If $N \geqslant K$, a customer joining the pool will not overtake any of the customers in the buffer who had joined before his entering to pool. At this time, FIFO is violated by atmost one successor in pool. Even this can be overcome by a slight modification by redefining the N-policy by resetting b in (i, j, b, h)

as zero at the time of p-transfer.

3.4 Performance measures

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{K} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{ijbh}$$

for $i \geqslant 1$ and

$$a_0 = x_{00} + \sum_{j=1}^{K} \sum_{h=1}^{m} x_{0j0h}.$$

2. The probability that there are j customers in the buffer (including the one in service) is

$$b_j = \sum_{h=1}^{m} x_{0j0h} + \sum_{i=1}^{\infty} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{ijbh}$$

for $1 \leqslant j \leqslant K$ and

$$b_0 = x_{00}.$$

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that a customer, on its arrival enters the pool is

$$\gamma_1\gamma_2 b_K + \delta \sum_{j=1}^{K-1} (1 - s_j) b_j.$$

6. The probability that an arriving customer enters service immediately is b_0.

7. The rate at which a customer enters the buffer is

$$\lambda b_0 + \lambda \sum_{j=1}^{K-1} s_j b_j.$$

8. The rate at which the customer who leave the system without service (mean number of customers not joining the system per unit time) is

$$\theta_{LOST} = \lambda(1-\delta) \sum_{j=1}^{K-1} (1 - s_j) b_j + \lambda(1-\gamma_1) b_K + \lambda\gamma_1(1-\gamma_2) b_K.$$

9. The rate at which pooled customers transfer in to the buffer is

$$\theta_{TR} = \sum_{i=1}^{\infty} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{i1bh} S_{h0} + \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=0}^{N-2} \sum_{h=1}^{m} x_{ijbh} p S_{h0}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0} + \sum_{i=1}^{\infty}\sum_{j=2}^{L}\sum_{h=1}^{m} x_{ijNh}pS_{ho}.$$

10. The rate at which pooled customers transfer under N-policy(mean number of transfers under N-policy per unit time) is

$$T_N = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0}.$$

11. Mean number of customers served out per unit time
$= (1 - b_o)\frac{1}{-\beta S^{-1}e}$.

3.5 Numerical results

We present some numerical results in order to illustrate the performance of the system. We reinterpret the probabilities γ_1, γ_2, s_x, δ as:

γ_1 : Customer's eagerness to join the pool when the buffer is full.

γ_2 : Server's interest on a new work to postpone when the buffer is full.

s_x : Server's interest on a new work to accept in to the buffer when x customers are present.

δ : Customer's special interest to the service station.

Take $\gamma_1 = \frac{Lp}{K} + \frac{1}{N}$ in order to bring out explicitly the dependence of γ_1 on the system parameters.

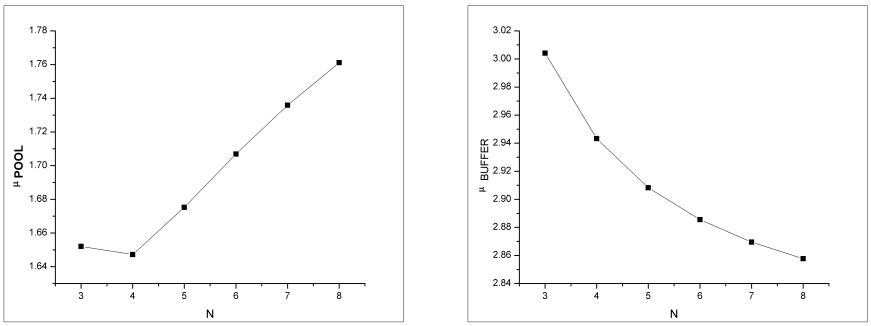

Fig 3.2: N versus μ_{POOL} and μ_{BUFFER}

This is justified as follows. Larger the L value, the customer encountering the buffer full will be inclined to join the pool with higher probability. Also same is the relationship of γ_1 with p. On the other hand, γ_1 inversely varies with K. The additional term $\frac{1}{N}$ comes through N-policy. Here as N increases γ_1 decreases so that γ_1 and N vary inversely. But the relationship is feasible for those values of L, p, K and N such that $0 \leqslant \gamma_1 \leqslant 1$. This is possible if $N \geqslant K$ and such a selection is consistent. However N can be made less than K by suitably selecting other variables so that $0 \leqslant \gamma_1 \leqslant 1$, and that can be considered as an incentive to customers joining the pool.

The impact of N on various measures of descriptors with $K = 6, L = 3, m = 2, \lambda = 7, p = 0.5, s_1 = 0.9, s_2 = 0.8, s_3 = 0.7, s_4 = 0.6, s_5 = 0.5, \gamma_1 = \frac{Lp}{K} + \frac{1}{N}\gamma_2 = 0.8, \delta_0.5$

$$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \quad S = \begin{bmatrix} -12.5 & 6.0 \\ 6.0 & -12.5 \end{bmatrix} \quad S^0 = \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix}$$

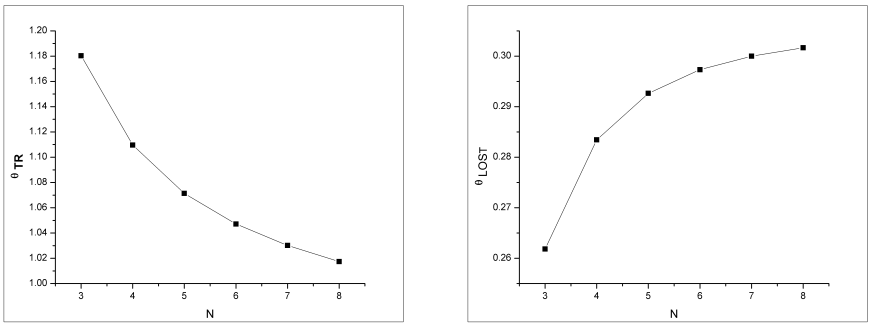

Fig 3.3: N versus θ_{LOST} and θ_{TR}

is shown in figure 3.2 and figure 3.3. As N increases μ_{BUFFER}, θ_{TR} decrease monotonically whereas θ_{LOST} increases monotonically; but μ_{POOL} decreases at first and then increases. This is due to the fact that by our assumption γ_1 varies inversely as N. So as N increases, customer's attraction to the pool decreases when the buffer is full. So the pool size decreases. Also transfer rate decreases. This will make the number in the buffer decrease. So the influence of δ increases which makes the number in the pool increase.

By keeping $K = 6, p = 0.5, m = 2, \lambda = 7, N = 5, \gamma_1 = \frac{Lp}{K} + \frac{1}{N}$ the effect of L on various measures is shown in figures 3.4 and 3.5. Here also μ_{BUFFER}, θ_{TR} are monotonically increasing as L increases, as expected. This will gradually makes the buffer full. So the influence of δ decreases which makes the pool size decreasing and the effect of γ_1 increases which makes θ_{LOST} monotonically decreasing. The measures are numerically computed for various values of p by keeping $L = 3$ and shown in table 3.1. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing in p as expected. For a lower L, increase in p

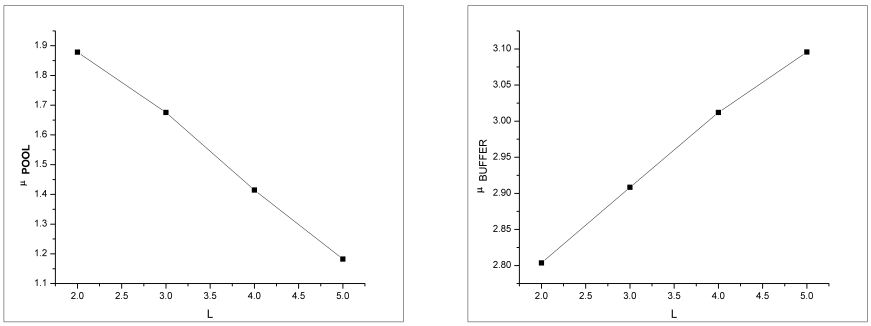

Fig 3.4: L versus μ_{POOL} and μ_{BUFFER}

does not contribute much towards increase in the buffer size, in contrast to the effect of δ.

p	μ_{POOL}	μ_{BUFFER}	θ_{TR}	θ_{LOST}
0.3	1.6284441	2.8322663	1.0228137	0.3299661
0.4	1.6445645	2.8706489	1.0458466	0.3119004
0.5	1.6752849	2.9082990	1.0713297	0.2926464
0.6	1.7203127	2.9454260	1.0992774	0.2722456
0.7	1.7798784	2.9821990	1.1297385	0.2507135

Table 3.1: $K = 6, L = 3, m = 2, \lambda = 7, N = 5, \gamma_1 = \frac{Lp}{K} + \frac{1}{N}$

Remark 3.5.1. If $s_x = 1, \forall x, x = 1, 2, ..., K-1$ and $\gamma_2 = 1$ then we get the model discussed in [**?**].

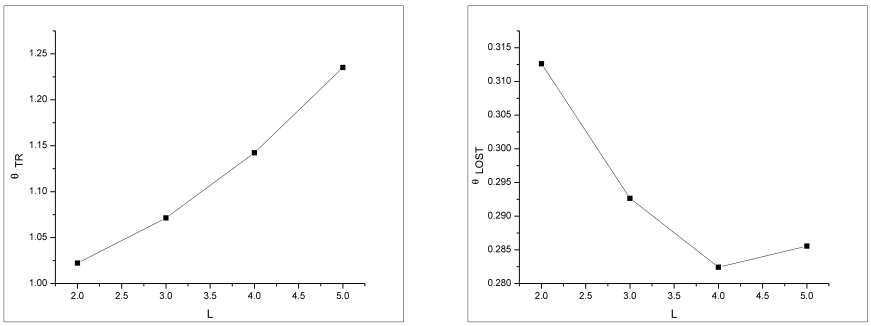

Fig 3.5: L versus θ_{TR} and θ_{LOST}

3.6 A Game Theoretic Approach

If the buffer is full with K customers, a newly arrived customer may join the pool with probability γ_1 or leave the system without joining the pool with probability $1-\gamma_1$. At the same time the server may permit a customer to join the pool with probability γ_2 or may decline admission to the pool with probability $1-\gamma_2$. The situation can be modelled by a two-person zero-sum game as follows.

Let the customer be treated as the player 1 and the system as the player 2. The player 1 has two alternatives: (1) join pool (2) leave the system without joining the pool. The player 2 has also two alternatives: (1) allow the customer to enter pool (2)does not allow the customer to enter pool.

Let the mixed strategy of the player 1 be $(\gamma_1, 1-\gamma_1)$ where $0 < \gamma_1 < 1$ and that of the player 2 be $(\gamma_2, 1-\gamma_2)$ where $0 < \gamma_2 < 1$ where γ_1 is the probability of the customer to join the pool and γ_2 is the probability of

the server to admit the customer to the pool. Then the pay-off matrix of player 1 is

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

where C_{11} is the gain to the customer when he decides to join the pool and the server admit him to the pool; C_{12} is the loss to the customer when he decides to join the pool and at the same time the server does not admit him to the pool; C_{21} is the gain to the customer when he decides not to join the pool but the server is ready to admit him to the pool; C_{22} is the gain to the customer when he decides not to join the pool and the server decides not to admit him to the pool. Now a valid assumption can be $C_{11} > C_{12}$, $C_{22} > C_{12}$ and $C_{22} > C_{21}$. Then

$$\gamma_1 = \frac{C_{22} - C_{21}}{(C_{11} + C_{22}) - (C_{21} + C_{12})},$$

$$\gamma_2 = \frac{C_{22} - C_{12}}{(C_{11} + C_{22}) - (C_{21} + C_{12})}$$

and the value of the game is

$$\gamma = \frac{C_{11}C_{22} - C_{12}C_{21}}{(C_{11} + C_{22}) - (C_{21} + C_{12})}.$$

Chapter 4

An $M/M/1$ Queue with Postponed work and service interruption under N-policy

Interruption in the service of a customer is a common phenomenon. Several reasons like server breakdowns or arrival of priority customers create interruption. In this chapter we introduce interruption in a system with postponed work. If a customer on arrival, finds the buffer having finite capacity full, he is allowed to join a pool of postponed work having infinite capacity with a specified probability. A customer will select such a facility, if he is not bothered about the waiting time and his importance lies in the

Some results of this chapter are included in the following paper.
1. A.Krishnamoorthy, C.B.Ajayakumar, An $M/M/1$ Queue with Postponed work and service interruption under N-policy (Communicated)

service of that particular service station. Such a postponed work will be tranferred to the buffer for immediate service with a specified probability, at a service completion epoch if the number of customers in the buffer at that time is less than a pre-assigned level. Also, If there is no customer left in the buffer at a service completion epoch, and atleast one is in the pool, the one at the head of the pool is transferred to the buffer with probability one for immediate service.

Here the postponed work is transferred to the buffer for immediate service due to the attaining of a pre-assigned low level in the buffer. But during the service of such a pooled customer, the buffer size may rise to a pre-assigned higher level and so the server will be compelled to preempt the service of that pool work. So the postponed work gets interruption and it is again postponed and wait at the head of the pool for next chance of transfer. Now we introduce an N-policy as follows. At the time of interruption, system starts to count the number of continuously served customers from the buffer. When it reaches a pre-assigned number N at a service completion epoch, the interrupted postponed work is again considered for immediate service and further interruption is not allowed for such a work in any circumstance. The interrupted work is assumed to be repeated after the interruption. A diagramatic representation of the model is given in figure 4.1.

The situation in this model is very common in real life situations. Some work may be postponed by the server to do it when he is about to be idle. When the server considers to deal with such a postponed work, it may be interrupted and again postponed due to the arrival of a number of emergency work.

4.1 Mathematical formulation

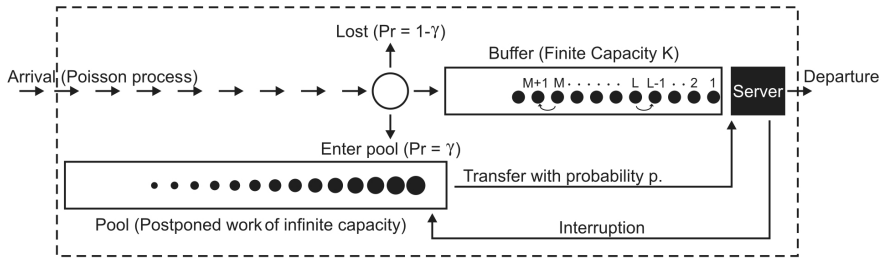

Fig 4.1: Postponed work with service interruption

Consider an $M/M/1$ queue with finite buffer of capacity K. If the buffer contains less than K customers including the one at server, newly arriving customer will join it. When the buffer is full with K customers, newly arriving jobs are not necessarily lost. They are offered the choice of leaving the system immediately or of being postponed until the system is less congested. That is, a customer can accept the offer of postponement with a probability γ $(0 \leqslant \gamma < 1)$. So he may join a pool of postponed work of infinite capacity. With probability $1 - \gamma$, he does not join the system. When at the end of a service, if there are postponed customers, the system operates as follows. If the buffer is empty, the one ahead of all waiting in the pool gets trasferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$ at a service completion epoch, then with a probability p, the head of the queue in pool is transferred to the finite buffer for immediate service. With probability $q = 1 - p$, no such transfer takes place. When at the end of a service, if the buffer is empty, and the pool has no work, the server becomes idle.

When the service of a pooled customer is going on, if the buffer size rises to a pre-assigned number $M+1$ such that $L \leqslant M \leqslant K-1$ at an arrival epoch, then the current customer is preempted with probability one. Thus this interrupted postponed work is further postponed and stay as the head of the queue in the pool for getting next chance of transfer. At that time, server will perform buffer work. At the time of interruption, the system starts to count the number of continuously served customers from the buffer. When it reaches N $(N > 0)$ at a service completion epoch, then the interrupted pooled customer gets transferred to the buffer for immediate service and no further interruption is allowed for such a customer. If the number N is not attained, the customer may be again get interrupted if he is considered again for immediate service due to the buffer size reaching $L-1$ or below.

Customers arrive according to a homogeneous Poisson process of rate λ. The duration of the successive services whether of regular or of postponed customers are independent and identically distributed exponential random variables with parameter μ. The model is studied as a quasi birth-death(QBD) process and matrix geometric solution is obtained (see [45] and [38]). We define the state space of the QBD and exhibit the structure of its infinitesimal generator.

The state space consists of all tuples of the form (i, j, b, r) where i denotes the number of postponed work in the pool having infinite capacity; j denotes the number of jobs in the finite buffer including the unit in service; b denotes the status of the system where

$$b = \begin{cases} 0 & , \text{ the buffer work is going on serving} \\ 1 & , \text{ the pool work is going on serving} \end{cases}$$

If $b = 0$, r denotes the number of continuously served customers from the buffer including the customer in service and if $b = 1$, r denotes the number of continuously served customers from the buffer only, during the period of interruption where $r \neq 0$. $r = 0$ indicates that the head of the pool work is not an interrupted one.

Consider the boundary level $i = 0$. We denote the empty system $(0, 0, 0, 0)$ by 0.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $r = 0$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $r = 0, 1, 2, ..., N$.

If $M + 1 \leqslant j \leqslant K$ and $b = 1$ then $r = N$. So the boundary level $i = 0$ constitute $MN + 2K + 1$ states. Now consider the level $i \neq 0$.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $r = 0, 1, 2, ..., N$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $r = 0, 1, 2, ..., N$.

If $M + 1 \leqslant j \leqslant K$ and $b = 1$ then $r = N$. So there are $2M(N + 1) + (K - M)(N + 2)$ states are there in the level $i \neq 0$.

The infinitesimal generator of the QBD describing the $M/M/1/K$ queue with postponed work and service interruption under N-policy is

of the form

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

where the matrix B_0 is of dimension $[MN + 2K + 1] \times [2M(N + 1) + (K - M)(N + 2)]$, B_1 is square matrix of order $MN + 2K + 1$ and B_2 is of dimension $[2M(N + 1) + (K - M)(N + 2)] \times [MN + 2K + 1]$. A_0, A_1 and A_2 are square of order $2M(N + 1) + (K - M)(N + 2)$. Each of these matrices is itself highly structured.

The matrix B_1 corresponds to the transition from the level 0 to 0 is given below.

$$B_1 = \begin{bmatrix} -\lambda & \lambda t_1 & & & & & & & \\ \mu t_2 & \Omega & \lambda I_{N+2} & & & & & & \\ & \mu t_2 & \Omega & \ddots & & & & & \\ & & \ddots & \ddots & \lambda I_{N+2} & & & & \\ & & & \mu t_2 & \Omega & \lambda t_3 & & & \\ & & & & \mu t_4 & \eta & \lambda I_2 & & \\ & & & & & \mu t_5 & \eta & \ddots & \\ & & & & & & \ddots & \ddots & \lambda I_2 \\ & & & & & & & \mu t_5 & \Theta \end{bmatrix}$$

where $\Omega = (-\lambda - \mu)I_{N+2}$ corresponds to the transition of buffer size from j to j where $j = 1, 2, ..., M$; $\eta = (-\lambda - \mu)I_2$ corresponds to the transition of buffer size from j to j where $j = M + 1, ..., K - 1$; $\Theta = (-\lambda\gamma - \mu)I_2$

corresponds to the transition of buffer size from K to K and all non specified entries are zeros. Also, t_1 is a row vector of dimension $N+2$, with first element 1 and all other entries are zeros. t_2 is a column vector of ones of dimension $N+2$.

$$t_3 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 1 \end{bmatrix}_{(N+2)\times 2} \qquad t_4 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}_{2\times(N+2)} \qquad t_5 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}_{2\times 2}$$

λI_{N+2} corresponds to the transition of buffer size from j to $j+1$ where $j = 1, 2, ..., M-1$; λt_3 corresponds to the transition of buffer size from M to $M+1$; λI_2 corresponds to the transition of buffer size from j to $j+1$ where $j = M+1, ..., K-1$; μt_2 corresponds to the transition of buffer size from j to $j-1$ where $j = 1, 2, ..., M$; μt_4 corresponds to the transition of buffer size from $M+1$ to M; μt_5 corresponds to the transition of buffer size from j to $j-1$ where $j = M+2, ..., K$.

The matrix B_0 corresponds to the transition from the level 0 to 1 such that the transition rate from the buffer size M to M is given by the block λV_0 representing interruption and the transition rate from the buffer size K to K is given by the block $\lambda\gamma t_7$ representing postponement and all other block entries are zero matrices where,

$$t_6 = \begin{bmatrix} \bar{0} & I_N \\ 0 & \bar{0} \end{bmatrix}_{(N+1)\times(N+1)} \qquad V_0 = \begin{bmatrix} \bar{0} & \bar{0} \\ t_6 & \bar{0} \end{bmatrix}_{(N+2)\times 2(N+1)}$$

$$t_7 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}_{2\times(N+2)}.$$

The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \bar{0} & diag \begin{bmatrix} \mu V_1, & \mu V_2, & \cdots, & \mu V_3, & \cdots, & \mu t_{10}, & \cdots, & \mu t_{10} \end{bmatrix} \end{bmatrix}$$

where $\bar{0}$ is a zero matrix of suitable dimension, μV_1 corresponds to the transition of buffer size from 1 to 1, μV_2 corresponds to the transition of buffer size from j to j where $j = 2, ..., L$, μV_3 corresponds to the transition of buffer size from j to j where $j = L+1, ..., M$, μt_{10} corresponds to the transition of buffer size from j to j where $j = M+1, ..., K$. where

$$t_8 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}_{(N+1)\times(N+1)} \qquad V_1 = \begin{bmatrix} \bar{0} & I_{N+1} \\ \bar{0} & t_8 \end{bmatrix}_{2(N+1)\times(N+2)}$$

$$V_2 = \begin{bmatrix} \bar{0} & F \\ \bar{0} & pt_8 \end{bmatrix}_{2(N+1)\times(N+2)} \qquad V_3 = \begin{bmatrix} \bar{0} & t_9 \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)\times(N+2)}$$

$$F = \begin{bmatrix} pI_N & \bar{0} \\ \bar{0} & 1 \end{bmatrix}_{(N+1)\times(N+1)}.$$

Also t_9 is an $(N+1)\times(N+1)$ matrix whose last entry is 1 and t_{10} is an $(N+2)\times 2$ matrix whose $[(N+1), 2]^{th}$ entry is 1 and all other elements in both the matrices are zeros.

