STOCHASTIC MODELLING AND ANALYSIS

ON (s,S) INVENTORY POLICY WITH/WITHOUT RETRIAL AND INTERRUPTION OF SERVICE/PRODUCTION

THESIS SUBMITTED TO THE
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
UNDER THE FACULTY OF SCIENCE

BY

SAJEEV S NAIR

Research Scholar (Reg. No. 2737)
Department of Mathematics
Cochin University of Science and Technology
Kochi-682022, INDIA

September 2011

CERTIFICATE

This is to certify that the thesis entitled **ON (s,S) INVENTORY POLICY WITH / WITHOUT RETRIAL AND INTERRUPTION OF SERVICE / PRODUCTION** is a bonafide record of the research work carried out by **Mr. Sajeev S Nair** under my supervision in the department of Mathematics . Cochin University of Science and Technology. The results embodied in the thesis have not been included in any other thesis submitted previously for the award of any degree or diploma.

Dr. A. Krishnamoorthy

(Supervising Guide)

Professor(Retired).

Department of Mathematics.

Kochi – 682 022.

Kochi - 22

2nd September 2011

DECLARATION

I, **Sajeev S Nair** hereby declare that this thesis entitled **ON (s,S) INVENTORY POLICY WITH / WITHOUT RETRIAL AND INTERRUPTION OF SERVICE / PRODUCTION** contains no material which had been accepted for any other Degree or Diploma 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 reference are made.

Sajeev S Nair
Research Scholar,
Registration No. 2737,
Department of Mathematics,
Cochin University of Science And Technology,
Kochi – 682022, Kerala.

Kochi - 22
2nd September 2011

ACKNOWLEDGEMENT

It is my pleasant duty to acknowledge the influence of my teachers who instilled in me the desire to pursue mathematical research. I have been very fortunate in having Professor A. Krishnamoorthy as my supervisor. He has been a constant source of inspiration, support and orientation. He always looked into the progress of my work, corrected and guided me with immense care and patience. I am greatly indebted to him. Dr. Viswanath C Narayanan has been close to this project from the very beginning and he was the co-author in all my papers. His remarkable knowledge and response with characteristic warmth benefited a lot. I owe a special debt to him.

The excellent facilities available at the department of Mathematics, Cochin University of Science and Technology were of great help in the realization of this doctoral thesis. I thank the various heads of the department. faculty members and administrative staff during the period of my research for providing such a congenial condition to work. I also thank the various fellow researchers from whom I benefited immensely. Among them Mr. B Gopakumar, Mr. M.K Sathyan. Mr. Varghese Jacob, Mr. C Sreenivasan , Dr. P.K Pramod , Mr.R Manikandan , Mr Ajayakumar ,Dr. K Lalitha , Ms. C.P Deepthi , Dr. K.P Jose, Dr.S.Babu , Ms T.Resmi , Ms Seema Varghese, Mr.Kirankumar. Mr.K.B Tony and Ms.Shiny Philip need special mention.

I thank the Principals and collegues of Govt.College, Kattappana, Govt.Victoria College, Palakkad and Govt. Women's Polytechnic College, Thrissur for their co-operation and encouragement during the course of the work. My sincere thanks to Mr.Ajithkumar Raja, Principal & Director, Mr. Joju Tharaken, Director, Sakthan Thampuran College of Mathematics and Arts, for providing me the facilities of the computer lab of the college ,for doing some of the numerical works in the thesis.

I express my heartfelt gratitude to all my friends and well wishers whose support and encouragement were always a source of inspiration for my achievement. Mr. M.P Rajan, Dr.V.K Krishnan, Prof. C. Govindankutty Menon, Mr. K Jayakumar, Mr. K.Sreekrishnan and several others are among them.

Over and above I would like to mention with affection about the moral support and encouragement given all throughout my work by my parents and my beloved wife Sandhya .

To

My

Parents and Teachers

CONTENTS

CHAPTER 1

Introduction

1.1 Description of queueing problems

In many real life situations customers have to wait in a queue for getting service. For examples customers wait in a bank counter, patients wait in a hospital, airplanes wait to take off or for landing etc. Queues may be reduced in size or prevented from being formed by providing additional service facilities which results in a drop in the profit. On the other hand excessively long queues may result in lost sales and loss of customers. Hence the problem is to achieve a balance between the cost associated with long queues and that associated with the reduction / prevention of waiting. Queueing theory is that branch of applied probability which studies such service systems and provides answers to the above problem.

Although there are many types of queueing systems, the following are the basic characteristics of any queueing process.

Arrival pattern of customers

The arrival pattern describes the manner in which customers arrive and join the queueing system. It is often measured in terms of average number of arrivals per unit time (mean arrival rate) or the average time between successive arrivals (mean inter arrival time). The arrival of customers is often expressed by means of a probability distribution of the number of arrivals or of the inter arrival time. Arrival may also occur in batches instead of one at a time.

If the queue is too long a customer may decide not to enter it upon arrival. Such a situation is called balking. On the other hand, a customer may enter the queue, but after some time lose patience and decide to leave. This is known as reneging. Another case is, when there is more than one queue, customers may switch from one to another which is called jockeying.

Service pattern of servers

The mode of service is represented by means of a probability distribution of the time required to serve a customer. Service may also be in single or in batches.

If the system is empty, the server is idle. The servers who become idle may leave the system for a random period called vacation. These vacations may be utilized to perform additional work assigned to the servers. However in retrial queues with no waiting space, it may be noted that each service is preceded and followed by an idle period.

Queuing discipline

The queuing discipline specifies the manner in which the customers are selected for service when a queue is formed. The most common disciplines are FIFO (First in First out) and LIFO (Last in First out).Another queue discipline is SIRO (Service in Random order). In some cases customers are given priorities upon entering the system. The ones with higher priorities are selected for service ahead of those with lower priorities.

Service channels

The number of service channels refers to the number of parallel service stations which can provide identical service to the customers.

Stages of service

A service may have several stages. That is a customer has to progress through a series of service stages prior to leaving the system. Such situations occur in tandem queues, network of queues etc.

1.2 Description of inventory systems

Inventory may be defined as stock of goods, commodities and other resources that are stored for the smooth conduct of business. In inventory models the availability of items is also to be considered in addition to the features in Queueing theory. If the time required to serve the items to the customers and time required to replenish the items (lead time) are both negligible then no queue is formed except in the case when order for replenishment is placed only when a number of back orders accumulate. If either service time or lead time or both are taken to be positive then a queue is formed; in the case of negligible service time with positive lead time, a queue of customers is formed provided backlog is permitted.

An (s, S) inventory policy is a policy according to which when inventory level drops to s then an order is placed. The order quantity is

Q = *S-s* so that the maximum inventory level is *S*. Such an inventory model is called *(s,S)* inventory model. In *(s, S)* policy, *s* and *S* are control variables with s, the reorder level and S the maximum number of items that can be held in the storage. Here we use *(s, S)* policy in the sense defined in Stanfel and Sivazlian [63]: the on hand inventory, on reaching the level *s*, an order for the fixed quantity *S-s* of the item is placed There are several other ordering policies: Order up to maximum *S* policy in which replenishment order is placed at levels smallest where as replenishment quantity is $S-i$ when inventory level is i $(0 \leq i \leq s)$ at replenishment epoch. In random order quantity policy, the order quantity can be any thing between $s+1$ and $S-s$. Yet another ordering policy is to place replenishment order when inventory level belongs to $\{0,1,\ldots,s\}$.

1.3 Some basic concepts

Stochastic process

A stochastic process is a collection of random variables $\{X(t), t\in T\}$. That is for each $t\in T$, X(t) is a random variable. The index t is often referred to as time and we refer to the possible values of X(t) as the state space of the process. The set T is called the index set of the process. If T is a countable set then the stochastic process is said to be a discrete (time) process. If T is an interval of the real line then the stochastic process is said to be a continuous (time) process. For instance, $\{X_n, n=0,1,\ldots..\}$ is a discrete time stochastic process indexed by the set of non negative integers, ,while $\{X(t), t \geq 0\}$ is a continuous time process indexed by non negative real numbers.

Markov Process

A stochastic process$\{X(t), t\in T\}$ is called a Markov process if it satisfies the condition

$Pr\{X(t_n) = x_n / X(t_{n-1})=x_{n-1}, X(t_{n-2}) = x_{n-2},.....X(t_0)=x_0\}=Pr\{X(t_n)=x_n / X(t_{n-1})=x_{n-1}\}$ for $t_0<t_1<.....<t_{n-1}<t_n$ and for every n ; $x_0, x_1,.... x_n$ are elements of the state space. This means that the distribution of any future occupancy depends only on the present state but not on the past.

Exponential distribution

A continuous random variable X is said to have an exponential distribution with parameter λ if its probability density function is given by $f(x) = \lambda e^{-\lambda x}$,$x\geq 0$ and $\lambda > 0$. This distribution has the memoryless property; that is $P[X>t+s / X>t] =P[X>s]$ for all t, s ≥ 0. Exponential distribution is relatively easy to work. In making a mathematical model for a real life phenomenon we often assume that certain random variables associated with the problem under study are exponentially distributed.

Renewal Process

Let $\{N(t), t \geq 0\}$be a counting process and X_n denote the time between the $(n-1)^{st}$ and n^{th} renewal. If the sequence of non negative random variables $\{X_1,X_2,.....\}$is independent and identically distributed then the counting process (the number of renewals up to time t), $\{N(t),t \geq 0\}$ is called a renewal process. Consider a renewal process having inter arrival times $X_1,X_2,....$with distribution function F. Set $S_n = \sum_{i=1}^{n} X_i, n \geq 1; S_0 = 0$.

Then we have $N(t) = \max\{n: S_n \le t\}$ and the distribution of N(t) is given by $P\{N(t) = n\} = F_n(t) - F_{n+1}(t)$ where F_n is the n-fold convolution of F with itself. The Poisson process is a renewal process where F is an exponential distribution.

Poisson Process

A renewal process $\{N(t), t \ge 0\}$ is said to be a Poisson process having rate λ if

(i) $N(0) = 0$.

(ii) The process has stationary and independent increments.

(iii) $P\{N(h)=1\} = \lambda h + o(h)$.

(iv) $P\{N(h) \ge 2\} = o(h)$.

It follows from the definition that for all $s,t \ge 0$,

$$P\{(N(t+s) - N(s)) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, n = 0,1,\dots.$$

For a Poisson process having parameter λ the inter arrival time has an exponential distribution with mean $1/\lambda$.

Continuous-time Phase type (PH) distributions

Consider a Markov chain Ω on the states $\{1,2,\dots,m+1\}$ with infinitesimal generator matrix $Q = \begin{bmatrix} T & T^0 \\ 0 & 0 \end{bmatrix}$ where the $m \times m$ matrix T satisfies $T_{ii} < 0$ for $1 \le i \le m$ and $T_{ij} \ge 0$ for $i \ne j$; T^0 is an $mx1$ coloumn matrix such that $Te + T^0 = 0$, where e is a column matrix of 1's of appropriate order. Let (α, α_{m+1}), where α is a $1xm$ dimensional row vector and α_{m+1} is a scalar such that $\alpha e + \alpha_{m+1} = 1$, be the initial probability vector of Ω. m+1 is an absorbing state for the Markov chain Ω .For eventual absorption into the absorbing state, starting from the initial state, it is necessary and sufficient that T is non singular. The probability distribution F(.) of time until absorption in the state $m+1$ corresponding to the initial probability vector (α, α_{m+1}) is given by $F(x)=1-\alpha e^{(Tx)} e$,$x \geq 0$.A probability distribution F(.) is a distribution of phase type if and only if it is the distribution of time until absorption of a finite Markov chain described above. The pair (α, T) is called a representation of F(.). The moments about origin are given by $E(X^k) = \mu_k' = (-1)^k k!(\alpha T^{-k} e)$ for $k \geq 0$.When m = 1 and $T=[-\lambda]$, the underlying PH-distribution is exponential.

PH-renewal process

A renewal process whose inter-renewal times have a PH distribution is called a PH-renewal process. To construct a PH-renewal process we consider a continuous time Markov chain with state space {1,2,…,m+1} having infinitesimal generator $Q = \begin{bmatrix} T & T^0 \\ 0 & 0 \end{bmatrix}$. The $m \times m$ matrix T is taken to be nonsingular so that absorption to the state $m+1$ occurs with probability1 from any initial state. Let (α,0) be the initial probability vector. When absorption occurs in the above chain we say a renewal has occurred. Then the process immediately starts anew in one of the states {1,2,…,m} according to the probability vector α . Continuation of this process gives a non terminating stochastic process called PH-renewal process.

Level Independent Quasi-Birth -Death (LIQBD) process

A level independent quasi birth and death process is a Markov chain on the state space E=$\{(i,j), i \geq 0, 1 \leq j \leq m\}$ with infinitesimal generator matrix Q given by

$$Q = \begin{bmatrix} B_0 & A_0 & & & & \\ B_1 & A_1 & A_0 & & & \\ & A_2 & A_1 & A_0 & & \\ & & A_2 & A_1 & A_0 & \\ & & & \cdot & & \cdot \\ & & & & & \cdot \\ & & & & & \cdot \end{bmatrix} \qquad (1.1)$$

The above matrix is obtained by partitioning the state space E into levels $\{\hat{0}, \hat{1}, \hat{2}, ...\}$, where $\hat{i} = \{(i,j), i \geq 0, 1 \leq j \leq m\}$. The states within the levels are called phases. The matrix B_0 denotes the transition rates within level $\hat{0}$, matrix B_1 denotes the transition rates from level $\hat{1}$ to level $\hat{0}$. A_2, A_1 and A_0 denote transition rates from level $\hat{i}$ to $(\hat{i-1}), \hat{i}$ and $(\hat{i+1})$ respectively.

Matrix Analytic Method

Matrix analytic approach to stochastic models was introduced by M.F Neuts to provide an algorithmic analysis for M/G/1 and GI/M/1 type queueing models. The following brief discussion gives an account of the method of solving an LIQBD using the matrix geometric method. For a detailed description, we refer to Neuts[56], Latouche and Ramaswami[52].

Let $x=(x_0,x_1,x_2,\ldots)$, be the steady state vector ,where x_i's are partitioned as $x_i=(x(i,0),x(i,1),x(i,2)\ldots,x(i,m))$, m being the number of phases with in levels.

Let $x_i = x_0R^i$, i≥1 .Then from xQ = 0 we get

$$x_0A_0 + x_1A_1 + x_2A_2 = 0$$

$$x_0A_0 + x_0RA_1 + x_0R^2A_2 = 0$$

$$x_0(A_0 + RA_1 + R^2A_2) = 0$$

Choose R such that $R^2A_2 + RA_1 + A_0 = 0$.

Also we have $x_0B_0 + x_1B_1 = 0$, which gives

$$x_0B_0 + x_0RB_1 = 0$$

i.e. $x_0(B_0 + RB_1) = 0$.

First we take x_0 as the steady state vector of $B_0 + RB_1$. Then x_i, for $i \geq 1$ can be found using the formulae; $x_i = x_0R^i$ for $i \geq 1$. Now the steady state probability distribution of the system is obtained by dividing each x_i, with the normalizing constant $[x_0 + x_1 + \ldots]\,e = x_0(I-R)^{-1}e$.

The above discussion leads to the following theorem.

Theorem The QBD with infinitesimal generator Q of the form (1.1) is positive recurrent if and only if the minimal non negative solution R of the matrix quadratic equation $R^2A_2+RA_1+A_0=0$ has all its eigen values inside the unit disc and the finite system of equations $x_0(B_0+RB_1)=0$, $x_0(I-R)^{-1}e=1$ has a unique solution x_0. If the matrix $A=A_0+A_1+A_2$ is irreducible, then $sp(R) <1$ if and only if $\pi A_0e < \pi A_2e$, where π is the stationary probability vector of A = $A_0+A_1+A_2$. The stationary probability vector $x=(x_0,x_1,\ldots)$ of Q is given by $x_i=x_0R^i$ for $i \geq 1$.

Level Dependent Quasi Birth Death (LDQBD) Process

A level dependent Quasi-Birth –Death process is a Markov process on a state space $E=\{(i,j), i \geq 0, 1 \leq j \leq n_i\}$ with infinitesimal generator matrix Q given by

$$Q = \begin{bmatrix} A_{10} & A_{00} & & & & \\ A_{21} & A_{11} & A_{01} & & & \\ & A_{22} & A_{12} & A_{02} & & \\ & & A_{23} & A_{13} & A_{03} & \\ & & & & \cdot & \cdot \\ & & & & & \cdot \\ & & & & & \cdot \end{bmatrix} \qquad 1.2$$

The generator matrix Q is obtained in the above form by partitioning the state space E into levels $\{\hat{0},\hat{1},\hat{2},\ldots.\}$. Here the transitions take place only to the immediately preceding and succeeding levels for $i \geq 1$. However the transition rate depends on the level i, unlike in the LIQBD, and therefore the spatial homogeneity of the associated process is lost.

A special class of LDQBD's is those which arise in retrial queueing models.

Neuts-Rao Truncation method

Since the repeating structure is lost in LDQBD, its analysis is much more involved. However Neuts and Rao [57] suggested a truncation procedure using which certain class of LDQBD's which include retrial models can be made to have a repeating structure from a certain level N^{th} , where N^{th} is sufficiently large. For giving a brief idea of their method, we assume that n_i =m for every i≥N so that each level ≥ N contains the same number of states. Note that this is the case in most of the retrial queueing models. To apply Nuets –Rao Truncation, we take $A_{1i}=A_{1N}$, $A_{2i}=A_{2N}$ and $A_{0i}=A_{0N}$ for all i ≥ N .In the case of the retrial queues this is equivalent to assuming that retrial rate remains constant after the number of orbital customers exceeds a certain limit N.

Define $A_N = A_{0N} + A_{1N} + A_{2N}$ and $\pi_N = (\pi_N(0,0), \pi_N(0,1), \pi_N(0,2),\ldots\ldots\ldots, \pi_N(0,m))$ be the steady state vector of the matrix A_N. Then the relations $\pi_N A_N = 0$ together with $\pi_N e = 1$ when solved give the various components of π_N . The truncated system is stable if and only if $\pi_N A_{2N} e > \pi_N A_{0N} e$ and the original system is stable if $\underset{n\to\infty}{Lim} \frac{\pi A_{0N} e}{\pi A_{2N} e} < 1$.

This thesis is on inventory with positive service time. We consider both classical and retrial queueing process associated with this. Also we discuss questions related to interruptions in service of customers. In addition production inventory with interruptions is also investigated. Further we bring in 'protection' of service and/production from interruption. Optimization problems related to these are extensively analyzed in this thesis.

Having described the tools for analysis, we move on to provide a review of the work done in the theme of the present thesis.

1.4 Review of related works

In all the studies on inventory systems prior to Berman et al [8], it was assumed that the serving of inventory is instantaneous. However, this is not the case in many practical situations. For example in a TV showroom, a customer usually spends some time with the salesperson before buying the TV or in a computer shop, after selecting the model, one will have to wait until all the required softwares are installed. In Berman et al [8] it is assumed that the amount of time taken to serve an item is constant. This leads to the analysis of a queue of demands formed in an inventory system. This study was followed by numerous studies by several researchers on many kinds of inventory models with positive service time. These studies include the papers Berman and Kim[9] and Berman and Sapna [10]. Among these, the first one takes a dynamic programming approach and the second one takes a Markov renewal theoretic approach.

More recently, Krishnamoorthy and his co-authors used Matrix geometric Method to study inventory models [13,27,33,35,39,43,44,45,69], where positive service time for providing the inventoried item is assumed. In Krishnamoorthy et. al. [43], and Deepak et. al. [14], an explicit product form solution for an inventory system with service time could be arrived at due to the assumption of zero lead time. It is worth mentioning that Schwarz et. al. [58] could obtain product form solution for the joint distribution of the number of customers in the system and the inventory level even in the case of positive lead time .This is achieved through the assumption that no customer joins the system when the inventory is out of stock; those who are already in the queue stay back. We refer to the survey paper by Krishnamoorthy et. al. [36] for more details on inventory models with positive service time.

In an (s, S) production inventory system, once the production process is switched on (at the level s, as the inventory falls from S to s), it is switched off only after the inventory level goes back to S, the maximum inventory level. This makes it distinct from (s, S) inventory system with positive lead-time, where once the order is placed (the moment at which the inventory level hits the re-order level s), the replenishment takes place by a quantity S-s after a random amount of time; usually the ordering quantity is taken such that the inventory level goes above s at the time the order materializes.

Krishnamoorthy and Viswanath [47] introduced the idea of positive service time in to a production inventory model by considering MAP arrivals and a correlated production process. This model being a very general one as far as the modeling parameters are concerned, only an algorithmic analysis of the model could be carried out there. In a very recent paper Krishnamoorthy and Viswanath [48], assuming all the underlying distributions as exponential, obtained a product form solution in the steady state for a production inventory model with positive service time. The above work was partly motivated by the paper by Schwarz et al. [58], where a product form solution has been obtained in an (s, S) inventory model with positive service time.

Because of the fast growing applicability in the communication and other fields, retrial queueing models are getting more and more attention. The literature on these type of queueing models is vast. We refer to the book by Falin and Templeton [16] and the very recent one by Atralejo and Gomez Corral [2] for an extensive analysis of both theory and applications on retrial queues.

The first study on inventory models with positive lead-time and unsatisfied customers thus created, going to an orbit and retry for inventory from there, was made by Artalejo et. al [6]. Analytical solution to the problem discussed there could be found in Ushakumari[42]. Following these, a number of papers on inventory models with retrial of unsatisfied customers emerged. A few among them are listed in what follows. The papers by Krishnamoorthy and Islam [30,31]: of which the first paper is on a production inventory model with retrial of customers and the second one analyses a production inventory model with random shelf times of the items with retrials of the orbiting customers. The papers by Krishnamoorthy et. al. [33,44], study inventory models with positive service time and retrial of customers from an orbit with an intermediate buffer of finite capacity . The paper by Krishnamoorthy and Jose [34] investigates and compares different (s, S) inventory models with an orbit of infinite capacity, having / not having a finite buffer.

Service interruption models studied in the literature include different types of service unavailability that may be due to server taking vacations, server breakdown, server interruptions, arrival of a priority customer etc. The paper by White and Christie [73] on an M/M/1 queuing model with exponentially distributed service interruption durations was the first one to introduce the concept of service interruption. Some of the later papers which analyze queuing models with service interruptions, assuming general distribution for the service and interruption times, are by Jaiswal [20,21], Gaver [18], Keilson [23], Avi-Itzhak and Naor [7] and Thiruvengadom [65]. In all these papers it is assumed that the arrival of a high priority customer interrupts the service of a lower priority customer. Some other papers on service interruption models include Ibe and Trivedi

[19], Federgruen and Green [17], Van Dijk [70], Takine and Sengupta [64], Masuyama and Takine [53]. Kulkarni and Choi[49] study two models of single server retrial queue with server breakdowns. In the first model, the customer whose service is interrupted, either leaves the system or rejoins the orbit; whereas in the second model the interrupted service is repeated after the repair is completed. Some other papers which study retrial queues with an unreliable server include Aissani and Artalejo [1], Artalejo and Gomez-Corral[3], Wang et al[71], Sherman and Kharoufeh[59], Sherman et.al.[60]. Marie and K.Trivedi [55] study the stability condition of an M/G/1 priority queue with two classes of jobs. Class 1 jobs have preemptive priority over class 2 jobs. They consider three different types of preemptions and the effects of possible work loss (due to preemption) on the stability condition for the queueing system.

The queueing model analyzed by Krishnamoorthy and Ushakumari [42], where disaster can occur to the unit undergoing service, the one by Wang et al [72] with disaster and unreliable server are also models with server interruptions.

In a recent paper Krishnamoorthy et. al. [37] study queues with service interruption and repair, where a decision on whether to repeat or resume the interrupted service is made according to whether a phase type distributed random clock that starts ticking the moment interruption strikes, realizes after or before the removal of the current interruption. Another paper by Krishnamoorthy et. al. [29] studies a queuing model where no damage to the server is assumed due to interruption, so that the server there needs no repair. A decision is to be made whether to restart or resume the interrupted service. Interruption being a random variable is determined by the competition of two exponential random variables.

For more detailed reports on queueing models with interruptions we refer to the survey paper by Krishnamoorthy et al [38]. Priority queueing models are not discussed by them since in such cases it is not server breakdown that causes interruption of service of lower priority customers.

There are numerous studies on inventory systems where interruption occurs due to unreliable suppliers. We refer to the papers by Tomlin [66,67] and the paper by Chen and Li [12] for details on such studies. Our work is in an entirely different direction from described above in [66,67,12] for the following reasons. First of all, in the above papers, interruption occurs due to an unreliable supplier, whereas in our models, we do not assume that the supplier is unreliable and it is the unreliable server who causes interruptions. Most importantly, in our models, interruptions occur in the middle of a service and there is no restriction on the number of possible interruptions during a service. Krishnamoorthy et al [41] can be considered as the first paper to introduce the concept of service interruption, which occurs in the middle of a service, in an inventory system. The steady state distribution has been obtained explicitly in product form in the above paper. In another paper [40] by the same authors, the above model has been extended by considering positive lead-time.

In a queueing system where the service process consists of certain number of phases, with service subject to interruptions, the concept of protecting a few phases of service (which may be so costly to afford an interruption) from interruption could be an important idea. Klimenok et al [25] studies a multi-server queueing system with finite buffer and negative customers where the arrival is BMAP and service is PH-type. They assume that a negative customer can delete an ordinary customer in service if the service of a customer goes on in any of the unprotected phases; whereas if the service process is in some protected phase, the service of the customer is protected from the effect of the negative customers. Klimenok and Dudin [24] extends the above paper by considering disciplines of complete admission and complete rejection. Further, Klimenok and Dudin [24] assumes an infinite buffer. Krishnamoorthy et al [28] introduces the idea of protection in a queueing system where the service process is subject to interruptions. They assume that the final *m-n* phases of the Erlang service process with *m* phases are protected from interruption. Whereas if the service process belongs to the first n phases, it is subject to interruption and an interrupted service is resumed/repeated after some random time. There is no reduction (removal) in the number of customers due to interruption and no bound was assumed on the number of interruptions that can possibly occur in the course of a service. In this way, this study differs from the earlier ones where atmost one interruption was possible during a service and where the customer whose service got interrupted is removed from the system. The interruption models that we discuss in this thesis fall under the category of type I counter. This amounts to saying that when a server is under going an interruption no further interruption can befall it.