In the matrix A_0, the transition rate from the buffer size M to M is given by the block λV_4 representing interruption and the transition rate

from the buffer size K to K is given by the block $\lambda\gamma I_{N+2}$ representing postponement and all other block entries are zero matrices where,

$$V_4 = \begin{bmatrix} \bar{0} & \bar{0} \\ t_6 & \bar{0} \end{bmatrix}_{2(N+1)\times 2(N+1)}.$$

The matrix A_1 is given by

$$A_1 = \begin{bmatrix} \Delta_1 & \Gamma_1 & & & & & & & & \\ \Phi_1 & \Delta_1 & \Gamma_1 & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \Phi_2 & \Delta_1 & \Gamma_1 & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & \Phi_2 & \Delta_1 & \Gamma_2 & & & \\ & & & & & \Phi_3 & \Delta_2 & \Gamma_3 & & \\ & & & & & & \Phi_4 & \Delta_2 & \ddots & \\ & & & & & & & \ddots & \ddots & \Gamma_3 \\ & & & & & & & & \Phi_4 & \Delta_3 \end{bmatrix}$$

where $\Delta_1 = (-\lambda - \mu)I_{2(N+1)}$ corresponds to the transition of buffer size from j to j where $j = 1, 2, ..., M$; $\Delta_2 = (-\lambda - \mu)I_{N+2}$ corresponds to the transition of buffer size from j to j where $j = M+1, ..., K-1$; $\Delta_3 = (-\lambda\gamma - \mu)I_{N+2}$ corresponds to the transition of buffer size from K to K; $\Gamma_1 = \lambda I_{2(N+1)}$ corresponds to the transition of buffer size from j to $j+1$ where $j = 1, 2, ..., M-1$; $\Gamma_2 = \lambda V_6$ corresponds to the transition of buffer size from M to $M+1$; $\Gamma_3 = \lambda I_{N+2}$ corresponds to the transition of buffer size from j to $j+1$ where $j = M+1, ..., K-1$; $\Phi_1 = q\mu V_5$ corresponds to the transition of buffer size from j to $j-1$ where $j =$

$1, 2, ..., L$; $\Phi_2 = \mu V_5$ corresponds to the transition of buffer size from j to $j-1$ where $j = L+1, ..., M$; $\Phi_3 = \mu V_7$ corresponds to the transition of buffer size from $M+1$ to M; $\Phi_4 = \mu V_8$ corresponds to the transition of buffer size from j to $j-1$ where $j = M+2, ..., K$.

$$t_{11} = \begin{bmatrix} 1 & 0 & \bar{0} \\ \bar{0} & \bar{0} & I_{N-1} \\ 0 & 0 & \bar{0} \end{bmatrix}_{(N+1)\times(N+1)} \qquad t_{12} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{(N+1)\times 1}$$

$$t_{13} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{1\times(N+1)}$$

$$V_5 = \begin{bmatrix} t_{11} & \bar{0} \\ t_8 & \bar{0} \end{bmatrix}_{2(N+1)\times 2(N+1)} \qquad V_6 = \begin{bmatrix} I_{N+1} & \bar{0} \\ \bar{0} & t_{12} \end{bmatrix}_{2(N+1)\times(N+2)}$$

$$V_7 = \begin{bmatrix} t_{11} & \bar{0} \\ t_{13} & \bar{0} \end{bmatrix}_{(N+2)\times 2(N+1)} \qquad V_8 = \begin{bmatrix} t_{11} & \bar{0} \\ t_{13} & \bar{0} \end{bmatrix}_{(N+2)\times(N+2)}$$

where $\bar{0}$ is zero matrix of appropriate order. All non specified entries are zeros. The matrix A_2 is given by

$$A_2 = diag \begin{bmatrix} \mu V_9, & \mu V_{10}, & \cdots, & \mu V_{11}, & \cdots, & \mu V_{11}, & \mu V_{12} & \cdots & \mu V_{12} \end{bmatrix}$$

where μV_9 corresponds to the transition of buffer size from 1 to 1, μV_{10} corresponds to the transition of buffer size from j to j where $j = 2, ..., L$, μV_{11} corresponds to the transition of buffer size from j to j where $j = L+1, ..., M$, μV_{12} corresponds to the transition of buffer size from j to j

where $j = M+1, ..., K$. Also,

$$V_9 = \begin{bmatrix} \bar{0} & I_{N+1} \\ \bar{0} & t_8 \end{bmatrix}_{2(N+1)\times 2(N+1)} \qquad V_{10} = \begin{bmatrix} \bar{0} & F \\ \bar{0} & pt_8 \end{bmatrix}_{2(N+1)\times 2(N+1)}$$

$$V_{11} = \begin{bmatrix} \bar{0} & t_9 \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)\times 2(N+1)} \qquad V_{12} = \begin{bmatrix} \bar{0} & t_{12} \\ \bar{0} & \bar{0} \end{bmatrix}_{(N+2)\times(N+2)}.$$

4.1.1 Stability criterion

Theorem 4.1.1. *The system is stable if and only if*

$$\lambda\gamma\left(\sum_{r=0}^{N}\pi_{K0r} + \pi_{K1N}\right) + \lambda\sum_{r=0}^{N-1}\pi_{M1r} < \frac{1}{\sum_{l=1}^{2M(N+1)+(K-M)(N+2)} m_{1l}}.$$

Proof. Let $G_{ll'}(k,x)$ be the conditional probability that the QBD process starting in the state $l = (i,j,b,h)$ (for $i > 1$) at time $t = 0$ reaches the state $l' = (i-1, j', b', h')$ for the first time, involving exactly k transitions (that is after exactly k service completions from the system) and completing before time x. Because of the structure of Q, the probability $G_{ll'}(k,x)$ does not depend on i. The matrix with elements $G_{ll'}(k,x)$ is denoted by $G(k,x)$.

Now introduce the transform matrix,

$$\hat{G}(z,\theta) = \sum_{k=1}^{\infty} z^k \int_0^{\infty} e^{-\theta x} dG(k,x)$$

for $|z| \leqslant 1$, $\theta > 0$. The matrix $\hat{G}(z,\theta)$ satisfies the matrix equation

$$\hat{G}(z,\theta) = z(\theta I - A_1)^{-1}A_2 + (\theta I - A_1)^{-1}A_0\hat{G}^2(z,\theta)$$

Use the notations $C_0(\theta) = (\theta I - A_1)^{-1}A_2$ and $C_2(\theta) = (\theta I - A_1)^{-1}A_0$. Now the transform matrix $\hat{G}(z,\theta)$ is equal to the minimal non negative solution of the matrix quadratic equation

$$X(z,\theta) = zC_0(\theta) + C_2(\theta)X^2(z,\theta)$$

and it is obtained by successive substitutions starting with the zero matrix. Also we have

$$\lim_{z\to 1,\theta\to 0} \hat{G}(z,\theta) = G(k,x) = [G_{ll'}(k,x)]$$

Suppose the matrix $A = A_0+A_1+A_2$ is irreducible. Then the necessary and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0 + A_1 + A_2$. That is it is the unique solution to $\pi A = 0$, $\pi e = 1$ and $A = A_0 + A_1 + A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. G is obtained as the minimal non negative solution to the equation $G = C_0+C_2G^2$ where $C_0 = (-A_1)^{-1}A_2$ and $C_2 = (-A_1)^{-1}A_0$. That is, G is the minimal non negative solution of the matrix quadratic equation $A_2 + A_1G + A_0G^2 = 0$.

Let $M_1 = [m_{1_l}]$ denotes the column vector of dimension $2M(N+1) + (K-M)(N+2)$ where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started

in the state l. Then,

$$M_1 = \left[-\frac{\partial}{\partial\theta}\hat{G}(z,\theta)e\right]_{\theta=0,z=1} = -(A_1 + A_0(I+G))^{-1}e.$$

For the system stability, the rate of drift from level i to level $i-1$ should be greater than that to level $i+1$. This means that the Markov Chain(MC) is stable if and only if $\pi A_2 e > \pi A_0 e$. The rate of drift from level i to the level $i+1$ is given by $\lambda\gamma\left(\sum_{r=0}^{N}\pi_{K0r} + \pi_{K1N}\right) + \lambda\sum_{r=0}^{N-1}\pi_{M1r}$.

It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to

$$\lambda\gamma\left(\sum_{r=0}^{N}\pi_{K0r} + \pi_{K1N}\right) + \lambda\sum_{r=0}^{N-1}\pi_{M1r} < \frac{1}{\sum_{l=1}^{2M(N+1)+(K-M)(N+2)} m_{1l}}.$$

□

So by an appropriate choice of γ , that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate is greater than service rate.

4.1.2 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of Q be partitioned by

the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{4.1}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = 0 \tag{4.2}$$

and the vectors x_0, x_1 are obtained by solving the eqution

$$x_0 B_1 + x_1 B_2 = 0 \tag{4.3}$$

$$x_0 B_0 + x_1 (A_1 + RA_2) = 0 \tag{4.4}$$

subject to the normalising condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{4.5}$$

To determine x, the rate matrix R should be computed. We use logarithmic reduction algorithm, as in section 2.2.2 in chapter 2, for this purpose.

We again partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0M}, x_{0(M+1)}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, ..., x_{iM}, x_{i(M+1)}, ..., x_{iK})$$

where $i \geqslant 1$ and x_{00} is a scalar and $x_{0j} = (x_{0j0}, x_{0j1})$, where if $1 \leqslant j \leqslant M$, then x_{0j0} are scalars and x_{0j1} are vectors of order $N + 1$ and if $M + 1 \leqslant$

$j \leqslant K$, then x_{0j0} and $x_{0j1}$are scalars. Also

$$x_{ij} = (x_{ij0}, x_{ij1})$$

where $i \geqslant 1$ and if $1 \leqslant j \leqslant M$, then x_{ij0} and x_{ij1} are vectors of order $N+1$ and if $M+1 \leqslant j \leqslant K$, then x_{ij0} are vectors of order $N+1$ but x_{ij1} are scalars.

4.2 Performance characteristics

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{M} \sum_{b=0}^{1} \sum_{r=0}^{N} x_{ijbr} + \sum_{j=M+1}^{K} (x_{ij1N} + \sum_{r=0}^{N} x_{ij0r})$$

for $i > 0$ and

$$a_0 = x_{00} + \sum_{j=1}^{M} (x_{0j00} + \sum_{r=0}^{N} x_{0j1r}) + \sum_{j=M+1}^{K} (x_{0j00} + x_{0j1N}).$$

2. The probability that there are j customers in the buffer (including the one in service)is

$$b_j = \begin{cases} x_{00}\ , & if \quad j = 0 \\ x_{0j00} + \sum_{r=0}^{N} x_{0j1r} + \sum_{i=1}^{\infty} \sum_{b=0}^{1} \sum_{r=0}^{N} x_{ijbr}, if\ 1 \leqslant j \leqslant M \\ x_{0j00} + x_{0j1N} + \sum_{i=1}^{\infty} \sum_{r=0}^{N} x_{ij0r} + \sum_{i=1}^{\infty} x_{ij1N}, if \quad M+1 \leqslant j \leqslant K \end{cases}$$

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that a customer, on its arrival enters the pool is γb_K.

6. The probability that an arriving customer enters service immediately is b_0.

7. The rate at which the customer who find the buffer full leave the system without service (mean number of customers not joining the system per unit time) is

$$\theta_{LOST} = \lambda(1 - \gamma) b_K.$$

8. The rate at which pooled customers transfer in to the buffer for immediate service is

$$\theta_{TR} = \sum_{i=1}^{\infty} \sum_{b=0}^{1} \sum_{r=0}^{N} x_{i1br} \mu + \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=0}^{1} \sum_{r=0}^{N-1} x_{ijbr} p \mu + \sum_{i=1}^{\infty} \sum_{j=1}^{K} x_{ij0N} \mu.$$

9. Interruption rate is

$$I_R = \sum_{i=0}^{\infty} \sum_{r=0}^{N-1} x_{iM1r}\lambda.$$

4.3 Numerical results

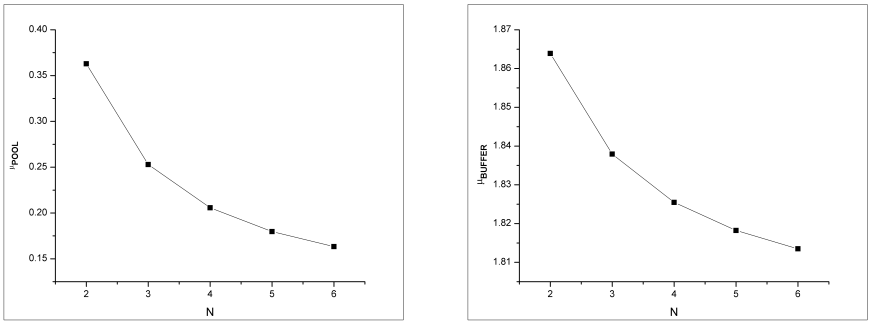

Fig 4.2: N versus μ_{POOL} and μ_{BUFFER}

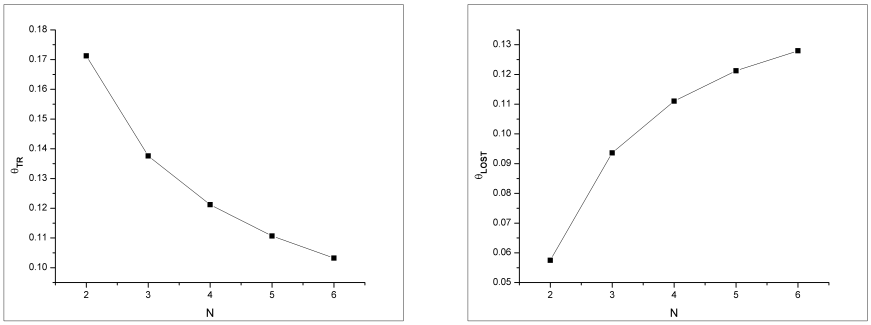

Fig 4.3: N versus θ_{TR} and θ_{LOST}

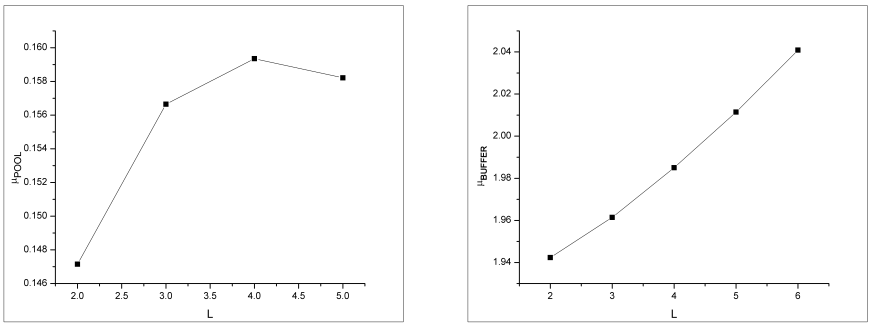

Fig 4.4: L versus μ_{POOL} and μ_{BUFFER}

In this section, we illustrate the performance of the system by considering some numerical results. A customer encountering the buffer full, will be inclined to join the pool with higher γ if the L and p values are larger. On the other hand γ inversely varies with K and N. To model this situation, we take $\gamma = \frac{Lp}{K} + \frac{1}{N}$. But the relationship is feasible for those values of L, p, K and N such that $0 \leqslant \gamma \leqslant 1$.

The impact of N on various measures of descriptors with $K = 6, L = 3, M = 4, \lambda = 5, \mu = 7, p = 0.5, \gamma = \frac{Lp}{K} + \frac{1}{N}$, is shown in figure 4.2 and figure 4.3. As N increases μ_{POOL}, μ_{BUFFER}, θ_{TR} decrease monotonically whereas θ_{LOST} increases monotonically. This is due to the fact that by our assumption γ varies inversely as N and as a result, loss rate increases and inflow rate to the pool decreases as N increases. So the transfer rate of the interrupted customer from the pool to the buffer decreases, and thus mean buffer size decreases.

By keeping $K = 6, L = 3, M = 4, \lambda = 5, \mu = 7, N = 3, \gamma = \frac{Lp}{K} + \frac{1}{N}$ the effect of p on various measures are numerically computed and shown in

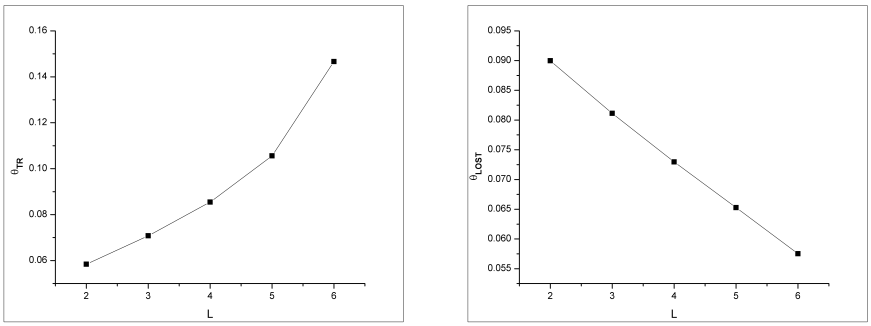

Fig 4.5: L versus θ_{TR} and θ_{LOST}

table 4.1. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing in p, as expected. The measures are computed for various values of L also by keeping $K = 7, p = 0.5, M = 6, \lambda = 5, \mu = 7, N = 4, \gamma = \frac{Lp}{K} + \frac{1}{N}$ and shown in figures 4.4 and 4.5. Here also μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing as expected, in L. All the above are true due to the fact that by our assuption, γ varies directly as p and L. As a result, loss rate decreases and inflow rate to the pool increases as p and L increases. This will make μ_{POOL} increasing. But after an initial increase, μ_{POOL} will decrease for lower arrival rate λ. This is due to the increasing of the transfer rate from the pool to the buffer. So mean buffer size increases. Further as M increases interruption rate decreases as shown in figure 4.6 which is also expected.

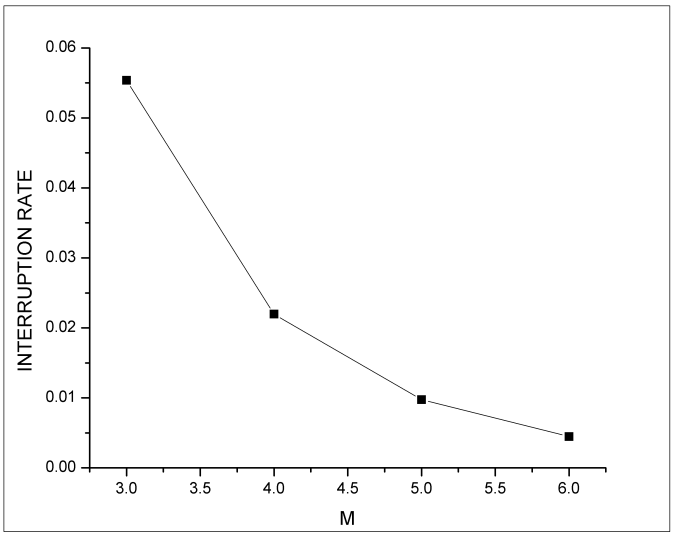

Fig 4.6: M versus interruption rate

p	μ_{POOL}	μ_{BUFFER}	θ_{TR}	θ_{LOST}
0.1	0.1765235	1.7919855	0.0841676	0.1322650
0.2	0.1941219	1.8023695	0.0969422	0.1228671
0.3	0.2126601	1.8134754	0.1100714	0.1133072
0.4	0.2322224	1.8253230	0.1236016	0.1035701
0.5	0.2529308	1.8379211	0.1375848	0.0936373
0.6	0.2749491	1.8512663	0.1520791	0.0834871
0.7	0.2984872	1.8653363	0.1671473	0.0730956
0.8	0.3238123	1.8800848	0.1828565	0.0624368
0.9	0.3512580	1.8954289	0.1992740	0.0514831

Table 4.1: $K = 6, M = 4, \lambda = 5, \mu = 7, N = 3, L = 3, \gamma = \frac{Lp}{K} + \frac{1}{N}$

Chapter 5

A Discrete time $Geo/PH_d/1$ Queue with Postponed work under N-policy

In the previous chapters we described continuous time level independent quasi birth-death processes of postponed work. We started with a basic model and assumptions that are gradually added to realize more practical situations. From this chapter on we consider some realistic models in discrete time. Now a days, discrete time queueing systems are widely applied in telecommunications and computer networks. This is the reason for its wide analysis by many researchers. As a beginning, in this chapter, we

Some results of this chapter are included in the following paper.
1. A.Krishnamoorthy, C.B.Ajayakumar, P.K.Pramod, A Discrete time $Geo/PH_d/1$ Queue with Postponed work under N-policy (Communicated)

consider the discrete time counter part of the model discussed in chapter 2.

5.1 Mathematical description

Consider a $Geo/PH_d/1$ queue with finite buffer of capacity K. If the buffer contains less than K customers including the one at server, newly arriving customer will join it. When the buffer is full with K customers newly arriving customers are offered the choice of leaving the system immediately with probability $1-\gamma$ or of being postponed with probability γ $(0 \leqslant \gamma < 1)$ until the system is less congested. When at the end of a service, if there are postponed customers, the system operates as described in chapter 2. That is, if the buffer is empty the one ahead of all waiting in the pool gets trasferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$, at a service completion epoch, then again the job at the head of the buffer starts service and with probability p, the head of the queue in pool is transferred to the finite buffer and positioned as the last among the waiting customers in the buffer. With probability $q = 1-p$, no such transfer takes place. If there is at least L customers in the buffer at a service completion epoch then no such transfer takes place. Also if the pool contain at least one postponed job, the continuously served customers from the buffer since the last transfer under N-policy is counted, at each service completion epoch. When it reaches a pre-assigned number N $(N > 0)$, then the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. At this time, system does not consider the p-transfer. The N-

policy introduced here differs from the classical N-policy as explained below. In the classical case, N customers are to queue up to start the new service cycle once the system becomes empty. However in the present case N-policy is applied to determine a priority service to be given to a customer from the pool.

In Continuous time case atmost one event can take place with positive probability in a short interval. But in discrete time queueing system, time axis is divided in to intervals of equal length called slots, and where all queueing activities takeplace at the slot boundaries. An arrival and a departure also can happen at a slot boundary. In other words, two or more distinct events can also take place at slot boundaries in the discrete time set up. Whereas in the continuous time case we were constructing the infinitesimal generator of the underlying Markov chain, in the discrete time set up, it is the transition probability matrix, with transitions taking place at slot boundaries is considered. Let the time axis be marked by $0, 1, 2, ..., n, ...$ For mathematical clarity, we assume that departures occur in the interval $(n-, n)$ and arrivals occur in the interval $(n, n+)$. That is, departures occur at the moment immediately before the slot boundaries and arrivals occur at the moment immediately after the slot boundaries.

In this chapter we assume that the time between two successive arrivals is governed by a geometrical law with parameter α and the service time distribution of each customer by a discrete phase-type distribution described by an irreducible PH-representation (β, S) of order m. So, probability of an arrival is α. The model is studied as a quasi birth-death(QBD) process and a solution of the classical matrix geometric type is obtained (see [45] and [38]). We define the state space of the QBD and exhibit the structure of its transition probability matrix.

The state space consists of all tuples of the form (i, j, b, h) with $i \geqslant 1$; $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant m$ where i is the number of postponed work, j is the number of work in the finite buffer including the unit in service, b is the number of continuously served customers from the buffer since the last transfer from the pool under the N-policy and h is the phase of the service in progress at any given epoch. For a given value of $i \neq 0$, $K(N+1)m$ states constitute the level i of the QBD. Now consider the boundary level $i = 0$. Here we denote the empty system $(0, 0, 0, 0)$ by 0. Also there are Km states of the form $(0, j, 0, h)$, $1 \leqslant j \leqslant K$; $1 \leqslant h \leqslant m$. This is due to the fact that when the pool has no customers, N-policy is suspended. These have the same significance as before, except that in these states, no postponed job is present, but there are jobs in the finite buffer. These $Km+1$ states make up the boundary level 0 of the QBD. At time n, the system can be described by the Markov process $\{X_n, n \geqslant 1\}$ with $X_n = (i, j, b, h)$.

The transition probability matrix is

$$P = \begin{bmatrix} B_1 & B_0 & & & & \\ B_2 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the matrix B_0 is of dimension $(Km+1) \times K(N+1)m$, B_1 is square matrix of order $Km+1$ and B_2 is of dimension $K(N+1)m \times (Km+1)$. A_0, A_1 and A_2 are square matrices of order $K(N+1)m$. Each of these matrices is itself highly structured.

Except for a single block $\alpha\gamma t_5 \otimes S$ at its south- east corner, the matrix B_0 is zero, where t_5 is a row vector of order $N+1$ with first element 1 and all other elements are zeros. The matrix B_1 corresponds to the transition from the level 0 to 0 is given below:

$$B_1 = \begin{bmatrix} 1-\alpha & \alpha\beta & & & & \\ \Delta_3 & \Delta_1 & \alpha S & & & \\ & \Delta_4 & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \Delta_1 & \alpha S \\ & & & & \Delta_4 & \Delta_2 \end{bmatrix}$$

where all non specified entries are zeros;

$\Delta_1 = (1-\alpha)S + \alpha S^0\beta$ corresponds to the transition of the buffer size from j to j for $j = 1, 2, ..., K-1$;

$\Delta_2 = (1-\alpha\gamma)S + \alpha S^0\beta$ corresponds to the transition of the buffer size from K to K;

$\Delta_3 = (1-\alpha)S^0$ corresponds to the transition of the buffer size from 1 to 0;

$\Delta_4 = (1-\alpha)S^0\beta$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., K$.

The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \Delta_6 & \Omega_6 & & & & & & & & \\ & \Delta_7 & \Omega_7 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \ddots & \ddots & & & & & \\ & & & & \Delta_7 & \Omega_7 & & & & \\ & & & & & \Delta_8 & \Omega_8 & & & \\ & & & & & & \ddots & \ddots & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & \Delta_8 & \Omega_8 \\ & & & & & & & & & \Delta_9 \end{bmatrix}$$

where

$\Delta_6 = (1-\alpha)t_6 \otimes S^0\beta$ corresponds to the transition of the buffer size from 1 to 1;

$\Omega_6 = \alpha t_6 \otimes S^0\beta$ corresponds to the transition of the buffer size from 1 to 2;

$\Delta_7 = (1-\alpha)t_7 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$;

$\Omega_7 = \alpha t_7 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$;

$\Delta_8 = (1-\alpha)t_8 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to j for $j = L+1, ..., K-1$;

$\Omega_8 = \alpha t_8 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j+1$ for $j = L+1, ..., K-1$;

$\Delta_9 = (1-\alpha\gamma)t_8 \otimes S^0\beta$ corresponds to the transition of the buffer size from K to K.

Also t_6 is a column vector of order $N+1$ with all entries 1, t_7 is a column vector of order $N+1$ with all elements p except 1 at $(N,1)^{th}$ position and t_8 is a column vector of order $(N+1)$ with $(N,1)^{th}$ element 1 and all other elements zero.

The matrix A_0 is zero except for a single block $\alpha\gamma I_{N+1} \otimes S$ at its south-east corner where I_{N+1} is the identity matrix of order $N+1$. The matrix A_2 is given by

$$A_2 = \begin{bmatrix} \Gamma_1 & \Omega_1 & & & & & & & & \\ & \Gamma_2 & \Omega_2 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \ddots & \ddots & & & & & \\ & & & & \Gamma_2 & \Omega_2 & & & & \\ & & & & & \Gamma_3 & \Omega_3 & & & \\ & & & & & & \ddots & \ddots & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & \Gamma_3 & \Omega_3 \\ & & & & & & & & & \Gamma_4 \end{bmatrix}$$

where

$\Gamma_1 = (1-\alpha)t_1 \otimes S^0\beta$ corresponds to the transition of the buffer size

from 1 to 1;

$\Gamma_2 = (1-\alpha)t_2 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$;

$\Gamma_3 = (1-\alpha)t_3 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to j for $j = L+1, ..., K-1$;

$\Gamma_4 = (1-\alpha\gamma)t_3 \otimes S^0\beta$ corresponds to the transition of the buffer size from K to K

$\Omega_1 = \alpha t_1 \otimes S^0\beta$ corresponds to the transition of the buffer size from 1 to 2;

$\Omega_2 = \alpha t_2 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$;

$\Omega_3 = \alpha t_3 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j+1$ for $j = L+1, ..., K-1$.

Also t_1 is a square matrix of order $N+1$ given by

$$t_1 = \begin{bmatrix} \bar{0} & I_N \\ 1 & \bar{0} \end{bmatrix}$$

where I_N is identity matrix of order N and $\bar{0}$ is zero matrix of appropriate

order. Again t_2 is a square matrix of order $N+1$ given by

$$t_2 = \begin{bmatrix} 0 & p & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & p & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & p & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ p & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

and t_3 is a square matrix of order $N+1$ with $(N, N+1)^{th}$ entry 1 and all other entries zero.

The matrix A_1 is given by

$$A_1 = \begin{bmatrix} \Gamma_5 & \Delta_5 & & & & & & & & \\ \Omega_4 & \Gamma_6 & \Delta_5 & & & & & & & \\ & \ddots & \ddots & \ddots & & & & & & \\ & & \ddots & \ddots & \ddots & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & \Omega_4 & \Gamma_6 & \Delta_5 & & & \\ & & & & & \Omega_5 & \Gamma_7 & \Delta_5 & & \\ & & & & & & \ddots & \ddots & \ddots & \\ & & & & & & & \ddots & \ddots & \ddots \\ & & & & & & & & \Omega_5 & \Gamma_7 & \Delta_5 \\ & & & & & & & & & \Omega_5 & \Gamma_8 \end{bmatrix}$$

where

$\Delta_5 = \alpha I_{N+1} \otimes S$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, 2, ..., K-1$;

$\Gamma_5 = I_{N+1} \otimes (1-\alpha)S$ corresponds to the transition of the buffer size from 1 to 1;

$\Gamma_6 = \alpha q t_4 \otimes S^0\beta + (1-\alpha)I_{N+1} \otimes S$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$;

$\Gamma_7 = \alpha t_4 \otimes S^0\beta + (1-\alpha)I_{N+1} \otimes S$ corresponds to the transition of the buffer size from j to j for $j = L+1, L+2, ..., K-1$;

$\Gamma_8 = \alpha t_4 \otimes S^0\beta + (1-\alpha\gamma)I_{N+1} \otimes S + \alpha\gamma t_3 \otimes S^0\beta$ corresponds to the transition of the buffer size from K to K;

$\Omega_4 = (1-\alpha)q t_4 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, ..., L$;

$\Omega_5 = (1-\alpha)t_4 \otimes S^0\beta$ corresponds to the transition of the buffer size from j to $j-1$ for $j = L+1, ..., K$.

Also t_4 is a square matrix of order $N+1$ which is given below.

$$t_4 = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

5.1.1 Stability criterion

Theorem 5.1.1. *The system is stable if and only if*

$$\alpha\gamma \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}$$

where π is the unique solution to $\pi A = \pi$; $\pi e = 1$ for $A = A_0 + A_1 + A_2$.

Proof. Let $G_{ll'}$ be the conditional probability that the QBD process starting in the state $l = (i, j, b, h)$ (for $i > 1$) where $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant m$ at time $t = 0$ reaches the state $l' = (i-1, j', b', h')$ where $1 \leqslant j' \leqslant K$; $0 \leqslant b' \leqslant N$; $1 \leqslant h' \leqslant m$, for the first time, in a finite time. That is

$$G_{ll'} = P[\tau < \infty : \chi(\tau) = l' | \chi(0) = l]$$

where τ is the first passage time from the level i to the level $i-1$. Because of the structure of Q, the probability $G_{ll'}$ does not depend on i. The matrix with elements $G_{ll'}$ is denoted by G.

Suppose the matrix $A = A_0 + A_1 + A_2$ is irreducible. Then the necessary and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0 + A_1 + A_2$. That is, it is the unique solution to $\pi A = \pi$, $\pi e = 1$ and $A = A_0 + A_1 + A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. G is obtained as the minimal non negative solution

to the matrix quadratic equation

$$G = A_2 + A_1 G + A_0 G^2.$$

This is obvious. On the left-hand side, G records the distribution of the first state visited in l' conditioned on the initial state being in l. On the right-hand side, these visits to l' are decomposed in to three groups; the first term corresponds to the case where the QBD directly moves from i to $i-1$ in one transition with probabilities recorded in A_2; as for the second term, with probabilities recorded in A_1, the QBD remains in l from where it still has to move eventually to l', with probabilities recorded in G; finally for the last term, with probabilities recorded in A_0, the QBD moves up to $i+1$ from where it still has to move eventually to l, with probabilities recorded in G and then to l' again with probabilities recorded in G.

Let $m_1 = [m_{1_l}]$ denotes the column vector of dimension $K(N+1)m$ where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started in the state l. We have $G = (I - A_1)^{-1}A_2 + (I - A_1)^{-1}A_0 G^2$. Consequently $m_1 = [I - A_1 - A_0(I + G)]^{-1}e$.

For the system stability, the rate of drift from level i to level $i-1$ should be greater than that to level $i+1$. The rate of drift from level i to the level $i+1$ is given by $\alpha\gamma \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh}$. It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to

$$\alpha\gamma \sum_{b=0}^{N} \sum_{h=1}^{m} \pi_{Kbh} < \frac{1}{\sum_{l=1}^{K(N+1)m} m_{1_l}}.$$

□

So by an appropriate choice of γ, that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate is greater than service rate.

5.1.2 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of P be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{5.1}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = R \tag{5.2}$$

and the vectors x_0, x_1 are obtained by solving the eqtions

$$x_0(B_1 - I) + x_1 B_2 = 0 \tag{5.3}$$

$$x_0 B_0 + x_1(A_1 - I + RA_2) = 0 \tag{5.4}$$

subject to the normalising condition

$$x_0 e + x_1(I - R)^{-1} e = 1 \tag{5.5}$$

From the above discussion it is clear that to determine x, a key step is the computation of the rate matrix R. For this, we use logarithmic reduction algorithm (see [38]). The important steps of this algorithm is given below.
Assign $H := (I - A_1)^{-1}A_0$; $L := (I - A_1)^{-1}A_2$; $G := L$; and $T := H$;
and repeat
$U := HL + LH$; $M := H^2$; $H := (I - U)^{-1}M$; $M := L^2$;
$L := (I - U)^{-1}M$; $G := G + TL$; $T := TH$
until $||1 - G.e||_\infty \leqslant \epsilon$.
Then $R = A_0(I - A_1 + A_0G)^{-1}$
We can partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, x_{i3}, ..., x_{iK})$$

where $i \geqslant 1$; x_{00} is a scalar and x_{0j}, $1 \leqslant j \leqslant K$ are vectors of order m and

$$x_{ij} = (x_{ij0}, x_{ij1}, ..., x_{ijN})$$

where $i \geqslant 1$; $1 \leqslant j \leqslant K$ and x_{ijb}, $0 \leqslant b \leqslant N$ are vectors of order m.

5.2 Computation of Expected values

In this section we derive the expected waiting time of a tagged customer (i) in the buffer and (ii) in the pool. Also we calculate the expectation of the number of FIFO violation.

5.2.1 Expected waiting time in buffer

We denote the mean waiting time of customers, who upon their arrival enter the buffer, by $E(W_1)$.

Case 1. $N \geqslant K$

In this case the tagged customer is not affected by the new arrivals in the buffer and in the pool. So we can calculate the waiting time by considering the system state at which the tagged customer enters. Hence

$E(W_1) = \sum_i \sum_j \sum_b \sum_h E(\text{waiting time of the customer who finds the system in state } (i,j,b,h))\ Pr(\text{system is in state } (i,j,b,h))$

$$E(W_1) = \sum_{j=1}^{k-1} \sum_{h=1}^{m} \beta(I-S)^{-1} e(j-1) x_{0j0h}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{b=0}^{N-1} \sum_{h=1}^{m} \beta(I-S)^{-1} e(j-1+\psi) x_{ijbh}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K-1} \sum_{h=1}^{m} \beta(I-S)^{-1} e(j-1) x_{ijNh} + \pi^*(I-S)^{-1} e$$

where $\pi^* P^* = \pi$, $\pi^* e = 1$ and $P^* = S + S^0\beta$ and

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], \quad 0 \leqslant b < N$$

where [y] denotes the greatest integer value of y. $\pi^*(I-S)^{-1}e$ is the additional time required to complete the service of the customer who is in service when the tagged person enters the buffer.

Case 2. $N < K$

In this case, the tagged customer in the buffer will be affected by the number of new arrivals in the pool and so the number of new arrivals in the buffer. So the waiting time of the tagged customer depends on the following susequent developments in the pool: one or more visits to zero level, and a finite number of customers joining the pool after the tagged customer. Because of the complexity of calculation, we may turn to computing an upper bound on the waiting time, by keeping in mind, the fact that only a maximum finite number K of persons in the pool will affect the tagged person. In the worst case we have $N = 1$ which represents service alternating between buffer and pool. So an upper bound for the waiting time of a customer who upon his arrival enters the buffer in the state (i, j, b, h), is

$$UB(W_1) = \sum_{j=1}^{k-1}\sum_{h=1}^{m} \beta(I-S)^{-1}e(j-1+[\tfrac{j}{N}])x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{b=0}^{N-1}\sum_{h=1}^{m} \beta(I-S)^{-1}e(j-1+\psi)x_{ijbh}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K-1}\sum_{h=1}^{m} \beta(I-S)^{-1}e(j-1+[\frac{j-1}{N}])x_{ijNh} + \pi^*(I-S)^{-1}e$$

where $\pi^* P^* = \pi$, $\pi^* e = 1$ and $P^* = S + S^0\beta$ and

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

$\pi^*(I-S)^{-1}e$ is the additional time required to complete the service of the customer who is at the server when the tagged person enter buffer.

5.2.2 Expected waiting time in pool

We denote the expected waiting time of a customer who upon his arrival enters the pool, by $E(W_2)$.

To find this, first we define the Markov process $\{X(t)\}$ as follows. $X(t) = (a, j, b, h)$ where a denotes the rank of the tagged customer entered pool, j denotes the number of customers in the buffer, b denotes the number of continuously served customers from buffer and h is the phase of the service process at time t. The rank a of the customer is assumed to be r if he joins as r^{th} customer in pool. His rank may decrease to 1 with the customers ahead of him transferred from pool to buffer. Since the customers who arrive after the tagged customer cannot change his rank, level changing transitions in $\{X(t)\}$ can takeplace only to one side of the diagonal. We arrange the statespace of $\{X(t)\}$ as

$$\{r, r-1, ..., 2, 1\} \times \{1, 2, ..., K\} \times \{0, 1, ..., N\} \times \{1, 2, ..., m\}$$

with absorbing state 0 in the sense that the tagged customer is either selected to be served under N-policy or placed in the buffer with probability p or to the server with probability 1 if the buffer size reduces to 0 at the end of a service. The infinitesimal generator of the process is

$$\tilde{P} = \begin{bmatrix} T & T^0 \\ \bar{0} & 0 \end{bmatrix}$$

where

$$T = \begin{bmatrix} A_1 & A_2 & & & & \\ & A_1 & A_2 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & A_1 & A_2 \\ & & & & & A_1 \end{bmatrix}$$

of order $rK(N+1)m$ and

$$T^0 = \begin{bmatrix} \bar{0} \\ \vdots \\ \bar{0} \\ B_2 \end{bmatrix}.$$

Now the expected absorption time of a particular customer is given by the column vector

$$E_w^{(r)} = \tilde{I}(I - T)^{-1}e$$

where $\tilde{I} = \begin{bmatrix} I_{K(N+1)m} & \bar{0} \end{bmatrix}$ having order $K(N+1)m \times rK(N+1)m$ and e is a column vector of ones of order $rK(N+1)m$. So the expected waiting time of the tagged customer is

$$W_L = \sum_{r=1}^{\infty} x_r E_w^{(r)}$$

where x_r is the steady state probability vector corresponding to $i = r$. W_L gives the waiting time of a customer in pool up to the epoch of his transfer to buffer.

Case 1. $N \geqslant K$

Expected waiting time in pool is

$$E(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L(x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+\sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh}(W_L + W^{(1)} p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$W^{(1)} = \sum_{j=1}^{L}\sum_{h=1}^{m} \beta(I-S)^{-1} e(j-1) x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m} \beta(I-S)^{-1} e(j-1+\psi) x_{ijbh}$$

and

$$\psi = 1 + \left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

Case 2. $N < K$

In this case we get an upperbound $UB(W_2)$ for the waiting time in pool.

$$UB(W_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh} W_L(x_{ij(N-1)h} s_{h0} + x_{i1bh} s_{h0})$$

$$+\sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh}(W_L + UB(W^{(1)}) p(\sum_{j=1}^{L} x_{ijbh} s_{h0}))$$

where

$$UB(W^{(1)}) = \sum_{j=1}^{L}\sum_{h=1}^{m}\beta(I-S)^{-1}e(j-1+\left[\frac{j-1}{N}\right])x_{0j0h}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{L}\sum_{b=0}^{N-1}\sum_{h=1}^{m}\beta(I-S)^{-1}e(j-1+\psi)x_{ijbh}$$

and

$$\psi = 1+\left[\frac{j-(N-b)}{N}\right], 0 \leqslant b < N.$$

5.2.3 Expected number of FIFO violation

It may be noted that the N-policy leads to violation of FIFO rule for customers in the pool. For example assume that there are two or more customers in the pool at a service completion epoch at which the number in the buffer droped to $L-1$ or below and the number of continuously served customers reached $N-1$. So the first in the pool may be selected under p-transfer and placed as the last in the buffer. When the next service is completed, the current head of the pool gets transferred to the buffer for immediate service there by violating the FIFO rule for pooled customers. Further it may be noted that this situation does not arise among the queued customers in the buffer.