1.5 An Outline of the Present Work

This thesis is divided into six chapters including the present introductory chapter.

In the second chapter, we consider a single server queueing system with inventory where customers arrive according to a Poisson process. Customers, finding the server busy upon arrival, join an orbit of infinite capacity from where they retry for service. Service times are exponentially distributed. Immediately after the completion of a service the server either picks a customer from the orbit with probability p for the next service, if there is item in the inventory, or remains idle. Inventory is replenished according to the (s, S) policy, with lead times following exponential distribution. Primary arrivals do not join the orbit while the inventory level is zero. Stability of the above system is analyzed and steady state vector is calculated using Neuts-Rao truncation. An extensive numerical study of various performance measures such as mean and variance of waiting time of an orbital customer is carried out.

In the third chapter, we consider a single server queuing system with inventory where customers arrive according to a Poisson process. Inventory is served according to an exponential distribution. Replenishment of inventory is according to the (s, S) policy with lead time also following an exponential distribution. The service process is subject to interruptions, with the occurrence of the latter constituting a Poisson process. The interrupted server is repaired, the repair time being exponentially distributed. We assume that during interruption the customer being served waits there until his service is completed and also that no inventory is lost due to interruption. We also assume that while the server is on interruption no arrival is entertained and replenishment order placed, if any, is cancelled. Further when the inventory level is zero no fresh customer is permitted to join the system. Stability of the above system is analyzed and steady state vector is calculated numerically. Several system performance measures including waiting time of a customer in the system are also studied numerically.

In chapter 4 we discuss a special case of chapter 3. In that chapter, we consider a single server queuing system with inventory where customers arrive according to a Poisson process. Inventory is served according to an exponential distribution provided there are customers. Inventory is replenished according to the (s, S) policy with zero lead-time. The service process is subject to interruptions, which occurs according to a Poisson process. The interrupted server is repaired with the repair time following an exponential distribution. We assume that during interruption, the customer being served waits there until his service is completed and also that no inventory is lost due to this interruption. Stability of the above system is analyzed and steady state vector is calculated explicitly. Explicit formulas for system performance measures such as expected number of customers in the system, expected inventory size, expected interruption rate, waiting time of a customer in the system are also obtained.

In the fifth chapter we consider a single server queueing system to which customers arrive according to a Poisson process each demanding exactly one unit of an inventoried item. Service time durations are exponentially distributed. Inventory is replenished according to (s,S) policy, with lead time following exponential distribution. The service may get interrupted according to a Poisson process and if so the service restarts after a time interval that is exponentially distributed. Customers, upon arrival, finding the server busy, leaves the service area and joins an orbit from where they retry for service. The interval between two successive repeated attempts is exponentially distributed.

We assume that while the server is on an interruption an arriving customer joins the system with a certain probability. We also make the assumption that while the server is on an interruption a retrying customer goes back to the orbit with a certain probability and otherwise leaves the system. Again no arrival or retrial is entertained when the inventory level is zero. Stability of the above system is analyzed and steady state vector is calculated using Neuts-Rao truncation. A thorough numerical study of various performance measures such as mean and variance of waiting time of an orbital customer is carried out.

In the sixth chapter we consider a production inventory system with positive service time, with time for producing each item following Erlang distribution. Customers arrive according to a Poisson process. When the inventory level falls to s, production process is switched on and it is switched off when inventory level reaches S. Service time to each customer also follows Erlang distribution. The service gets interrupted according to a Poisson process and if so the service is repeated after an exponentially distributed time. The final few phases of the service process are assumed to be protected in the sense that the service will not be interrupted while being in these phases. The same is the case with the production process.

We assume that no inventory is lost due to a service interruption and that the customer being served waits there until his service is completed .On the other hand in the case of interruption to production process we assume that the item being produced is lost. Stability of the above system is analyzed and steady state vector is calculated numerically. A thorough numerical study of various performance measures is carried out.

CHAPTER 2

An Inventory Model With Retrial And Orbital Search *

2.1 Introduction

Because of the fast growing applicability in the communication and other fields, retrial queueing models are getting more and more attention. The literature on these type of queueing models is vast. (We refer to the books by Falin and Templeton [16] and Atralejo and Gomez Corral [2], for an extensive analysis of both theory and applications on retrial queues).

The first study on inventory models with positive lead time and unsatisfied customers thus created going to an orbit and retries for inventory from there was done by Artalejo et. al [6]. After their work, a number of papers on inventory models with retrial of unsatisfied customers emerged. A few among them can be listed as follows. The papers by Krishnamoorthy and Islam [30,31]; of which the first paper is on a production inventory model with retrial of customers and the second one analyses a production inventory model with random times for the shelf life of the items as well as for the retrials the

* The results in this chapter was presented as a paper at the international symposium on probability theory and stochastic process in honour of Professor S.R.S Varadhan FRS held at the Cochin university of Science and Technology from February 06-09,2009. It was also published in the Bulletin of Kerala Mathematics Association, Special Issue. 47-65, October 2009; Guest Editor: S. R. S. Varadhan FRS .

of orbiting customers. The papers by Krishnamoorthy et. al. [33,44], study inventory models with positive service times, retrial of customers from an orbit, and an intermediate buffer of finite capacity to store the commodity. The paper by Krishnamoorthy and Jose [34] compares different (s, S) inventory models with an orbit of infinite capacity, having and not having a finite buffer.

One peculiarity of the classical retrial queueing models is that, every service is sandwiched between two idle periods of the server. Deviating from this, Artalejo et. al. [5] analyzed an M/G/1 queueing model with retrial of orbital customers, where a service completion may be followed by beginning of a new service. They accomplish this by introducing an entity called 'orbital search' done by the server immediately after a service completion. Precisely, they assumed that immediately after a service completion, the server, with some probability makes an instantaneous search for an orbital customer for the next service. This model was generalized by Dudin et. al. [15], by assuming that the arrival process is BMAP and also that the search time is not negligible but is a random variable with a general distribution that depends on the number of customers in the orbit. An M/G/1 retrial queue with orbital search and non-persistant customers was studied by Krishnamoorthy et. al. [26]. Chakravarthy et. al. [11], analyzed a multi-server retrial queueing model with Poisson arrival process and orbital search. Wuchner et. al. [74] introduces an orbital search in finite-source retrial queues and uses MOSEL-2 tool for their analysis. A recent paper by Krishnamoorthy et. al. [46] on a MAP/PH/1 retrial queue with service interruption applies the idea of orbital search to a queueing system where the server becomes free either by service completion or by an interruption. Search time was assumed to be negligible in [11,26,46,74].

This chapter is on an (s, S) inventory system, where a positive lead-time for replenishment and a positive time for meeting the demand is assumed. Those customers, encountering an idle server and positive inventory, are immediately taken into service and customers who at the time of arrival find an idle server with zero inventory are considered lost. In this case those who are already present will stay back. A customer who finds the server busy, joins an orbit of infinite capacity and from there retries for service, with inter-retrial times exponentially distributed. For decreasing the waiting of orbital customers we introduce the orbital search. Hence at a service completion epoch, the server, with probability p, makes a search in the orbit and picks a customer, if any, randomly from the orbit, provided there is at least one item left in the inventory for the next service. Only at service completion epochs and not at arbitrary time points in an idle period, does the server makes a search for orbital customers . The search time is assumed to be negligible.

Orbital search was introduced with the hope that it would decrease the length of server idle period. However studies (see, Artalejo et al) [5] on retrial queueing models how that the search probability p has no effect on the steady state probability that the server is idle. But as p increases, the expected number of customers in the orbit and hence the expected waiting time of orbiting customers, decrease. Hence in our model, we study the waiting time of an orbital customer by approximating it with the waiting time in the corresponding model with finite orbital capacity. The approximation procedure is similar to that carried out in Artalejo and Gomez Corral [4]. It may be noted that the search of orbital customers brings down the expected waiting time of orbital customers.

This chapter is arranged as follows. In section 2.2, we describe the mathematical model under study. In section 2.3, a necessary and sufficient condition for the stability of the system is obtained and steady state distribution is found. Section 2.4 is devoted to some system performance measures like the expected waiting time of an orbital customer. Finally in section 2.5 we provide some results of the numerical experiments carried out for analyzing different aspects of the system under study.

2.2 Mathematical Model

The model under study is described as follows. Customers arrive to a single server counter according to a Poisson process of rate λ where inventory is served. Service times are iid exponential random variables with parameter μ. Inventory is replenished according to (s,S) policy, the replenishment time being exponentially distributed with parameter η. An arriving customer, finding the server busy, enters an orbit from where it retries for service. The interval between two successive repeated attempts is exponentially distributed with rate $j\theta$, given that the number of customers in the orbit is j. Immediately after a service the server goes for a search of customers in the orbit and picks a customer from the orbit with probability p or remains idle with probability

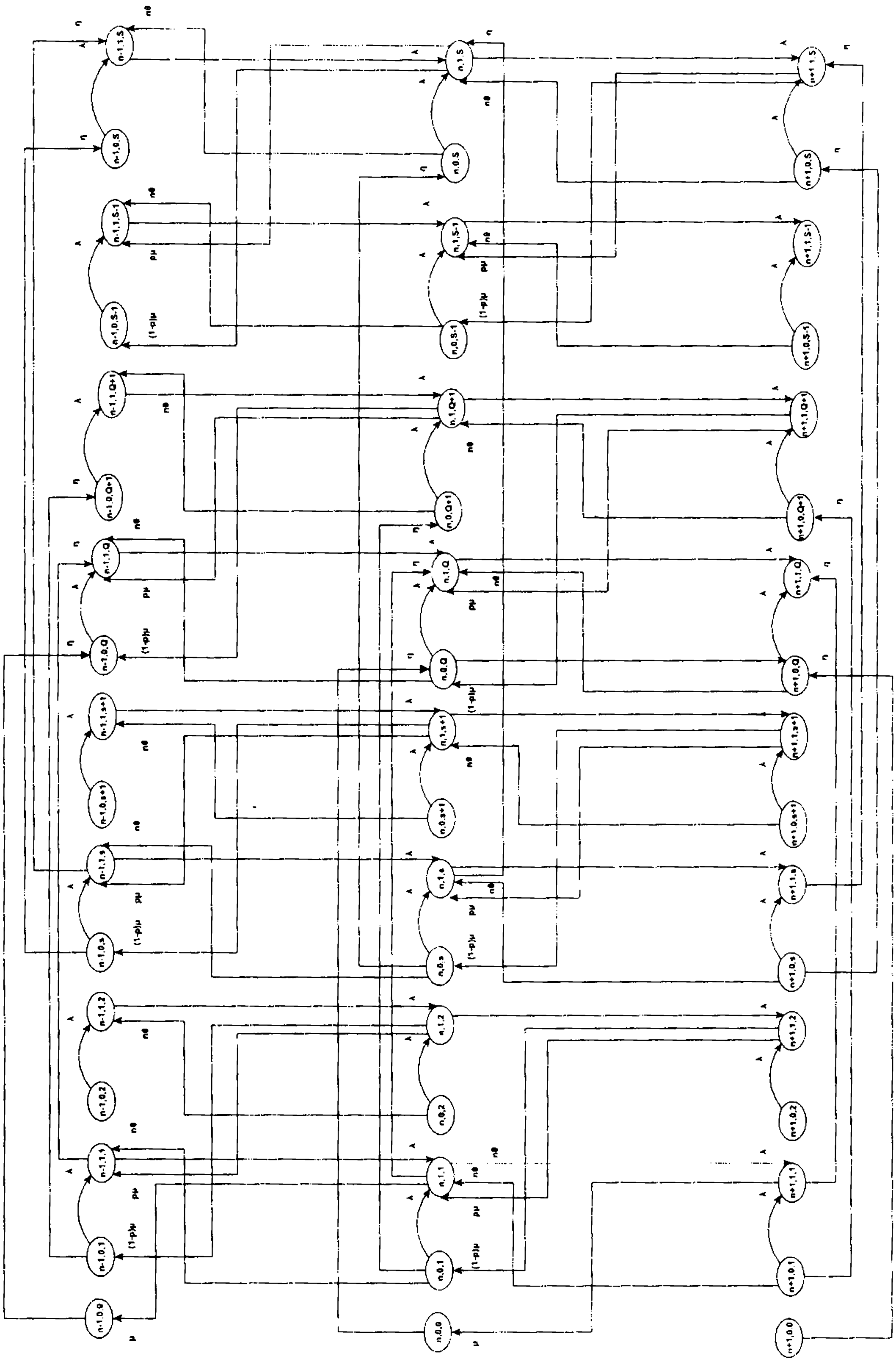
1- p. When the inventory level is zero no arrival or retrial is entertained. In the sequel, I denotes an identity matrix and e denotes a column vector of 1's of appropriate orders.

Let $N(t)$ be the number of customers in the orbit and $L(t)$ be the inventory level at time t. Also let $C(t)=\begin{cases} 1, & \text{if the server is busy} \\ 0, & \text{if the server is idle} \end{cases}$ be the server state. Then

$\Omega = \{X(t), t \geq 0\} = \{(N(t), C(t), L(t)), t \geq 0\}$ is a Markov chain on the state space $((Z_+ \cup \{0\}) \times \{0,1\} \times \{1,2,3,...,S\}) \cup ((Z_+ \cup \{0\}) \times \{0\} \times \{0\})$.

The state space of the Markov chain is partitioned in to levels $\hat{i}$ defined as

$\hat{i}$ = { (i,0,0), (i,0,1), (i,0,2),..., (i,0,s), (i,0,s+1),..., (i,0,Q), (i,0,Q+1),..., (i,0,S), (i,1,1), (i,1,2),..., (i,1,s), (i,1,s+1),..., (i,1,Q), (i,0,Q+1),.... (i,0,S) }, $i \geq 0$ and $Q=S-s$.



STATE TRANSITION DIAGRAM

This makes the Markov chain under consideration, a level dependent quasi birth death process with infinitesimal generator matrix

$$\hat{Q} = \begin{bmatrix} A_{10} & A_0 & 0 & 0 & 0 & 0 \\ A_{21} & A_{11} & A_0 & 0 & 0 & 0 \\ 0 & A_{22} & A_{12} & A_0 & 0 & 0 \\ 0 & 0 & A_{23} & A_{13} & A_0 & 0 \\ \cdot & \cdot & & & & \\ & & & \cdot & \cdot & \cdot \end{bmatrix},$$

where each entry is a square matrix of order (2S +1).

The transition from level $i \rightarrow i+1$ is represented by the matrix

$$A_0 = \begin{bmatrix} 0 & 0 \\ 0 & \lambda I \end{bmatrix}$$

The transition from level $i \rightarrow i\text{-}1$ is represented by the matrix ,

$A_{2j} = \begin{bmatrix} 0 & B_{2j} \\ 0 & C_{2j} \end{bmatrix}$, where B_{2j} (k+1,S+1+k)= jθ and C_{2j} (S+1+k,S+k)= pμ for k = 1to S .

The transition $i \rightarrow i$ is represented by the matrices

$$A_{1j} = \begin{bmatrix} D_1 & 0 & D_2 & D_3 & 0 & 0 \\ 0 & D_4 & 0 & 0 & D_5 & 0 \\ 0 & 0 & D_6 & 0 & 0 & D_7 \\ D_8 & 0 & 0 & D_9 & 0 & D_{10} \\ D_{11} & D_{12} & 0 & 0 & D_{13} & 0 \\ 0 & D_{14} & D_{15} & 0 & 0 & D_{16} \end{bmatrix}$$

S+1
S-2s-1 2S+3
s+1
s
S-2s+1
s+1

where,

D_1 is an (s+1)x(s+1) matrix whose non-zero entries are given by $D_1(1, 1) = -\eta$ and $D_1(k, k) = -(\eta+\lambda+j\theta)$, for $k \neq 1$.

D_2 is an (s+1)x(s+1) matrix whose non-zero entries are given by $D_2(k, k) = \eta$.

D_3 is an (s+1)x s matrix whose non-zero entries are given by $D_3(k+1,k) = \lambda$, for $k=1$ to s.

D_4 is an (S-2s-1)x(S-2s-1) matrix whose non-zero entries are given by $D_4(k,k) = -(\lambda+j\theta)$.

D_5 is an (S-2s-1)x(S-2s-1) matrix whose non-zero entries are given by $D_5(k,k) = \lambda$.

D_6 is an (s+1)x(s+1) matrix whose non-zero entries are given by $D_6(k,k) = -(\lambda+j\theta)$.

D_7 is an (s+1)x(s+1) matrix whose non-zero entries are given by $D_7(k,k) = \lambda$.

D_8 is an sx(s+1) matrix with non-zero entries given by $D_8(k,k) = (1-p)\mu$, for $k \neq 1$ and $D_8(1,1) = \mu$.

D_9 is an sxs matrix whose non-zero entries are given $D_9(k,k) = -(\eta+\lambda+\mu)$.

D_{10} is an sx(s+1) matrix whose non-zero entries are given $D_{10}(k,k+1) = \eta$.

D_{11} is an (S-2s+1)x(s+1) matrix whose non-zero entries are given by

$D_{11}(1,s+1) = (1-p)\mu$.

D_{12} is an (S-2s-1)x(S-2s-1) matrix with non-zero entries given by

$D_{12}(k+1,k) = (1-p)\,\mu$.

D_{13} is an (S-2s-1)x(S-2s-1) matrix with non-zero entries given by

$D_{13}(k, k) = -(\lambda + \mu)$.

D_{14} is an (s+1)x(S-2s-1) matrix with non-zero entries given by

$D_{14}(1, S-2s-1) = (1-p)\mu$.

D_{15} is an (s+1)x(s+1) matrix whose non-zero entries are given by

$D_{15}(k+1,k) = (1-p)\mu$.

D_{16} is an (s+1)x(s+1) matrix whose non-zero entries are given

$D_{16}(k,k) = -(\lambda + \mu)$.

2.3 Analysis of the Model

In this section we perform the steady state analysis of the model by first deriving the stability condition of the model under study.

2.3.1 Stability Condition

For finding the stability condition for the system under study, we apply Neuts-Rao truncation first. Suppose $A_{1i} = A_{1N}$ and $A_{2i} = A_{2N}$ for all i $\geq$N. Then the generator matrix of the truncated system will look like this:

$$\tilde{Q} = \begin{bmatrix} A_{10} & A_0 & 0 & 0 & 0 & & & & & & \\ A_{21} & A_{11} & A_0 & 0 & 0 & 0 & & & & & \\ 0 & A_{22} & A_{12} & A_0 & 0 & 0 & 0 & & & & \\ 0 & 0 & A_{23} & A_{13} & A_0 & 0 & 0 & 0 & & & \\ & & & & .. & .. & .. & & & & \\ & & & & & & .. & .. & .. & & \\ & & & & & A_{2N} & A_{1N} & A_0 & & & \\ & & & & & 0 & A_{2N} & A_{1N} & A_0 & 0 & \\ & & & & & & & .. & & .. & \\ & & & & & & & & .. & .. & .. \\ & & & & & & & & & & \end{bmatrix}$$

Define $A_N = A_0 + A_{1N} + A_{2N}$ and $\pi_N = (\pi_N(0,0), \pi_N(0,1), \pi_N(0,2), \ldots\ldots, \pi_N(0,S), \pi_N(1,1), \pi_N(1,2), \ldots\ldots, \pi_N(1,S))$ with $\pi_N(i,j) \geq 0$ and $\pi_N(0,0) + \pi_N(0,1) + \ldots\ldots + \pi_N(1,S) = 1$. Then the relations $\pi_N A_N = 0$ and $\pi_N e = 1$ when solved gave the various components of π_N as

$$\pi_N(1,1) = \frac{\eta}{\mu}\,\pi_N(0,0)$$

$$\pi_N(1,i) = \frac{\eta(\eta+\mu)^{i-1}}{\mu^i}\left[\frac{\eta+\lambda+N\theta}{\eta p+\lambda+N\theta}\right]^{i-1}\pi_N(0,0); \text{ for } i = 2 \text{ to } s+1$$

$$\pi_N(1,s+1) = \pi_N(1,s+2) = \ldots\ldots = \pi_N(1,Q)$$

$$\pi_N(1,Q+j) = \pi_N(1,Q) - \pi_N(1,j)\;;\text{ for } j = 1 \text{ to } s$$

π_N (1,Q+2) = π_N (1,Q)- π_N (1, 2)

$$\pi_N(0,i) = \frac{\eta(\eta+\mu)^i}{\mu^{i+1}}(1\text{-p})\,\mu\,\frac{(\eta+\lambda+N\theta)^{i-1}}{(\eta p+\lambda+N\theta)^i}\,\pi_N(0,0)\ \text{;for i=1 to s}$$

$$\pi_N(0,s+1) = \frac{\eta(\eta+\mu)^s}{\mu^{s+1}}(1\text{-p})\,\mu\,\frac{(\eta+\lambda+N\theta)^s}{(\eta p+\lambda+N\theta)^s(\lambda+N\theta)}\,\pi_N(0,0)$$

π_N (0,s+1)= π_N (0,s+2)=.......= π_N (0,Q-1)

$$\pi_N(0,Q+j)=(1\text{-p})\frac{\mu}{(\lambda+N\theta)}\pi_N(1,Q+j+1)+\frac{\eta}{(\lambda+N\theta)}\pi_N(0,j)\ \text{;for j= 0 to s-1}$$

$$\pi_N(0,S)=\frac{\eta}{(\lambda+N\theta)}\pi_N(0,s).$$

Let

SA=

$$\left[\frac{N\theta s(1-p)\mu}{\lambda+N\theta}+\frac{N\theta(S-2s-1)(1-p)\mu}{\lambda+N\theta}+\frac{N\theta(1-p)\mu}{\lambda+N\theta}+p\mu(s-1)+p\mu(Q-s)+p\mu\right]$$

and

$$SB=\frac{N\theta\eta(1-p)\mu}{\mu(\lambda+N\theta)}-\frac{N\theta\eta}{\lambda+N\theta}+\eta\,p.$$

The truncated system is stable if and only if $\pi_N A_{2N} e > \pi_N A_0 e$. That is iff

$$\frac{\eta(\eta+\mu)^s}{\mu^{s+1}}\left[\frac{(\eta+\lambda+N\theta)}{(\eta p+\lambda+N\theta)}\right]^s\pi_N(0,0)\,SA-\pi_N(0,0)SB >$$

$$\lambda Q\,\frac{\eta(\eta+\mu)^s}{\mu^{s+1}}\left[\frac{(\eta+\lambda+N\theta)}{(\eta p+\lambda+N\theta)}\right]^s\pi_N(0,0)$$

As N tends to infinity, this reduces to $\frac{\lambda}{\mu}<1$.

Thus we have the following theorem for stability of the system under study.

Theorem 2.1

The Markov Chain Ω is stable if and only if $\frac{\lambda}{\mu} < 1$.

The above theorem shows that the stability of the inventory system under study is independent of the search probability p.

2.3.2 Computation of Steady State Vector

We find the steady state vector of Ω, by approximating it with the steady state vector of the truncated system. Let $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$, be the steady state vector, where each $\pi_i = \pi_i(j, k)$, j = 0,1 and k=$1,2,\ldots,S$.

Suppose $A_{1i} = A_{1N}$ and $A_{2i} = A_{2N}$ for all i ≥ N. Let $\pi_{N+r} = \pi_{N-1}\, R^{r+1}$, for $r \geq 0$, then from $\pi Q = 0$ we get

$$\pi_{N-1}\ A_0 + \pi_N\ A_{1N} + \pi_{N+1}\ A_{2N} = 0$$

$$\pi_{N-1}\ A_0 + \pi_{N-1}\ R\ A_{1N} + \pi_{N-1}\ R^2\ A_{2N} = 0$$

$$\pi_{N-1}\ (A_0 + R\ A_{1N} + R^2\ A_{2N}) = 0$$

Choose R such that

$$A_0 + R\, A_{1N} + R^2 A_{2N} = 0.$$

We call this R as R_N. Also we have

$$\pi_{N-2}\ A_0 + \pi_{N-1}\ A_{1N-1} + \pi_N\ A_{2N} = 0.$$

$$\pi_{N-2}\ A_0 + \pi_{N-1}\ (A_{1N-1} + R_N\ A_{2N}) = 0$$

$$\pi_{N-1} = -\pi_{N-2}\, A_0 (A_{1N-1} + R_N\ A_{2N})^{-1}.$$

$$= \pi_{N-2}\, R_{N-1}, \text{ where}$$

$R_{N-1} = - A_0 (A_{1N-1} + R_N \ A_{2N})^{-1}$ Also

$\pi_{N-3} A_0 + \pi_{N-2} A_{1N-2} + \pi_{N-1} A_{2N-1} = 0.$

$\pi_{N-3} A_0 + \pi_{N-2} (A_{1N-2} + R_{N-1} A_{2N-1}) = 0$

$\pi_{N-2} = -\pi_{N-3} A_0 (A_{1N-2} + R_{N-1} A_{2N-1})^{-1}.$

$= \pi_{N-3} R_{N-2},$ where

$R_{N-2} = - A_0 (A_{1N-2} + R_{N-1} A_{2N-1})^{-1}$ and so on. Finally

$\pi_0 A_{10} + \pi_1 A_{21} = 0$ becomes

$\pi_0 (A_{10} + R_1 A_{21}) = 0$

First we take π_0 as the steady state vector of $A_{10} + R_1 A_{21}$. Then π_i, for $i \geq 1$ can be found using the recursive formulae;

$$\pi_i = \pi_{i-1} R_i, \text{ for } 1 \leq i \leq N-1.$$

Now the steady state probability distribution of the truncated system is obtained by dividing each π_i, with the normalizing constant

$[\pi_0 + \pi_1 + \ldots\ldots\ldots\ldots]e = [\pi_0 + \pi_1 + \ldots\ldots\ldots + \pi_{N-2} + \pi_{N-1} (I - R_N)^{-1}]e$.

2.4 System Performance Measures

2.4.1 Waiting time analysis of an orbital customer

Since no queue is formed in the orbit, customers independent of each other try to access the service .Therefore computation of the waiting time distribution is extremely complex. Hence we limit ourselves to the computation of expected waiting time.

We mentioned in the introduction that the search probability p has no effect on the server idle probability; but it brings down the number of customers in the orbit and hence their waiting time. So here we give some numerically tractable approximation formulae for calculating the moments of the waiting time of an orbital customer. Though we can find the expected waiting time using Little's Law, the second moment and variance of the waiting time are not easy to find. These moments are found by approximating the waiting time in the system under study by those in a corresponding system with finite orbit capacity.

Let $E(W_L)$ be the expected waiting time of an orbital customer in the system under study and $E(W_L^{(N)})$ be that in the corresponding system with finite orbit capacity N. Then

$$E(W_L) = \underset{N\to\infty}{Lim}\, E(W_L^{(N)}).$$

For the system with finite orbit capacity, $W_L^{(N)}$ can be found as the time until absorption in a Markov chain $\{X(t), t \geq 0\} = \{(N(t), C(t), L(t)), t \geq 0\}$, if the tagged customer is in the orbit ,where

$N(t)$=number of customers in the orbit including the tagged customer

$$C(t)=\begin{cases}1, & \text{if the server is busy}\\ 0, & \text{if the server is idle}\end{cases}$$

$L(t)$=inventory level at time t and $X(t)=\Delta$,if the tagged customer gets service. The state space of $X(t)$ is $\{\Delta\}\cup(i,j,k)$, $i=1,2,\ldots,N$; $j=1,2$; k varies from 0 to S if j=0 and k varies from 1to S if j=1 .The generator matrix of $X(t)$ is

$$\hat{Q}_W(N) = \begin{bmatrix} T & T^0 \\ 0 & 0 \end{bmatrix},$$

where T^0 is an N (2S+1) x 1 matrix given by

T^0((i-1) (2S+1)+j,1) = θ, j = *2 to S+1 ; i=1to N,*

T^0((i-1) (2S+1)+j,1) = $\frac{p\mu}{i}$, j = *S+3 to 2S+1 ; i=1to N* and

$$T = \begin{bmatrix} A_{11} & A_0 & & & & & \\ \tilde{A}_{21} & A_{12} & A_0 & & & & \\ 0 & \tilde{A}_{22} & A_{13} & A_0 & & & \\ 0 & 0 & \tilde{A}_{23} & A_{14} & A_0 & & \\ & & & & & & \\ & & & & & \tilde{A}_{2(N-1)} & \tilde{A}_{2N} \end{bmatrix};$$

where

$$\tilde{A}_{2j} = \begin{bmatrix} 0 & \tilde{B}_{2j} \\ 0 & \tilde{C}_{2j} \end{bmatrix}$$

with

$\tilde{C}_{2j}$ (S+1+k,S+k) = $\frac{j}{j+1}$ pμ for *k=1 to S*

$\tilde{B}_{2j}$ (k+1,S+1+k) = jθ for *k=1 to S*

and all other matrices are as defined in the generator matrix Q. Thus

$$E(W_L^{(N)}) = -\ \alpha\ T^{-1} e,$$

where $\alpha = \pi_L = (\pi_{L0}, \pi_{L1}, \pi_{L2}, \ldots\ldots \pi_{LN})$; $\pi_{Li} = \pi_i$ with entries corresponding to server is idle states taken as zero. It was verified numerically that for large N, $E(W_L^{(N)})$ converges according to Little's theorem.

2.4.2 Other Performance Measures

The following system performance measures are calculated numerically.

1. The probability that server is busy is given by

$$P(\beta)= \sum_{i=0}^{\infty}\sum_{j=0}^{S}\pi(i,1,j)$$

2. The expected number of customers in the system is given by

$$E(\sigma)= \sum_{i=0}^{\infty}\sum_{j=0}^{S}\{i\pi(i,1,j)+i\pi(i,0,j)\}.$$

3. The effective search rate is given by

$$EFSR=\sum_{i=1}^{\infty}\sum_{J=0}^{S}p\mu\pi(i,1,j)$$

4. The expected inventory level is given by

$$E(\omega)= \sum_{i=0}^{\infty}\sum_{j=0}^{S}j\{\pi(i,0,j)+\pi(i,1,j)\}$$

5. The expected number of successful retrials is given by

$$E(s\tau)= \sum_{i=0}^{\infty}\sum_{j=0}^{S}i\theta\pi(i,0,j)$$

6. The expected replenishment rate is given by

$$EFRR= \sum_{i=0}^{\infty}\sum_{j=0}^{s}\eta\pi(i,0,j)+\eta\pi(i,1,j)$$

7. The probability that inventory level is zero is given by

P(L=0)= $\sum_{i=0}^{\infty} \pi(i,0,0)+\pi(i,1,0)$

8. The probability that inventory level is greater than s is given by

P(L>s)= $\sum_{i=0}^{\infty} \sum_{j=s+1}^{S} \pi(i,0,j)+\pi(i,1,j)$

2.5 Numerical Illustration

In this section we provide numerical illustration of the system performance as underlying parameters vary.

2.5.1 System behavior as different parameters vary

Effect of search probability on various performance measures

Tables 1 and 2 show that the search probability has only a narrow effect on server busy probability; even that, we suspect may be due to approximation errors. We are not yet able to find an analytic expression for the probability of server busy, though we strongly believe that it will be independent of the search probability p .The behavior of measures like the expected number of customers in the orbit, and effective search rate as p increases, is as expected; where, as the first measure decreases, the second one increases. Tables 3 and 4 show that expected waiting time as well as the variance decreases with increase in p ; which are clear indicators of the fact that search mechanism increases the performance of the system. Another interesting observation that we can get from Tables 1 and 2 is that the expected rate of successful retrials, E(sτ) decrease with increase in p. This shows that introduction of search increases the number of unsatisfied retrials which may lead us to wonder whether to increase the search probability. All these phenomenon can be visualized from figures 1,2,3 and 4.

Effect of replenishment rate η on various performance measures

Table 5 shows that an increase in the parameter η makes an increase in measures like server busy probability, effective search rate, rate of successful retrials and expected inventory level; whereas the expected number of customer in the orbit decreases. From Table 6 one can see that as η increases, initially there is a comparatively high decrease in expected waiting time and variance which seems to be stabilizing as η increases further.

Effect of the service rate μ on various performance measures

From Table 7, we observe that as the service rate increases all the measures like server busy probability, effective search rate, effective rate of successful retrials and expected number of customers, decreases. Where the decrease in the effective search rate can be attributed to the decrease in server busy probability. The decrease in effective rate of successful retrials, in spite of an increase in server idle probability, may be due to the simultaneous drop in expected number of orbital customers. Table 8 shows the decrease in mean and variance of waiting time, as μ increases. The comparatively heavy drop in the variance as the service rate changes from 1.2 to 1.6 (arrival rate being 1), shows a more stabilized system can be achieved increasing the service rate a bit.

Effect of the reorder level s on various performance measures

Table 9 shows the increase in s with other parameters fixed, makes the effective replenishment rate to increase, which is expected because as s increases, more orders will be placed. Same is the reason behind the increase in expected inventory in the system. From the table it is also evident that reorder level does not too much vary other performance measures.

Effect of the maximum inventory level S on various performance measures

Table 10 shows the behavior of system performance measures with increase in S is same as that with increase in s; except for the measure effective replenishment rate, which decreases. This is because of the delay, caused by increase in S, in placing a new order.

Effect of retrial rate on various performance measures

Table 11 shows the behavior of system performance measures with increase in θ. As the retrial rate increases the expected number of successful retrials also increases. Hence the expected number of customers in the system decreases and so the effective search rate also decreases as expected. The other performance measures are not much affected by the retrial rate. .

2.6 COST ANALYSIS

For finding an optimal value for p and other parameters, we introduce a cost function C = CRP*EFRR + CN* $E(\sigma)$ + CI* $E(\omega)$ + CSR*EFSR +

CIDL* (1-P(ß)), where CRP is the cost of inventory procurement, CN is the cost of holding customers, CI is the cost of holding inventory, CSR is the search cost and CIDL is the cost per unit time due to an idle server. For various values of the parameters we saw that some of the performance measures increases while the others decrease. As an example as p increases E(σ) decreases, EFSR increases whereas other performance measures doesn't have any significant change. Hence we were able to get a concave shape for the cost curve. The problem of optimizing the cost for various parameter values was carried out. Few illustrations are given below.

Figures 5,6,7 and 8 show an optimum value in terms of the cost function C, for the parameters p, S, s and μ respectively. Here we wish to point out that these optimum values may depend on the particular costs taken.

Table 2.1. Effect of p on the various performance measures

λ=1, μ=1.5, θ=4, η=0.1, s=5, S=15

p	P(ß)	P(L>s)	EFRR	E(ω)	E(σ)	E(sτ)	EFSR
0.1	0.41156	0.40169	0.06173	4.13544	2.04305	0.36878	0.04278
0.2	0.41151	0.40171	0.06173	4.13638	1.99077	0.32684	0.08467
0.3	0.41145	0.40173	0.06172	4.13731	1.93959	0.28578	0.12567
0.4	0.4114	0.40174	0.06171	4.13822	1.8895	0.24561	0.16579
0.5	0.41135	0.40176	0.0617	4.13912	1.8405	0.20632	0.20503
0.6	0.4113	0.40178	0.0617	4.14001	1.79257	0.16789	0.24341
0.7	0.41126	0.40179	0.06169	4.14089	1.74571	0.13032	0.28094
0.8	0.41121	0.40181	0.06168	4.14175	1.69999	0.0936	0.31761
0.9	0.41117	0.40182	0.06167	4.14259	1.65513	0.05771	0.35349

Table 2.2 Effect of p on the various performance measures

λ=1, μ=1.5, θ=4, η=1, s=5, S=15

p	P(ß)	P(L>s)	EFRR	EFSR	E(σ)	E(sτ)	E(ω)
0.1	0.66481	0.91959	0.09972	0.07344	1.78043	0.59138	9.5014
0.2	0.66479	0.91928	0.09972	0.14524	1.72639	0.51954	9.50143
0.3	0.66476	0.91898	0.09971	0.21544	1.67356	0.44933	9.50147
0.4	0.66474	0.91867	0.09971	0.28402	1.62193	0.38072	9.5015
0.5	0.66471	0.91837	0.09971	0.35102	1.5715	0.3137	9.50153
0.6	0.66469	0.91808	0.0997	0.41644	1.52225	0.24825	9.50157
0.7	0.66466	0.91778	0.0997	0.4803	1.47417	0.18436	9.5016
0.8	0.66464	0.91749	0.0997	0.54262	1.42724	0.12201	9.50164
0.9	0.66461	0.9172	0.09969	0.60343	1.38146	0.06118	9.50168

Table 2.3 Variation in waiting time with search probability p

$\lambda = 1$, $\eta = 0.1$, $\theta = 4$, μ =1.5, s =5, S =15

p	E(W_L)	E(σ)	$E(W_L^2)$	V(W_L)
0.2	1.9901	1.9907	38.9522	34.9914
0.4	1.8889	1.8895	36.3992	32.831
0.6	1.7921	1.7925	34.0213	30.8096
0.8	1.6995	1.6999	31.8082	28.9199
1	1.611	1.6113	29.7498	27.1543

Table 2.4. Variation in waiting time with search probability p

λ=1, μ=1.5, θ=4, η=0.1, s=8, S=15

p	$E(W_L)$	$E(\sigma)$	$E(W_L^2)$	$V(W_L)$
0.1	2.0642	2.0649	43.810	39.549
0.2	2.0121	2.0127	42.3525	38.3037
0.3	1.9611	1.9617	40.9446	37.0984
0.4	1.9112	1.9117	39.5850	35.9321
0.5	1.8624	1.8629	38.2724	34.8039
0.6	1.8146	1.815	37.0054	33.7126
0.7	1.7678	1.7683	35.7826	32.6573
0.8	1.7221	1.7225	34.6029	31.637
0.9	1.6774	1.6778	33.4647	30.6507
1	1.6338	1.6342	32.3668	29.6974

Table 2.5 Effect of η on various performance measures

λ=1, μ=1.5, θ=4, p= 0.1, s=5, S=15

η	P(ß)	P(L>s)	EFRR	EFSR	$E(\sigma)$	$E(s\tau)$	$E(\omega)$
0.1	0.4115	0.4016	0.6173	0.0427	2.043	0.3687	4.1354
0.3	0.6142	0.7185	0.0918	0.066	1.8378	0.5464	7.397
0.5	0.6501	0.8284	0.0975	0.0711	1.7972	0.579	8.5371
0.7	0.6604	0.8801	0.0990	0.0726	1.7856	0.5877	9.0801
0.9	0.6639	0.9093	0.0995	0.0732	1.7814	0.5906	9.3913
1.1	0.6653	0.9279	0.0998	0.0732	1.7797	0.5918	9.5917
1.3	0.6659	0.9405	0.0999	0.0735	1.7789	0.5923	9.731
1.5	0.663	0.9497	0.09994	0.0736	1.7785	0.5925	9.8334
1.7	0.6664	0.9566	0.09997	0.0737	1.7782	0.5926	9.9118
1.9	0.6665	0.9620	0.09998	0.0738	1.7781	0.5927	9.9736

Table 2.6. Effect of the replenishment rate η on the waiting time

λ=1, θ=4, μ=1.5, p=0.1, s=5, S=15

η	$E(W_L)$	E(σ)	$E(W_L^2)$	$V(W_L)$
0.1	2.0424	2.043	40.2976	36.1261
0.5	1.7967	1.7972	13.2469	10.0185
0.9	1.7809	1.7814	12.5669	9.395
1.3	1.7784	1.7789	12.4739	9.3112
1.7	1.7777	1.7782	12.4525	9.292

Table 2.7 Effect of service rate μ on various performance measures

μ	P(ß)	P(L>s)	EFRR	EFSR	E(σ)	E(sτ)	E(ω)
1.2	0.8308	0.91	0.0997	0.3462	4.898	0.4846	9.5014
1.6	0.6231	0.9207	0.0997	0.2676	1.2814	0.3555	9.5015
2	0.4985	0.9268	0.0997	0.217	0.6427	0.2814	9.5015
2.4	0.4154	0.9308	0.0997	0.1823	0.3991	0.233	9.5015
2.8	0.356	0.9336	0.0997	0.1571	0.2772	0.1989	9.5015
3.2	0.3115	0.9357	0.0997	0.138	0.2064	0.1735	9.5015

Table 2.8 Effect of the service rate μ on the waiting time

λ=1, θ=4, η=1, p=0.5, s=5, S=15

μ	$E(W_L)$	E(σ)	$E(W_L^2)$	$V(W_L)$
1.2	4.519	4.7729	72.649	52.2275
1.6	1.2393	1.2394	6.6186	5. 826
2.	0.6175	0.6175	1.8549	1.4735
2.4	0.3811	0.3811	0.7941	0.6488
2.8	0.2631	0.2631	0.4225	0.3533

Table 2.9. Effect of reorder level s on various performance measures

λ=1, η =1, θ = 4, p = 0.4, μ =1.5, S = 31

s	P(ß)	P(I>s)	EFRR	EFSR	E(σ)	E(sτ)	E(ω)
5	0.6659	0.9686	0.0384	0.285	1.6201	0.3809	17.509
7	0.6664	0.966	0.0416	0.2854	1.6193	0.3809	18.501
9	0.6666	0.9629	0.0454	0.2856	1.6191	0.381	19.5
11	0.6666	0.9592	0.05	0.2856	1.6191	0.381	20.499
13	0.6666	0.9546	0.0555	0.2856	1.619	0.381	21.499
15	0.6666	0.949	0.0625	0.2856	1.619	0.381	22.49

Table 2.10. Effect of Maximum inventory level S on various performance measures

λ =1, η = 1, θ =4, p = 0.4, μ =1.5, s =5

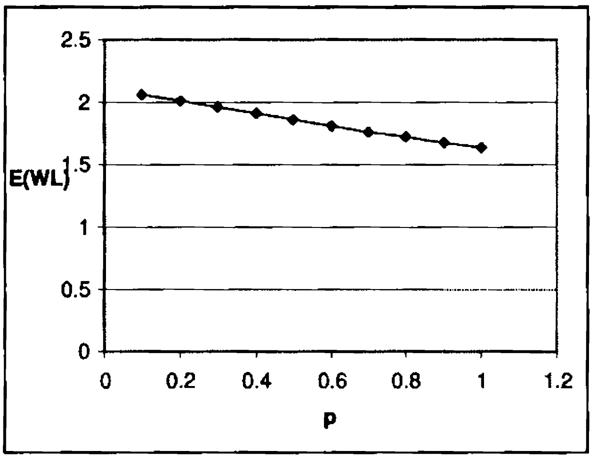
S	P(ß)	P(L>s)	EFRR	EFSR	E(σ)	E(sτ)	E(ω)
12	0.6639	0.8839	0.1422	0.2833	1.623	0.3805	7.9956
14	0.6645	0.9096	0.1107	0.2838	1.622	0.3806	9
16	0.6649	0.926	0.0906	0.2841	1.6216	0.3807	10.002
18	0.6651	0.9374	0.0767	0.2844	1.6212	0.3807	11.004
20	0.6653	0.9457	0.0665	0.2845	1.6209	0.3808	12.005
22	0.6655	0.9521	0.0587	0.2846	1.6207	0.3808	13.006

Table 2.11. Effect of Retrial rate on various performance measures

λ =1, η = 1, μ =4, p = 0.5, S=15, s =5

θ	P(ß)	P(L>s)	EFRR	EFSR	E(σ)	E(sτ)	E(ω)
1	0.24926	0.939	0.09971	0.15967	0.20402	0.0896	9.50174
1.2	0.24926	0.93897	0.09971	0.155	0.1887	0.0937	9.50171
1.4	0.24926	0.93893	0.09971	0.15221	0.17691	0.0970	905017
1.6	0.24926	0.93889	0.09971	0.14955	0.1675	0.0997	9.5018
1.8	0.24926	0.93886	0.09971	0.14736	0.1599	0.1019	9.5016
2	0.24926	0.93882	0.09971	0.14552	0.1535	0.1037	9.50166

Figure 2.1. Search probability versus Expected Waiting tii



•

Figure 2.2. Search probability versus effective search rate

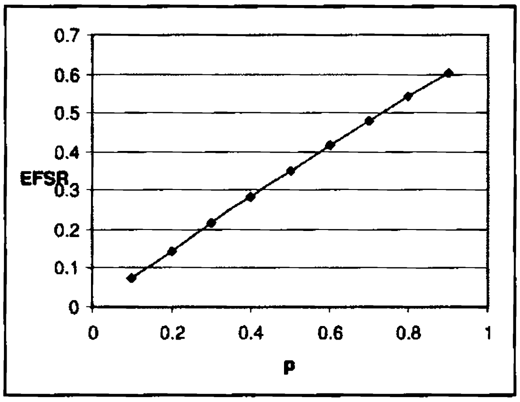


Figure 2.3 Search probability verses expected number of successful retrials

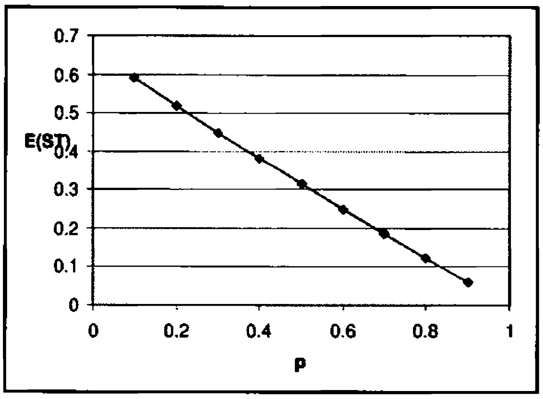


Figure 2.4 Search probability verses expected number of customers

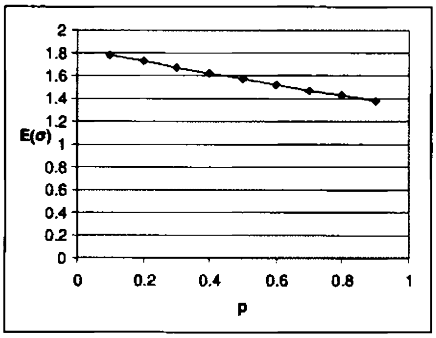


Figure 2.5. Effect of search probability on cost function C
CRP = 1500, CN = 100, CI = 250, CSR = 400, CIDL = 850
λ=1, μ=2, θ=0.5, η=1, s=5, S=15

p	C
0.1	3201
0.2	3200.98
0.3	3200.95
0.4	3200.94
0.5	3200.93
0.6	3200.93

0.7	3200.94
0.8	3200.94
0.9	3200.96

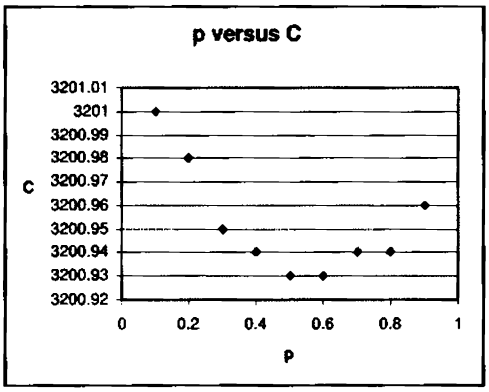


Figure 2.6. Effect of S on cost function C
CRP =1500, CN=100, CI=15, CSR=400, CIDL=850.
λ =1, μ=2, θ =1.5, η = 0.9, p = 0.6, s = 5

S	Cost
15	897.56
16	891.43
17	887.57
18	885.45
19	884.69
20	885.64
21	886.28
22	888.25
23	890.83
24	893.94
25	897.48

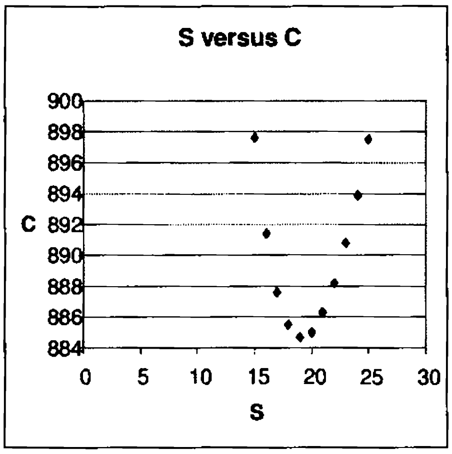


Figure 2.7 Effect of s on cost function C
CRP = 150, CN = 600, CI = 10, CSR = 100, CIDL = 150
λ = 6, μ =7, θ =1.5, η = 0.9, p = 0.4, S = 31

s	C
6	7730.93
7	7726.51
8	7723.4
9	7721.5
10	7720.71
11	7720.98
12	7722.243
13	7724.44
14	7727.56
15	7731.6
16	7736.39

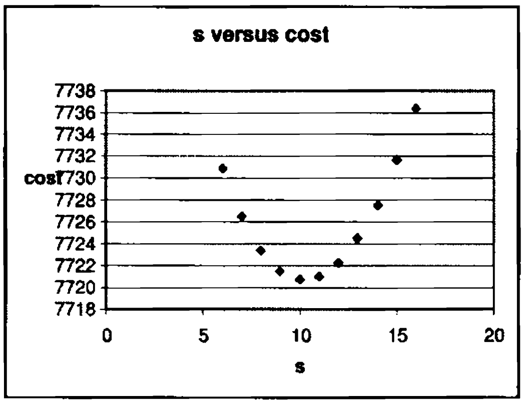
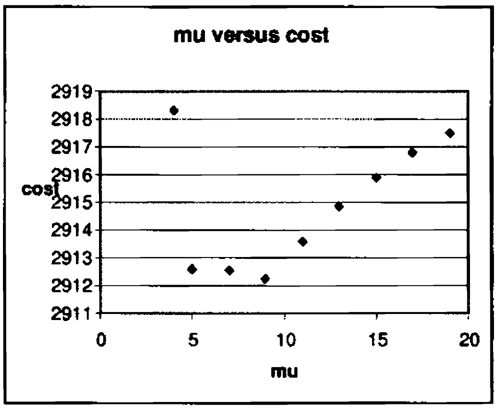


Figure 2.8 Effect of service rate μ on Cost function C
CRP = 500, CN = 400, CI = 250, CSR = 450, CIDL= 500
λ=1 θ=4 η=1 p=0.6 s=5 S=15

μ	COST
4	2918.3
5	2912.6
7	2912.5
9	2912.2
11	2913.6
13	2914.9
15	2915.9
17	2916.8
19	2917.5



CHAPTER 3

An Inventory Model with Server Interruptions And Positive Lead Time

3.1. Introduction

In the previous chapter we concentrated on retrial inventory with orbital search. Now we focus service and production interruption in inventory with positive service time. Service interruption models studied in the literature include different types of service unavailability that may be due to server taking vacations, server breakdown, server interruptions, arrival of a priority customer etc.

Unlike in Marie and Trivedi [55] in this chapter the interrupted service is assumed to resume on completion of interruption. The queueing model analyzed by Krishnamoorthy and Ushakumari [42], where disaster can occur to the unit undergoing service, the one by Wang et al [72] with disaster and unreliable server are also models with server interruptions.

A recent paper by Krishnamoorthy et. al. [37] studies queues with service interruption and repair, where a decision is to be made on whether to repeat or resume the interrupted service according to whether a phase type distributed random clock that starts ticking the moment interruption strikes, realizes after or before the removal of the current interruption. Another paper by Krishnamoorthy et. al. [29] studies a queuing model where no damage to the server is assumed due to interruption, so that the server there needs no repair. But after interruption, a decision is to be made whether to restart or resume the interrupted service, which requires a random time determined from the competition of two exponential random variables.

The first study on inventory systems where a processing time is required for serving the inventory is due to Berman et. al.[8], which was a deterministic model. Berman and Kim[9] and Berman and Sapna [10] are the first to discuss inventory with positive service time (exponential distribution) and that with arbitrary distribution, respectively. Whereas [9] is concerned with infinite waiting room case with zero lead time, [10] considers a finite waiting space only. Among these, the first one takes a dynamic programming approach and the second one takes a Markov renewal theoretic approach.

More recently, Krishnamoorthy and his co-authors used Matrix Analytic Methods to study a few inventory models [13,27,33,35, 39,43,44,45,69], where a service time for providing the inventoried item is assumed. In Krishnamoorthy et. al. [43], and Deepak et. al. [14], an explicit product form solution for an inventory system with service time could be arrived at due to the assumption of zero lead time. It is worth mentioning that Schwarz et. al. [58] could obtain product form solution for the joint distribution of the number of customers the system and the inventory level even in the case of positive lead time, at least with assumption that no customer joins the system when the inventory is out of stock.