We compute the expectation of the indicator random variable defined as FIFO violation in pool. Its expectation is the probability for FIFO

violation in pool which is given by

$$P_{FIFO} = \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=N-j+1}^{N-1} x_{ijbh} pS_{h0}.$$

The FIFO may be violated by more than one customers who join the pool after the tagged customer joins the buffer when $N < L$. However this can be overcome by making N larger than L. If $N \geqslant K$, a customer joining the pool will not overtake any of the customers in the buffer who had joined before his entering the pool. At this time, FIFO is violated by atmost one successor in pool. Even this can be overcome by a slight modification by redefining the N-policy by resetting b in (i, j, b, h) as zero at the time of p-transfer.

5.3 Performance characteristics

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{K} \sum_{b=0}^{N} \sum_{h=1}^{m} x_{ijbh}$$

for $i \geqslant 1$ and

$$a_0 = x_{00} + \sum_{j=1}^{K} \sum_{h=1}^{m} x_{0j0h}.$$

2. The probability that there are j customers in the buffer (including

the one in service) is

$$b_j = \sum_{h=1}^{m} x_{0j0h} + \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{ijbh}$$

for $1 \leqslant j \leqslant K$ and

$$b_0 = x_{00}.$$

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that a customer, on its arrival enters the pool is γb_K.

6. The probability that an arriving customer enters service immediately is b_0.

7. The probability that the customer who find the buffer full leave the system without service (mean number of customers not joining the system per unit time) is

$$\theta_{LOST} = \alpha(1 - \gamma) b_K.$$

That is

$$\theta_{LOST} = \alpha(1-\gamma)(\sum_{h=1}^{m} x_{0K0h} + \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{iKbh}).$$

8. The rate at which pooled customers transfer in to the buffer is

$$\theta_{TR} = \sum_{i=1}^{\infty}\sum_{b=0}^{N}\sum_{h=1}^{m} x_{i1bh}S_{h0} + \sum_{i=1}^{\infty}\sum_{j=2}^{L}\sum_{b=0}^{N-2}\sum_{h=1}^{m} x_{ijbh}pS_{h0}$$

$$+\sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0} + \sum_{i=1}^{\infty}\sum_{j=2}^{L}\sum_{h=1}^{m} x_{ijNh}pS_{ho}.$$

9. The rate at which pooled customers transfer under N-policy(mean number of transfers under N-policy per unit time) is

$$T_N = \sum_{i=1}^{\infty}\sum_{j=1}^{K}\sum_{h=1}^{m} x_{ij(N-1)h}S_{h0}.$$

10. Mean number of customers served out per unit time is

$$\mu_{SERVED} = (1-b_o)\,\frac{1}{\beta(I-S)^{-1}e}.$$

5.4 Numerical results

We present some numerical results in order to illustrate the performance of the system. Take $\gamma = \frac{Lp}{K} + \frac{1}{N}$ in order to bring out explicitly the dependence of γ on the system parameters.

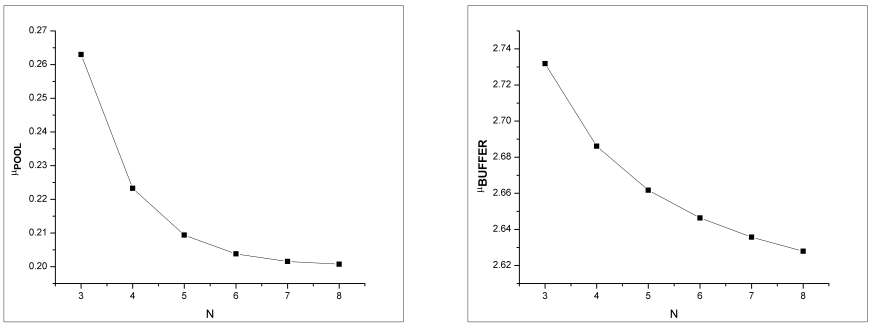

Fig 5.1: N versus μ_{POOL} and μ_{BUFFER}

L	μ_{POOL}	μ_{BUFFER}	θ_{TR}	θ_{LOST}
2	0.1490699	2.6181972	0.0079576	0.0241934
3	0.2093819	2.6617301	0.0102942	0.0220709
4	0.2621410	2.7166486	0.0131344	0.0200996
5	0.2811960	2.7723031	0.0173686	0.0188443

Table 5.1: $K = 6, p = 0.5, m = 2, \alpha = 0.4, N = 5, \gamma = \frac{Lp}{K} + \frac{1}{N}$

This is justified as follows. Larger the L value, the customer encountering the buffer full, will be inclined to join the pool with higher probability. Also same is the relationship of γ with p. On the other hand, γ inversely varies with K. The additional term $\frac{1}{N}$ comes through N-policy. Here as N increases γ decreases so that γ and N vary inversely. But the relationship is feasible for those values of L, p, K and N such that $0 \leqslant \gamma \leqslant 1$. This is possible if $N \geqslant K$ and such a selection is highly consistent. But N can be made less than K by suitably selecting other variables so that $0 \leqslant \gamma \leqslant 1$, and that can be considered as an incentive to customers joining the pool.

The impact of N on various measures of descriptors with $K = 6, L =$

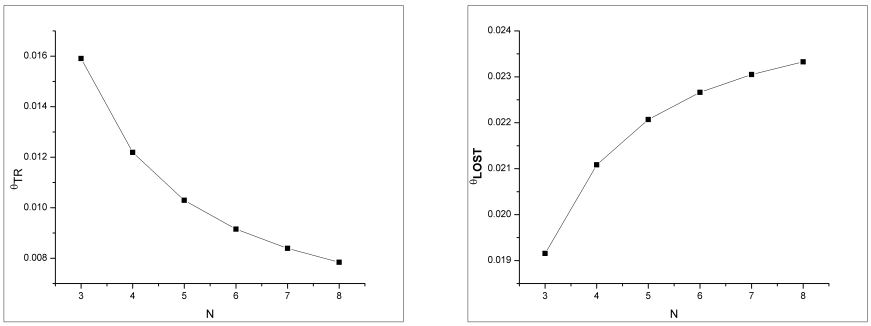

Fig 5.2: N versus θ_{TR} and θ_{LOST}

$3, m = 2, \alpha = 0.4, p = 0.5, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \quad S = \begin{bmatrix} 0.3 & 0.2 \\ 0.4 & 0.2 \end{bmatrix} \quad S^0 = \begin{bmatrix} 0.5 \\ 0.4 \end{bmatrix}$$

is shown in figures 5.1 and 5.2. As N decreases μ_{POOL}, μ_{BUFFER}, θ_{TR} increase monotonically whereas θ_{LOST} decrease monotonically. This is due to the fact that by our assumption γ varies inversely as N and as a result, loss rate decreases and inflow rate to the pool increases as N increases. As N decreases, transfer rate from pool to buffer increases, and thus mean buffer size increases.

By keeping $K = 6, L = 3, m = 2, \alpha = 0.4, N = 5, \gamma = \frac{Lp}{K} + \frac{1}{N}$ the effect of p on various measures is shown in figures 5.3 and 5.4. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing in p, as expected. The measures are numerically computed for various values of L and shown in table 5.1. Here also μ_{POOL}, μ_{BUFFER}, θ_{TR} are monotonically increasing and θ_{LOST} is monotonically decreasing

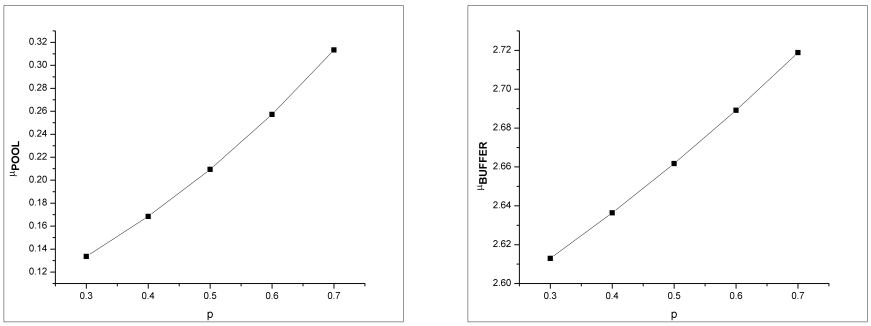

Fig 5.3: p versus μ_{POOL} and μ_{BUFFER}

as expected, in L. All the above are true due to the fact that by our assuption, γ varies directly as p and L. As a result, loss rate decreases and inflow rate to the pool increases as p and L increases. This will make μ_{POOL} increasing. Also transfer rate from pool to buffer increases as p and L increases. So mean buffer size increases.

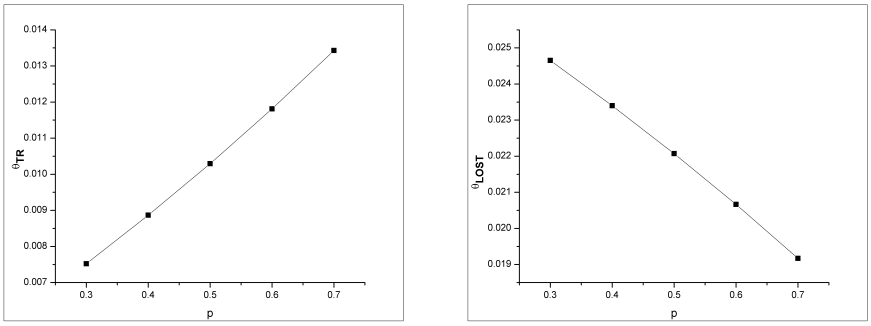

Fig 5.4: p versus θ_{TR} and θ_{LOST}

Chapter 6

Discrete time $Geo/E_d/1$ Queues with Postponed work and Protected stages

In this chapter, we study a discrete time $Geo/E_d/1$ queue with postponed work with the service of each customer having n stages of which first v stages are unprotected. Till now we assumed that all the arriving customers are alike. But here we categorise the customers in to high and low priority customers. If the buffer has at least one customer, the low priority ones are postponed and high priority ones wait in the buffer. When the buffer is full, the system will not permit further arrivals of high pri-

Some results of this chapter are included in the following papers.
1. A.Krishnamoorthy, C.B.Ajayakumar, Discrete time $Geo/E_d/1$ Queues with Postponed work and Protected stages (Communicated).

ority customers. But at that time, a low priority customer can enter the pool with a specified probability or it is also lost from the system with complementary probability.

This is highly practical since in many cases, time is a constraint for the existence of finite capacity queues and the server will be interested to make maximum gain. So he naturally turns to high priority customers and the srvice of low priority customers will be postponed. If the buffer is full, no further high priority customers join it. At that time, even the low priority customer may not accept the offer of postponement. So the priority based postponement is desirable from the system point of view.

However the postponed work are transferred to the buffer for immediate service with a specified probability at a service completion epoch if the number in the buffer at that time is less than a pre-assigned lower level. But during the service of that postponed work, the buffer size may rise to a pre-assigned higher level and so the server will be compelled to preempt the service of the low priority customer. But the server cannot preempt the work if it is on protected stages of service. We discuss two models in this chapter. The preempted work from unprotected stages will be lost for ever from the system in model-1. This is considered as a negative arrival. In queues with negative arrivals, it is assumed that customer(s) in service is removed due to such an arrival. However in this model, a new arrival hitting a pre-assigned higher level in the buffer takes the role of negative arrival since it decreases a low priority customer in the system. This is also common in practical cases. Actually the undergoing work gets damaged without completing the service. Sudden death of a patient in an operation theatre is an example of such a negative arrival.

In model-2, the preempted work from unprotected stages is considered as an interruption and such an interrupted work is again postponed and wait at the head of the pool for the next chance of transfer. But in the interruption period, if the number of continuously served higher priority work from the buffer attains a pre-assigned number at a service completion epoch, the interrupted customer is transferred to the buffer for immediate service and no further interruption is allowed to that customer in service. The service to the interrupted customer is repeated when it is taken for service again.

6.1 Model-1: With negative arrivals

6.1.1 Mathematical formulation

Consider a $Geo/E_d/1$ queue with finite buffer of capacity K. The time between two successive arrivals is governed by a geometrical law with parameter α and the service time of each customer is ruled by a discrete Erlang distribution having n stages of which first v stages are unprotected and the remaining $n-v$ stages are protected. It is described by an irreducible PH-representation (β, S) of order n where

$$\beta = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{1\times n} \quad ;$$

$$S = \begin{bmatrix} s_{11} & s_{12} & & \\ & s_{22} & s_{23} & \\ & & \ddots & \ddots \\ & & & \end{bmatrix}_{n\times n} \quad ; \quad S^0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ s_{n0} \end{bmatrix}_{n\times 1} .$$

Arriving customers are classified in to two categories; high priority customers with arrival probability p_1 and low priority customers with arrival probability p_2. If a higher priority customer on arrival finds the buffer not full, he joins the same. Otherwise he leaves the system permanently. If a low priority customer on arrival sees the buffer empty, he enters the buffer for immediate service. If the buffer has at least one customer, he proceeds to a pool of postponed work having infinite capacity. But if the buffer is full, he joins the pool only with a specified probability γ $(0 \leqslant \gamma < 1)$. With probability $1-\gamma$, such customers do not join the system. So clearly the pool will occupy only low priority customers. But the buffer will be occupied by high priority customers and atmost one transferred low priority customer.

When at the end of a service, if there are postponed customers, the system operates as follows. If the buffer is empty, the one ahead of all waiting in the pool gets transferred to the buffer for immediate service. If the buffer contains y jobs, where $1 \leqslant y \leqslant L-1$; $2 \leqslant L \leqslant K-1$, at a service completion epoch, then with probability p, the head of the queue in the pool is transferred to the buffer for immediate service. With probability $q = 1-p$, no such transfer takes place. When at the end of a service, if the buffer is empty, and the pool has no work, the server becomes idle.

When a pool work is on service in unprotected stages, if the buffer

size rises to a pre-assigned number $M+1$ such that $L \leqslant M \leqslant K-1$ at an arrival epoch, the server will preempt the service of the lower priority customer. The preempted work from unprotected stages will be lost for ever from the system. Here the event which causes to decrease the number of customers by 1, without an actual service completion, is the arrival of a high priority customer to rise the buffer level from M to $M+1$. So here this is the negative arrival(event). We emphasize that if the pool work at server is on protected stages, such an arrival does not act as a negative arrival as there is no preemption for the low priority customer. Following a negative arrival, the server will perform the buffer work. A diagrammatic representation of model-1 is given in figure 6.1.

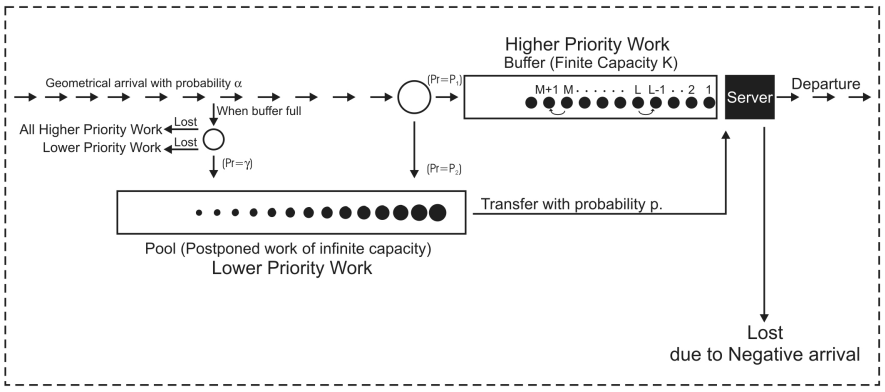

Fig 6.1: $Geo/E_d/1$ queue with postponed work and Negative arrival

In this discrete time queueing system, time axis is divided in to intervals of equal length called slots, and where all queueing activities take place at the slot boundaries. An arrival and a departure also can happen at a slot boundary. In other words, two or more distinct events can also take place at slot boundaries in the discrete time set up. Let the time axis

be marked by $0, 1, 2, ..., m, ...$ For mathematical clarity, we assume that departures occur in the interval $(m-, m)$ and arrivals occur in the interval $(m, m+)$. That is, departures occur at the moment immediately before the slot boundaries and arrivals occur at the moment immediately after the slot boundaries. The model is studied as a quasi birth-death(QBD) process and a solution of the classical matrix geometric type is obtained (see [45] and [38]). We define the state space of the QBD and exhibit the structure of its transition probability matrix.

The state space consists of all tuples of the form (i, j, b, h) where i denotes the number of postponed work in the pool having infinite capacity; j denotes the number of jobs in the finite buffer including the unit in service; b denotes the status of the system where

$$b = \begin{cases} 0 & , \quad \text{buffer work is in progress} \\ 1 & , \quad \text{pool work is being served} \end{cases}$$

and h denotes the stage of service in progress at that instant.

Consider the boundary level $i = 0$. We denote the empty system $(0, 0, 0, 0)$ by 0.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $h = 1, 2, ..., n$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $h = 1, 2, ..., n$.

If $M + 1 \leqslant j \leqslant K$ and $b = 1$ then $h = v + 1, ..., n$. So the boundary level $i = 0$ constitute $N_1 = 1 + 2Mn + (K - M)(2n - v)$ states.

Now consider the level $i \neq 0$.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $h = 1, 2, ..., n$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $h = 1, 2, ..., n$.

If $M + 1 \leqslant j \leqslant K$ and $b = 1$ then $h = v + 1, ..., n$. So there are $N_2 = 2Mn + (K - M)(2n - v)$ states are there in the level $i \neq 0$.

The transition probability matrix is

$$P = \begin{bmatrix} B_1 & B_0 & & & & \\ B_2 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the matrix B_0 is of dimension $N_1 \times N_2$, B_1 is square matrix of order N_1 and B_2 is of dimension $N_2 \times N_1$. A_0, A_1 and A_2 are square of order N_2. Each of these matrices is itself highly structured.

We use the following matrices in the sequel.

$$E = S^0\beta = \begin{bmatrix} \bar{0} & \bar{0} \\ s_{n0} & \bar{0} \end{bmatrix}_{n \times n} ;$$

$$t_1 = \begin{bmatrix} \alpha p_1 \beta & \alpha p_2 \beta \end{bmatrix}_{1 \times 2n} \quad ; \quad t_2 = \begin{bmatrix} (1-\alpha)S^0 \\ (1-\alpha)S^0 \end{bmatrix}_{2n \times 1} ;$$

$$V_1 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \alpha p_2 E \\ \alpha p_1 E & (1-\alpha)S + \alpha p_2 E \end{bmatrix}_{2n\times 2n} ;$$

$$V_2 = \begin{bmatrix} S & \bar{0} \\ \bar{0} & S \end{bmatrix}_{2n\times 2n} ; \qquad V_3 = \begin{bmatrix} E & \bar{0} \\ \bar{0} & E \end{bmatrix}_{2n\times 2n} ;$$

$$V_4 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 E & (1-\alpha)S \end{bmatrix}_{2n\times 2n} ;$$

$$V_5 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 E + \alpha p_1 t_3 & (1-\alpha)S \end{bmatrix}_{2n\times 2n} ; \quad t_3 = \begin{bmatrix} e_v & \bar{0} \\ \bar{0} & \bar{0} \end{bmatrix}_{n\times n}$$

where e_v is a column vector of ones of order v;

$$t_4 = \begin{bmatrix} s_{(v+1)(v+1)} & s_{(v+1)(v+2)} & & \\ & s_{(v+2)(v+2)} & \ddots & \\ & & \ddots & \\ & & & s_{nn} \end{bmatrix}_{(n-v)\times(n-v)} ;$$

$$t_5 = \begin{bmatrix} \bar{0} \\ t_4 \end{bmatrix}_{n\times(n-v)} ; \qquad t_6 = \begin{bmatrix} \bar{0} & \bar{0} \\ s_{n0} & \bar{0} \end{bmatrix}_{(n-v)\times n}$$

$$V_6 = \begin{bmatrix} S & \bar{0} \\ \bar{0} & t_5 \end{bmatrix}_{2n\times(2n-v)} \quad ; \qquad V_7 = \begin{bmatrix} E & \bar{0} \\ t_6 & \bar{0} \end{bmatrix}_{(2n-v)\times 2n} \quad ;$$

$$V_8 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 t_6 & (1-\alpha)t_4 \end{bmatrix}_{(2n-v)\times(2n-v)} \quad ;$$

$$V_9 = \begin{bmatrix} S & \bar{0} \\ \bar{0} & t_4 \end{bmatrix}_{(2n-v)\times(2n-v)} \quad ; \qquad V_{10} = \begin{bmatrix} E & \bar{0} \\ t_6 & \bar{0} \end{bmatrix}_{(2n-v)\times(2n-v)} \quad ;$$

$$V_{11} = \begin{bmatrix} (1-\alpha p_2\gamma)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 t_6 & (1-\alpha p_2\gamma)t_4 \end{bmatrix}_{(2n-v)\times(2n-v)} \quad ;$$

$$V_{12} = \begin{bmatrix} \bar{0} & E \\ \bar{0} & E \end{bmatrix}_{2n\times 2n} \quad ; V_{13} = \begin{bmatrix} (1-\alpha)S & \alpha p_2 E \\ \bar{0} & (1-\alpha)S + \alpha p_2 E \end{bmatrix}_{2n\times 2n} \quad ;$$

$$V_{14} = \begin{bmatrix} (1-\alpha)S + q\alpha p_1 E & p\alpha p_2 E \\ q\alpha p_1 E & (1-\alpha)S + p\alpha p_2 E \end{bmatrix}_{2n\times 2n} .$$

The matrix B_1 corresponds to the transition from the level 0 to 0 is

given below:

$$B_1 = \begin{bmatrix} 1-\alpha & t_1 & & & & & & & & \\ t_2 & V_1 & \Phi_1 & & & & & & & \\ & \Delta_1 & V_2 & \ddots & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & & V_5 & \Phi_2 & & & & \\ & & & & \Delta_2 & V_8 & \Phi_3 & & & \\ & & & & & \Delta_3 & \ddots & \ddots & & \\ & & & & & & \ddots & \ddots & & \\ & & & & & & & & V_8 & \Phi_3 \\ & & & & & & & & \Delta_3 & V_{11} \end{bmatrix}_{N_1 \times N_1}$$

where $\Phi_1 = \alpha p_1 V_2$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, 2, ..., M-1$, $\Phi_2 = \alpha p_1 V_6$ corresponds to the transition of the buffer size from M to $M+1$ and $\Phi_3 = \alpha p_1 V_9$ corresponds to the transition of the buffer size from j to $j+1$ for $j = M+1, M+2, ..., K-1$. $\Delta_1 = (1-\alpha)V_3$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., M$, $\Delta_2 = (1-\alpha)V_7$ corresponds to the transition of the buffer size from $M+1$ to M and $\Delta_3 = (1-\alpha)V_{10}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$. Also V_1 corresponds to the transition of the buffer size from 1 to 1, V_4 corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., M-1$, V_5 corresponds to the transition of the buffer size from M to M, V_8 corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and V_{11} corresponds to the transition of the buffer size from K to K.