There are numerous studies on inventory systems where interruption occurs due to an unreliable supplier. We refer to the papers by Tomlin [66, 67] and the paper by Chen and Li [12] for more details on such studies. Our work is in an entirely different direction from the above works, due to few reasons. First of all, in the above papers, interruption occurs due to an unreliable supplier, whereas in our model, we do not assume that the supplier is unreliable and it is the unreliable server who causes interruptions. Most importantly, in our model, interruptions occur in the middle of a service and there is no restriction on the number of possible interruptions during a service.

This chapter introduces the concept of interruption to an inventory system where the processing of inventory requires a random time, which leads to a queue of customers waiting for inventory. The arrival process is assumed to be Poisson and service time follows an exponential distribution. During the processing of inventory, the service may be interrupted due to breakdown of the server. The failure time of a busy server is assumed as exponentially distributed and the failed server is taken for repair immediately, where the repair time also follows an exponential distribution. Inventory is managed according to an (s,S) policy with zero lead time. In the papers by Nicola , Kulkarni and Trivedi [51] as well as Marie and Trivedi [55] different policies for the lost work upon service interruption namely namely. prs (preventive resume), pri (preventive repeat identical) and prd (preventive repeat different) have been considered. Since in our model the service times are exponentially distributed we need not consider these cases, other than in "repeat same", separately. Here we consider the case of repeat different only.

As in [43] and [14], the assumption of instantaneous replenishment leads to an explicit steady state analysis under the stability condition. The optimal values for reorder level s and maximum inventory level S is also analyzed based on a cost function.

This chapter is arranged as follows. In section 3.2, we provide the mathematical modeling of the above system; in section 3.3 we obtain the stability condition and steady state distribution is found. Section 3.4 is devoted to some performance measures like expected waiting time of a customer and their behavior is analyzed with variation of parameters in section 3.5. A cost function is also constructed in that section and its nature studied numerically.

3.2. Mathematical Model

The system is described as under. Customers arrive to a single server counter according to a Poisson process of rate λ where inventory is served. Service times are iid exponential random variables with parameter μ. Inventory is replenished according to (s,S) policy, the replenishment time being exponential random variable with parameter η. While the server serves a customer, the service may be interrupted, the interruption time being exponential random variable with parameter δ_1. Following a service interruption the service restarts according to an exponentially distributed time with parameter δ_2.

For the model under discussion, we make the following assumptions:

- No inventory is lost due to service interruption.
- The customer being served when interruption occurs waits there until the interruption is repaired .
- No arrival is entertained when the inventory level is zero.
- While the server is on an interruption, an order placed if any, is cancelled.

Let $N(t)$ be the number of the customers in the system including the one being served (if any), $L(t)$ be the inventory level and

$$S(t)=\begin{cases}0 & \text{if the server is idle}\\ 1 & \text{if the server is busy}\\ 2 & \text{if the server is on interruption}\end{cases}.$$

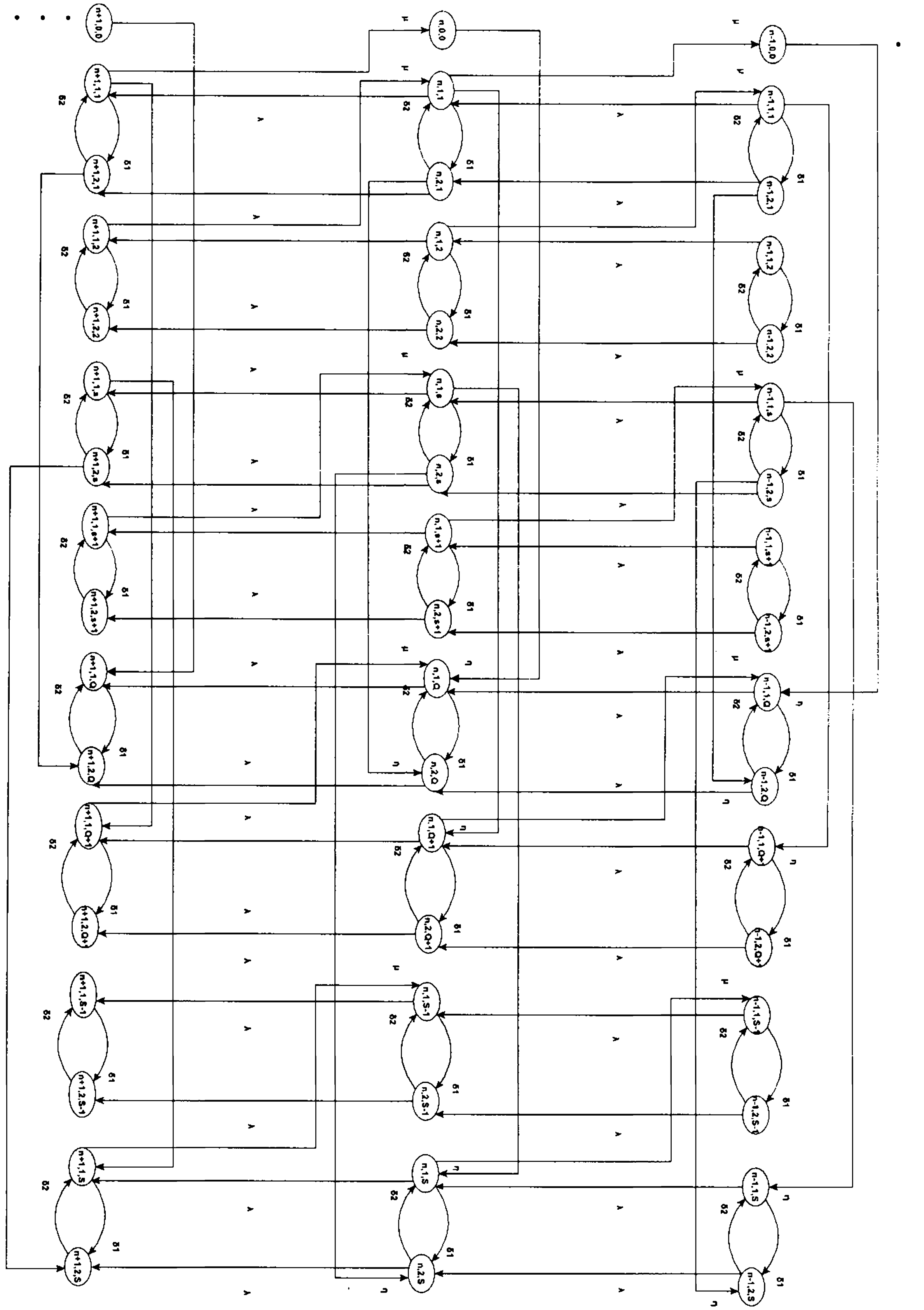
Then $\Omega=\{X(t), t\geq 0\}=\{(N(t),S(t),L(t)), t\geq 0\}$ is a Markov chain with state space $E=\{(0,0,k)\,|\,0\leq k\leq S\}\cup\{(i,0,0)\,|\,i\geq 1\}\cup\{(i,j,k)\,|\,i\geq 1, j=1,2, 1\leq k\leq S\}$.

The state space of the Markov chain is partitioned into levels $\tilde{i}$ defined as

$\tilde{0}=\{(0,0,0),(0,0,1),...,(0,0,S)\}$, and

$\tilde{i}=\{(i,0,0),(i,1,1),...,(i,1,S),(i,2,1),...,(i,2,S)\}$, for $i\geq 1$.

This makes the Markov chain under consideration, a level independent Quasi Birth Death (QBD) process. In the sequel, $S-s = Q$, I_n denotes identity matrix of order n, I denotes an identity matrix of appropriate order and e denotes a column vector of 1's of appropriate order.



STATE TRANSITION DIAGRAM

The infinitesimal generator of the process Ω is

$$H = \begin{bmatrix} B_0 & B_1 & 0 & 0 & 0 & 0 & 0 \\ B_2 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & & & \\ & & & & & & \\ & & & & & & . \end{bmatrix}$$

$$B_0 = \begin{bmatrix} C_0 & 0 & C_1 \\ 0 & C_2 & 0 \\ 0 & 0 & C_3 \end{bmatrix}, \text{ where } C_0 = \begin{bmatrix} -\eta & 0 \\ 0 & -(\lambda+\eta)I_s \end{bmatrix}, \ C_1 = \eta I_{(s+1)\times(s+1)}, \ C_2 = -\lambda I_{(S-2s-1)}$$

$C_3 = -\lambda I_{(s+1)}$.

$$B_1 = \begin{bmatrix} D_0 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & D_1 & 0 \end{bmatrix} \text{ where } D_1 = \lambda I_{(s+1)}, \ D_2 = \lambda I_{(S-2s-1)}, \ D_0 = \begin{bmatrix} 0 & 0 \\ 0 & \lambda I_s \end{bmatrix}_{(s+1)\times(s+1)}$$

$$B_2 = \begin{bmatrix} E_1 & 0 & 0 \\ E_3 & E_2 & 0 \\ 0 & E_4 & E_1 \\ 0 & 0 & 0 \end{bmatrix} \text{where}$$

$$E_1 = \begin{bmatrix} 0 & 0 \\ \mu I_s & 0 \end{bmatrix}_{(s+1)\times(s+1)}, \ E_2 = \begin{bmatrix} 0 & 0 \\ \mu I_{S-2s-2} & 0 \end{bmatrix}_{(S-2s-1)\times(S-2s-1)},$$

$$E_3 = \begin{bmatrix} 0 & \mu \\ 0 & 0 \end{bmatrix}_{(S-2s-1)\times(s+1)}, \ E_4 = \begin{bmatrix} 0 & \mu \\ 0 & 0 \end{bmatrix}_{(s+1)\times(S-2s-1)}.$$

$$A_1 = \begin{bmatrix} G_1 & 0 & G_2 & G_3 & 0 & 0 \\ 0 & G_4 & 0 & 0 & G_5 & 0 \\ 0 & 0 & G_6 & 0 & 0 & G_7 \\ G_8 & 0 & 0 & G_9 & 0 & G_{10} \\ 0 & G_{11} & 0 & 0 & G_{12} & 0 \\ 0 & 0 & G_{13} & 0 & 0 & G_{14} \end{bmatrix}, \text{ where } G_1 = \begin{bmatrix} -\eta & 0 \\ 0 & -(\eta+\lambda+\mu+\delta_1)I_s \end{bmatrix}$$

$G_2 = \eta I_{(s+1)}$, $G_3 = \begin{bmatrix} 0 \\ \delta_1 I_s \end{bmatrix}_{(s+1)\times s}$, $G_4 = -(\lambda+\mu+\delta_1)I_{(S-2s-1)}$, $G_5 = \delta_1 I_{(S-2s-1)}$,

$G_6 = -(\lambda+\mu+\delta_1)I_{(s+1)}$, $G_7 = \delta_1 I_{(s+1)}$, $G_8 = [0 \quad \delta_2 I_s]_{s\times(s+1)}$, $G_9 = -(\lambda+\eta+\delta_2)I_s$,

$G_{10} = [0 \quad \eta I_s]_{s\times(s+1)}$, $G_{11} = \delta_2 I_{(S-2s-1)}$, $G_{12} = -(\lambda+\delta_2)I_{(S-2s-1)}$, $G_{13} = \delta_2 I_{(s+1)}$,

$$G_{14} = -(\lambda+\delta_2)I_{(s+1)}, \quad A_0 = \begin{bmatrix} 0 & 0 \\ 0 & \lambda I_{2S} \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} E_1 & 0 & 0 & 0 \\ E_3 & E_2 & 0 & 0 \\ 0 & E_4 & E_1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

3.3 Analysis of the model

3.3.1 Stability condition

Define $A = A_0 + A_1 + A_2$ and let $\pi = (\pi(0,0), \pi(1,1), \pi(1,2), \ldots, \pi(1,S), \pi(2,1), \ldots, \pi(2,S))$ be the steady state vector of the generator matrix A. Then πA=0 gives the following equations.

$$
\begin{aligned}
&-\eta\pi(0,0)+\mu\pi(1,1)=0\\
&-(\eta+\mu+\delta_1)\pi(1,i)+\mu\pi(1,i+1)+\delta_2\pi(2,i)=0;1\le i\le s\\
&-(\mu+\delta_1)\pi(1,i)+\mu\pi(1,i+1)+\delta_2\pi(2,i)=0;s+1\le i\le Q-1\\
&\eta\pi(0,0)-(\mu+\delta_1)\pi(1,Q)+\mu\pi(1,Q+1)+\delta_2\pi(2,Q)=0\\
&\eta\pi(1,i)-(\mu+\delta_1)\pi(1,Q+i)+\mu\pi(1,Q+i+1)+\delta_2\pi(2,Q+i)=0;1\le i\le s-1\\
&\eta\pi(1,s)-(\mu+\delta_1)\pi(1,S)+\delta_2\pi(2,S)=0
\end{aligned}
\qquad (3.1.a)
$$

$$
\begin{aligned}
&\delta_1\pi(1,i)-(\eta+\delta_2)\pi(2,i)=0;1\le i\le s\\
&\delta_1\pi(1,i)-\delta_2\pi(2,i)=0;s+1\le i\le Q\\
&\delta_1\pi(1,Q+i)+\eta\pi(2,i)-\delta_2\pi(2,Q+i)=0;1\le i\le s
\end{aligned}
\qquad (3.1.b)
$$

Adding all the equations in (3.1.b) we get

$$\pi(2,1)+\pi(2,2)+....+\pi(2,S)=\frac{\delta_1}{\delta_2}\left(\pi(1,1)+\pi(1,2)+....+\pi(1,S)\right) \qquad (3.1.c)$$

By a known theorem, the QBD process with generator H is stable if and only if $\pi A_0 e<\pi A_2 e$ (see Neuts [56]).

That is, if and only if

$$\lambda\left(\pi(1,1)+...+\pi(1,S)+\pi(2,1)+...+\pi(2,S)\right)<\mu\left(\pi(1,1)+...+\pi(1,S)\right).$$

Applying 3.1.c, the above inequality reduces as

$$\lambda<\frac{\delta_2\mu}{(\delta_1+\delta_2)}.$$

Thus, we have the following theorem for stability of the system under study.

Theorem 3.1

The Markov chain Ω is stable if and only if $\lambda<\dfrac{\delta_2\mu}{(\delta_1+\delta_2)}$

3.3.2 Computation of steady state vector

We find the steady state vector of Ω numerically. Let $\pi = (\pi_0, \pi_1, \pi_2, ...)$ be the steady state vector, where π_0 is partitioned as $\pi_0 = (\pi_0(0,0), \pi_0(0,1), ..., \pi_0(0,S))$ and π_i 's are partitioned as $\pi_i = (\pi_i(0,0), \pi_i(1,1), \pi_i(1,2), ..., \pi_i(1,S), \pi_i(2,1), \pi_i(2,2), ..., \pi_i(2,S))$. Since Ω is a level independent QBD process, its steady state vector is given by $\pi_i = \pi_1 R^{i-1}$, $i \geq 1$ (see Neuts [56]), where R is the minimal non-negative solution of the matrix-quadratic equation $R^2 A_2 + RA_1 + A_0 = 0$. For finding the boundary vectors π_0 and π_1, we have from $\pi H = 0$,

$$\pi_0 B_1 + \pi_1 A_1 + \pi_2 A_2 = 0$$

ie. $\pi_0 B_1 + \pi_1 (A_1 + RA_2) = 0$

ie. $\pi_1 = -\pi_0 B_1 (A_1 + RA_2)^{-1}$

$= \pi_0 W$, where $W = -B_1 (A_1 + RA_2)^{-1}$.

Further, $\pi_0 B_0 + \pi_1 B_2 = 0$

ie. $\pi_0 (B_0 + WB_2) = 0$

First we take π_0 as the steady state vector of the generator matrix $B_0 + WB_2$. Then π_i, for $i \geq 1$ can be found using the formulae: $\pi_1 = \pi_0 W$, $\pi_i = \pi_1 R^{i-1}$, $i \geq 2$. Finally, the steady state probability distribution of the system under study is obtained by dividing each π_i with the normalizing constant

$\pi_0 e + (\pi_1 + \pi_2 + ...) e = \pi_0 \left(I + W(I - R)^{-1} \right) e$.

3.4 System Performance measures

3.4.1 Expected waiting time of a customer in the queue

For computing the expected waiting time in the queue of a tagged customer, who joins as r[th] customer in the queue, we consider the Markov process $\Psi = \left(\hat{N}(t), S(t), L(t)\right)$, where $\hat{N}(t)$ denotes the rank, which is the position of the customer in the queue; $S(t)$ and $L(t)$ have the same definition as in section 2. The state space of the Markov chain Ψ is given by $\hat{E} = \{(i,0,0), 1 \le i \le r-1\} \cup \{(i,j,k), 1 \le i \le r; j = 1,2; 1 \le k \le S\} \cup \Delta$,

where Δ is an absorbing state which corresponds to the tagged customer being taken for service. The infinitesimal generator matrix of the process Ψ is given by $\hat{Q} = \begin{bmatrix} T & T^0 \\ 0 & 0 \end{bmatrix}$, where T^0 is an $\left(r(2S+1)-1\right) \times 1$ matrix such that $T^0(i,1) = \mu$, for $2 \le i \le S+1$ and $T = \begin{bmatrix} B & 0 & 0 & . & & & 0 \\ A_2 & B & 0 & . & & & 0 \\ 0 & A_2 & B & 0 & & . & 0 \\ & & A_2 & B & & & \\ & & & . & . & . & \\ & & & & . & . & \\ & & & & & \hat{A}_2 & \hat{B} \end{bmatrix}$, where

$$B=\begin{bmatrix} B_1 & 0 & B_2 & B_3 & 0 & 0 \\ 0 & B_4 & 0 & 0 & B_5 & 0 \\ 0 & 0 & B_6 & 0 & 0 & B_7 \\ B_8 & 0 & 0 & B_9 & 0 & B_{10} \\ 0 & B_{11} & 0 & 0 & B_{12} & 0 \\ 0 & 0 & B_{13} & 0 & 0 & B_{14} \end{bmatrix}; \qquad B_1=\begin{bmatrix} -\eta & 0 \\ 0 & -(\eta+\mu+\delta_1)I_s \end{bmatrix}, \; B_2=\eta I_{s+1},$$

$$B_3=\begin{bmatrix} 0 \\ \delta_1 I_s \end{bmatrix}_{(s+1)\times s},$$

$B_4=-(\mu+\delta_1)I_{S-2s-1}$ $B_5=\delta_1 I_{S-2s-1}$, $B_6=-(\mu+\delta_1)I_{s+1}$,

$B_7=\delta_1 I_{s+1}; B_{11}=\delta_2 I_{S-2s-1}$, $B_{13}=\delta_2 I_{s+1}$, $B_{12}=-\delta_2 I_{S-2s-1}$, $B_{14}=-\delta_2 I_{s+1}$,

$B_9=-(\eta+\delta_2)I_s$, $B_8=\begin{bmatrix} 0 & \delta_2 I_s \end{bmatrix}_{s\times(s+1)}$, $B_{10}=\begin{bmatrix} 0 & \eta I_s \end{bmatrix}_{s\times(s+1)}$,

$\hat{B}(i,j)=B(i+1,j+1);\ 1\le i,j\le 2S$,

$\hat{A}_2(i,j)=A_2(i+1,j);\ 1\le i\le 2S,\ 1\le j\le 2S+1$.

Now, the waiting time W^r of the tagged customer who joins as the r^th^ customer in the queue, which is the time until absorption in the Makov process Ψ is given by the column vector $W^r=\hat{I}_{2S}(-T^{-1})e$, where $\hat{I}_{2S}=\begin{bmatrix} 0 & I_{2S} \end{bmatrix}_{(2S)\times(r(2S+1)-1)}$. Hence, the expected waiting time of a general customer is given by

$$E(\mathrm{W_L})=\sum_{r=1}^{\infty}\hat{\pi}_r W^r,$$

where $\hat{\pi}_r$ is a $1\times 2S$ dimensional row vector defined by $\hat{\pi}_r(i)=\pi_r(i+1)$, $1\le i\le 2S$. In a similar manner, we can find the second moment of the waiting time of an orbital customer as

$$E(\mathrm{W_L}^2)=\sum_{r=1}^{\infty}\hat{\pi}_r W_2^r \text{ , where } W_2^r=2\hat{I}_{2S}(T^{-2})e \text{ (see Neuts [56]).}$$

The variance of the waiting time of a customer is given by $V(W_L) = E(W_L^2)-(E(W_L))^2$.

3.4.2 Other performance measures

1. The probability that server is busy is given by $P_\beta = \sum_{i=1}^{\infty}\sum_{j=1}^{S}\pi_i(1,j)$.

2. The probability that server is on interruption is given by $P_\alpha = \sum_{i=1}^{\infty}\sum_{j=1}^{S}\pi_i(2,j)$.

3. The probability that server is idle is given by $P_\gamma = 1 - P_\alpha - P_\beta$.

4. The expected inventory level is given by

$$EIL = \sum_{j=1}^{S} j\pi_0(0,j) + \sum_{i=1}^{\infty}\sum_{j=1}^{S} j\{\pi_i(1,j)+\pi_i(2,j)\}.$$

5. The expected number of customers in the system is given by

$$EN = \sum_{i=1}^{\infty}\sum_{j=1}^{S}\{i\pi_i(1,j)+i\pi_i(2,j)\} + \sum_{i=1}^{\infty} i\pi_i(0,0)$$

6. The expected rate of ordering is given by $E_{OR} = \sum_{i=1}^{\infty}\mu\pi_i(1,s+1)$

7. The expected interruption rate is given by $E_{INTR} = \delta_1\sum_{i=1}^{\infty}\sum_{j=1}^{S}\pi_i(1,j)$.

8. The expected replenishment rate is given by

$$EFRR = \sum_{i=0}^{\infty}\sum_{j=0}^{s}\eta\pi(i,0,j)+\eta\pi(i,1,j)+\eta\pi(i,2,j).$$

9. Loss rate due to no item in the inventory is given by $LZI = \lambda\sum_{i=0}^{\infty}\pi_i(0,0)$

3.5 Numerical Illustration

In this section, we provide numerical illustration of the system performance by studying the effect of different parameters on the system performance measures.

3.5.1 Effect of the Interruption Rate δ_1

In Table 1, we see that as the interruption rate δ_1 increases, there is a considerable increase in the expected number of customers, which is expected since the overall service rate decreases with increase in δ_1. At the same time, note the high expected inventory level in the system and hence the low loss rate of customers from the system. As we expect an increase in the queue length as a product of low service completion rate, we also expect an increase in the loss rate also. Since the loss rate itself is low, in the table, the increase in loss rate is also narrow. Similarly, the increase in the expected inventory level is also very small. Note that the server idle probability $1-P_\alpha-P_\beta$ is decreasing considerably with increase in δ_1 and is equal to 0.09, when δ_1 is 3.2. However, the fraction of time the server remains active is also decreasing, which shows how severe is the effect of interruption. The narrow decrease in the server busy probability can be attributed to a similar increase in the loss rate of customers. This narrow decrease in P_β and the high values for the expected number of customers shows that the queue of unsatisfied customers is building up rapidly with increasing interruption rate. Table 2 shows an increase in the waiting time of a customer in the queue with an increase in the interruption rate; but this is as expected in lights of discussion in the above paragraph. Also, notice the high variance, which is increasing with δ_1. All these indicate the severe effect of interruption on the waiting time.

Table 1: Effect of the interruption rate δ_1 on various performance measures

λ=2, μ=5, η=1, δ_2=2.5, s=10, S=31

δ_1	P_β	P_α	*EIL*	*EFRR*	*EN*	*LZI*
2	0.3991	0.3193	19.0046	0.0714	3.4854	0.00463523
2.2	0.3990	0.3512	19.0048	0.0688	4.1674	0.00478141
2.4	0.3990	0.3831	19.0049	0.0662	5.0515	0.00492933
2.6	0.3990	0.4149	19.0051	0.0635	6.2431	0.00507770
2.8	0.3990	0.4468	19.0052	0.0608	7.9364	0.00522505
3	0.3989	0.4787	19.0054	0.0580	10.5327	0.00536621
3.2	0.3989	0.5106	19.0055	0.0552	15.0173	0.00546784

Table 2: Effect of the interruption rate δ_1 on waiting time of a customer

λ=2, μ=5, η=1, δ_2=2.5, s=5, S=15

δ_1	$E(W_L)$	$V(W_L)$	*EN*
2	1.9495	9.5718	3.4839
2.2	2.3327	13.2140	4.1658
2.4	2.7952	18.1585	5.0498
2.6	3.3320	24.6826	6.2413
2.8	3.9179	32.8502	7.9345
3	4.4746	42.0954	10.5309

3.5.2 Effect of the Repair Rate δ_2

As the repair rate increases, we expect an increase in the service completion rate and hence a decrease in the expected number of customers in the system, which can be viewed in Table 3. The decrease in the fraction of time the server remains interrupted is also as expected. Note that in the Table when δ_2 is less than or equal to 2.4, this fraction is larger than the fraction of time the server is active and this is reversed as δ_2 increases. The high values for the expected inventory level can be seen to be the reason behind the low values for the loss rate. Since the loss rate itself is small, so is its decrease with increase in δ_2. Since the expected inventory level is decreasing, the effective replenishment rate can be seen to be increasing.

Table 4 studies the effect of the repair rate δ_2 on the waiting time of a customer in the queue and it shows that the expected waiting time decreases with increase in δ_2. Tables 3 and 4 suggest that increasing the repair rate can reduce the severe effect of interruption; but how far one can do this may depend on the particular situation to handle.

Table 3: Effect of the repair rate δ_2 on various performance measures

λ=2, μ =5, η=1, δ_1=2.5, s =10, S =31

δ_2	P_β	P_α	*EIL*	*EFRR*	*EN*	*LZI*
2	0.3988	0.4985	19.0060	0.0578	13.9993	0.00595635
2.2	0.3989	0.4533	19.0055	0.0610	8.7155	0.00551963
2.4	0.3990	0.4156	19.0051	0.0637	6.3481	0.00515544
2.6	0.3990	0.3837	19.0049	0.0659	5.0161	0.00486830
2.8	0.3991	0.3563	19.0046	0.0678	4.1677	0.00463955
3	0.3991	0.3326	19.0044	0.0695	3.5830	0.00445517
3.2	0.3991	0.3118	19.0043	0.0710	3.1573	0.00417948

Table 4: Effect of the repair rate δ_2 on waiting time of a customer

λ=2, μ =5, η=1, δ_1=2.5, s =5, S =15

δ_2	$E(W_L)$	$V(W_L)$	EN
2	4.7555	49.1395	13.9986
2.2	4.0965	35.9026	8.7142
2.4	3.3662	25.1951	6.3464
2.6	2.7840	17.9824	5.0143
2.8	2.3474	13.2485	4.1658
3	2.0206	10.0948	3.5811

3.5.3 Effect of the Re-Order Level s

Table 5 shows that as the re-order level s increases the expected inventory level in the system also increases. Since the orders are being placed early with an increase in s, this is expected. The increase in the effective replenishment rate *EFRR* and the effective re-order rate E_{OR} is also obvious. The expected number of customers seems not much affected by the increases in the re-order level and the reason could be the assumption of disallowing customers to join the system with zero inventories. Since the inventory level is increasing, the loss rate can be seen to be decreasing. The narrow increase in the expected number of customers may be due to the decrease in the loss rate. Table 5 also shows a narrow increase in the server active as well as the server interrupted probabilities, which can be attributed to the possible increase in the number of services due to increase in the expected inventory level.

Table 6 shows a decrease in the expected waiting time of a customer in the queue with an increase in s. For an explanation note that when the re-order level increases, with the maximum inventory level being fixed, the time between two order placements decreases. Hence it becomes less probable that a customer encounters shortage of inventory while waiting, even if he joins the system when the inventory is in lower levels, which ultimately leads to a decrease in his/her waiting time. The decrease in the waiting time variance with increase in s is also in favor of the system performance.

Table 5: Effect of the re-order level s on various performance measures

λ=2, μ =5, η=1, δ_1=2, δ_2=2.5, S =31

s	P_β	P_α	*EIL*	*EFRR*	*EN*	E_{OR}	*LZI*
5	0.3952	0.3162	16.6131	0.0597	3.484965	0.0760	0.023812
6	0.3966	0.3173	17.0686	0.0613	3.485058	0.0793	0.017141
7	0.3975	0.3180	17.5401	0.0633	3.485158	0.0828	0.012342
8	0.3982	0.3186	18.0222	0.0657	3.485260	0.0866	0.008893
9	0.3987	0.3190	18.5112	0.0684	3.485344	0.0906	0.006415
10	0.3991	0.3193	19.0046	0.0714	3.485411	0.0950	0.004635
11	0.3993	0.3195	19.5008	0.0749	3.485470	0.0998	0.003354

Table 6: Effect of the re-order level *s* on waiting time of a customer

λ=2, μ =5, η=1, δ_1=2, δ_2=2.5, S=20

s	$E(W_L)$	$V(W_L)$	EN
4	1.9513	9.5950	3.4844
5	1.9495	9.4846	3.4845
6	1.9475	9.3980	3.4846
7	1.9456	9.3292	3.4847
8	1.9584	9.4904	3.4849

3.5.4 Effect of the Maximum Inventory Level *S*

In Table 7, one see that as the maximum inventory level S increases, there is an increase in the expected inventory level but the effective replenishment rate *EFRR* and the effective re-order rate E_{OR} decreases. Hence, the performance of the measures *EFRR* and E_{OR} with an increase in the maximum inventory level S is just reverse to that with an increase in the re-order level s. This is because of the increase in the difference S-s and hence a possible reduction in the number of order placing, with an increase in S. As in the case of the re-order level s, here also the increase in the inventory level leads to a decrease in the loss rate *LZI* and to a related increase, though narrow, in the expected number of customers. As in the case of the re-order level s, the server active as well as the server-interrupted fractions is increasing with increase in S.

Table 8 shows that unlike in the case of the re-order level s, there is an increase in the waiting time of customers with an increase in S. Here note that when the maximum inventory level increases, with the re-order level being fixed, the time between two order placements increases. Hence, it becomes more probable that a customer, who joins the system while the inventory is in lower levels, encounter a shortage of inventory while waiting in the queue, which leads to an increase in his/her waiting time. Though there is an increase in the waiting time, the table shows that the margin of increase is narrow. This, together with the decrease in the variance suggests that an increase in the maximum inventory level favors system performance.

Table 7: Effect of the maximum inventory level S on various performance measures

$\lambda=2,\ \mu=5,\ \eta=1,\ \delta_1=2,\ \delta_2=2.5,\ s=10$

S	P_β	P_α	*EIL*	*EFRR*	*EN*	E_{OR}	*LZI*
23	0.3985	0.3188	14.9925	0.1152	3.485248	0.1533	0.007475
24	0.3986	0.3189	15.4948	0.1070	3.485279	0.1424	0.006943
25	0.3987	0.3190	15.9967	0.0999	3.485311	0.1329	0.006482
26	0.3988	0.3190	16.4985	0.0937	3.485334	0.1246	0.006078
27	0.3989	0.3191	17.0000	0.0882	3.485347	0.1173	0.005722
28	0.3989	0.3191	17.5013	0.0833	3.485369	0.1108	0.005405
29	0.3990	0.3192	18.0025	0.0789	3.485385	0.1050	0.005121

Table 8: Effect of the maximum inventory level S on waiting time of a customer

$\lambda=2,\ \mu=5,\ \eta=1,\ \delta_1=2,\ \delta_2=2.5,\ s=5$

S	$E(W_L)$	$V(W_L)$	EN
10	1.9562	10.1594	3.4821
11	1.9564	10.0265	3.4827
12	1.9572	9.93840	3.4831
13	1.9581	9.87650	3.4834
14	1.9591	9.83110	3.4836
15	1.9601	9.7970	3.4839
20	1.9645	9.7103	3.4845

3.5.5 Effect of the Replenishment Rate η

Table 9 shows that the parameter that affects the expected inventory level besides the re-order level s and the maximum inventory level S is the replenishment rate η. As η increases, the expected inventory level and the effective replenishment rate, both increases. As the expected inventory level increases, the loss rate decreases; but because of the high-expected inventory level in the system, the loss rate is narrow. The increase in the server busy as well as the server interruption probabilities and the narrow increase in the expected number of customers, with increase in η, has the same reasoning as in the case of increase in the re-order level s.

Table 9: Effect of the replenishment rate η on various performance measures

$\lambda=2$, $\mu=5$, $\delta_1=2$, $\delta_2=2.5$, $s=10$, $S=31$

η	P_β	P_α	*EIL*	*EFRR*	*EN*	*LZI*
1.0	0.39907	0.31926	19.0	0.0714	3.4854	0.004635
1.2	0.39954	0.31964	19.3	0.0721	3.4856	0.002277
1.5	0.39982	0.31986	19.7	0.0732	3.4857	0.000895
2.0	0.39995	0.31996	20.0	0.0748	3.4857	0.000242
2.5	0.39999	0.31999	20.2	0.0762	3.4857	0.000082
3.0	0.39999	0.31999	20.3	0.0775	3.4857	0.000032
6.0	0.40000	0.32000	20.7	0.0825	3.4857	0.0000008
10.0	0.40000	0.32000	20.8	0.0860	3.4857	0.0000002

3.6 Cost Analysis

In subsections 5.3 and 5.4, we noted that the increase in the re-order as well as the maximum inventory levels is in favor of the system performance. Now, for checking the existence of optimal value for the levels s and S, we introduce a cost function

$$\Phi_{cost} = CI \times EIL + CN \times EN + CR \times E_{INTR} + (K + (S - s)\ K_1) \times E_{OR} + CL \times LZI\ ,$$

where CI is the cost of holding inventory, CN is the cost of holding customers, CR is the cost incurred due to interrupted service, K is the fixed cost of ordering , K_1 is the cost of a single inventory and CL is the cost incurred due to the loss of customers when the inventory level falls to zero.

3.6.1 Optimality of the Maximum Inventory Level *S*

Among the measures involved in the cost function, the effective re-order rate E_{OR} and the loss rate LZI are decreasing and all other measures are increasing. Hence, we expect a concave shape for the cost function from which we can get the optimal value for the maximum inventory level S. However, Figure 1 (a) shows a linear cost function, which indicates that the nature of the cost function depends on the costs involved. Figure 1 (b) shows that if the unit time cost CL incurred due to loss of customers is increased so as to catch the decrease in the loss rate LZI, we can get a concave cost function, which gives S=12 as the optimal value for the maximum inventory level S.

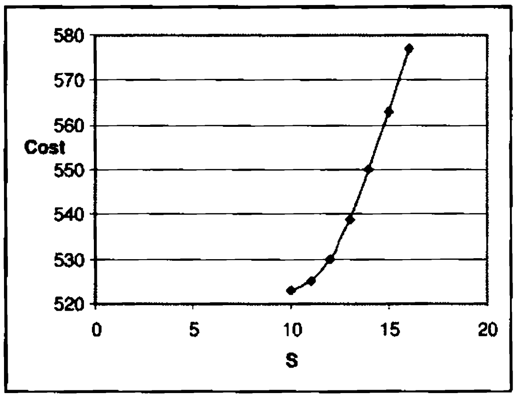
Figure 1: Effect of maximum inventory level S on the Cost function.

CI=40, CN=30, CR=75, K=400, K_I=35;

λ=2, μ=10, δ_1=2, δ_2=2.5, η=3, s =4

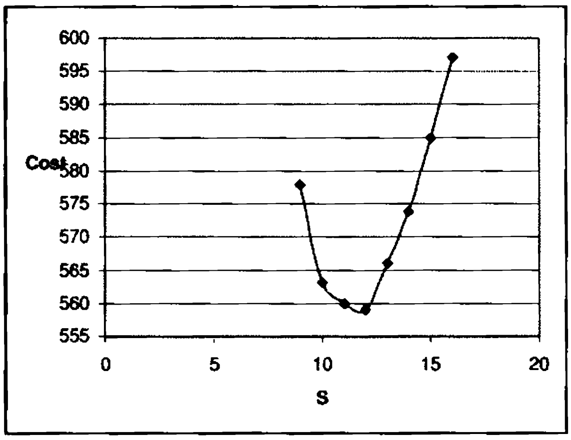
CL=100

S	Cost
10	523
11	525
12	530
13	539
14	550
15	563
16	577



(b) CL=4600

S	Cost
9	578
10	563
11	560
12	**559**
13	566
14	574
15	585
16	597



3.6.2 Optimality of the Re-Order Level *s*

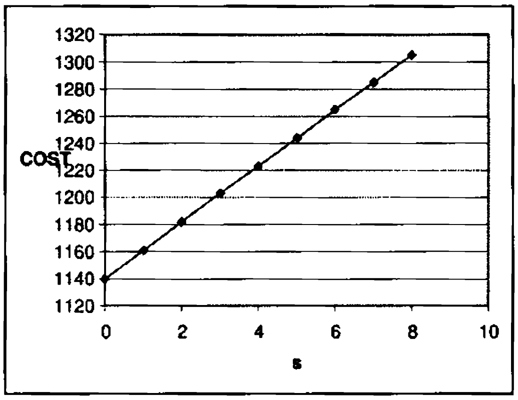
In the case of increase in the re-order level *s*, loss rate is the only decreasing measure, which is involved in the cost function. Hence, the cost function will be linear unless we select the cost *CL* to capture the decrease in the loss rate. Figure 2(a) shows a linear cost function, whereas Figure 2 (b) shows that by increasing the cost *CL* from 100 to 7500, we get a concave cost function, which gives 4 as the optimal value for the re-order level *s*.

Figure 2: Effect of the reorder level *s* on the Cost function.

CI=40, *CN*=30, *CR*=75, *K*=500, K_l=35; λ=2, μ=10, δ_1=2, δ_2=2.5, η=3, *S*=20

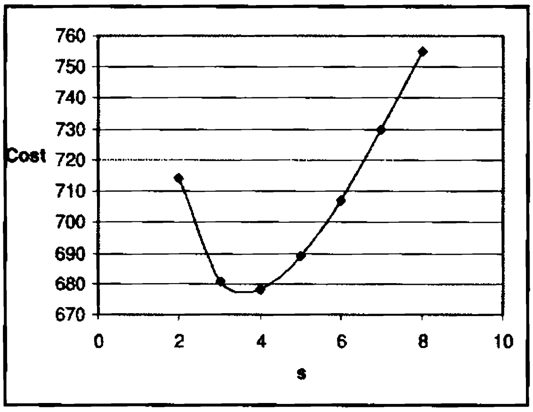
(a) CL=100

s	Cost
2	1182
3	1203
4	1223
5	1244
6	1265
7	1285
8	1305



(b) CL=7500

s	Cost
2	714
3	681
4	**678**
5	689
6	707
7	730
8	755



CHAPTER 4

An Inventory Model with Server Interruptions*

4.1. Introduction

The results in this chapter turn out to be a particular case of what was discussed in chapter3. Nevertheless the sharper assumption of zero lead time has produced several stronger results in this chapter. One consequence is the explicit expression for the stability of the system and the closed form expression for the system state distribution. Further our investigation of the optimal reorder level (s) and the maximum number(S) of items that could be stored in the inventory could be made analytical.

This chapter introduces the concept of interruption to an inventory system where the processing of inventory requires a random time, which leads to a queue of customers waiting for inventory. The arrival process is assumed to be Poisson and service time follows an exponential distribution. During the processing of inventory, the service may be interrupted due to breakdown of the server. The failure time of a busy server is assumed as exponentially distributed and the failed server is taken for repair immediately, where the repair time also follows an exponential distribution. Inventory is managed according to an (s,S) policy with zero lead time. As in [43] and [14], the assumption of instantaneous replenishment leads to an explicit steady state analysis under the stability

* The results in this chapter was presented as a paper in the 5th International Conference on Queueing Theory and Network Applications; July 24-26, 2010, Beijing, China . It is also published in the ACM Digital Library, Proceedings of the 5th International Conference on Queueing Theory and Network Applications ,Pages 132-139,doi>10.1145/1837856.1837876

condition. The optimal values for reorder level s and maximum inventory level S is also analyzed based on a cost function. This chapter is arranged as follows. In section 4.2, we do the mathematical modeling of the above system; in section 4.3 we obtain the stability condition and the explicit steady state probability vector under stability. Explicit expressions for several important performance measures are obtained in section 4.4 and their behavior, as different parameters vary, is discussed in section 4.5. A cost function is also constructed in that section and its nature studied numerically.

4.2 Mathematical Model

The system is described as under. Customers arrive to a single server counter according to a Poisson process of rate λ where inventory is served. Duration of service are iid exponential random variables with parameter μ. Inventory is replenished according to (s,S) policy, the replenishment being instantaneous. Further no shortage is permitted. While the server serves a customer, the service can be interrupted, the inter occurrence time of interruption being exponentially distributed with parameter δ_1. Following a service interruption the service restarts according to an exponentially distributed time with parameter δ_2.

For the model under discussion, we make the following assumptions:

1. No inventory is lost due to service interruption.
2. The customer being served when interruption occurs, waits there until his service is completed.

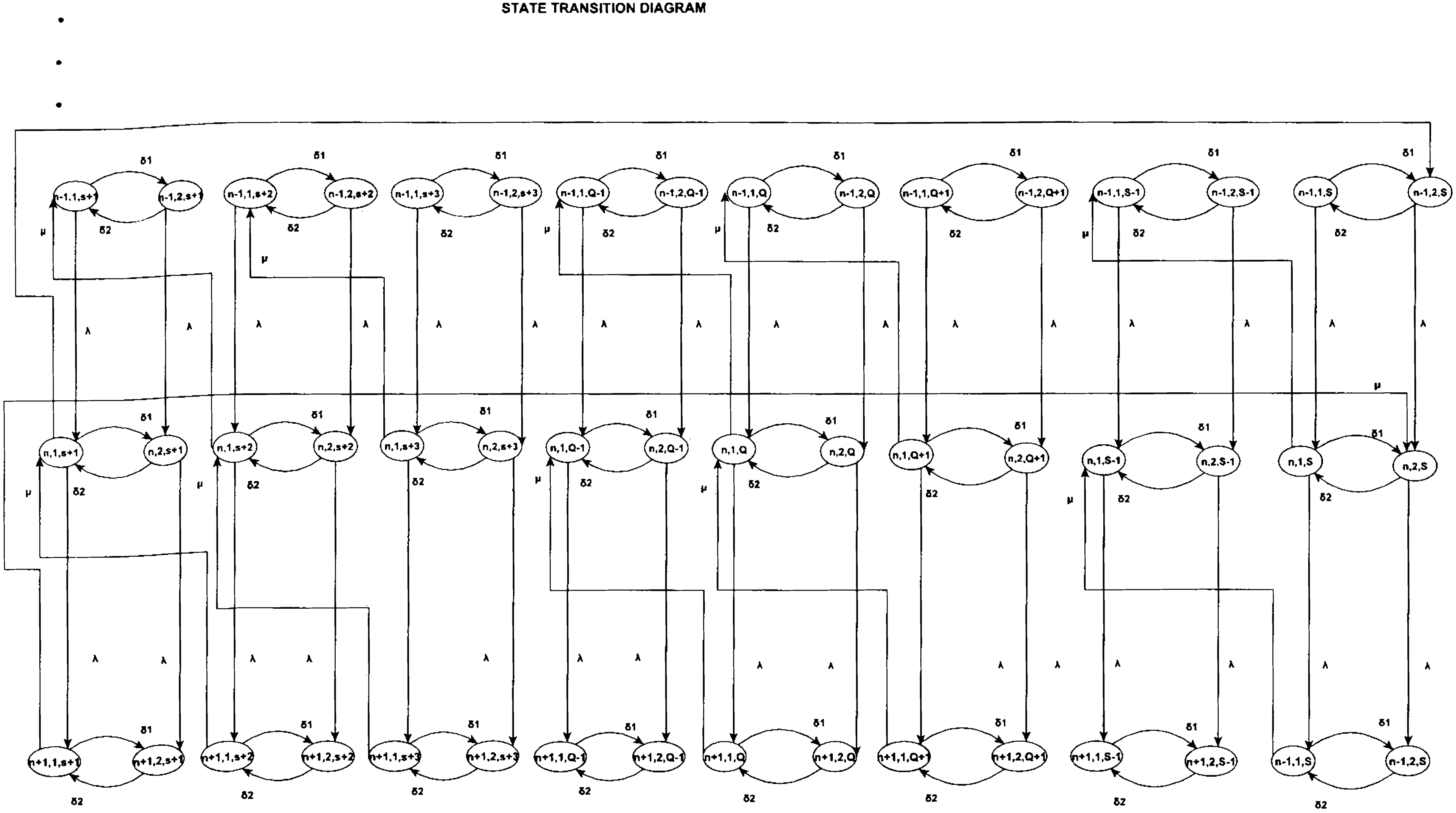
At time t let N(t) be the number of the customers in the system including the one being served, L(t) be the inventory level and set

$$S(t) = \begin{cases} 0 \text{ if the server is idle} \\ 1 \text{ if the server is busy} \\ 2 \text{ if the server is on interruption} \end{cases}.$$

Then $\Omega = \{ X(t), t \geq 0 \} = \{(N(t),S(t),L(t)), t \geq 0\}$ will be a Markov chain with state space $E = \{(0,0,k) \mid s \leq k \leq S-1\} \cup \{(i,j,k) \mid i \geq 1, j = 1,2, s+1 \leq k \leq S\}$.

The state space of the Markov chain is partitioned into levels $\tilde{i}$ defined as $\tilde{0} = \{(0,0,s),\ldots,(0,0,S-1)\}$, and $\tilde{i} = \{ (i,1,s+1), \ldots\ldots,(i,1,S), (i,2,s+1), \ldots\ldots, (i,2,S)\}$, for $i \geq 1$. This makes the Markov chain under consideration, a level independent Quasi Birth Death (QBD) process. In the following sequel, Q stands for S-s, I_n denotes an identity matrix of order n and e denotes a column matrix of 1's of appropriate order.

STATE TRANSITION DIAGRAM



Now the infinitesimal generator matrix of the process is

$$T=\begin{bmatrix} B_0 & B_1 & 0 & 0 & 0 & 0 & 0 \\ B_2 & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & & & \\ & & & & & & \\ & & & & & & . \end{bmatrix}, \text{ where}$$

$$B_0=-\lambda I_Q; B_1=\begin{bmatrix} D_1 & 0 \end{bmatrix}_{Q\times 2Q}; \quad D_1=\begin{bmatrix} 0 & 0 & & & \lambda \\ \lambda & 0 & . & & 0 \\ 0 & \lambda & 0 & & . \\ . & . & & . & . \\ 0 & 0 & & \lambda & 0 \end{bmatrix}_{Q\times Q}; B_2=\begin{bmatrix} \mu I \\ 0 \end{bmatrix}_{2Q\times Q}$$

$$A_1=\begin{bmatrix} -(\lambda+\mu+\delta_1)I & \delta_1 I \\ \delta_2 I & -(\lambda+\delta_2)I \end{bmatrix}_{2Q\times 2Q}$$, where each block is a $Q\times Q$ matrix

$A_0=\lambda I_{2Q}$ and $A_2=\begin{bmatrix} D_2 & 0 \\ 0 & 0 \end{bmatrix}_{2Q\times 2Q}$, where each block is a $Q\times Q$ matrix.

4.3. Analysis of the Model

4.3.1 Stability condition

Define $A=A_2+A_1+A_0$ and let $\pi = (\pi(1,s+1),\ldots,\pi(1,S),$ $\pi(2,s+1),\ldots,\pi(2,S))$ be the steady state vector of the generator matrix A. The relations $\pi A=0$ and $\pi e=1$ when solved result in the values of various components of π as

$\pi(1,s+1)=\ldots\ldots= \pi(1,S)= \dfrac{\delta_2}{Q(\delta_1+\delta_2)}$ and $\pi(2,s+1)=\ldots\ldots\ldots= \pi(2,S)=$ $\dfrac{\delta_1}{Q(\delta_1+\delta_2)}$.

The QBD process with generator T is stable if and only if the rate of drift to the left is larger than the rate of drift of the level to the right; that is $\pi A_0 e < \pi A_2 e$ (see Neuts[56]), that is if and only if

$$\lambda < \frac{\delta_2 \mu}{(\delta_1+\delta_2)}.$$

Thus we have the following theorem for stability of the system under study.

Theorem 4.1

The Markov chain Ω is stable if and only if $\dfrac{\lambda}{\mu}\dfrac{(\delta_1+\delta_2)}{\delta_2} < 1$.

Note: Since the lead-time is assumed as zero, the absence of the inventory parameters s and S is expected. The quantity $\dfrac{(\delta_1+\delta_2)}{\mu\delta_2}$ is actually the expected duration of an effective service (this has been derived in section 4.2), which is subject to interruptions at a rate δ_1 and to repairs at rate δ_2. Therefore, $\frac{\lambda}{\mu}\frac{(\delta_1+\delta_2)}{\delta_2}$ is the number of arrivals during a service, which should be less than 1 for stability of the system under study.

4.3.2 Computation of steady state vector

We find the steady state vector of Ω explicitly. Let $\pi=(\pi_0, \pi_1, \pi_2, \ldots)$, be the steady state vector ,where π_0 is partitioned as $\pi_0=(\pi_0(0,s+1),\ldots, \pi_0(0,S))$ and π_i's are partitioned as $\pi_i=(\pi_i(1,s+1),\ldots,\pi_i(1,S),\ \pi_i(2,s),\ \pi_i(2,s+1),\ldots,\ \pi_i(1,S))$.

Then from $\pi T=0$ and $\pi e=1$ we get

$$-\lambda\pi_0(0,j)+\mu\pi_1(1,j+1)=0, s\le j\le S-1 \qquad 3.2.1$$

$$\lambda\pi_0(0,j)-(\lambda+\mu+\delta_1)\pi_1(1,j)+\delta_2\pi_1(2,j)+\mu\pi_2(1,j+1)=0,\ s+1\le j\le S-1$$
$$\lambda\pi_0(0,s)-(\lambda+\mu+\delta_1)\pi_1(1,S)+\delta_2\pi_1(2,S)+\mu\pi_2(1,s+1)=0, \qquad 3.2.2$$

$$\lambda\pi_i(1,j)-(\lambda+\mu+\delta_1)\pi_{i+1}(1,j)+\delta_2\pi_{i+1}(2,j)+\mu\pi_{i+2}(1,j+1)=0,\ s+1\le j\le S-1$$
$$\lambda\pi_i(1,S)-(\lambda+\mu+\delta_1)\pi_{i+1}(1,S)+\delta_2\pi_{i+1}(2,S)+\mu\pi_{i+2}(1,s+1)=0\ ,\ i\ge 1 \qquad 3.2.3$$

$$\delta_1\pi_1(1,j)-(\lambda+\delta_2)\pi_1(2,j)=0,\ s+1\le j\le S-1 \qquad 3.2.4$$
$$\lambda\pi_i(2,j)+\delta_1\pi_{i+1}(1,j)-(\lambda+\delta_2)\pi_{i+1}(2,j)=0\ ,\ i\ge 1,\ s+1\le j\le S-1. \qquad 3.2.5$$

For solving the above system of equations, we first consider an M/PH/1 queue with arrival process Poisson with parameter λ and service time for each customer having PH distribution with representation (α, K), where the initial probability vector is α=(1,0) and the matrix K=$\begin{bmatrix} -(\mu+\delta_1) & \delta_1 \\ \delta_2 & -\delta_2 \end{bmatrix}$. Then the generator matrix of this system (namely, the M/PH/1queue) has the form:

$$\hat{T} = \begin{bmatrix} -\lambda & \lambda\alpha & 0 & 0 & 0 & . & . & . \\ K^0 & K-\lambda I & \lambda I & 0 & 0 & . & . & . \\ 0 & K^0\alpha & K-\lambda I & \lambda I & 0 & . & . & . \\ 0 & 0 & K^0\alpha & K-\lambda I & \lambda I & 0 & . & . \\ & & & & . & . & . & . \\ & & & & . & . & . & . \\ & & & & & . & . & . \\ & & & & & & . & . \end{bmatrix}, \text{ where } K^0 = \begin{bmatrix} \mu \\ 0 \end{bmatrix}.$$

Let $x = (x(0), x(1), x(2), ...)$ be the steady state vector of $\hat{T}$. Partitioning $x(i)$ s' as $x(0) = x(0,0)$, $x(i) = (x(i,1), x(i,2)), i \geq 1$, the steady state relation $x\hat{T} = 0$, gives us the following equations.

$$-\lambda x(0,0) + \mu x(1,1) = 0 \qquad 3.2.a$$

$$\lambda x(0,0) - (\lambda + \mu + \delta_1)x(1,1) + \delta_2 x(1,2) + \mu x(2,1) = 0 \qquad 3.2.b$$

$$\delta_1 x(1,1) - (\lambda + \delta_2)x(1,2) = 0 \qquad 3.2.c$$

$$\lambda x(i,1) - (\lambda + \mu + \delta_1)x(i+1,1) + \delta_2 x(i+1,2) + \mu x(i+2,1) = 0, \ i \geq 1 \qquad 3.2.d$$

$$\lambda x(i,2) + \delta_1 x(i+1,1) - (\lambda + \delta_2)x(i+1,2) = 0 \ i \geq 1 \qquad 3.2.e$$

If we assume that

$$\left.\begin{aligned} &\pi_0(0,s) = \pi_0(0,s+1) = ... = \pi_0(0,S-1) and \\ &\pi_i(j,s+1) = \pi_i(j,s+2) = ... = \pi_i(j,S), i \geq 1, j = 1,2 \end{aligned}\right\} \qquad (3.2.I),$$

then the S-s equations in 3.2.1 for each value of I will be the same as the single equation 3.2.a and similarly equations 3.2.2 to 3.2.5 reduce to 3.2.b to 3.2.e respectively and therefore the probabilities $\pi_i(j,k)$ can be obtained from the corresponding $x(i,j)$ as

$$\left.\begin{aligned} &\pi_0(0,k)=\frac{1}{Q}x(0,0), s\le k\le S-1 \\ &\pi_i(j,k)=\frac{1}{Q}x(i,j), j=1,2; s+1\le k\le S \end{aligned}\right\} \quad (3.2.\text{II})$$

The intuition behind the assumption 3.2.I is that, since replenishment is instantaneous, in the steady state, there will be equal chance for each inventory level to be visited. It can be verified that the $\pi_i(j,k)$'s obtained from 3.2.II, satisfies the steady state equations 3.2.1 to 3.2.5 and so are the unique steady state probabilities of the system under the stability condition.