The matrix B_0 is given by

$$B_0 = \begin{bmatrix} & & & & & & & \\ \Phi_4 & & & & & & & \\ \Delta_4 & \ddots & & & & & & \\ & \ddots & \Phi_4 & & & & & \\ & & \Delta_5 & \Phi_5 & & & & \\ & & & \Delta_6 & \ddots & & & \\ & & & & \ddots & \Phi_5 & & \\ & & & & & \Delta_6 & \Phi_6 & \end{bmatrix}_{N_1 \times N_2}$$

where $\Phi_4 = \alpha p_2 V_2$ corresponds to the transition of the buffer size from j to j for $j = 1, 2, ..., M$, $\Phi_5 = \alpha p_2 V_9$ corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and $\Phi_6 = \alpha p_2 \gamma V_9$ corresponds to the transition of the buffer size from K to K. Also $\Delta_4 = \alpha p_2 V_3$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., M$, $\Delta_5 = \alpha p_2 V_7$ corresponds to the transition of the buffer size from $M+1$ to M and $\Delta_6 = \alpha p_2 V_{10}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$.

The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \Delta_7 & \Phi_7 & & & & \\ & \Delta_8 & \Phi_8 & & & \\ & & \ddots & \ddots & & \\ & & & \Delta_8 & \Phi_8 & \\ & & & & & \end{bmatrix}_{N_2 \times N_1}$$

where $\Phi_7 = \alpha p_1 V_{12}$ corresponds to the transition of the buffer size from

1 to 2 and $\Phi_8 = p\alpha p_1 V_{12}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$. Also $\Delta_7 = (1-\alpha)V_{12}$ corresponds to the transition of the buffer size from 1 to 1 and $\Delta_8 = (1-\alpha)pV_{12}$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$.

The matrix A_1 is given by

$$A_1 = \begin{bmatrix} V_{13} & \Phi_9 & & & & & & & & \\ \Delta_9 & V_{14} & \ddots & & & & & & & \\ & \ddots & \ddots & & & & & & & \\ & & \Delta_{10} & V_4 & \ddots & & & & & \\ & & & \ddots & \ddots & \ddots & & & & \\ & & & & & V_5 & \Phi_{10} & & & \\ & & & & & \Delta_{11} & V_8 & \Phi_{11} & & \\ & & & & & & \Delta_{12} & \ddots & \ddots & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & & V_{11} \end{bmatrix}_{N_2 \times N_2}$$

where V_{13} corresponds to the transition of the buffer size from 1 to 1, V_{14} corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$, V_4 corresponds to the transition of the buffer size from j to j for $j = L+1, L+2, ..., M-1$, V_5 corresponds to the transition of the buffer size from M to M, V_8 corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$, V_{11} corresponds to the transition of the buffer size from K to K.$\Phi_9 = \alpha p_1 V_2$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, 2, ..., M-1$. $\Phi_{10} = \alpha p_1 V_6$ corresponds to the transition of the buffer size from M to $M+1$, $\Phi_{11} = \alpha p_1 V_9$ corresponds to the transition of the buffer size from j

to $j+1$ for $j = M+1, M+2, ..., K-1$. Also $\Delta_9 = (1-\alpha)qV_3$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., L$, $\Delta_{10} = (1-\alpha)V_3$ corresponds to the transition of the buffer size from j to $j-1$ for $j = L+1, ..., M$, $\Delta_{11} = (1-\alpha)V_7$ corresponds to the transition of the buffer size from $M+1$ to M and $\Delta_{12} = (1-\alpha)V_{10}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$.

The matrix A_0 is given by

$$A_0 = \begin{bmatrix} \Phi_{12} & & & & & & & & \\ \Delta_{13} & \ddots & & & & & & & \\ & \ddots & & & & & & & \\ & & \Delta_{14} & \ddots & & & & & \\ & & & \ddots & & & & & \\ & & & & \Delta_{15} & \Phi_{13} & & & \\ & & & & & \Delta_{16} & \ddots & & \\ & & & & & & \ddots & & \\ & & & & & & & \Delta_{16} & \Phi_{14} \end{bmatrix}_{N_2 \times N_2}$$

where

The matrix A_2 is given by

$$A_2 = \begin{bmatrix} \Delta_{17} & \Phi_{15} & & & & \\ & \Delta_{18} & \Phi_{16} & & & \\ & & \ddots & \ddots & & \\ & & & \Delta_{18} & \Phi_{16} & \\ & & & & & \end{bmatrix}_{N_2 \times N_2}$$

where $\Phi_{15} = \alpha p_1 V_{12}$ corresponds to the transition of the buffer size from 1 to 2 and $\Phi_{16} = p\alpha p_1 V_{12}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$. Also $\Delta_{17} = (1-\alpha)V_{12}$ corresponds to the transition of the buffer size from 1 to 1 and $\Delta_{18} = (1-\alpha)pV_{12}$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$.

6.1.2 Stability criterion

Theorem 6.1.1. *The system is stable if and only if*

$$\alpha p_2 \sum_{j=1}^{K-1} \sum_{b=0}^{1} \sum_{h=1}^{n} \pi_{jbh} + \alpha p_2 \gamma \sum_{b=0}^{1} \sum_{h=1}^{n} \pi_{Kbh} < \frac{1}{\sum_{l=1}^{N_2} m_{1l}}$$

where $N_2 = 2Mn + (K - M)(2n - v)$ *and* π *is the unique solution to* $\pi A = \pi$; $\pi e = 1$ *for* $A = A_0 + A_1 + A_2$.

Proof. Let $G_{ll'}$ be the conditional probability that the QBD process, starting in the state $l = (i, j, b, h)$ (for $i > 1$) where $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant m$ at time $t = 0$ reaches the state $l' = (i-1, j', b', h')$ where $1 \leqslant j' \leqslant K$; $0 \leqslant b' \leqslant N$; $1 \leqslant h' \leqslant m$, for the first time, in a finite time. That is

$$G_{ll'} = P[\tau < \infty : \chi(\tau) = l' | \chi(0) = l]$$

where τ is the first passage time from the level i to the level $i-1$. Because of the structure of Q, the probability $G_{ll'}$ does not depend on i. The matrix with elements $G_{ll'}$ is denoted by G.

Suppose the matrix $A = A_0 + A_1 + A_2$ is irreducible. Then the necessary

and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0 + A_1 + A_2$. That is, it is the unique solution to $\pi A = \pi$, $\pi e = 1$ and $A = A_0 + A_1 + A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. G is obtained as the minimal non negative solution to the matrix quadratic equation

$$G = A_2 + A_1 G + A_0 G^2.$$

This is obvious. On the left-hand side, G records the distribution of the first state visited in l' conditioned on the initial state being in l. In the right-hand side, these visits to l' are decomposed in to three groups; the first term corresponds to the case where the QBD directly moves from i to $i-1$ in one transition with probabilities recorded in A_2; as for the second term, with probabilities recorded in A_1, the QBD remains in l from where it still has to move eventually to l', with probabilities recorded in G; finally for the last term, with probabilities recorded in A_0, the QBD moves up to $i+1$ from where it still has to move eventually to l, with probabilities recorded in G and then to l' again with probabilities recorded in G.

Let $m_1 = [m_{1_l}]$ denote the column vector of dimension $K(N+1)m$ where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started in the state l. We have $G = (I - A_1)^{-1} A_2 + (I - A_1)^{-1} A_0 G^2$. Consequently $m_1 = [I - A_1 - A_0(I + G)]^{-1} e$.

For the system stability, the rate of drift from level i to level $i-1$ should be greater than that to level $i+1$. The rate of drift from level

i to the level $i+1$ is given by $\alpha p_2 \sum_{j=1}^{K-1} \sum_{b=0}^{1} \sum_{h=1}^{n} \pi_{jbh} + \alpha p_2 \gamma \sum_{b=0}^{1} \sum_{h=1}^{n} \pi_{Kbh}$. It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to the given stability criterion.

□

So by an appropriate choice of γ, that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate is greater than service rate.

6.1.3 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of P be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{6.1}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2 A_2 = R \tag{6.2}$$

and the vectors x_0, x_1 are obtained by solving the eqtions

$$x_0(B_1 - I) + x_1 B_2 = 0 \tag{6.3}$$

$$x_0 B_0 + x_1(A_1 - I + RA_2) = 0 \tag{6.4}$$

subject to the normalising condition

$$x_0 e + x_1 (I - R)^{-1} e = 1 \tag{6.5}$$

From the above discussion it is clear that to determine x, a key step is the computation of the rate matrix R. For this purpose, we use logarithmic reduction algorithm as in section 5.1.2 in chapter 5. We again partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0M}, x_{0(M+1)}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, ..., x_{iM}, x_{i(M+1)}, ..., x_{iK})$$

where $i \geqslant 1$ and x_{00} is a scalar and $x_{0j} = (x_{0j0}, x_{0j1})$, where if $1 \leqslant j \leqslant M$, then both x_{0j0} and x_{0j1} are vectors of order n and if $M+1 \leqslant j \leqslant K$, then x_{0j0} is a vector of order n and x_{0j1} is a vector of order $n - v$. Also

$$x_{ij} = (x_{ij0}, x_{ij1})$$

where $i \geqslant 1$ and if $1 \leqslant j \leqslant M$, then both x_{ij0} and x_{ij1} are vectors of order n and if $M + 1 \leqslant j \leqslant K$, then x_{ij0} are vectors of order n and x_{ij1} are vectors of order $n - v$.

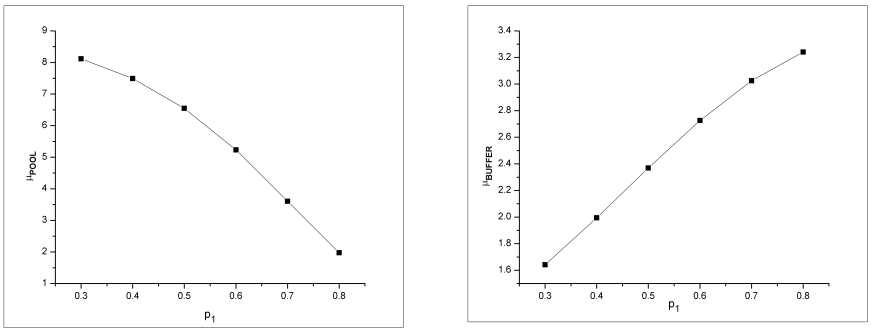

Fig 6.2: p_1 versus μ_{POOL} and μ_{BUFFER}

6.1.4 Performance characteristics

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{M}\sum_{b=0}^{1}\sum_{h=1}^{n} x_{ijbh} + \sum_{j=M+1}^{K}\left(\sum_{h=1}^{n} x_{ij0h} + \sum_{h=v+1}^{n} x_{ij1h}\right)$$

for $i > 0$ and

$$a_0 = x_{00} + \sum_{j=1}^{M}\sum_{b=0}^{1}\sum_{h=1}^{n} x_{ojbh} + \sum_{j=M+1}^{K}\left(\sum_{h=1}^{n} x_{0j0h} + \sum_{h=v+1}^{n} x_{0j1h}\right).$$

2. The probability that there are j customers in the buffer (including the one in service)is

$$b_j = \begin{cases} x_{00} \ , & if \quad j = 0 \\ \sum\limits_{b=0}^{1}\sum\limits_{h=1}^{n} x_{0jbh} + \sum\limits_{i=1}^{\infty}\sum\limits_{b=0}^{1}\sum\limits_{h=1}^{n} x_{ijbh} & if \ \ 1 \leqslant j \leqslant M \end{cases}$$

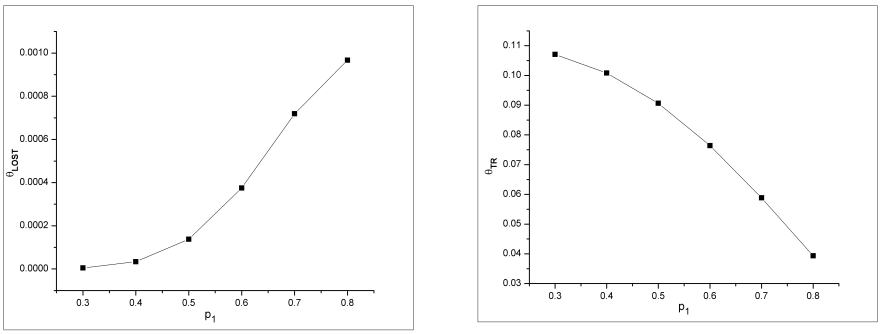

Fig 6.3: p_1 versus θ_{LOST} and θ_{TR}

and for $M+1 \leqslant j \leqslant M$,

$$b_j = \sum_{h=1}^{n} x_{0j0h} + \sum_{h=v+1}^{n} x_{0j1h} + \sum_{i=1}^{\infty} (\sum_{h=1}^{n} x_{ij0h} + \sum_{h=v+1}^{n} x_{ij1h}).$$

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that an arriving customer enters service immediately is b_0.

6. The rate at which the lower priority customer who find the buffer

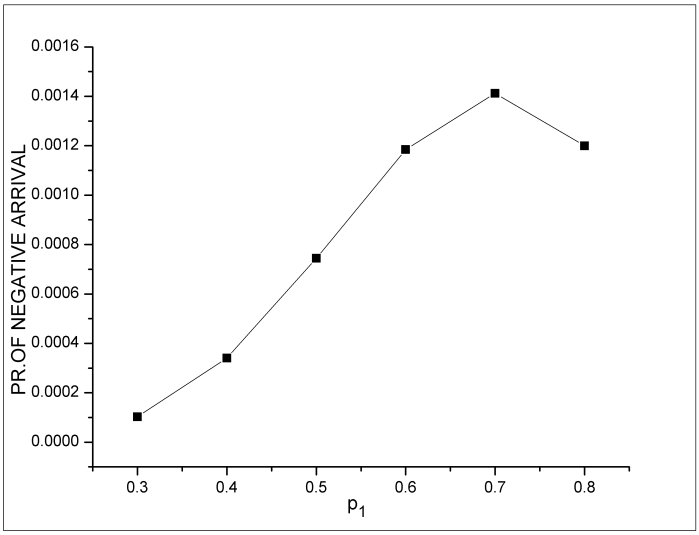

Fig 6.4: p_1 versus the probability of negative arrival

full leave the system without entering pool is

$$\theta_{LOST} = \alpha p_2 (1-\gamma) b_K.$$

7. The rate at which pooled customers transfer in to the buffer for immediate service is

$$\theta_{TR} = \sum_{i=1}^{\infty} \sum_{b=0}^{1} x_{i1bn} s_{n0} + \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=0}^{1} x_{ijbn} p s_{n0}.$$

8. Probability for a negative arrival (the rate at which negative arrival

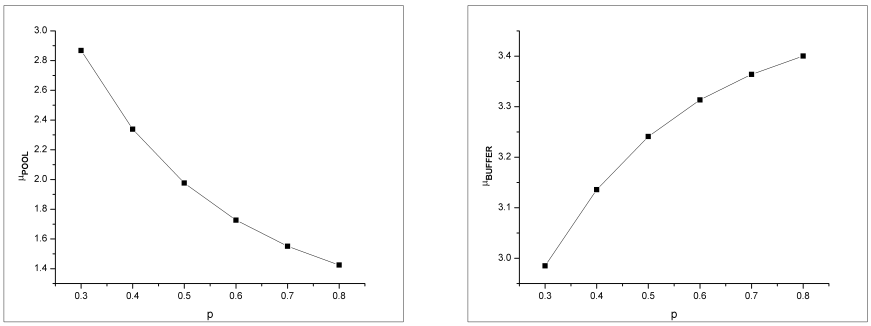

Fig 6.5: p versus μ_{POOL} and μ_{BUFFER}

occurs) is

$$N_R = \sum_{i=0}^{\infty} \sum_{h=1}^{v} x_{iM1h} \alpha p_1.$$

6.1.5 Numerical results

To illustrate the performance of the system, we present the following numerical results. A lower priority customer encountering the buffer full, will be inclined to join the pool with higher value of γ if the value of L and p are larger. On the other hand, γ inversely varies with K. Based on this, we can take $\gamma = \frac{Lp}{K}$. The impact of p_1 (the probability of higher priority customer) on various measures with $K = 7, L = 4, M = 5, n = 4, v = 2, \alpha = 0.24, p = 0.5, \gamma = \frac{Lp}{K}$,

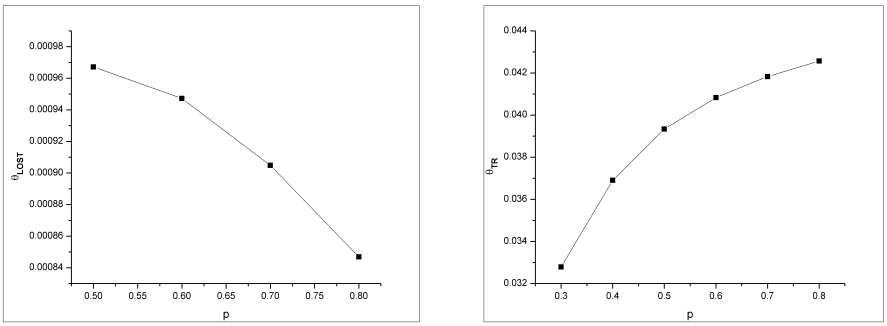

Fig 6.6: p versus θ_{LOST} and θ_{TR}

$$S = \begin{bmatrix} 0.001 & 0.999 & & \\ & 0.001 & 0.999 & \\ & & 0.0015 & 0.9985 \\ & & & 0.001 \end{bmatrix} \text{ and } S^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.999 \end{bmatrix}$$

are numerically computed and shown in figures 6.2, 6.3 and 6.4. As p_1 increases the mean pool size decreases due to the decrease of lower priority customers. At the same time, the mean buffer size increases. Hence the transfer rate from the pool to the buffer decreases. As the buffer becomes full, loss rate of lower priority customers starts to increase. As the buffer size approaches M, probability of a negative arrival increases at first and then decreases when it rises above M.

The effect of p on various measures with $K = 7, L = 4, M = 5, n = 4, v = 2, \alpha = 0.24, p_1 = 0.8, \gamma = \frac{Lp}{K}$ and for the same S and S^0 mentioned above, is computed and shown in figures 6.5, 6.6 and 6.7. Here also as p increases, transfer rate increases. So the mean pool size decreases for a high value of p_1. Then the buffer size increases. Also probability of loss

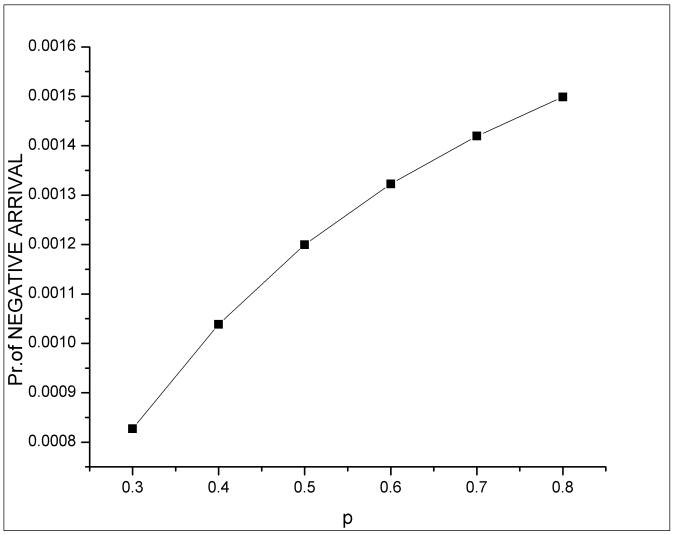

Fig 6.7: p versus the probability of negative arrival

of lower priority customers decreases due to the effect of the dependence of p on γ. As p increases probability of a negative arrival increases as expected.