Now for the steady state probabilities $x(i,j)$, we have results available for the standard M/PH/1 queue (see Neuts[56]), which gives

$x(i)=x(1)R^{i-1}, i\ge 1$, where

$$R=\begin{bmatrix} \dfrac{\lambda}{\mu} & \dfrac{\lambda\delta_1}{\mu(\lambda+\delta_2)} \\ \dfrac{\lambda}{\mu} & \dfrac{\lambda(\mu+\delta_1)}{\mu(\lambda+\delta_2)} \end{bmatrix} \text{ and}$$

$$x(1)=x(0)\begin{bmatrix} \dfrac{\lambda}{\mu} & \dfrac{\lambda\delta_1}{\mu(\lambda+\delta_2)} \end{bmatrix}; \quad x(0)=x(0,0)=1-\frac{\lambda(\delta_1+\delta_2)}{\mu\delta_2}$$

4.4 System Performance Measures

4.4.1 Expected Number of interruptions encountered by a customer

For computing expected number of interruptions encountered by a customer we proceed in the same line as in Krishnamoorthy et. al. [29] by considering a Markov process $\{X_1(t), t \geq 0\} = \{(N_1(t), S_1(t)), t \geq 0\}$, where $N_1(t)$ denotes the number of interruptions that has occurred up to time t ; $S_1(t)=0$ or 1 according as the service is under interruption or not at time t . The Markov process $\{X_1(t),\ t \geq 0\}$ has state space $\{0,1,2,....\} \times \{0,1\} \cup \{\Delta\}$, where Δ is an absorbing state which denotes service completion. The infinitesimal generator of the process is the same as in [20]:

$$\hat{U} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & . & . & . \\ \hat{B}_{00} & \hat{A}_{00} & \hat{A}_{01} & 0 & 0 & & . & . \\ \hat{A}_2 & 0 & \hat{A}_1 & \hat{A}_0 & 0 & & . & . \\ \hat{A}_2 & 0 & 0 & \hat{A}_1 & \hat{A}_0 & . & . & \ddots \\ \hat{A}_2 & 0 & 0 & 0 & \hat{A}_1 & \hat{A}_0 & . & . \\ & & . & & & & . & . \\ & & & & . & . & . & . \\ & & & & & & . & . \end{bmatrix}$$

, where in the present case,

$$\hat{B}_{00} = [\mu],\ \hat{A}_{00} = [-(\mu+\delta_1)],\ \hat{A}_{01} = [\delta_1 \quad 0],\ \hat{A}_2 = \begin{bmatrix} 0 \\ \mu \end{bmatrix},\ \hat{A}_1 = \begin{bmatrix} -\delta_2 & \delta_2 \\ 0 & -(\mu+\delta_1) \end{bmatrix}$$

and $\hat{A}_0 = \begin{bmatrix} 0 & 0 \\ \delta_1 & 0 \end{bmatrix}$.

If y_k is the probability that absorption occurs with exactly k interruptions, then

$y_0= -\hat{A}_{00}^{-1}\hat{B}_{00}=\dfrac{\mu}{\mu+\delta_1}$

$y_k=(-\hat{A}_{00}^{-1}\hat{A}_{01})\ (-\hat{A}_1^{-1}\hat{A}_0)^{k-1}(-\hat{A}_1^{-1}\hat{A}_2)=\dfrac{\mu}{\mu+\delta_1}(\dfrac{\delta_1}{\mu+\delta_1})^k$, k=1,2,3,...

The expected number of interruptions before absorption is given by $E_I=\sum_{k=0}^{\infty} ky_k = (-\hat{A}_{00}^{-1}\hat{A}_{01})\ \left[I_2-(-\hat{A}_1^{-1}\hat{A}_0)\right]^{-1} e=\dfrac{\delta_1}{\mu}$.

4.4.2 Expected duration of an interrupted service

Here we calculate the average duration of an interrupted service. The procedure for this is again similar to that in [29]. The service process with interruption can be viewed as a Markov process with two transient states 0 and 1, which denote whether the server is interrupted or is busy respectively, and a single absorption state Δ. The absorption state Δ denotes the completion of the service after the intervening interruptions and repairs. The process can be represented by $\hat{X}(t)=\{0,1,\Delta\}$. Let T be the time until absorption in the process $\hat{X}(t)$. The infinitesimal generator matrix of the process is given by

$\hat{H}=[\hat{B},\hat{B}_0]$, where $\hat{B}=\begin{bmatrix}-\delta_2 & \delta_2\\ \delta_1 & -(\mu+\delta_1)\end{bmatrix}$ and $\hat{B}_0=\begin{bmatrix}0\\ \mu\end{bmatrix}$.

The probability distribution F(.) of T is given by F(x)=1- ξ exp($\hat{B}$ x)e,x$\geq$0. Its density function F'(x) in $(0,\infty)$ is given by F'(x) = ξ exp($\hat{B}$ x) $\hat{B}_0$. The Laplace-Stieltjes transform f(s) of F(.) is f(s)= ξ(sI- $\hat{B}$)$^{-1}$ $\hat{B}_0$. The expected time E_s for service completion is

$E_s=\xi\,(-\hat{B})^{-1}e = \dfrac{\delta_1+\delta_2}{\mu\delta_2}$

4.4.3 Expected amount of time a customer is served during his (possibly interrupted) service

In section 4.4.2, we derived the expected number of interruptions as $E_1 = \frac{\delta_1}{\mu}$. Since each interruption has a repair time with mean $\frac{1}{\delta_2}$, the total expected time of repair during a service is $E_{rt} = \frac{\delta_1}{\mu\delta_2}$. The expected duration of time a customer is served during the service process is then given by $E_S - E_{rt} = \frac{1}{\mu}$.

4.4.4 Analysis of waiting time

Though we can find the expected waiting time using Little's formulae we do so otherwise and verify the result obtained using the above.. For any M/G/1 queue the mean waiting time of a customer in the system is given by E{W}= E{W_Q+s} = E(s)+E{W_Q} [see 52]

$$= E_s + \frac{\lambda}{2(1-\rho)}E(s^2) = \frac{\delta_1+\delta_2}{\mu\delta_2} + \frac{\lambda}{2\{1-\frac{\lambda(\delta_1+\delta_2)}{\mu\delta_2}\}} 2\xi(-\hat{B})^{-2}e$$

$$= \frac{\delta_1+\delta_2}{\mu\delta_2} + \frac{\lambda}{2\{1-\frac{\lambda(\delta_1+\delta_2)}{\mu\delta_2}\}}\frac{2}{\mu^2\delta_2^{\,2}}\{\mu\delta_1+(\delta_1+\delta_2)^2\} = \frac{\delta_2^2+\delta_1\lambda+\delta_1\delta_2}{\delta_2\{\mu\delta_2-\lambda(\delta_1+\delta_2)\}}$$

.We have obtained the expression for the expected number of customers in the system as EN= $\frac{\lambda}{\delta_2}.\frac{\delta_2^2+\delta_1\lambda+\delta_1\delta_2}{\mu\delta_2-\lambda(\delta_1+\delta_2)}$ .Hence Little's theorem is verified.

4.4.5 Busy Period

We have the expected duration of a busy period T is given by [54]

$$E(T)= E(s)+\frac{\lambda}{2(1-\rho)}E(s^2) = \frac{\delta_1+\delta_2/\mu\delta_2}{\{1-\lambda(\delta_1+\delta_2)/\mu\delta_2\}} = \frac{\delta_1+\delta_2}{\mu\delta_2-\lambda(\delta_1+\delta_2)}$$

4.4.6 Other performance measures

1. Probability that server is busy is given by $P_\beta = \sum_{i=1}^{\infty}\sum_{j=s+1}^{S}\pi_i(1,j) = \frac{\lambda}{\mu}$.

2. Probability that server is on interruption is given by

$$P_\alpha = \sum_{i=1}^{\infty}\sum_{j=s+1}^{S}\pi_i(2,j) = \frac{\delta_1}{\delta_2}\frac{\lambda}{\mu}.$$

3. Probability that server is idle is given by

$$P_\gamma = 1-P_\alpha-P_\beta = 1-\frac{\lambda}{\mu}\left(1+\frac{\delta_1}{\delta_2}\right).$$

4. Expected inventory level is given by

$$EIL = \sum_{j=s+1}^{S}\pi_0(0,j)+\sum_{i=1}^{\infty}\sum_{j=s+1}^{S}\pi_i(1,j)+\pi_i(2,j) = \frac{s+S}{2}.$$

5. Expected number of customers in the system is given by

$$EN = \sum_{i=1}^{\infty} \sum_{j=s+1}^{S} \{i\pi_i(1, j) + i\pi_i(2, j)\} = \pi_1(I - R)^{-2} e = \frac{\lambda}{\delta_2} \cdot \frac{\delta_2^2 + \delta_1\lambda + \delta_1\delta_2}{\mu\delta_2 - \lambda(\delta_1 + \delta_2)}.$$

6. Expected rate of ordering is given by $E_{OR} = \sum_{i=1}^{\infty} \mu\pi_i(1, s+1) = \frac{\lambda}{Q}$

7. Expected interruption rate is given by $E_{INTR} = \delta_1 \sum_{i=1}^{\infty} \sum_{j=s+1}^{S} \pi_i(1, j) = \frac{\delta_1\lambda}{\mu}$.

4.5 System behavior with variations in parameters

The explicit expressions for all the system performance measures make the analysis of their dependence on various parameters more transparent.

The maximum inventory level S and the reorder level s affects the expected inventory level and expected reorder rate only. The other performance measures are independent of s and S. This can be attributed to the fact that replenishment is instantaneous.

The expression for server interruption probability P_α shows that, if we take $\delta_1 = \delta_2$, the probability P_α is just $\frac{\lambda}{\mu}$ which is independent of both δ_1 and δ_2.

The expected number of customers in the system increases with increase in arrival rate λ and decreases with decrease in service rate μ; both these facts are clear from the expression for expected number of customers. However, since the effect of the parameters δ_1 and δ_2 on the expected number of customers is not that clear from the expression for EN, we studied this numerically. Table 1(a) and 1(b) show the effect of δ_1 and δ_2 respectively on

EN. Table 1(a) shows that as δ_1, the interruption rate increases, EN also increases ,which is expected as the interruptions become more frequent, the effective service time of a customer increases and this leads to an increase in the queue length. Table 1(b) shows that an increase in the repair rate δ_2, results in a decrease in the expected number of customers in the system. This is also expected as the repair rate increases; the server becomes active in a shorter time after an interruption which leads to an increase in the service completion rate and hence the queue length also.

δ1	EN
2	1.417
2.2	1.561
2.4	1.722
2.6	1.902
2.8	2.104
3	2.33
3.2	2.595
3.4	2.897
3.6	3.25
3.8	3.667
4	4.167

(a)

δ2	EN
2	2.5
2.2	2.129
2.4	1.869
2.6	1.678
2.8	1.532
3	1.416
3.2	1.324
3.4	1.248
3.6	1.184
3.8	1.13
4	1.083

(b)

Table1: Effect of δ_1 and δ_2 on E (σ) with $\lambda=1$, $\mu=3$, s=6, S=20; we have taken for table 1(a), $\delta_2=3$, and for table 1(b), $\delta_1=2$.

4.6 Cost Analysis

For computing optimal values for s and other parameters, we introduce a cost function $C=CI^* EIL+CN^* EN+CR^* E_{INTR}+\{K+(S-s)K_1\}^* E_{OR}$, where CI is the cost of holding inventory ,CN is the cost of holding customers, CR is the cost incurred due to interrupted service, K is the fixed cost of ordering and K_1 is the cost of a single item in inventory . The problem of optimizing the cost for various parameter values is carried out. A few illustrations are given below.

EIL and E_{OR} are the only measures involved in the cost function, which are affected by the reorder level *s*. As values of these measures increase with increase in s, an increase in the cost function with s is expected; figure 1 shows this. Thus the optimal value for the reorder level can be concluded to be 0. This can also be explained heuristically. Since lead time is zero and no shortage is permitted, it is optimal to order for replenishment of inventory when the level falls to zero at a service completion epoch. Further if at a service completion epoch the inventory level has fallen to zero and no customer is left behind in the system, then it is optimal to place order for replenishment at the arrival epoch of the first customer to the idle system. This results in drastic reduction of holding cost. These explain why the optimal value s* of s is zero.

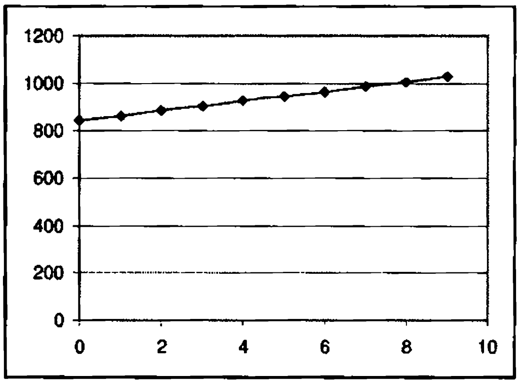
In the case of the maximum inventory level S, *EIL* increases with increase in S but E_{OR} decreases with increase in S. Hence depending on the nature of the costs attached to these measures, the cost function either increases with S, which is the case with figure 2, or shows a convex nature as in figure 3. For figure 2, the fixed cost K was taken as 35 whereas for figure 3, a much larger value 750 for K is assumed, which brings the convex (non linear) nature to the graph.

As the arrival rate λ increases or the interruption rate δ_1 increases, the measures involved in the cost function either increases or remain fixed; as a result the cost function also increases with an increase in the parameters λ and δ_1. These results can be visualized in figures 4 and 5 respectively. As compared to parameters λ and δ_1, the parameters μ and δ_2 have a reverse effect on the performance measures involved in the cost function; that is these measures decrease as μ and δ_2 increase. Hence the curve corresponding to the cost function has a negative slope in the cases of the parameters μ and δ_2 which can be visualized from figures 6 and 7 respectively.

Conclusion:

We analysed an (s,S) inventory problem with service interruption. Incidence of interruption forms a Poisson process; service times and removal of interruptions are independent exponentially distributed random variables. Lead-time is assumed to be zero and further since no shortage is permitted, the inventory level probability turns out to be discrete uniform. Explicit expression for the rate matrix could be arrived at. Several measures of performance for optimal system design have been computed. Convexity/monotonicity of cost function in S is numerically arrived at. Also the optimal reorder level is shown to be zero. The computation of the effective service time is achieved through the matrix analytic method.

s	Cost
0	844.25
1	864.34
2	884.44
3	904.56
4	924.69
5	944.83
6	965
7	985.19
8	1005.42
9	1025.68

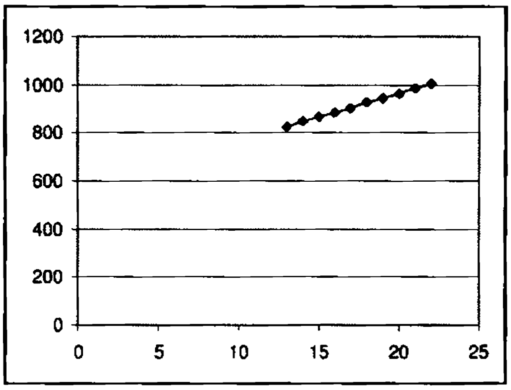


Reorder level s versus Cost

Figure1: Effect of the reorder level s on the Cost function

$CI=40, CN=30, CR=75, K=35, K_1=350, \lambda=1, \mu=3, \delta_1=2, \delta_2=3, S=20$

S	Cost
13	827.5
14	846.88
15	866.39
16	886
17	905.68
18	925.42
19	945.19
20	965
21	984.83
22	1004.69

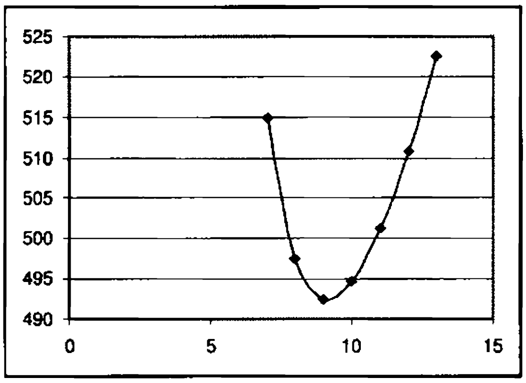


Maximum inventory level S versus Cost

Figure 2: Effect of maximum inventory level S on Cost function.

CI=40,CN=30,CR=75,K=35,K_1=350,λ=1,μ=3,δ_1=2,δ_2=3,s=6

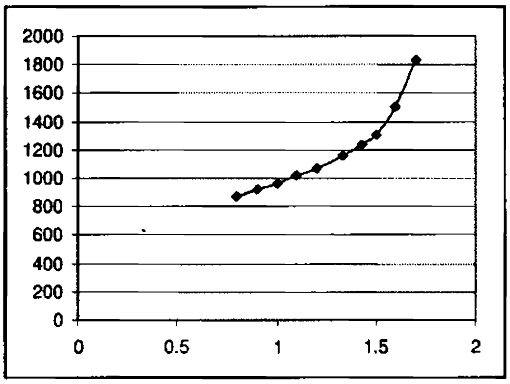
S	Cost
7	515
8	497.5
9	**492.5**
10	494.64
11	501.25
12	510.83
13	522.5



Maximum inventory level S versus Cost

Figure 3: Effect of S on Cost. CI=40, CN=30, CR=75, K=750, K_1=35, λ=1, μ=3, δ_1=2, δ_2=3, s=3

λ	Cost
0.8	868.56
0.9	915.85
1	965
1.1	1016.81
1.2	1072.6
1.33	1155.27
1.43	1233.63
1.5	1303.75
1.6	1500.37
1.7	1829.85

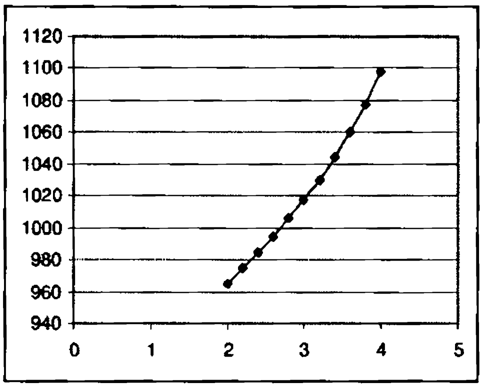


Arrival rate λ versus Cost

Figure 4:Effect of λ on Cost.

CI=40,CN=30,CR=75,K=35,K_1=350,μ=3,δ_1=2,δ_2=3,s=6,S=20

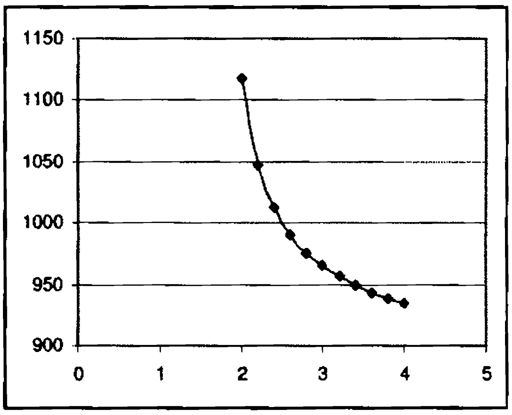
δ_1	Cost
2	965
2.2	974.34
2.4	984.17
2.6	994.56
2.8	1005.63
3	1017.5
3.2	1030.36
3.4	1044.42
3.6	1060
3.8	1077.5
4	1097.5



Interruption rate δ_1 versus Cost

Figure 5:Effect of $\delta_1$on Cost.
CI=40,CN=30,CR=75,K=35,K_1=350,λ=1,μ=3,δ_2=3,s=6,S=20

μ	Cost
2	1117.50
2.2	1046.93
2.4	1012.27
2.6	990.91
2.8	976.07
3	965.00
3.2	956.33
3.4	949.31
3.6	943.48
3.8	938.54
4	934.29

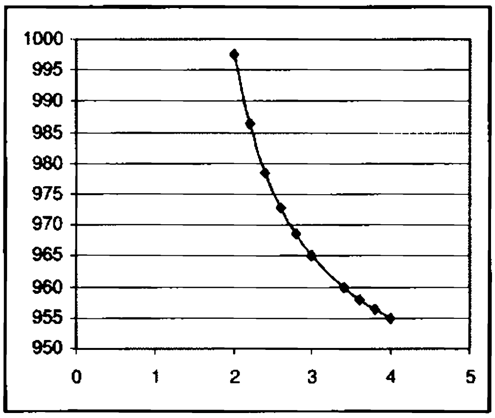


Service rate μ versus Cost

Figure 6:Effect of μ on Cost.

CI=40,CN=30,CR=75,K=35,K_1=350,λ=1,δ_1=2,δ_2=3,s=6,S=20

δ_2	Cost
2	997.50
2.2	986.36
2.4	978.57
2.6	972.83
2.8	968.45
3	965.00
3.2	962.22
3.4	959.93
3.6	958.01
3.8	956.39
4	955.00



Repair rate δ_2 versus Cost

Figure 7:Effect of δ_2 on Cost. CI=40, CN=30, CR=75, K=35, K_1=350,λ=1,μ=3,δ_1=2,s=6,S=20

CHAPTER 5

An Inventory Model with Server Interruptions and Retrials *

5.1 Introduction

In the previous chapter we assumed that replenishment is instantaneous and no shortage is permitted (shortage cost infinity).In this chapter we assume that replenishment of inventory does not materialize immediately as placement of order ,rather takes a random amount of time for order materialization .Thus we have now a finite shortage cost situation. Another salient feature of this chapter is that the QBD we develop here is level dependent. The reason behind this is the state dependent retrial rate of orbital customers. LDQBD's are much more complex than LIQBD's since we do not get a repeating pattern for the entries of the infinitesimal generator of the process. However we make it level independent through a process called truncation.

The first study on inventory models with positive lead time, with unsatisfied customers thus created going to an orbit to try again for inventory from there, was by Artalejo et. al [6]. Whereas their approach is algorithmic, Ushakumari [68] produces analytical solution to the same model. Following these, a number of papers on inventory models with retrial of unsatisfied customers emerged. One may refer to Krishnamoorthy and Islam [30,31] Two other papers where an inventory model with retrial of demands is considered

* The results of this chapter was presented as a paper at the8th International Workshop on Retrial Queues; July 27-29, 2010, Beijing, China .It has also been accepted for publication in the journal 'Operational Research'(article No: DOI: 10.1007/s12351-011-0120-8)

are by Sivakumar [61,62]; where the first one considers an (s, S) perishable inventory system in which demands occur from a finite source and those demands that arrive in a stock-out period, are sent to an orbit. The discussion in [62] is on a two-commodity system where customers, encountering both commodities out-of-stock, proceed to an orbit of infinite capacity. In Krishnamoorthy and Jose [34], the authors analyse and compare different (s, S) inventory models with an orbit of infinite capacity. They consider situations where a finite waiting station/no waiting station is provided for fresh/retrial customers. In all these models, the presence of the retrying customers results in an in-flow and out-flow pattern that is distinct from those where only a queue of unsatisfied demands has been considered. In this regard the paper by Krishnamoorthy and Islam [32], where an (s, S) inventory system with a finite pool of unsatisfied demands is studied, has an in-flow and out-flow pattern that is distinct from both the above type of models.

Different kinds of interruptions in service such as server breakdown, server going on vacation, arrival of priority customers, being common real life phenomena, analysis of queueing models with these features partially or completely incorporated, is important. An M/M/1 queueing model with service interruption was first studied by White and Christie [73]. The service interruption of such customers is assumed to be due to arrival of priority customers. The policy adopted is preemptive repeat. However, it may be noted that in the case of exponentially distributed service time, repeat or resumption of interrupted service do not make any difference. Krishnamoorthy et. al [29,37, 38] provide a glimpse of earlier work on queues with service interruption and provide several new results.

In a very recent paper, Krishnamoorthy et. al [41] considered an (s,S) inventory model with positive service time and instantaneous replenishment, where the service process is subject to interruptions. In the present chapter, we extend the above model by assuming that replenishment of items requires a random amount of lead-time. As in [41], the service in the present case is also subject to interruptions. Now since the replenishment is not instantaneous, an increase in waiting time is probable which motivates us to introduce a retrial queueing model here. It may be noted that the interruption process assumed here has the property that, at a time only one interruption is encountered by the server. In our model, whenever the service is restarted after interruption, it is assumed that the entire service is repeated from the beginning and further it is prd. This is in contast to Nicola ,Kulkarni and Trivedi [51] as well as Marie and Trivedi [55].