6.2 Model-2: With service interruptions under N-policy

Here we discuss the discrete time version of the model discussed in chapter 4, with several additional features such as protected stages of service and priority of customers.

6.2.1 Mathematical formulation

In model-1, when a pooled customer is on service, if the buffer size rises to a pre-assigned number $M+1$ such that $L \leqslant M \leqslant K-1$, at an arrival epoch, the server will preempt the pool work in progress in unprotected stages and the preempted work will be lost for ever from the system. But in this model, the preempted work is considered to get interrupted. This interrupted pool work is postponed and stay as the head of the queue in the pool for getting next chance of transfer. From the epoch of interruption, the server will serve customers from the buffer and the counting of the number of continuously served customers from the buffer starts. When it reaches N $(N > 0)$ at a service completion epoch, the interrupted pooled customer gets transferred to the buffer for immediate service and further interruption is not allowed for such a work. The server will repeat the interrupted work when it is considered again. All other assumptions are same as that of model-1. A diagrammatic representation of the model-2 is given in figure 6.8.

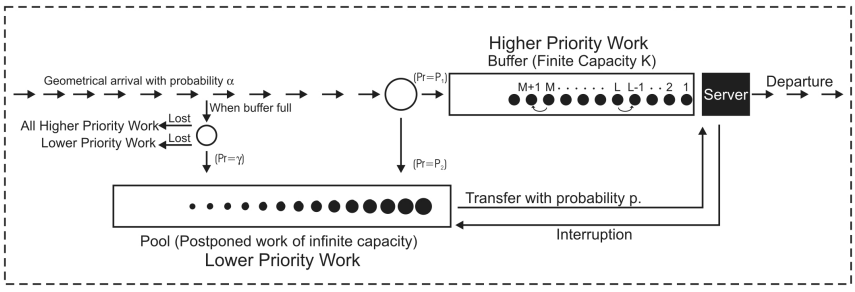

Fig 6.8: $Geo/E_d/1$ queue with postponed work and Service interruption

The state space consists of all tuples of the form (i, j, b, r, h) where i

denotes the number of postponed work in the pool having infinite capacity; j denotes the number of jobs in the finite buffer including the unit in service; b denotes the status of the system where

$$b = \begin{cases} 0 & , \quad \text{buffer work is in progress} \\ 1 & , \quad \text{pool work being served} \end{cases}$$

If $b = 0$, r denotes the number of continuously served customers from the buffer including the work at server and If $b = 1$, r denotes the number of continuously served customers from the buffer only, during the period of interruption with $r \neq 0$; $r = 0$ indicates that the head of the pool work is not an interrupted one; h denotes the stage of service in progress at that epoch.

Consider the boundary level $i = 0$. We denote the empty system $(0, 0, 0, 0)$ by 0.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $r = 0$ and $h = 1, 2, ..., n$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $r = 0, 1, 2, ..., N$, and $h = 1, 2, ..., n$.

If $M+1 \leqslant j \leqslant K$, $b = 1$ and $r = 0, 1, 2, ..., N-1$ then $h = v+1, ..., n$.

If $M + 1 \leqslant j \leqslant K$, $b = 1$ and $r = N$ then $h = 1, 2, ..., n$.

So the boundary level $i = 0$ constitute $\aleph_1 = 1 + M(N + 2)n + (K - M)[2n + N(n - v)]$ states.

Now consider the level $i \neq 0$.

If $1 \leqslant j \leqslant K$ and $b = 0$ then $r = 0, 1, 2, ..., N$ and $h = 1, 2, ..., n$.

If $1 \leqslant j \leqslant M$ and $b = 1$ then $r = 0, 1, 2, ..., N$ and $h = 1, 2, ..., n$.

If $M+1 \leqslant j \leqslant K$, $b = 1$ and $r = 0, 1, 2, ..., N-1$ then $h = v+1, ..., n$.

If $M+1 \leqslant j \leqslant K$, $b = 1$ and $r = N$ then $h = 1, 2, ..., n$.

So there are $\aleph_2 = 2Mn(N+1) + (K-M)[N(n-v) + (N+2)n$ states are there in the level $i \neq 0$.

The transition probability matrix is

$$P = \begin{bmatrix} B_1 & B_0 & & & & \\ B_2 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}$$

where the matrix B_0 is of dimension $\aleph_1 \times \aleph_2$, B_1 is square matrix of order $\aleph_1$ and B_2 is of dimension $\aleph_2 \times \aleph_1$. A_0, A_1 and A_2 are square of order $\aleph_2$. Each of these matrices is itself highly structured.

We use the following matrices in the sequel. β, S, S^0 are all same as that of model-1 and $E = S^0\beta$.

$$S^* = \begin{bmatrix} s_{(v+1)(v+1)} & s_{(v+1)(v+2)} & & \\ & s_{(v+2)(v+2)} & \ddots & \\ & & \ddots & \\ & & & s_{nn} \end{bmatrix}_{(n-v)\times(n-v)} ;$$

$$u_1 = \begin{bmatrix} \beta & \bar{0} \end{bmatrix}_{1\times(N+1)n} \quad ; \quad u_2 = e_{N+1} \otimes S^0;$$

$$u_3 = \begin{bmatrix} E & \bar{0} \end{bmatrix}_{n\times(N+1)n} \quad ; \quad u_4 = e_{N+1} \otimes E;$$

$$u_5 = \begin{bmatrix} (1-\alpha)S + \alpha p_2 E & & \cdots & \\ \alpha p_2 E & (1-\alpha)S & & \\ \vdots & & \ddots & \\ \alpha p_2 E & & & (1-\alpha)S \end{bmatrix}_{(N+1)n\times(N+1)n};$$

$$u_6 = I_{N+1} \otimes S \quad ; \quad u_7 = \begin{bmatrix} S & \bar{0} \end{bmatrix}_{n\times(N+1)n};$$

$$u_8 = \begin{bmatrix} e_{N+1} \otimes E & \bar{0} \end{bmatrix}_{(N+1)n\times(N+1)n};$$

$$u_9 = \begin{bmatrix} \bar{0} \\ S^* \end{bmatrix}_{n\times(n-v)} \quad ; \quad u_{10} = \begin{bmatrix} I_N \otimes u_9 & \bar{0} \\ \bar{0} & S \end{bmatrix}_{(N+1)n\times[N(n-v)+n]};$$

$$u_{11} = \begin{bmatrix} \bar{0} & I_N \otimes F \\ 0 & \bar{0} \end{bmatrix}_{(N+1)n\times(N+1)n} \quad ; \quad u_{12} = \begin{bmatrix} \bar{0} & \bar{0} \\ s_{n0} & \bar{0} \end{bmatrix}_{((n-v)\times n} ;$$

$$u_{13} = \begin{bmatrix} e_N \otimes u_{12} \\ E \end{bmatrix}_{[N(n-v)+n]\times n};$$

$$u_{14} = \begin{bmatrix} I_N \otimes S^* & \bar{0} \\ \bar{0} & S \end{bmatrix}_{[N(n-v)+n]\times[N(n-v)+n]};$$

$$u_{15} = \begin{bmatrix} e_N \otimes u_{12} & \bar{0} \\ E & \bar{0} \end{bmatrix}_{[N(n-v)+n]\times(N+1)n};$$

$$u_{16} = \begin{bmatrix} \bar{0} & \bar{0} \\ \bar{0} & E \end{bmatrix}_{(N+1)n\times(N+1)n}; \quad u_{17} = \begin{bmatrix} \bar{0} & \bar{0} \\ \bar{0} & E \end{bmatrix}_{(N+1)n\times[N(n-v)+n]};$$

$$u_{18} = I_{N+1} \otimes E \quad ; \quad u_{19} = \begin{bmatrix} I_N \otimes pE & \bar{0} \\ \bar{0} & E \end{bmatrix}_{(N+1)n\times(N+1)n};$$

$$u_{20} = \begin{bmatrix} E & \bar{0} & \bar{0} \\ \bar{0} & \bar{0} & I_{N-1} \otimes E \\ \bar{0} & \bar{0} & \bar{0} \end{bmatrix}_{(N+1)n\times(N+1)n};$$

$$H_1 = \begin{bmatrix} \alpha p_1 \beta & \alpha p_2 u_1 \end{bmatrix}_{1\times(N+2)n}; \quad H_2 = \begin{bmatrix} S^0 \\ t_2 \end{bmatrix}_{(N+2)n\times 1};$$

$$H_3 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \alpha p_2 u_3 \\ \alpha p_1 u_4 & u_5 \end{bmatrix}_{(N+2)n\times(N+2)n};$$

$$H_4 = \begin{bmatrix} \alpha p_1 S & \bar{0} \\ \bar{0} & \alpha p_1 u_6 \end{bmatrix}_{(N+2)n\times(N+2)n}; H_5 = \begin{bmatrix} E & \bar{0} \\ u_4 & \bar{0} \end{bmatrix}_{(N+2)n\times(N+2)n};$$

$$H_6 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 u_4 & (1-\alpha)u_6 \end{bmatrix}_{(N+2)n\times(N+2)n};$$

$$H_7 = \begin{bmatrix} \alpha p_1 S & \bar{0} \\ \bar{0} & \alpha p_1 u_{10} \end{bmatrix}_{(N+2)n\times(N+2)n};$$

$$H_8 = \begin{bmatrix} E & \bar{0} \\ u_{13} & \bar{0} \end{bmatrix}_{[N(n-v)+2n]\times(N+2)n};$$

$$H_9 = \begin{bmatrix} (1-\alpha)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 u_{13} & (1-\alpha)u_{14} \end{bmatrix}_{[N(n-v)+2n]\times[N(n-v)+2n]};$$

$$H_{10} = \begin{bmatrix} \alpha p_1 S & \bar{0} \\ \bar{0} & \alpha p_1 u_{14} \end{bmatrix}_{[N(n-v)+2n]\times[N(n-v)+2n]};$$

$$H_{11} = \begin{bmatrix} E & \bar{0} \\ u_{13} & \bar{0} \end{bmatrix}_{[N(n-v)+2n]\times[N(n-v)+2n]};$$

$$H_{12} = \begin{bmatrix} (1-\alpha p_2\gamma)S + \alpha p_1 E & \bar{0} \\ \alpha p_1 u_{13} & (1-\alpha p_2\gamma)u_{14} \end{bmatrix}_{[N(n-v)+2n]\times[N(n-v)+2n]};$$

$$H_{13} = \begin{bmatrix} \bar{0} & u_{18} \\ \bar{0} & u_8 \end{bmatrix}_{2(N+1)n\times(N+2)n};$$

$$H_{14} = \begin{bmatrix} (1-\alpha)u_6 & \alpha p_2 u_{18} \\ \bar{0} & (1-\alpha)u_6 + \alpha p_2 u_8 \end{bmatrix}_{2(N+1)n\times 2(N+1)n};$$

$$H_{15} = \begin{bmatrix} u_6 & \bar{0} \\ \bar{0} & u_6 \end{bmatrix}_{2(N+1)n\times 2(N+1)n}; \quad H_{16} = \begin{bmatrix} \bar{0} & u_{19} \\ \bar{0} & pu_8 \end{bmatrix}_{2(N+1)n\times(N+2)n};$$

$$H_{17} = \begin{bmatrix} u_{20} & \bar{0} \\ u_8 & \bar{0} \end{bmatrix}_{2(N+1)n\times 2(N+1)n};$$

$$H_{18} = \begin{bmatrix} (1-\alpha)u_6 + q\alpha p_1 u_{20} & p\alpha p_2 u_{18} \\ q\alpha p_1 u_8 & (1-\alpha)u_6 + p\alpha p_2 u_8 \end{bmatrix}_{2(N+1)n\times 2(N+1)n};$$

$$H_{19} = \begin{bmatrix} \bar{0} & u_{16} \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)n\times(N+2)n}; \; H_{20} = \begin{bmatrix} \bar{0} & u_{17} \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)n\times[N(n-v)+2n]};$$

$$H_{21} = \begin{bmatrix} (1-\alpha)u_6 + \alpha p_1 u_{20} & \alpha p_2 u_{16} \\ \alpha p_1 u_8 & (1-\alpha)u_6 \end{bmatrix}_{2(N+1)n\times 2(N+1)n};$$

$$H_{22} = \begin{bmatrix} u_6 & \bar{0} \\ \bar{0} & u_{10} \end{bmatrix}_{2(N+1)n\times[(N+2)n+N(n-v)]};$$

$$H_{23} = \begin{bmatrix} \alpha p_2 u_6 & \bar{0} \\ \alpha p_1 t_{11} & \alpha p_2 u_6 \end{bmatrix}_{2(N+1)n\times 2(N+1)n};$$

$$H_{24} = \begin{bmatrix} \bar{0} & u_{17} \\ \bar{0} & \bar{0} \end{bmatrix}_{[(N+1)n+N(n-v)+n]\times[n(n-v)+2n]};$$

$$H_{25} = \begin{bmatrix} u_{20} & \bar{0} \\ u_{15} & \bar{0} \end{bmatrix}_{[(N+1)n+N(n-v)+n]\times[2(N+1)n]};$$

$$H_{26} = \begin{bmatrix} (1-\alpha)u_6 + \alpha p_1 u_{20} & \alpha p_2 u_{17} \\ \alpha p_1 u_{15} & (1-\alpha)u_{14} \end{bmatrix}$$

having order $[(N+1)n + N(n-v) + n] \times [(N+1)n + N(n-v) + n]$;

$$H_{27} = \begin{bmatrix} u_6 & \bar{0} \\ \bar{0} & u_{14} \end{bmatrix}_{[(N+1)n+N(n-v)+n]\times[(N+1)n+N(n-v)+n]};$$

$$H_{28} = \begin{bmatrix} u_{20} & \bar{0} \\ u_{15} & \bar{0} \end{bmatrix}_{[(N+1)n+N(n-v)+n]\times[(N+1)n+N(n-v)+n]};$$

$$H_{29} = \begin{bmatrix} (1-\alpha p_2\gamma)u_6 & \alpha p_2\gamma u_{17} \\ \alpha p_1 u_{15} & (1-\alpha p_2\gamma)u_{14} \end{bmatrix}$$

having order $[(N+1)n+N(n-v)+n]\times[(N+1)n+N(n-v)+n]$;

$$H_{30} = \begin{bmatrix} u_7 & \bar{0} \\ \bar{0} & u_6 \end{bmatrix}_{(N+2)n\times 2(N+1)n}; \quad H_{31} = \begin{bmatrix} u_3 & \bar{0} \\ u_8 & \bar{0} \end{bmatrix}_{(N+2)n\times 2(N+1)n};$$

$$H_{32} = \begin{bmatrix} \bar{0} & \bar{0} \\ u_{11} & \bar{0} \end{bmatrix}_{(N+2)n\times 2(N+1)n}; \quad H_{33} = \begin{bmatrix} u_3 & \bar{0} \\ u_{15} & \bar{0} \end{bmatrix}_{[(N(n-v)+2n]\times 2(N+1)n};$$

$$H_{34} = \begin{bmatrix} u_7 & \bar{0} \\ \bar{0} & u_{14} \end{bmatrix}_{[N(n-v)+2n]\times[(N+1)n+N(n-v)+n]};$$

$$H_{35} = \begin{bmatrix} u_3 & \bar{0} \\ u_{15} & \bar{0} \end{bmatrix}_{[(N(n-v)+2n]\times[(N+1)n+N(n-v)+n]};$$

$$H_{36} = \begin{bmatrix} \bar{0} & u_{18} \\ \bar{0} & u_8 \end{bmatrix}_{2(N+1)n \times 2(N+1)n};$$

$$H_{37} = \begin{bmatrix} \bar{0} & u_{19} \\ \bar{0} & pu_8 \end{bmatrix}_{2(N+1)n \times 2(N+1)n}; \quad H_{38} = \begin{bmatrix} \bar{0} & u_{16} \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)n \times 2(N+1)n};$$

$$H_{39} = \begin{bmatrix} \bar{0} & u_{17} \\ \bar{0} & \bar{0} \end{bmatrix}_{2(N+1)n \times [(N+1)n+N(n-v)+n]};$$

$$H_{40} = \begin{bmatrix} \bar{0} & u_{17} \\ \bar{0} & \bar{0} \end{bmatrix}_{[(N+1)n+N(n-v)+n] \times [(N+1)n+N(n-v)+n]}.$$

The matrix B_1 corresponds to the transition from the level 0 to 0 is given below:

$$B_1 = \begin{bmatrix} 1-\alpha & H_1 & & & & & & & \\ \Omega_1 & H_3 & H_4 & & & & & & \\ & \Omega_2 & H_6 & \ddots & & & & & \\ & & \ddots & \ddots & H_7 & & & & \\ & & & \Omega_3 & H_9 & H_{10} & & & \\ & & & & \Omega_4 & \ddots & \ddots & & \\ & & & & & \ddots & \ddots & & \\ & & & & & & & H_9 & H_{10} \\ & & & & & & & \Omega_4 & H_{12} \end{bmatrix}_{\aleph_1 \times \aleph_1}$$

where $\Omega_1 = (1-\alpha)H_2$ corresponds to the transition of the buffer size from 1 to 0, $\Omega_2 = (1-\alpha)H_5$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., M$, and $\Omega_3 = (1-\alpha)H_8$ corresponds to the transition of the buffer size from $M+1$ to M, $\Omega_4 = (1-\alpha)H_{11}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$. Also H_1 corresponds to the transition of the buffer size from 0 to 1, H_4 corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, 2, ..., M-1$, H_7 corresponds to the transition of the buffer size from M to $M+1$, H_{10} corresponds to the transition of the buffer size from j to $j+1$ for $j = M+1, M+2, ..., K-1$ and H_3 corresponds to the transition of the buffer size from 1 to 1, H_6 corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., M$, H_9 corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$, H_{12} corresponds to the transition of the buffer size from K to K.

The matrix B_0 is given by

$$B_0 = \begin{bmatrix} & & & & & & & \\ \Theta_1 & & & & & & & \\ \Omega_5 & \ddots & & & & & & \\ & \ddots & \Theta_2 & & & & & \\ & & \Omega_6 & \Theta_3 & & & & \\ & & & \Omega_7 & \ddots & & & \\ & & & & \ddots & \ddots & & \\ & & & & & \Omega_7 & \Theta_4 & \end{bmatrix}_{\aleph_1 \times \aleph_2}$$

where $\Omega_5 = \alpha p_2 H_{31}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., M$, $\Omega_6 = \alpha p_2 H_{33}$ corresponds to the transition

of the buffer size from $M+1$ to M and $\Omega_7 = \alpha p_2 H_{35}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$. Also $\Theta_1 = \alpha p_2 H_{30}$ corresponds to the transition of the buffer size from j to j for $j = 1, 2, ..., M-1$, $\Theta_2 = \alpha p_2 H_{30} + \alpha p_2 H_{32}$ corresponds to the transition of the buffer size from M to M,$\Theta_3 = \alpha p_2 H_{34}$ corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and $\Theta_4 = \alpha p_2 \gamma H_{34}$ corresponds to the transition of the buffer size from K to K.