This chapter is arranged as follows. In section 5.2, we describe the mathematical model under study. In section 5.3, a necessary and sufficient condition for the stability of the system is obtained and steady state distribution is computed. Section 5.4 is devoted to some system performance measures like the expected waiting time of an orbital customer. Finally in section 5.5 we provide some results of the numerical experiments carried out for analyzing different aspects of the system under study. Concluding remarks are given in section 6.

5.2 Mathematical Model

Before proceeding with the modeling of the problem under investigation, we introduce a few notations and assumptions that are used in the sequel:

- Arrival of primary customers -- Poisson process of rate λ
- service time -- exponential random variable with parameter μ
- s is the reorder level and S is the maximum number of items that can be stored ($S > 2s$)
- lead time (the time elapsed from placing an order for replenishment of the item until it is delivered) -- exponentially distributed with parameter η
- service interruption process-- Poisson process of rate δ_1
- interruption duration -- exponential random variable with parameter δ_2
- retrial rate -- $j\theta$, when there are j customers in the orbit (thus retrial rate turns out to be level dependent)
- probability of a primary customer joining the system when the server is in interrupted state -- p
- probability of an orbital customer quitting the system after an unsuccessful retrial due to server in interrupted state -- q
- identity matrix of order n -- I_n
- identity matrix of appropriate order -- I
- column vector of 1's of appropriate order -- e

The model under study is described as follows: Customers arrive to a single server counter according to a Poisson process of rate λ where inventory is served. Service times are iid exponentially distributed random variables with parameter μ. Inventory is replenished according to (s, S) policy, with the lead time distribution exponential with parameter η.

While the server serves a customer, the service may get interrupted with the interruption process governed by a Poisson process of rate δ_1. It is assumed that while the server is under interruption, no further interruption can befall the server. On completion of an interruption the service restarts, with the duration of an interruption exponentially distributed with parameter δ_2. No waiting space is provided for customers, other than for the one whose service gets interrupted. An arriving customer, finding the server busy, leaves the service area and joins an orbit of infinite capacity from where it retries for service. The duration of the interval between two successive repeated attempts is exponentially distributed with parameter $j\theta$ when the number of customers in the orbit is j. While the server is on an interruption, an arriving customer (primary) joins the system with probability p and a retrying customer goes back to the orbit with probability $(1-q)$. With complementary probabilities the customer leaves the system in both cases. When the inventory level is zero no primary arrival or retrial is entertained (primary arrivals have to leave the system without getting admission to orbit and retrial customers stay put in the orbit).

Let $N(t)$ be the number of customers in the orbit and $L(t)$ be the inventory level at time t. Also let

$$C(t)=\begin{cases}0, & \text{if the server is idle}\\ 1, & \text{if the server is busy}\\ 2, & \text{if the server is on interruption}\end{cases}$$

be the server status. Then $\Omega = \{X(t);t \ge 0\} = \{(N(t),C(t),L(t));t \ge 0\}$ is a Markov chain on the state space $((Z_+ \cup \{0\}) \times \{0,1,2\} \times \{1,2,3,\ldots,S\}) \cup ((Z_+ \cup \{0\}) \times \{0\} \times \{0\})$.

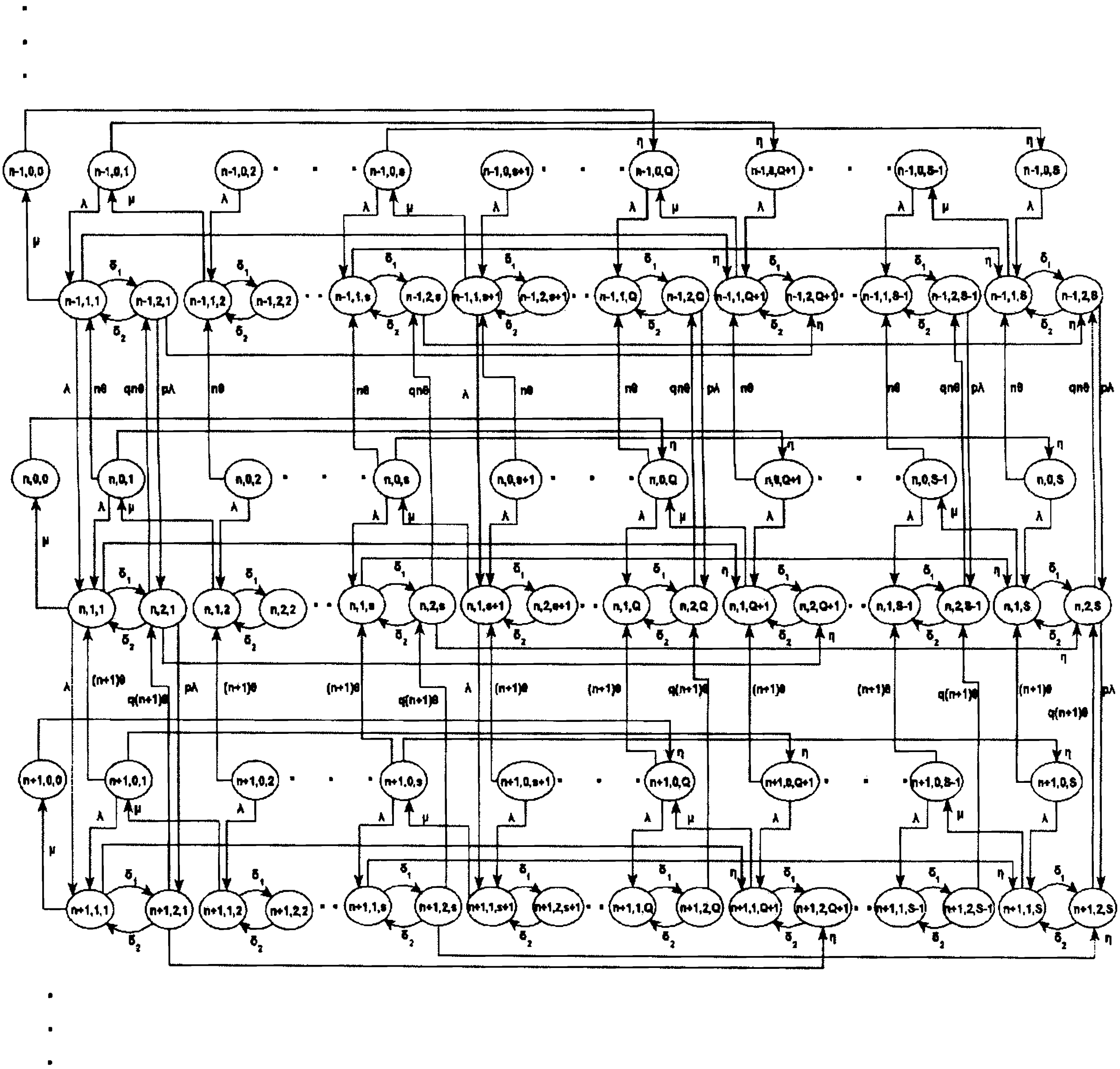
The state space of the Markov chain is partitioned in to levels $\hat{i}$ defined as $\{(i,0,j);0\leq j\leq S\}\cup\{(i,1,j);1\leq j\leq S\}\cup\{(i,2,j);1\leq j\leq S\}$, $i\geq 0$ and Q = S-s.

5.2.1 A typical illustration of the transitions of the Markov chain is as given below:

For i≥0 ,

$$(i,1,k)\xrightarrow{\lambda}(i+1,1,k)$$
$$(i,2,k)\xrightarrow{p\lambda}(i+1,2,k)$$
$$(i,0,k)\xrightarrow{i\theta}(i-1,1,k);1\leq k\leq S$$
$$(i,2,k)\xrightarrow{qi\theta}(i-1,2,k)$$

$$(i,0,k)\xrightarrow{\lambda}(i,1,k);1\leq k\leq S$$
$$(i,0,k)\xrightarrow{\eta}(i,0,k+Q);0\leq k\leq s$$
$$(i,1,k)\xrightarrow{\eta}(i,1,k+Q);1\leq k\leq s$$
$$(i,2,k)\xrightarrow{\eta}(i,2,k+Q);1\leq k\leq s$$
$$(i,1,k)\xrightarrow{\mu}(i,0,k-1);1\leq k\leq S$$
$$(i,1,k)\xrightarrow{\delta_1}(i,2,k)$$
$$(i,2,k)\xrightarrow{\delta_2}(i,1,k)$$



The Markov chain Ω in which the above transitions occur, is a level-dependent quasi birth and death process (LDQBD) with infinitesimal generator matrix

$$\hat{Q} = \begin{bmatrix} A_{10} & A_0 & 0 & 0 & 0 & 0 \\ A_{21} & A_{11} & A_0 & 0 & 0 & 0 \\ 0 & A_{22} & A_{12} & A_0 & 0 & 0 \\ & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot \\ & & & & & \cdot \end{bmatrix},$$

where each entry is a $(3S+1)$x $(3S+1)$ matrix, which is explained in detail below:

$A_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda I_S & 0 \\ 0 & 0 & p\lambda I_S \end{bmatrix}$ represents transitions from level i to i+1 due to arrival of a customer; note that by assumption, arrival rate is $p\lambda$, when the server is interrupted.

$A_{2j} = \begin{bmatrix} 0 & B_{2j} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & qj\theta I_S \end{bmatrix}$ represents transitions from level j to j-1, where

$B_{2j} = \begin{bmatrix} 0 \\ j\theta I_S \end{bmatrix}_{(S+1)\times S}$ represents transitions from level j to j-1 due to retrial of an orbital customer becoming successful and $qj\theta I_S$ represents those due to an orbital customer leaving the system after an unsuccessful retrial finding the interrupted server.

$A_{1j} = \begin{bmatrix} D_1 & D_2 & 0 \\ D_4 & D_5 & D_6 \\ 0 & D_8 & D_9 \end{bmatrix}$ represents transitions within the level j, where the sub-matrices are explained as follows.

$$D_1 = \begin{bmatrix} D_{11} & 0 & \eta I_{(s+1)} \\ 0 & -(\lambda + j\theta) I_{(S-2s-1)} & 0 \\ 0 & 0 & -(\lambda + j\theta) I_{(s+1)} \end{bmatrix}, \text{ with}$$

$$D_{11} = \begin{bmatrix} -\eta & 0 \\ 0 & -(\eta + \lambda + j\theta) I_s \end{bmatrix}_{(s+1)\times(s+1)}$$

represents transitions within level j, where the server status remains as idle; note that the only transition here is that due to replenishment of the items represented by $\eta I_{(s+1)}$.

$D_2 = \begin{bmatrix} 0 \\ \lambda I_S \end{bmatrix}_{(S+1)\times S}$ represents transitions within level j, where the server status changes from idle to busy due to arrival of a customer.

D_4 is an $S \times (S+1)$ matrix which represents transitions within level j, where the server status changes from busy to idle due to a service completion; further the non-zero entries of D_4 are given by $(D_4)_{i,i} = \mu$, $1 \le i \le S$.

$$D_5 = \begin{bmatrix} -(\eta + \lambda + \mu + \delta_1) I_s & 0 & C \\ 0 & -(\lambda + \mu + \delta_1) I_{(S-2s-1)} & 0 \\ 0 & 0 & -(\lambda + \mu + \delta_1) I_{(s+1)} \end{bmatrix} \text{ and}$$

$$D_9 = \begin{bmatrix} -(\eta + p\lambda + \delta_2 + jq\theta) I_s & 0 & C \\ 0 & -(p\lambda + \delta_2 + jq\theta) I_{(S-2s-1)} & 0 \\ 0 & & -(p\lambda + \delta_2 + jq\theta) I_{(s+1)} \end{bmatrix}$$

with $C = [0 \;\; \eta I_s]_{s\times(s+1)}$, represents transitions within level j, where the server status remains as busy and interrupted respectively. Note that the only transition here is that due to replenishment of the items represented by the matrix C.

$D_6 = \delta_1 I_S$ represents transitions within level j, due to server interruption and $D_8 = \delta_2 I_S$ those due to server status changing from interrupted to busy.

5.3 Analysis of the Model

In this section, using Matrix Analytic Methods (for details on Matrix Analytic Methods, see Neuts [56]), we perform the steady state analysis. First let us look at the stability of the system.

5.3.1 Stability Condition

For investigating the stability condition of the system under study, first we apply Neuts-Rao [57] truncation to the LIQBD. To this end suppose that $A_{1i} = A_{1N}$ and $A_{2i} = A_{2N}$ for all i ≥N. The generator matrix of the truncated system Ω_N will look as under:

$$\tilde{Q}_N = \begin{bmatrix} A_{10} & A_0 & 0 & 0 & 0 & & & & \\ A_{21} & A_{11} & A_0 & 0 & 0 & 0 & & & \\ 0 & A_{22} & A_{12} & A_0 & 0 & 0 & 0 & & \\ & & \cdot & & \cdot & & & & \\ & & & & \cdot & \cdot & & & \\ & & & & A_{2N} & A_{1N} & A_0 & & \\ & & & & 0 & A_{2N} & A_{1N} & A_0 & 0 \\ & & & & & & \cdot & \cdot & \\ & & & & & & & & \cdot \end{bmatrix}$$

Define $A_N = A_0 + A_{1N} + A_{2N}$ and let $\pi_N = (\pi_N(0,0), \pi_N(0,1), \pi_N(0,2), \ldots, \pi_N(0,S), \pi_N(1,1), \pi_N(1,2), \ldots, \pi_N(1,S), \pi_N(2,1), \pi_N(2,2), \ldots, \pi_N(2,S))$ be the steady state vector of A_N.

From the well-known results of Matrix Analytic Methods (see Neuts [56]), it follows that the truncated system, which is a level-independent quasi birth death process, is stable if and only if $\pi_N\ A_{2N}\, e > \pi_N\ A_0\, e$, that is, if and only if

$$N\theta\left[\pi_N(0,1)+\pi_N(0,2)+...+\pi_N(0,S)\right]+qN\theta\left[\pi_N(2,1)+\pi_N(2,2)+...+\pi_N(2,S)\right]>$$
$$\lambda\left[\pi_N(1,1)+\pi_N(1,2)+...+\pi_N(1,S)\right]+\ p\lambda\left[\pi_N(2,1)+\pi_N(2,2)+...+\pi_N(2,S)\right].$$

This reduces to the system being stable if and only if

$$\{\frac{N\theta\mu}{\lambda+N\theta}+\frac{qN\theta\delta_1}{\delta_2}\}[\pi_N(1,1)+\pi_N(1,2)+...+\pi_N(1,S)]>$$

$$\{\lambda+\frac{p\lambda\delta_1}{\delta_2}\}[\pi_N(1,1)+\pi_N(1,2)+...+\pi_N(1,S)],$$

which on further simplification yields that the system is stable if and only if

$$\lambda+\frac{p\lambda\delta_1}{\delta_2}<\frac{N\theta\mu}{\lambda+N\theta}+\frac{qN\theta\delta_1}{\delta_2}.$$

Because of the second factor on the right hand side of the above inequality, we see that the system is stable whenever the probability q is greater than zero. Now, when q = 0, taking the limit in the above inequality as $N\to\infty$, it reduces to:

$$\lambda+\frac{p\lambda\delta_1}{\delta_2}<\mu.$$

Thus we have the following theorem for stability of the system under study:

Theorem 5.1

When the probability q that a retrying customer leaves the orbit after an unsuccessful retrial, is greater than zero, the Markov Chain Ω is stable irrespective of the other system parameters and when q = 0, it is stable if and only if $\lambda + \frac{p\lambda\delta_1}{\delta_2} < \mu$.

5.3.2 Computation of Steady State Probability Vector

We find the steady state vector of Ω, by approximating it with the steady state vector of the truncated system, Ω_N with generator matrix $\tilde{Q}_N$. Let $\pi^{(N)} = (\pi_0, \pi_1, \pi_2, \ldots)$, be the steady state vector of Ω_N where each π_i is a row vector consisting of 3S+1 elements represented as

π_i=(π(i,0,0),π(i,0,1),π(i,0,2),...,π(i,0,S),π(i,1,1),π(i,1,2),...,π(i,1,S),π(i,2,1),π(i,2,2),...,π(i,2,S))

.

Then from known results of Matrix Analytic Methods (see Neuts [56]), it follows that

$$\pi_{N+r} = \pi_{N-1}\ (R_N)^{r+1},\ \text{for}\ r \geq 0,$$

where R_N is the minimal non-negative solution of the matrix quadratic equation,

$$(R_N)^2\ A_{2N} + R_N\ A_{1N} + A_0 = 0, \text{ and}$$

$$\pi_{N-i} = \pi_{N-i-1} R_{N-i},\ \text{for}\ 1 \leq i \leq N-1,\ \text{where}$$

$$R_{N-i} = -A_0\left(A_{1N-i} + R_{N-i+1} A_{2,N-i+1}\right)^{-1}.$$

Now for computing π_0, we have the equation π_0 (A_{10} + R_1 A_{21}) = 0. First we take π_0 as the steady state vector of the generator matrix A_{10} + R_1 A_{21}. Then π_i, for $1 \leq i \leq N-1$, can be found using the recursive formulae; $\pi_i = \pi_{i-1} R_i$. The steady state probability distribution of the truncated system is then obtained by dividing each π_i, with the normalizing constant

$[\pi_0 + \pi_1 + \ldots]e = [\pi_0 + \pi_1 + \ldots + \pi_{N-2} + \pi_{N-1}\,(I - R_N)^{-1}]e$.

5.4. System Performance Measures

5.4.1 Waiting Time Analysis of an Orbital Customer

Since no queue is formed in the orbit, customers, independently of each other, try to access the server. Therefore computation of the waiting time distribution becomes extremely complex though it has been achieved in some special cases (see books [2, 16] for details). Hence we restrict ourselves to the computation of the moments of the waiting time. Though we can find the expected waiting time using Little's Law, the second moment and variance of the waiting time are not easy to find. These moments are found by approximating the waiting time in the system under study by those in a corresponding system with finite orbit capacity.

Let $E(W_L)$ be the expected waiting time of an orbital customer in the system under study and $E(W_L^{(N)})$ be that in the corresponding system with finite orbit capacity N. Then $E(W_L) = \lim_{N\to\infty} E(W_L^{(N)})$.

For the system with finite orbit capacity N, $W_L^{(N)}$ can be found as the time until absorption in a Markov chain $\{\hat{X}(t), t \geq 0\}$, where $\{\hat{X}(t), t \geq 0\} = \{(\hat{N}(t), C(t), L(t)), t \geq 0\}$, if the tagged customer is in the orbit and $\hat{X}(t) = \Delta$, if either the tagged customer gets service or quits the system. In the above, $\hat{N}(t)$ denotes the number of customers in the orbit including the tagged customer, C(t) and L(t) are as defined in section 2. Since the orbit capacity is N, we have $1 \le \hat{N}(t) \le N$. The state space of the process $\{\hat{X}(t), t \ge 0\}$ is

$$\{\Delta\} \cup \{(i,0,k) \mid 1 \le i \le N, 0 \le k \le S\} \cup \{(i,j,k) \mid 1 \le i \le N, 1 \le j \le 2, 1 \le k \le S\},$$

where Δ is an absorbing state. The generator matrix of this process is

$$\hat{Q}_W(N) = \begin{bmatrix} T & T^0 \\ 0 & 0 \end{bmatrix},$$

where T^0 is an $N(3S+1) \times 1$ matrix given by

$T^0((i-1)(3S+1)+j,1) = \theta$, j = 2 to S+1 ; i=1 to N,

$T^0((i-1)(3S+1)+j,1) = q\theta$, j = 2$S$+2 to 3$S$+1 ; i =1 to N

and the matrix T is given by

$$T = \begin{bmatrix} A_{11} & A_0 & & & & & \\ \hat{A}_{21} & A_{12} & A_0 & & & & \\ 0 & \hat{A}_{22} & A_{13} & A_0 & & & \\ & & & & & & \\ & & & & \cdot & & \\ & & & & \hat{A}_{2(N-2)} & A_{1(N-1)} & A_0 \\ & & & & & \hat{A}_{2(N-1)} & \hat{A}_{1N} \end{bmatrix}, \text{ where}$$

$$\hat{A}_{2j} = \begin{bmatrix} 0 & \hat{B}_{2j} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (j-1)q\theta I_S \end{bmatrix}$$

with

$$\hat{B}_{2j} = \begin{bmatrix} 0 & 0 \\ (j-1)\theta I_S & 0 \end{bmatrix}_{(S+1)\times(S+1)} \quad \text{and} \quad \hat{A}_{1N} = A_{1N} + A_0$$

and all other matrices are as defined in the generator matrix Q. Thus $E(W_L^{(N)}) = -\alpha\ T^{-1}e$ (see Neuts [56], page 46),

where $\alpha = \pi_L = (\pi_{L0}, \pi_{L1}, \pi_{L2}, \ldots, \pi_{LN})$; $\pi_{Li} = \pi_i$ with entries corresponding to server is idle states taken as zero. It has been verified numerically that for large N, $E(W_L^{(N)})$ converges according to Little's theorem.

In a similar manner, we can find the second moment of the waiting time of an orbital customer as

$$E(W_L^2) = \lim_{N \to \infty} E\left(\left(W_L^{(N)}\right)^2\right),$$

where $E\left(\left(W_L^{(N)}\right)^2\right)$ is the corresponding second moment in the truncated system and is given by

$$E\left(\left(W_L^{(N)}\right)^2\right) = 2\ \alpha\ T^{-2}e \quad \text{(see Neuts [56], page 46).}$$

Finally, the variance of the waiting time of an orbital customer is given by

$$V(W_L) = E(W_L^2) - (E(W_L))^2.$$

The conditional probability that a customer leaves the system without taking service given that he arrives while the server is busy is given by

$$P_{WS} = -\alpha\ T^{-1}\hat{T} \quad \text{(see Neuts [56], page 46),}$$

where $\hat{T}$ is an N(3S+1) x 1 matrix whose non zero entries are given by

$\hat{T}$ ((i-1) (3S+1)+j,1) = $q\theta$, j = 2S+2 to 3S+1 ; i=1to N

5.4.2 Other Performance Measures

The following system performance measures are calculated numerically.

1. The probability that server is busy is given by $P_\beta = \sum_{i=0}^{\infty}\sum_{j=1}^{S}\pi(i,1,j)$.

2. The probability that server is on interruption is given by

   $P_\alpha = \sum_{i=0}^{\infty}\sum_{j=1}^{S}\pi(i,2,j)$.

3. The probability that server is idle is given by $P_\gamma = 1 - P_\alpha - P_\beta$

4. The expected number of customers in the orbit is given by

   $$E(\sigma) = \sum_{i=0}^{\infty}\sum_{j=0}^{S} i\pi(i,0,j) + \sum_{i=0}^{\infty}\sum_{j=1}^{S} i\{\pi(i,1,j)+\pi(i,2,j)\}.$$

5. The expected inventory level is given by

   $$E(\omega) = \sum_{i=0}^{\infty}\sum_{j=0}^{S} j\pi(i,0,j) + \sum_{i=0}^{\infty}\sum_{j=1}^{S} j\{\pi(i,1,j)+\pi(i,2,j)\}$$

6. The effective rate of successful retrials is given by $E(s\tau) = \sum_{i=0}^{\infty}\sum_{j=0}^{S} i\theta\pi(i,0,j)$

7. The effective replenishment rate is given by

   $$EFRR = \sum_{i=0}^{\infty}\sum_{j=0}^{s}\eta\pi(i,0,j) + \sum_{i=0}^{\infty}\sum_{j=1}^{s}\eta\{\pi(i,1,j)+\pi(i,2,j)\}$$

8. The probability that inventory level is zero is given by

   $$P(L=0) = \sum_{i=0}^{\infty}\pi(i,0,0).$$

9. The probability that inventory level is greater than s is given by

   $$P(L>s) = \sum_{i=0}^{\infty}\sum_{j=s+1}^{S}\{\pi(i,0,j)+\pi(i,1,j)+\pi(i,2,j)\}.$$

10. The effective interruption rate is given by $E_{INTR} = \delta_1 \sum_{i=0}^{\infty} \sum_{j=1}^{S} \pi(i,1,j)$.

11. The effective repair rate is given by $E_{REP} = \delta_2 \sum_{i=0}^{\infty} \sum_{j=1}^{S} \pi(i,2,j)$.

12. The effective loss rate of orbital customers after seeing an interrupted server on retrial is given by $ER_{LOSS} = \sum_{i=0}^{\infty} \sum_{j=1}^{S} qi\theta\pi(i,2,j)$.

13. The effective rate at which arriving customers are lost on seeing an interrupted server $EA_{LOSS} = (1-p)\lambda \sum_{i=0}^{\infty} \sum_{j=1}^{S} \pi(i,2,j)$

14. The effective rate at which customers are lost finding the inventory level as zero $EO_{LOSS} = \sum_{i=0}^{\infty} \lambda\pi(i,0,0)$

15. The effective rate at which orders are placed $ER_{OR} = \sum_{i=0}^{\infty} \mu\pi(i,1,s+1)$

16. The expected rate at which customers are lost

$E_{LOSS}=ER_{LOSS}+EA_{LOSS}+EO_{LOSS}$

5.4.3. The Cost Function

To investigate whether an optimal value exists for the re-order level s, we studied the following cost function.

$COST = CINTR*E_{INTR}+CLOSS*E_{LOSS}+CN* E(\sigma) +CI* E(\omega)$,

where CINTR is the cost per interruption per unit time, CLOSS is the unit time cost assigned when a customer is lost, CN is the holding cost per customer per unit time and CI is the inventory holding cost.

5.5 Numerical Illustration

In this section, we provide numerical illustration of the system performance as the underlying parameters vary.