The matrix B_2 is given by

$$B_2 = \begin{bmatrix} \Omega_8 & \Theta_5 & & & & & & & & \\ & \Omega_9 & \Theta_6 & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \Omega_{10} & \Theta_7 & & & & & \\ & & & & \ddots & \ddots & & & & \\ & & & & & & \Theta_8 & & & \\ & & & & & & \Omega_{11} & \Theta_9 & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & \ddots & \end{bmatrix}_{\aleph_2 \times \aleph_1}$$

where $\Omega_8 = (1-\alpha)H_{13}$ corresponds to the transition of the buffer size from 1 to 1, $\Omega_9 = (1-\alpha)H_(16)$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$, and $\Omega_{10} = (1-\alpha)H_{19}$ corresponds to the transition of the buffer size from j to j for $j = L+1, L+2, ..., M$, $\Omega_{11} = (1-\alpha)H_{24}$ corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and $\Omega_{12} = (1-\alpha)p_2\gamma H_{24}$ corresponds to the

transition of the buffer size from K to K. Also $\Theta_5 = \alpha p_1 H_{13}$ corresponds to the transition of the buffer size from 1 to 2, $\Theta_6 = \alpha p_1 H_{16}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$, $\Theta_7 = \alpha p_1 H_{19}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = L+1, L+2, ..., M-1$, $\Theta_8 = \alpha p_1 H_{20}$ corresponds to the transition of the buffer size from M to $M+1$ and $\Theta_9 = \alpha p_1 H_{24}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = M+1, M+2, ..., K-1$.

$$A_1 = \begin{bmatrix} H_{14} & \Theta_{10} & & & & & & & & \\ \Omega_{12} & H_{18} & \ddots & & & & & & & \\ & \ddots & \ddots & & & & & & & \\ & & \Omega_{13} & H_{21} & \ddots & & & & & \\ & & & \ddots & \ddots & \Theta_{10} & & & & \\ & & & & & H_{21} & \Theta_{11} & & & \\ & & & & & \Omega_{14} & H_{26} & \Theta_{12} & & \\ & & & & & & \Omega_{15} & & \ddots & \\ & & & & & & & \ddots & \ddots & \Theta_{12} \\ & & & & & & & & \Omega_{15} & H_{29} \end{bmatrix}_{\aleph_2 \times \aleph_2}$$

where $\Omega_{12} = (1-\alpha) q H_{17}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., L$, $\Omega_{13} = (1-\alpha) H_{17}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = L+1, L+2, ..., M$ and $\Omega_{14} = (1-\alpha) H_{25}$ corresponds to the transition of the buffer size from $M+1$ to M, $\Omega_{15} = (1-\alpha) H_{28}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$. Also $\Theta_{10} = \alpha p_1 H_{15}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 1, 2, ..., M-1$,

$\Theta_{11} = \alpha p_1 H_{22}$ corresponds to the transition of the buffer size from M to $M+1$, $\Theta_{12} = \alpha p_1 H_{27}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = M+1, M+2, ..., K-1$, H_{14} corresponds to the transition of the buffer size from 1 to 1, H_{18} corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$, H_{21} corresponds to the transition of the buffer size from j to j for $j = L+1, L+2, ..., M$, H_{26} corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$, H_{29} corresponds to the transition of the buffer size from K to K.

$$A_0 = \begin{bmatrix} \Theta_{13} & & & & & & & & \\ \Omega_{16} & \ddots & & & & & & & \\ & \ddots & & & & & & & \\ & & \Omega_{17} & \ddots & & & & & \\ & & & \ddots & V_{23} & & & & \\ & & & & \Omega_{18} & \Theta_{14} & & & \\ & & & & & \Omega_{19} & \ddots & & \\ & & & & & & & \Theta_{14} & \\ & & & & & & & \Omega_{19} & \Theta_{15} \end{bmatrix}_{\aleph_2 \times \aleph_2}$$

where $\Omega_{16} = q\alpha p_2 H_{17}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = 2, 3, ..., L$, $\Omega_{17} = \alpha p_2 H_{17}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = L+1, L+2, ..., M$, $\Omega_{18} = \alpha p_2 H_{25}$ corresponds to the transition of the buffer size from $M+1$ to M and $\Omega_{19} = \alpha p_2 H_{28}$ corresponds to the transition of the buffer size from j to $j-1$ for $j = M+2, M+3, ..., K$. Also $\Theta_{13} = \alpha p_2 H_{15}$ corresponds to the transition of the buffer size from j to j for $j = 1, 2, ..., M-1$, H_{23} corresponds to the

transition of the buffer size from M to M, $\Theta_{14} = \alpha p_2 H_{27}$ corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and $\Theta_{15} = \alpha p_2 \gamma H_{27}$ corresponds to the transition of the buffer size from K to K.

$$A_2 = \begin{bmatrix} \Omega_{20} & \Theta_{16} & & & & & & & & \\ & \Omega_{21} & \Theta_{17} & & & & & & & \\ & & \ddots & \ddots & & & & & & \\ & & & \Omega_{22} & \Theta_{18} & & & & & \\ & & & & & \ddots & & & & \\ & & & & & \ddots & \Theta_{19} & & & \\ & & & & & & \Omega_{23} & \Theta_{20} & & \\ & & & & & & & \ddots & \ddots & \\ & & & & & & & & \Omega_{24} & \end{bmatrix}_{\aleph_2 \times \aleph_2}$$

where $\Omega_{20} = (1-\alpha)H_{36}$ corresponds to the transition of the buffer size from 1 to 1, $\Omega_{21} = (1-\alpha)H_(37)$ corresponds to the transition of the buffer size from j to j for $j = 2, 3, ..., L$, and $\Omega_{22} = (1-\alpha)H_{38}$ corresponds to the transition of the buffer size from j to j for $j = L+1, L+2, ..., M$, $\Omega_{23} = (1-\alpha)H_{40}$ corresponds to the transition of the buffer size from j to j for $j = M+1, M+2, ..., K-1$ and $\Omega_{24} = (1-\alpha)p_2\gamma H_{40}$ corresponds to the transition of the buffer size from K to K. Also $\Theta_{16} = \alpha p_1 H_{36}$ corresponds to the transition of the buffer size from 1 to 2, $\Theta_{17} = \alpha p_1 H_{37}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = 2, 3, ..., L$, $\Theta_{18} = \alpha p_1 H_{38}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = L+1, L+2, ..., M-1$, $\Theta_{19} = \alpha p_1 H_{39}$ corresponds to the transition

of the buffer size from M to $M+1$ and $\Theta_{20} = \alpha p_1 H_{40}$ corresponds to the transition of the buffer size from j to $j+1$ for $j = M+1, M+2, ..., K-1$.

6.2.2 Stability criterion

Theorem 6.2.1. *The system is stable if and only if*

$$\alpha p_2 \left(\sum_{j=1}^{K-1} \sum_{r=0}^{N} \sum_{h=1}^{n} \pi_{j0rh} + \sum_{j=1}^{M} \sum_{r=0}^{N} \sum_{h=1}^{n} \pi_{j1rh} \right)$$

$$+\alpha p_2 \sum_{j=M+1}^{K-1} \left(\sum_{r=0}^{N-1} \sum_{h=v+1}^{n} \pi_{j1rh} + \sum_{h=1}^{n} \pi_{j1Nh} \right)$$

$$+\alpha p_2 \gamma \left(\sum_{r=0}^{N-1} \sum_{h=V+1}^{n} \pi_{K1rh} + \sum_{h=1}^{n} \pi_{K1Nh} + \sum_{r=0}^{N} \sum_{h=1}^{n} \pi_{K0rh} \right)$$

$$+\alpha p_1 \sum_{r=0}^{N-1} \sum_{h=1}^{v} \pi_{M1rh} < \frac{1}{\sum_{l=1}^{\aleph_2} m_{1l}}$$

where $\aleph_2 = 2Mn(N+1) + (K-M)[N(n-v) + (N+2)n$ *and* π *is the unique solution to* $\pi A = \pi$*;* $\pi e = 1$ *for* $A = A_0 + A_1 + A_2$.

Proof. Let $G_{ll'}$ be the conditional probability that the QBD process starting in the state $l = (i, j, b, h)$ (for $i > 1$) where $1 \leqslant j \leqslant K$; $0 \leqslant b \leqslant N$; $1 \leqslant h \leqslant m$ at time $t = 0$ reaches the state $l' = (i-1, j', b', h')$ where $1 \leqslant j' \leqslant K$; $0 \leqslant b' \leqslant N$; $1 \leqslant h' \leqslant m$, for the first time, in a finite time. That is

$$G_{ll'} = P[\tau < \infty : \chi(\tau) = l' | \chi(0) = l]$$

where τ is the first passage time from the level i to the level $i-1$. Because of the structure of Q, the probability $G_{ll'}$ does not depend on i. The matrix with elements $G_{ll'}$ is denoted by G.

Suppose the matrix $A = A_0+A_1+A_2$ is irreducible. Then the necessary and sufficient condition for the positive recurrence of the process is that the matrix G is stochastic. For this, the condition $\pi A_2 e > \pi A_0 e$ must be satisfied where π is the stationary probability vector associated with $A = A_0 + A_1 + A_2$. That is, it is the unique solution to $\pi A = \pi$, $\pi e = 1$ and $A = A_0+A_1+A_2$. The quantity $\rho = \frac{\pi A_0 e}{\pi A_2 e}$ is called the traffic intensity of the QBD process. G is obtained as the minimal non negative solution to the matrix quadratic equation

$$G = A_2 + A_1 G + A_0 G^2.$$

This is obvious. On the left-hand side, G records the distribution of the first state visited in l' conditioned on the initial state being in l. In the right-hand side, these visits to l' are decomposed in to three groups; the first term corresponds to the case where the QBD directly moves from i to $i-1$ in one transition with probabilities recorded in A_2; as for the second term, with probabilities recorded in A_1, the QBD remains in l from where it still has to move eventually to l', with probabilities recorded in G; finally for the last term, with probabilities recorded in A_0, the QBD moves up to $i+1$ from where it still has to move eventually to l, with probabilities recorded in G and then to l' again with probabilities recorded in G.

Let $m_1 = [m_{1_l}]$ denotes the column vector of dimension $K(N+1)m$ where m_{1_l} denotes the mean first passage time from the level i $(i > 1)$ to the level $i-1$ given that the first passage time started in the state

l. We have $G = (I - A_1)^{-1}A_2 + (I - A_1)^{-1}A_0G^2$. Consequently $m_1 = [I - A_1 - A_0(I + G)]^{-1}e$.

For the system stability, the rate of drift from level i to level $i - 1$ should be greater than that to level $i + 1$. It follows that the condition $\pi A_0 e < \pi A_2 e$ is equivalent to the given stability criterion.

□

So by an appropriate choice of γ, that is by postponing a fraction of overflowing customers, one can obtain a stable system even if arrival rate is greater than service rate.

6.2.3 Stationary distribution

Since the model is studied as a QBD process, its stationary distribution, if it exists, has a matrix geometric solution. Assume that the stability criterion is satisfied. Let the stationary vector x of P be partitioned by the levels in to subvectors x_i for $i \geqslant 0$. Then x_i has the matrix geometric form

$$x_i = x_1 R^{i-1} \tag{6.6}$$

for $i \geqslant 2$ where R is the minimal non negative solution to the matrix equation

$$A_0 + RA_1 + R^2A_2 = R \tag{6.7}$$

and the vectors x_0, x_1 are obtained by solving the eqution

$$x_0(B_1 - I) + x_1 B_2 = 0 \tag{6.8}$$

$$x_0 B_0 + x_1(A_1 - I + RA_2) = 0 \tag{6.9}$$

subject to the normalising condition

$$x_0 e + x_1(I - R)^{-1} e = 1 \tag{6.10}$$

From the above discussion it is clear that to determine x, a key step is the computation of the rate matrix R. Here also we use logarithmic reduction algorithm as in section 5.1.2 in chapter 5. We again partition x_i by sublevels as

$$x_0 = (x_{00}, x_{01}, x_{02}, ..., x_{0M}, x_{0(M+1)}, ..., x_{0K})$$

and

$$x_i = (x_{i1}, x_{i2}, ..., x_{iM}, x_{i(M+1)}, ..., x_{iK})$$

where $i \geqslant 1$ and x_{00} is a scalar and $x_{0j} = (x_{0j0}, x_{0j1})$, where if $1 \leqslant j \leqslant M$, then x_{0j0} are vectors of order n and x_{0j1} are vectors of order $(N+1)n$ and if $M+1 \leqslant j \leqslant K$, then x_{0j0} are vectors of order n and x_{0j1} are vectors of order $N(n-v)+n$. Also

$$x_{ij} = (x_{ij0}, x_{ij1})$$

where $i \geqslant 1$ and if $1 \leqslant j \leqslant M$, then x_{ij0} and x_{ij1} are vectors of order $(N+1)n$ and if $M+1 \leqslant j \leqslant K$, then x_{ij0} are vectors of order $(N+1)n$ but x_{ij1} are vectors of order $N(n-v)+n$.

6.2.4 Performance characteristics

1. The probability that there are i customers in the pool is

$$a_i = \sum_{j=1}^{M}\sum_{b=0}^{1}\sum_{r=0}^{N}\sum_{h=1}^{n} x_{ijbrh}$$

$$+ \sum_{j=M+1}^{K}\left(\sum_{r=0}^{N}\sum_{h=1}^{n} x_{ij0rh} + \sum_{r=0}^{N-1}\sum_{h=v+1}^{n} x_{ij1rh} + \sum_{h=1}^{n} x_{ij1Nh}\right)$$

for $i > 0$ and

$$a_0 = x_{00} + \sum_{j=1}^{M}\left(\sum_{h=1}^{n} x_{0j00h} + \sum_{r=0}^{N}\sum_{h=1}^{n} x_{0j1rh}\right)$$

$$+ \sum_{j=M+1}^{K}\left(\sum_{h=1}^{n} x_{0j00h} + \sum_{r=0}^{N-1}\sum_{h=v+1}^{n} x_{0j1rh} + \sum_{h=1}^{n} x_{0j1Nh}\right).$$

2. The probability that there are j customers in the buffer (including the one in service)is

$$b_j = \sum_{h=1}^{n} x_{0j00h} + \sum_{r=0}^{N}\sum_{h=1}^{n} x_{0j1rh} + \sum_{i=1}^{\infty}\sum_{b=0}^{1}\sum_{r=0}^{N}\sum_{h=1}^{n} x_{ijbrh}$$

for $1 \leqslant j \leqslant M$,

$$b_j = \sum_{h=1}^{n}\left(x_{0j00h} + x_{0j1Nh}\right) + \sum_{r=0}^{N-1}\sum_{h=v+1}^{n} x_{0j1rh}$$

$$+ \sum_{i=1}^{\infty}\left(\sum_{r=0}^{N}\sum_{h=1}^{n} x_{ij0rh} + \sum_{r=0}^{N-1}\sum_{h=v+1}^{n} x_{ij1rh} + \sum_{h=1}^{n} x_{ij1Nh}\right)$$

for $M+1 \leqslant j \leqslant K$ and $b_0 = x_{00}$.

3. The mean number of pooled customers is

$$\mu_{POOL} = \sum_{i=1}^{\infty} i a_i = x_1 (I - R)^{-2} e.$$

4. The mean buffer size is

$$\mu_{BUFFER} = \sum_{j=1}^{K} j b_j.$$

5. The probability that an arriving customer enters service immediately is b_0.

6. The rate at which the lower priority customer who find the buffer full leave the system without entering pool (mean number of customers not joining the system per unit time) is

$$\theta_{LOST} = \alpha p_2 (1 - \gamma) b_K.$$

7. The rate at which pooled customers transfer in to the buffer for immediate service is

$$\theta_{TR} = \sum_{i=1}^{\infty} \sum_{b=0}^{1} \sum_{r=0}^{N} x_{i1brn} s_{n0} + \sum_{i=1}^{\infty} \sum_{j=2}^{L} \sum_{b=0}^{1} \sum_{r=0}^{N-1} x_{ijbrn} p s_{n0}$$

$$+ \sum_{i=1}^{\infty} \sum_{j=1}^{K} x_{ij0Nn} s_{n0}.$$

8. Interruption rate is

$$I_R = \sum_{i=0}^{\infty} \sum_{r=0}^{N-1} \sum_{h=1}^{v} x_{iM1rh} \alpha p_1.$$

6.2.5 Numerical results

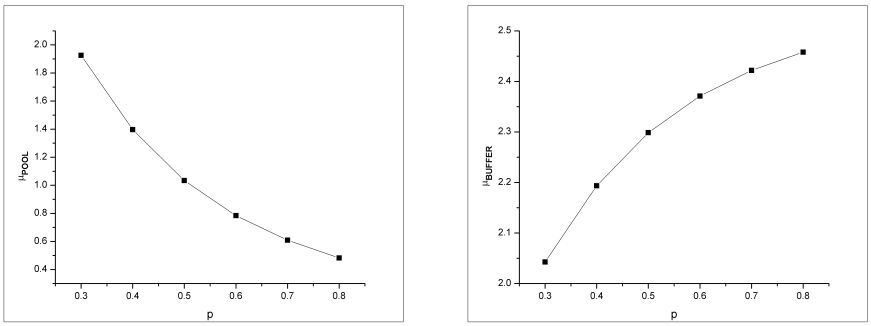

Fig 6.9: p versus μ_{POOL} and μ_{BUFFER}

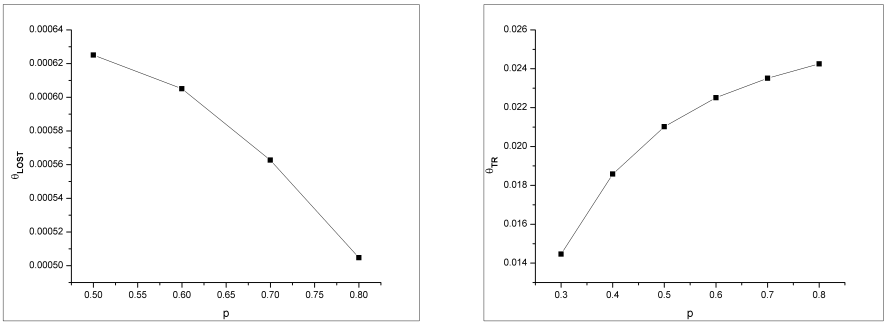

Fig 6.10: p versus θ_{LOST} and θ_{TR}

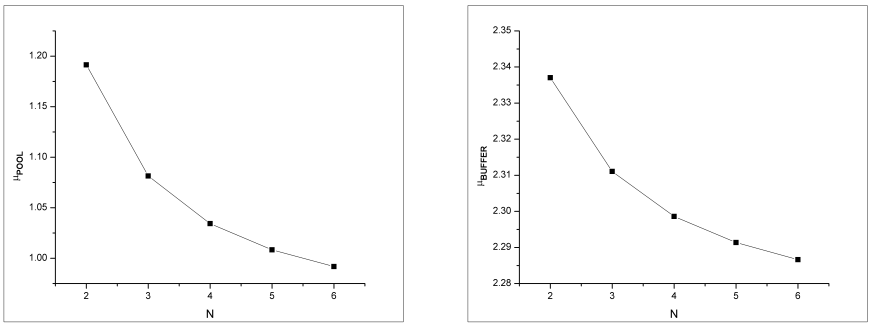

Fig 6.11: N versus μ_{POOL} and μ_{BUFFER}

In this section, we illustrate the performance of the system by considering some numerical results. A lower priority customer encountering the buffer full, will be inclined to join the pool with higher γ if the L and p values are larger. On the other hand γ inversely varies with K and N. To model this situation, we take $\gamma = \frac{Lp}{K} + \frac{1}{N}$. But the relationship is feasible for those values of L, p, K and N such that $0 \leqslant \gamma \leqslant 1$.

The effect of p on various measures with $K = 7, L = 4, M = 5, n = 4, N = 3, v = 2, \alpha = 0.2, p_1 = 0.8, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$S = \begin{bmatrix} 0.001 & 0.999 & & \\ & 0.001 & 0.999 & \\ & & 0.0015 & 0.9985 \\ & & & 0.001 \end{bmatrix} \text{ and } S^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.999 \end{bmatrix}$$

is computed and shown in figures 6.9 and 6.10. As p increases, transfer rate increases. So the mean pool size decreases for a high value of p_1. Then the buffer size increases. Also probability of loss of lower priority customers decreases due to the effect of the dependence of p on γ.

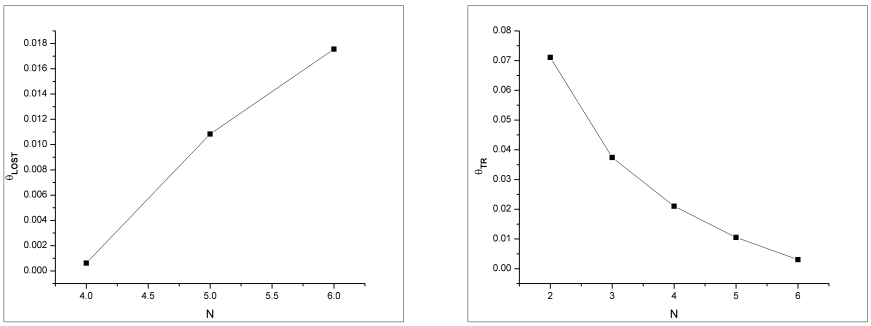

Fig 6.12: N versus θ_{LOST} and θ_{TR}

The impact of N on various measures with $K = 7, L = 4, M = 5, n = 4, v = 2, \alpha = 0.2, p_1 = 0.8, p = 0.5, \gamma = \frac{Lp}{K} + \frac{1}{N}$ and for the same S and S^0 mentioned above, is shown in figures 6.11 and 6.12. As N increases μ_{POOL}, μ_{BUFFER}, θ_{TR} decrease monotonically whereas θ_{LOST} increases monotonically. This is due to the fact that by our assumption γ varies inversely as N and as a result, loss rate increases and inflow rate to the pool decreases as N increases. So the transfer rate of the interrupted customer from the pool to the buffer decreases, and thus the mean buffer size decreases.

Chapter 7

A Comparison study and Conclusion

In this chapter we compare, wherever possible, the models described in chapters 2 to 6. In these models, eventhough the objective is to minimize the loss of customers due to the overflow of finite capacity buffer by means of postponement, we included the situations involving interruption, priority, protection and negative arrivals. So each model has its own importance. But the comparison will help to understand the relative performance of the models. For comparison, we consider the three continuous time models in chapters 2, 3 and 4 and the three discrete time models in chapters 5 and 6 separately. We call the model described in chapter 1 by model-I, the model described in chapter 2 by model-II and the model described in chapter 3 by model-III.

We start to compare the model-I and the model-II. In the model-I, a

	μ_{POOL}		μ_{BUFFER}	
N	$model-I$	$model-II$	$model-I$	$model-II$
3	4.3481455	1.6520991	4.1272182	3.0041764
4	2.8674562	1.6472070	3.9493656	2.9432521
5	2.3480656	1.6752849	3.8571582	2.9082990
6	2.0938632	1.7069173	3.8007855	2.8855472
7	1.9466079	1.7358847	3.7627959	2.8695617
8	1.8520240	1.7611248	3.7354820	2.8577244

Table 7.1: Effect of N on μ_{POOL} and μ_{BUFFER} in models I and II

	θ_{TR}		θ_{LOST}	
N	$model-I$	$model-II$	$model-I$	$model-II$
3	1.1983539	1.1803304	0.8487673	0.2618262
4	0.9203455	1.1096535	0.9117374	0.2834340
5	0.7786083	1.0713297	0.9418701	0.2926464
6	0.6934710	1.0471536	0.9592767	0.2973242
7	0.6359809	1.0301353	0.9705053	0.2999944
8	0.5944450	1.0173856	0.9783031	0.3016568

Table 7.2: Effect of N on θ_{TR} and θ_{LOST} in models I and II

newly arriving customer will join the buffer if it has a vacancy. It will make the buffer size of model-I larger than that in the model-II. This is due to the restriction in terms of probability to enter the buffer in the model-II. So the loss rate of the model-II will be always lower than that of the model-I. By keeping $K = 6, L = 3, m = 2, \lambda = 7, p = 0.5, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$\beta = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix} \quad S = \begin{bmatrix} -12.5 & 6.0 \\ 6.0 & -12.5 \end{bmatrix} \quad S^0 = \begin{bmatrix} 6.5 \\ 6.5 \end{bmatrix}$$

for the model-I and $K = 6, L = 3, m = 2, \lambda = 7, p = 0.5, s_1 = 0.9, s_2 =$

	μ_{POOL}		μ_{BUFFER}	
p	$model-I$	$model-II$	$model-I$	$model-II$
0.3	1.2279803	1.6284441	3.6858897	2.8322663
0.4	1.7021447	1.6445645	3.7669792	2.8706489
0.5	2.3480656	1.6752849	3.8571582	2.9082990
0.6	3.2551939	1.7203127	3.9574442	2.9454260
0.7	4.5879235	1.7798784	4.0678077	2.9821990

Table 7.3: Effect of p on μ_{POOL} and μ_{BUFFER} in models I and II

	θ_{TR}		θ_{LOST}	
p	$model-I$	$model-II$	$model-I$	$model-II$
0.3	0.5633762	1.0228137	1.0290717	0.3299661
0.4	0.6669288	1.0458466	0.9869518	0.3119004
0.5	0.7786083	1.0713297	0.9418701	0.2926464
0.6	0.8999512	1.0992774	0.8935239	0.2722456
0.7	1.0324645	1.1297385	0.8411404	0.2507135

Table 7.4: Effect of p on θ_{TR} and θ_{LOST} in models I and II

$0.8, s_3 = 0.7, s_4 = 0.6, s_5 = 0.5, \gamma_1 = \frac{Lp}{K} + \frac{1}{N}, \gamma_2 = 0.8, \delta_{=}0.5$ with same β, S and S^0 for the model II, various measures of descriptors are shown in the tables 7.1, 7.2, 7.3 and 7.4. As N increases, eventhough μ_{POOL} and μ_{BUFFER} of model-II are less than that of the model-I, θ_{LOST} of model-II is not larger than that for the model-I. This is because of the chance of vacancy in the buffer for a new arrival due to the probability restriction in the model-II. This makes the buffer less congested and so the transfer rate is higher for it. The same situation can be seen as p increases. So the model in chapter 3 is superior to the model in chapter 2.

The model-III discussed in chapter 4 is entirely different from the previous two models. In model-III, we transfer the pool work to the buffer for

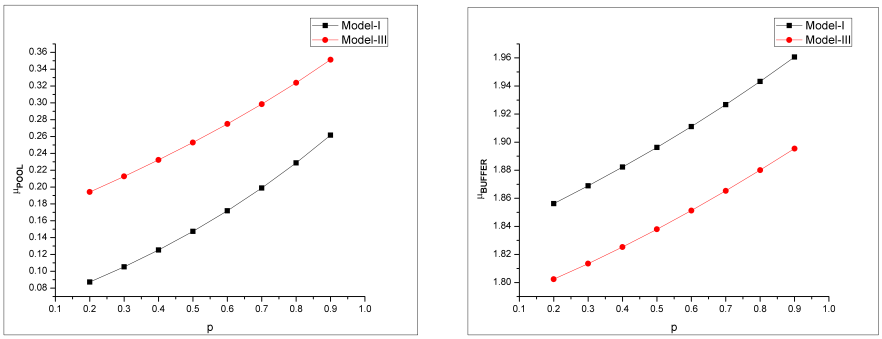

Fig 7.1: p versus μ_{POOL} and μ_{BUFFER} in models I and III

immediate service. But at that time, if the buffer size rises to M, the pool work at server is interrupted and postponed again. So clearly it will ensure that the buffer not full. This will reduce the loss rate. We do not go for a comparison of model-III with model-II as the former concentrates on the arrival process and the latter concentrates on the service process. As we compare the model-III with model-I, we can see that the loss rate of model-III is much smaller than that of model-I. So preemption in the model-III when the number of customers in the buffer rises to a pre-assigned level, reduces the rate of loss. Also θ_{TR} and μ_{BUFFER} are less and μ_{POOL} is higher for the model-III than that for the model-I. Comparison of model-I by fixing $K = 6, L = 3, \lambda = 5, m = 1, N = 3, (-\beta S^{-1}e)^{-1} = 7$, and $\gamma = \frac{Lp}{K} + \frac{1}{N}$ is done with model-III by fixing $K = 6, L = 3, \lambda = 5, \mu = 7, M = 4, N = 3$ and $\gamma = \frac{Lp}{K} + \frac{1}{N}$ as shown figures 7.1, 7.2, 7.3, and 7.4. From these observations, we can say that model-III is superior to model-I.

Now consider the three discrete time models in chapters 5 and 6. We call the model described in chapter 5 by model-IV, the first model described in chapter 6 by model-V and the second model by model-VI. We

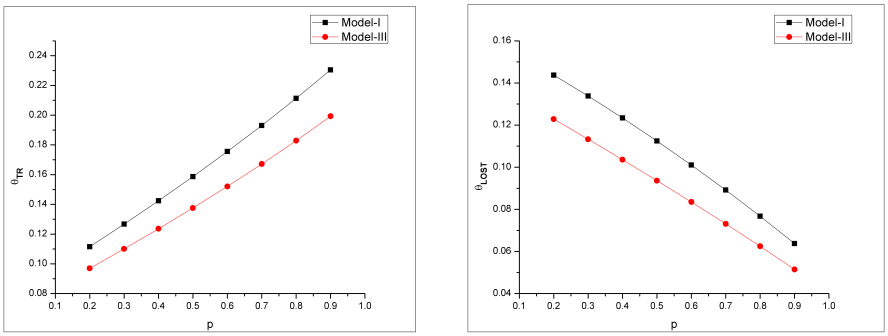

Fig 7.2: p versus θ_{LOST} and θ_{TR} in models I and III

start to compare model-IV and model-V. The priority based postponement of model-V makes its buffer size smaller than that in model-IV. So the loss rate of the model-V will be always lower than that of the model-IV. These measures are computed for various values of p by fixing, $K = 7, L = 4, \alpha = 0.24, m = 4, N = 3, \gamma = \frac{Lp}{K} + \frac{1}{N}$,

$$S = \begin{bmatrix} 0.001 & 0.999 & & \\ & 0.001 & 0.999 & \\ & & 0.0015 & 0.9985 \\ & & & 0.001 \end{bmatrix} \text{ and } S^0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0.999 \end{bmatrix}$$

for model-IV and $K = 7, L = 4, M = 5, n = 4, v = 2, \alpha = 0.24, p_1 = 0.8, \gamma = \frac{Lp}{K}$ with same S and S^0 for model-V, and shown in figure 7.5. From these observations it is clear that model-V is superior to model-IV. Comparison of model-IV is done with model-VI for various values of N by fixing $K = 7, L = 4, \alpha = 0.24, m = 4, N = 3$ in model-IV and $K = 7, L = 4, M = 5, n = 4, v = 2, \alpha = 0.24, p_1 = 0.8, N = 3$ in model-VI and shown in the figure 7.6. Here also we can see that loss rate of

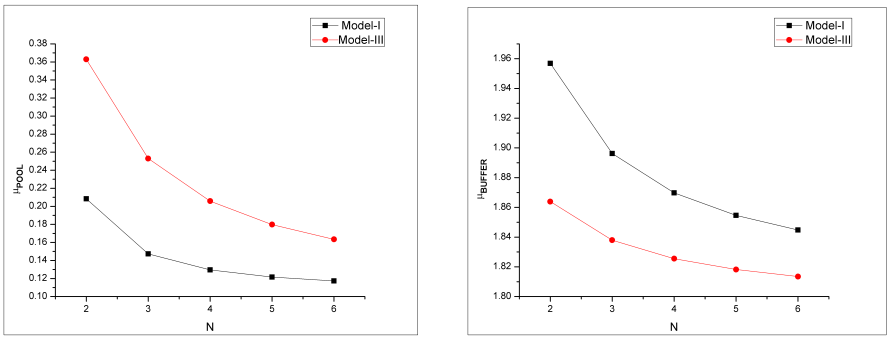

Fig 7.3: N versus μ_{POOL} and μ_{BUFFER} in models I and III

model-VI is much smaller than that of model-IV. This makes model-VI superior to model-IV. This is due to the effect of priority based postponement of the model-VI. Model-IV is a discrete time counterpart of model-I. We have not obtained any surprising results in model-IV to compare it with model-I.

As a conclusion, this thesis studied queues with postponed work under N-policy. It proposed finite buffer models with infinite capacity pool of postponed work. It suggested methods to minimize overflow jobs in finite capacity queues. It analysed various features in such a system and presented some numerical computation formulas. This can be considered as an extended study of the well-known area of finite capacity queues.

The models discussed in this thesis can be extended in several directions. A game theoretic approach is desirable in many cases. In discrete time, it is also possible to have arbitrarily distributed service time. In interruption models with postponed work, instead of N-policy, we can also analyse the effect of T-policy.

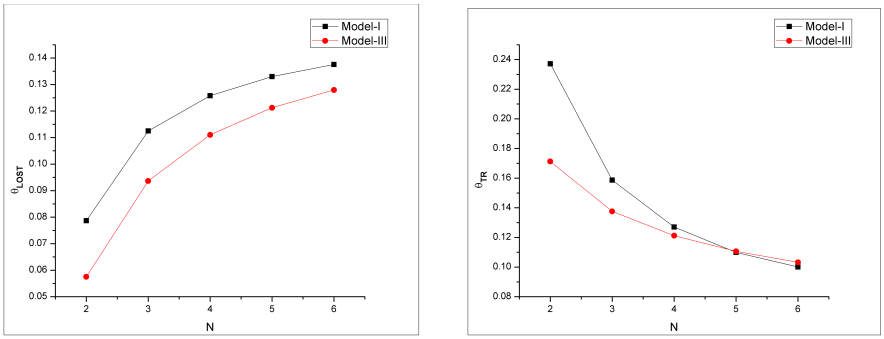

Fig 7.4: N versus θ_{LOST} and θ_{TR} in models I and III

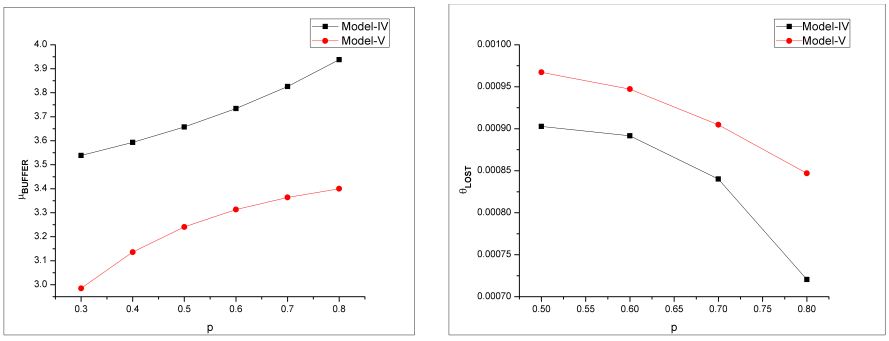

Fig 7.5: p versus μ_{BUFFER} and θ_{LOST} in models IV and V

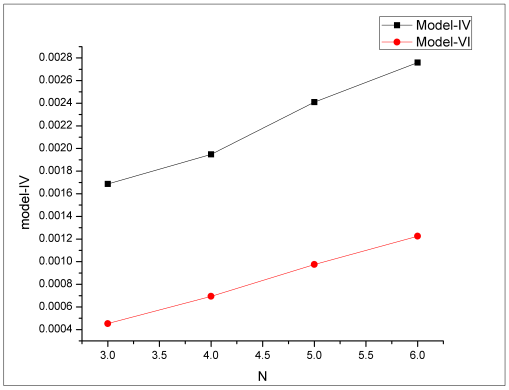

Fig 7.6: N versus θ_{LOST} in models IV and VI

Bibliography

[1] Alfa A.S. (2007) Discrete time queues and matrix-analytic methods,TOP,Springer publication, 147-185.

[2] Arivarignan G., Sivakumar B., and Jayaraman R. (2009) Stochastic Modelling of inventory systems with Postponed Demands and Multiple Server Vacations, International Journal of Applied Mathematics, Vol.1, 1-19.

[3] Atentia I. and Moreno P. (2006) A Discrete-Time Geo/G/1 retrial queue with the server subject to starting failures, Annals of OR, 141:85-107.

[4] Atentia I. and Moreno P. (2008) A discrete-time retrial queue with multiplicative repeated attempts, J Appl Math Comput (2008) 27: 6375, DOI 10.1007/s12190-008-0042-7.

[5] Baccelli F., Boyer P. and Hebuterne G. (1984) Single-server queue with impatient customers. Adv. Appl. Probab., 16:887-905.

[6] Barrer D.Y. (1957) Queueing with impatient customers and indifferent clerks, Operations Research, 5:644-649.

[7] Barrer D.Y. (1957) Queuing with impatient customers and ordered service, Operations Research, 650-656.

[8] Bocharov P.P., Pavlora O.I. and D. A. Puzikova D.A. (1999) M/G/1/r Retrial queuing system with priority of primary customers, TOP, 30(3-4): 89-98.

[9] Breuer L., Dudin A.N. and Klimenok V.I. (2002) A retrial BMAP/PH/N system, Queueing System, 40: 433-457.

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

[11] Ching Wai-Ki, Choi Sin-Man and Huang Min (2010) Optimal Service Capacity in a Multiple-Server Queueing System: A Game Theory Approch, Journal of Industrial and Management Optimization, Volume 6, Number 1, 73-102.

[12] Choi B.D. and Chang Y. (1999) Single server retrial queues with priority calls. Mathematical and Computer Modelling, 30 (3-4).

[13] Cinlar E. (1975) Introduction to stochastic processes, New Jersy: Prentice-Hall.

[14] Daley D.I. (1965) General Customer impatience in the queue G1/G/1, Journal of Applied probability, 2:186-205.

[15] Deepak T. G. (2001) Analysis of some queueing models related to N-policy, Ph.D thesis submitted at Cochin University of Science and Technology.

[16] Deepak T.G., Joshua V.C. and Krishnamoorthy A.(2004) Queues with postponed work, sociedad de Estdistica e Investigacion Operativa, Top vol 12, No. 2, pp. 375-398.

[17] Deepak T.G., Krishnamoorthy A., and Viswanath C. Narayan and Vineetha K. (2008) Inventary with service time and transfer of customers and/or inventory, Annals of Oper.Res 160:191-213.

[18] De Kok A.G. and Tijms H.C. (1985) A queueing system with impatient customers, Journal of Applied Probability, 22: 688-696.

[19] Diamond J.E. and Alfa A.S. (1999) Matrix analytic methods for a multi-server retrial queue with buffer, TOP, 7(2):249-266.

[20] Dudin A.N. and Klimenok V.I. (2000) A retrial BMAP/SM/1 system with linear repeated requests, Queueing Systems, 34:47-60.

[21] Dudin A.N., Klimenok V.I., and Tsarenkov G.V. (2002) Charateristics calculation for a single server queueing system with the batch Markovian arrival process,semi-markovian service and finite buffer., Automation and remote control, 8:87-101.

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

[23] Erlang A.K. (1909) The Theory of Probabilities and Telephone Conversations, Nyt Tidsskrift Matematik B.20, 33-39.

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

[25] Gail H.R., Hantler S.L. and Taylor B.A. (1998) Analysis of a non-preemptive priority multiserver queue, Adv. Appl. Probab, 20:852-879.

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

[27] Gendenko B.V. and Kovalenko I.N. (1989) Introduction to Queueing Theory, Birkhauser Boston Inc., Boston, 2nd edition.

[28] Gomez-corral A., Krishnamoorthy A. and Viswanath C. Narayan (2005) The impact of self generation of priorities on multi server queues with finite capacity Stochastic models 21:427-447.

[29] Gross D. and Harris C.M.(1988) Fundamentals of Queueing Theory, JohnWiley and Sons, New York.

[30] He Qi Ming, Neuts M. F.(2002) Two $M/M/1$ queues with transfer of customers. Kluwer Academic publishers, manufactured in The Netherlands.

[31] Jaiswal N.K. (1968) Priority Queues, Academic Press, New York.

[32] Karlin Samuel,Taylor H.M.(1975) A first course in Stochastic Processes, Academic press, Newyork.

[33] Karlin Samuel,Taylor H.M.(1981) A second course in Stochastic Processes, Academic press, Newyork.

[34] Krishnamoorthy A., Deepak T.G. and Viswanath C. Narayan and Vineetha.K.(2006) Effective utilization of idle time in an (s, S) in-

ventory with positive service time, Hindawi publishing corporation, Journal of Applied mathematics and stochastic analysis.

[35] Krishnamoorthy A., Gopakumar B. and Viswanath C. Narayan (2009) A queueing model with interruption resumption/restart and reneging, Bulletin of kerala mathematics association, Special issue, p.p 29-45.

[36] Krishnamoorthy A. and Islam M.E. (2004) Inventory system with postponed demands, Stoch. Anal. Appl. 22(3), 827-842.

[37] Krishnamoorthy A., Pramod P.K. and Deepak T.G. (2009) On a queue with interruptions and repeat/resumption of service, Non linear Analysis, Theory, Methods and Applications, Elsevier.

[38] Latouche G. and Ramaswami V.(1999) Introduction to Matrix Analytic Methods in Stochastic Modellings, ASA-SIAM, Philadelphia.

[39] Li.Jihong and Tian.Naishuo (2007) The $M/M/1$ queue with working vacations and vacation interruptions, Systems Engineering Society of China and Springer-Verlag.

[40] Langaris C. (1993) Waiting time analysis of a two-stage queueing system with priorities. Queueing Systems, 14:457-473.

[41] Leemanns H. (2000) Probable bounds for the mean queue lengths in a hetrogeneous priority queue. Queueing Systems, 36: 269-286.

[42] Medhi J.(2003)Stochastic models in queueing theory, Academic press, An imprint of Elsevier,USA.

[43] Medhi J.(1984) Stochastic Processes, New age international, NewDelhi.

[44] McKinsey J.C.C. (1952) Introduction to the Theory of Games, The RAND corporation, McGraw-Hill Book Company, New York.

[45] Neuts M.F.(1994) Matrix Geometric Solutions in Stochastic Models- An Algorithmic Approach, Dover Publications, New York.

[46] Paul Manuel, Sivakumar B. and Arivarignan G. (2007) Perishable inventory system with Postponed Demands and Negative Customers, Journal of Applied Mathematics and Decision Sciences, 1-12.

[47] Pramod P.K. (2010) On queues with interruptions and repeat or resumption of service, Ph.D thesis, Cochin University of Science and Technology.

[48] Sivakumar B. and Arivarignan G. (2008) An inventory system with Postponed Demands, Stochastic Analysis and Applications, Vol.22, No.3.,827-842.

[49] Stanford D.A. (1997) Waiting and interdeparture time in priority queues with Poisson-and general arrival-streams. Oper. Res., 45:725-735.

[50] Tijms H. C. (1986) Stochastic modelling and analysis- A computational approach, John wiley and sons.

[51] Takacs L. (1964) Priority queues. Operations Research, 12:63-74.

[52] Takacs L.(1974) A single server queue with limited virtual waiting time, Journal of Applied Probability, 11:612-617.

[53] Takine T. (1999) The non preemptive priority MAP/G/1 queue, Oper. Res., 47:917-927.

[54] Wang Q. (2004) Modelling and analysis of high risk patient queues. European J. Oper. Res., 155:502-515.

List of publications

- A.Krishnamoorthy, C.B.Ajayakumar, P.K.Pramod, An $M/PH/1$ Queue with Postponed work under N-policy (Communicated).
- A.Krishnamoorthy, C.B.Ajayakumar, Modified $M/PH/1$ Queue with Postponed work under N-policy (Communicated).
- A.Krishnamoorthy, C.B.Ajayakumar, An $M/M/1$ Queue with Postponed work and service interruption under N-policy (Communicated)
- A.Krishnamoorthy, C.B.Ajayakumar, P.K.Pramod, A Discrete time $Geo/PH_d/1$ Queue with Postponed work under N-policy (Communicated).
- A.Krishnamoorthy, C.B.Ajayakumar, Discrete time $Geo/E_d/1$ Queues with Postponed work and Protected stages (Communicated).