5.5.1 Effect of the Retrial Rate θ

Table 1(a) shows that as the retrial rate θ increases, the loss rate of retrying customers ER_{LOSS} increases; the main reason for this is the high value for the system quitting probability q (= 0.6). This increase in the loss rate leads to a decrease in the expected number of customers E(σ) and hence a decrease in the server busy probability P_β and server interruption probability P_α. Note that the decrease in the server interruption probability may be occurring because of the possible decrease in the number of services due to customer loss from the orbit. Hence the decrease in P_α, considering the corresponding decrease in P_β, should not be viewed as a gain to the system under study. Also the decrease in the interruption probability P_α may be taken as the reason for the decrease in the loss rate EA_{LOSS} of arriving customers. From the Table, one can infer that the idle probability of the server is increasing with θ; but this does not imply an increase in the number of successful retrials E($s\tau$). This decrease in E($s\tau$) with increase in θ may be due to the decrease in the number of customers in the system. There is a slight increase in the expected inventory level and a narrow decrease in the effective replenishment rate EFRR; the reason for this could be the decrease in server busy probability. Because of the increase in the expected inventory level, the loss rate due to zero inventory EO_{LOSS} must be increasing; but the Table displays a constant EO_{LOSS}, which indicates that the change may be too small. Table 1(b) shows a decrease in the expected waiting time of an orbital customer with increase in retrial rate θ; but one can see in the same Table that the conditional probability that a customer may quit the system without receiving any service is increasing. So the decrease in the waiting time does not favor the orbital customers.

Table 5.1 (a): Effect of retrial rate θ on various performance measures

λ=2 , μ =4, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, s=10, S=31

θ	P_β	P_α	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	$E(\sigma)$	$E(s\tau)$	$E(\omega)$
3.0	0.3332	0.2666	0.0635	0.2666	0.4798	0.0004	0.7449	0.5334	19.6672
3.2	0.3325	0.2660	0.0633	0.2660	0.5108	0.0004	0.7065	0.5284	19.6698
3.4	0.3319	0.2655	0.0632	0.2655	0.5417	0.0004	0.6727	0.5236	19.6723
3.6	0.3313	0.2651	0.0631	0.2651	0.5725	0.0004	0.6426	0.5191	19.6747
3.8	0.3308	0.2646	0.0630	0.2646	0.6033	0.0004	0.6157	0.5148	19.6770
4.0	0.3302	0.2642	0.0629	0.2642	0.6340	0.0004	0.5915	0.5108	19.6791
4.2	0.3297	0.2638	0.0628	0.2638	0.6647	0.0004	0.5696	0.5070	19.6811
4.4	0.3292	0.2634	0.0627	0.2634	0.6954	0.0004	0.5496	0.5034	19.6830

Table 5.1 (b): Effect of retrial rate θ on waiting time

λ=2, μ =4, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, s=4, S=10

θ	$E(W_L)$	$V(W_L)$	P_{WS}	$E(\sigma)$
3.0	0.2985	0.2042	0.1728	0.7517
3.2	0.2858	0.1883	0.1762	0.7133
3.4	0.2745	0.1748	0.1794	0.6794
3.6	0.2644	0.1631	0.1825	0.6493
3.8	0.2552	0.1529	0.1854	0.6223
4.0	0.2469	0.144	0.1882	0.5981
4.2	0.2393	0.1361	0.1909	0.5761
4.4	0.2323	0.1291	0.1935	0.5561

5.5.2 Effect of the Interruption Rate δ_1

It follows from Table 2(a) that, as the interruption rate δ_1 increases, the probability that the server being interrupted P_α increases and the server busy probability P_β decreases; but the server busy probability is high compared to the server interruption probability and the reason for this may be the high repair rate compared to that of interruption rate. Note that as the interruption rate increases, this gap between P_α and P_β diminishes with P_α dominating P_β. As frequent interruptions can cause lengthier services, loss rates EA_{LOSS} and ER_{LOSS} increases with increase in δ_1. Note that the expected inventory level is increasing with interruption rate, which may be due to the decrease in the server busy probability and so less inventory may be served. Same reasoning can be made for the decrease in the effective replenishment rate EFRR. The increase in the expected inventory level points to a decrease in the probability that the inventory level in the system is 0, and hence a decrease in the loss rate EO_{LOSS}. The increase in the loss rate explains the decrease in the expected number of orbital customers. From the Table, one can infer that the idle probability is decreasing and this in turn leads to a decrease in the expected number of successful retrials E(sτ); another reason for this could be the increased loss rate of orbital customers. In Table 2(b), one can see that the expected waiting time of a customer in the orbit is decreasing with increase in the interruption rate and at the same time the conditional P_{WS} probability that an orbital customer leaves the system without opting for service increases, which together points to the fact that the decrease in the waiting time is not in favor of the customer. This happens despite a low number of customers in the orbit, which indicates the harm, which interruptions can cause.

Table 5.2(a): Effect of the interruption rate δ_1 on various performance measures

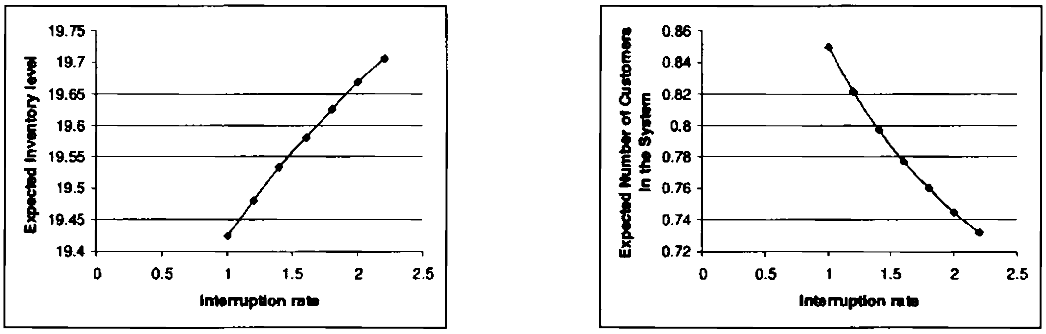
λ=2, μ =4, θ=3, η=1, δ_2=2.5, p=0.5, q=0.6, s=10, S=31

δ_1	P_β	P_α	E_{INTR}	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	E(σ)	E($s\tau$)	E(ω)
1.0	0.3941	0.1577	0.3941	0.0751	0.1577	0.4414	0.0010	0.8498	0.6831	19.4234
1.2	0.3798	0.1823	0.4557	0.0723	0.1823	0.4962	0.0008	0.8214	0.6457	19.4808
1.4	0.3667	0.2053	0.5134	0.0698	0.2053	0.5453	0.0007	0.7976	0.6128	19.5332
1.6	0.3547	0.2270	0.5676	0.0676	0.2270	0.5896	0.0006	0.7774	0.5836	19.5813
1.8	0.3435	0.2473	0.6184	0.0654	0.2474	0.6300	0.0005	0.7600	0.5572	19.6258
2.0	0.3332	0.2666	0.6664	0.0635	0.2666	0.6670	0.0004	0.7449	0.5335	19.6672
2.2	0.3236	0.2847	0.7118	0.0616	0.2847	0.7011	0.0004	0.7316	0.5118	19.7057
3.0	0.2966	0.3487	0.8717	0.0553	0.3487	0.4885	0.0002	0.6917	0.4413	19.8377
6.0	0.2124	0.5098	1.2748	0.0405	0.5098	0.6405	0.0000	0.6239	0.2942	20.1503

Table 5.2(b): Effect of the interruption rate δ_1 on waiting time

λ=2, μ =4, θ=3, η=1, δ_2=2.5, p=0.5, q=0.6, s=4, S=10

δ_1	E(W_L)	V(W_L)	E(σ)	P_{WS}
1.0	0.3052	0.2509	0.8635	0.1027
1.2	0.3025	0.2374	0.8332	0.1185
1.4	0.3007	0.2266	0.8078	0.1333
1.6	0.2995	0.2179	0.7863	0.1473
1.8	0.2988	0.2104	0.7677	0.1604
2.0	0.2985	0.2042	0.7516	0.1728



(a) (b)

Figure 5.1(a), (b). Impact of the interruption rate δ_1 on the expected number of customers in the system E(σ) and on the expected inventory level E(ω) with parameters λ=2 , μ =4, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, s=10, S=31 : The Figures show an opposite behavior of the measures E(σ) and E(ω) with increase in the interruption rate; where the expected number of customers is decreasing due to loss and the expected inventory increases because of a possible drop in the number of service completions. The Figures reflect the harm brought to the system by interruption.

5.5.3 Effect of the Repair Rate δ_2

In Table 3(a), one sees that, the increase in the repair rate leads to a decrease in the server interruption probability and to an increase in the server busy probability. This is expected; as the repair rate increases, the span of interruption period must be decreasing. The reason for increase in the effective interruption rate E_{INTR} is the increase in the server busy probability. Here note that the interruption rate δ_1 and the repair rate δ_2 are the only two parameters whose changes make the server busy probability P_β and the server interruption probability P_α to vary in opposite directions; that is, if one probability increases, the other probability decreases. Now as these probabilities vary in opposite directions, the effective interruption rate varies in the same direction as server busy probability. Coming back to the repair rate δ_2, from the Table 3(a), one observes that as the repair rate increases, the loss rate of the retrying customers ER_{LOSS} decreases and hence the expected orbit size E(σ) increases. This is expected, since fast repairs allow the server to render service to more customers. Now more number of services leads to a decrease in the inventory level, an increase in the effective replenishment rate EFRR and a narrow increase in the loss rate of customers due to zero inventory EO_{LOSS}. From Table 3(a), one can infer that the server idle probability is increasing, which is also due to an increase in the number of fast service completions and after each service completion, the server becomes idle as we are considering a retrial queue. The increase in the server idle probability then leads to a decrease in the loss rate of arriving customers EA_{LOSS} and to an increase in the expected number of successful retrials. In Table 3(b), one can see that the expected waiting time of an orbital customer is decreasing with increase in δ_2, which is also due to fast service completions. Note also that the conditional probability that an orbital customer may leave the system without receiving service is decreasing when the repair becomes faster.

Table 5.3(a): Effect of the repair rate δ_2 on various performance measures

λ=2, μ =4, θ=3, η=1, δ_1=2.5, p=0.5, q=0.6, s=10, S=31

δ_2	P_β	P_α	E_{INTR}	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	$E(\sigma)$	$E(s\tau)$	$E(\omega)$
1.5	0.2807	0.3743	0.5614	0.0535	0.3743	0.6737	0.0001	0.6973	0.4335	19.8771
1.7	0.2942	0.3461	0.5884	0.056	0.3461	0.6230	0.0002	0.7082	0.4580	19.8232
1.9	0.3059	0.3220	0.6117	0.0583	0.3220	0.5795	0.0002	0.7183	0.4799	19.7765
2.1	0.3161	0.3010	0.6322	0.0602	0.3010	0.5419	0.0002	0.7277	0.4995	19.7356
2.3	0.3251	0.2827	0.6503	0.0619	0.2827	0.5089	0.0002	0.7365	0.5172	19.6994
2.5	0.3332	0.2666	0.6664	0.0635	0.2666	0.4798	0.0003	0.7449	0.5334	19.6672
2.7	0.3405	0.2522	0.6809	0.0648	0.2522	0.4539	0.0003	0.7528	0.5483	19.6382

Table 5.3(b): Effect of the repair rate δ_2 on waiting time

λ=2, μ =4, θ=3, η=1, δ_1=2, p=0.5, q=0.6, s=4, S=10

δ_2	$E(W_L)$	$V(W_L)$	P_{WS}	$E(\sigma)$
1.5	0.3061	0.1879	0.2437	0.702
1.7	0.3037	0.1915	0.225	0.7133
1.9	0.3019	0.1949	0.209	0.7239
2.1	0.3005	0.1982	0.1953	0.7337
2.3	0.2994	0.2012	0.1833	0.7429
2.5	0.2985	0.2042	0.1728	0.7517
2.7	0.2978	0.2070	0.1636	0.7599

5.5.4 Effect of the Re-Order Level *s*

Tables 4(a) and (b) describes the effect of the re-order level *s* on various system performance measures. As the re-order level *s* increases, expected inventory level increases and hence the probability for a loss due to zero inventories in the system decreases. Note that the rate at which the retrying customers and arriving customers are lost due to interruption increases with increase in *s*. The reason for this may be the increase in the server busy probability P_β and a corresponding increase in the server interruption probability P_α. The decrease in the expected number of customers in the orbit also has the same reason. The Table shows that as *s* increases, there is a slight decrease in the rate of successful retrials; which can be attributed to the slight decrease in the server idle probability. Obviously, the effective replenishment rate EFRR has to increase with increase in *s* but the lower values for EFRR as well as the high values for the expected inventory level together suggests that replenishment is not frequently occurring in the system. Despite of the high service rate (twice as much as the arrival rate), the less frequent replenishments points to the severe effect of interruption on the system behavior. Table 4(b) shows a narrow decrease in the waiting time of a customer in the orbit with an increase in *s*. Note that one has to wait for some time besides a low expected number of customers in the orbit; and that there is a high probability that one may choose to leave the system with out opting for service.

Table 4(a) shows that the expected inventory level in the system is high even when the re-order level *s* is small. This made us to investigate whether an optimal value for *s* can be found. For this we studied the cost function defined in section 4.2. In the cost function, the measures E_{INTR} and $E(\omega)$ shows an increase with increase in *s*, while the measures E_{LOSS} and $E(\sigma)$ shows a decrease. Here note that the increase in the expected inventory is significant as compared to the changes in the other measures and therefore the cost will ultimately be increasing with *s*, which points to the optimal value zero for *s*. What we want to capture is the decrease in the measures E_{LOSS} and $E(\sigma)$; and hence get a convex nature for the cost function and an optimal value other than zero for *s*. For doing this, we assume a comparatively large cost for the loss of customers CLOSS, which is also reasonable in many practical situations. As expected, this assumption leads us to an optimal value for s (s = 5), as one can infer from Table 4(c). This Table also shows that if the cost CLOSS is not very high, the cost function is linearly increasing. These results can be more easily verified from Figures 3(a) and (b).

Table 5.4(a): Effect of the re-order level *s* on various performance measures

λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, S=31

s	P_β	P_α	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	$E(\sigma)$	$E(s\tau)$	$E(\omega)$
5	0.3323	0.2658	0.0511	0.2658	0.4785	0.0058	0.7457	0.540	17.19
6	0.3327	0.2662	0.0532	0.2662	0.4791	0.0034	0.7454	0.537	17.68
7	0.3329	0.2664	0.0555	0.2664	0.4794	0.0020	0.7451	0.535	18.17
8	0.3331	0.2665	0.0579	0.2665	0.4796	0.0012	0.745	0.534	18.67
9	0.3332	0.2665	0.0606	0.2665	0.4797	0.0007	0.7449	0.534	19.17
10	0.3332	0.2666	0.0635	0.2666	0.4798	0.0004	0.7449	0.533	19.67
11	0.3332	0.2666	0.0666	0.2666	0.4799	0.0002	0.7449	0.533	20.17
12	0.3333	0.2666	0.0702	0.2666	0.4799	0.0001	0.7448	0.533	20.67

Table 5.4(b): Effect of the re-order level s on waiting time
$\lambda=2$, $\mu=4$, $\theta=3$, $\eta=1$, $\delta_1=2$, $\delta_2=2.5$, $p=0.5$, $q=0.6$, $S=15$

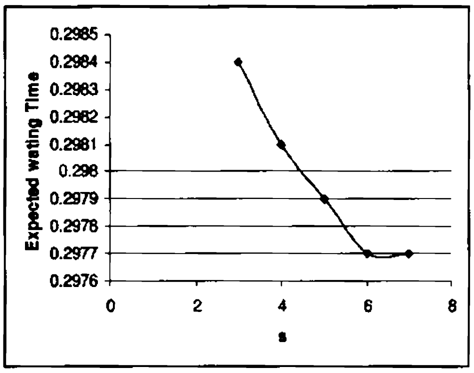
s	$E(W_L)$	$V(W_L)$	P_{WS}	$E(\sigma)$
3	0.2984	0.2022	0.1733	0.7508
4	0.2981	0.1952	0.1746	0.7486
5	0.2979	0.1910	0.1754	0.7472
6	0.2977	0.1884	0.1759	0.7463
7	0.2977	0.1869	0.1761	0.7458

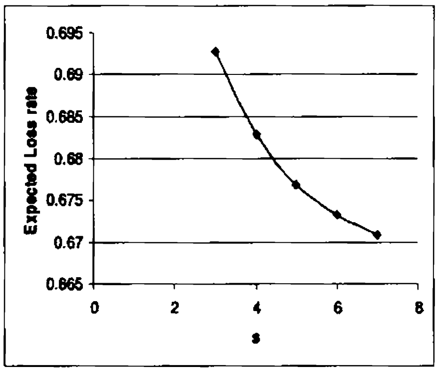
Table 5.4(c): Effect of the re-order level s on the cost function

$\lambda=2,\mu=4$, $\theta=3$, $\eta=1$, $\delta_1=2$, $\delta_2=2.5$, $p=0.5$, $q=0.6$,

S=25,CN=50,CI=60,CINTR=40

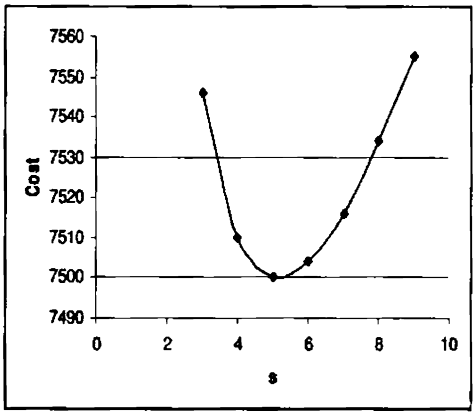
S	COST (C_{LOSS}=10000)	COST (C_{LOSS}=500)
3	7546	1075
4	7510	1094
5	7500	1117
6	7504	1140
7	7516	1164
8	7534	1189
9	7555	1213





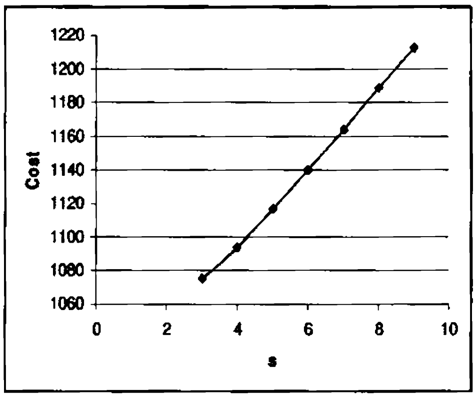
(a) (b)

Figure 5.2 (a), (b). Impact of the re-order level *s* on expected waiting time and on the loss rate of customers with parameters λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, S=15 : Both the waiting time and the loss rate are decreasing with increase in *s*. An increase in the expected inventory in the system brought by the increase in *s* can be thought of as the reason for the decrease in the total loss rate. A reference to table 4(a) shows that with an increase in *s*, there is a slight increase in the loss rate due to customers (both orbital and external) seeing an interrupted server. This together with the increase in the expected inventory level can be thought of as the reason behind the decrease in the waiting time. Note that the waiting time of an orbital customer may end with a quit from the system without receiving service.



(a)

CLOSS=10000



(b)

CLOSS=500

Figure 5.3 (a), (b). Investigation of an optimal value for the re-order level *s* with parameters λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, S=25, CN=50, CI=60, CINTR=40:

The cost curve in Figure (a), where the cost incurred due to customer loss is high (=10000), shows a convex nature for the cost function and an optimal value 5 for the re-order level *s*; while in Figure (b), as the loss cost is not very high (=500), the cost function is linearly increasing with increase in *s*.

5.5.5 Effect of the Maximum Inventory Level *S*

Table 5(a) shows that as the maximum inventory level *S* has only very little effect on majority of the system performance measures. The decrease in the effective replenishment rate EFRR and the increase in the expected inventory level in the system is quiet natural with increase in *S*. Because the inventory in the system is increasing, the loss due to zero inventories is decreasing. There is a very narrow increase in the server interruption probability, which may be due to the narrow increase in the number of orbital customers.

Table 5.5(a): Effect of the maximum inventory level *S* on various performance

measures with λ=2 , μ =4, θ=3, η=1, δ_2=2.5, δ_1 =2, p=0.5, q=0.6, s=10

S	P_β	P_α	E_{INTR}	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	E(σ)	E(sτ)	E(ω)
23	0.3332	0.2665	0.6663	0.1025	0.2665	0.4798	0.0007	0.7449	0.5338	15.666
24	0.3332	0.2665	0.6663	0.0952	0.2665	0.4798	0.0006	0.7449	0.5337	16.166
25	0.3332	0.2665	0.6664	0.0888	0.2665	0.4798	0.0006	0.7449	0.5336	16.666
26	0.3332	0.2665	0.6664	0.0833	0.2665	0.4798	0.0005	0.7449	0.5336	17.166
27	0.3332	0.2665	0.6664	0.0784	0.2665	0.4798	0.0005	0.7449	0.5336	17.666
28	0.3332	0.2665	0.6664	0.074	0.2666	0.4798	0.0005	0.7449	0.5335	18.166
29	0.3332	0.2666	0.6664	0.0701	0.2666	0.4798	0.0005	0.7449	0.5335	18.666
30	0.3332	0.2666	0.6664	0.0666	0.2666	0.4798	0.0004	0.7449	0.5335	19.166

Table5. 5(b): Effect of the maximum inventory level *S* on waiting time

λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, p=0.5, q=0.6, s=4

S	$E(W_L)$	$V(W_L)$	P_{WS}	E(σ)
10	0.2985	0.2042	0.1728	0.7517
11	0.2984	0.2014	0.1734	0.7507
12	0.2983	0.1993	0.1738	0.7500
13	0.2982	0.1976	0.1741	0.7494
14	0.2981	0.1963	0.1744	0.7490
15	0.2981	0.1952	0.1746	0.7486

5.5.6 Effect of the Joining Probability *p*

Table 6(a) and (b) studies the effect of the joining probability p of an arriving customer on the system behavior. Quiet naturally, the loss rate of arriving customers decreases with increase in the joining probability *p*. As more customers join the system, the expected number of customers E(σ), increases; the increase in E(σ) then leads to more retrials and therefore an increase in both the successful number of retrials E(sτ) and in the loss rate after retrials ER_{LOSS}. The increase in the number of orbital customers makes the server busier; but as the number of services increases, the probability of seeing an interrupted server P_α also increases. The possible increase in the number of services leads to an increase, though narrow, in the effective replenishment rate EFRR. For similar reasons, there is a decrease in the expected inventory level in the system and an increase in the loss rate of customers due to zero inventory EO_{LOSS}. Both these changes, especially that in EO_{LOSS}, are narrow which reflects the interruption factor affecting the system performance. Table 6(b) shows a slight decrease in the expected waiting time of a customer in the orbit; also note that the conditional probability that an orbital customer quits the system without receiving service is decreasing with increase in *p*. This is expected because when the joining probability *p* increases, the server busy probability shows a significant increase compared to the increase in the server interruption probability; and the customer loss, on arrival, takes place with probability 1-*p* only when the sever is interrupted. Thus increase in *p* leads to more service completions and this favors the system performance.

Table 5.6(a): Effect of the joining probability *p* on various performance measures

λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, q=0.6, s=10, S=31

p	P_β	P_α	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	E(σ)	E(sτ)	E(ω)
0	0.3186	0.2549	0.0607	0.5097	0.4588	0.0003	0.5180	0.4220	19.7257
0.2	0.3244	0.2596	0.0618	0.4153	0.4672	0.0004	0.6052	0.4666	19.7022
0.4	0.3303	0.2642	0.0629	0.3171	0.4756	0.0004	0.6971	0.5112	19.6788
0.6	0.3361	0.2689	0.0640	0.2151	0.4840	0.0004	0.7939	0.5556	19.6550
0.8	0.3419	0.2735	0.0651	0.1094	0.4923	0.0005	0.8955	0.5998	19.6324
1	0.3476	0.2781	0.0662	0	0.5006	0.0005	1.0020	0.6436	19.6094

Table 5.6(b): Effect of the joining probability *p* on waiting time

λ=2, μ =4, θ=3, η=1, δ_1=2, δ_2=2.5, q=0.6, s=4, S=10

p	E(W_L)	V(W_L)	P_{WS}	E(σ)
0	0.3067	0.2214	0.182	0.5251
0.2	0.3015	0.2114	0.1769	0.6122
0.4	0.2990	0.2058	0.1738	0.7040
0.6	0.2985	0.2033	0.1722	0.8006
0.8	0.2995	0.2030	0.1717	0.9020
1	0.3017	0.2044	0.1721	1.0080

5.5.7 Effect of the System Quitting Probability After an Unsuccessful Retrial *q*

After studying the effect of the joining probability *p*, now in Tables 7(a) and (b), we focus our attention on the effect of the system quitting probability *q* of an orbiting customer after an unsuccessful retrial. The increase in the loss rate after retrials ER_{LOSS} and a resulting decrease in the expected number of orbital customers E(σ) with increase in the loss probability *q* is obvious. The decrease in the number of customers results in a decrease in the server busy probability P_β and also in the server interruption probability P_α. Now this decrease in the probability P_α leads to a decrease in the loss rate of an arriving customer EA_{LOSS}. Note that the decrease in the server busy probability leads to an increase in the expected inventory level in the system and hence to a decrease in the effective replenishment rate EFRR and in the customer loss rate due to zero inventory in the system, EO_{LOSS}. Table (b) shows that an increase in *q* implies a decrease in the expected waiting time of an orbital customer; again this is not in favor of the customers as the probability that the customer quits the system before taken in to service is increasing.

Table 5.7(a): Effect of the system quitting probability *q* on various performance measures with $\lambda=2$, $\mu=4$, $\theta=3$, $\eta=1$, $\delta_2=2.5$, $\delta_1=2$, p=0.5, s=10, S=31

q	P_β	P_α	EFRR	EA_{LOSS}	ER_{LOSS}	EO_{LOSS}	E(σ)	E(sτ)	E(ω)
0	0.4163	0.3331	0.0793	0.3331	0	0.0170	3.6393	1.7600	19.335
0.2	0.3593	0.2874	0.0684	0.2874	0.1724	0.0006	1.2689	0.7324	19.563
0.4	0.3423	0.2739	0.0652	0.2739	0.3286	0.0005	0.9104	0.6031	19.631
0.6	0.3332	0.2666	0.0635	0.2666	0.4798	0.0004	0.7449	0.5334	19.667
0.8	0.3274	0.2619	0.0624	0.2619	0.6286	0.0004	0.6475	0.4891	19.690
1	0.3234	0.2587	0.0616	0.2587	0.776	0.0003	0.5829	0.4583	19.707

Table 5.7(b): Effect of the system quitting probability *q* on waiting time
λ=2, μ =4, 0=3, η=1, δ_1=2, δ_2=2.5, p=0.5, s=4, S=10

q	$E(W_L)$	$V(W_L)$	P_{WS}	$E(\sigma)$
0	1.3443	4.375	0	3.6609
0.2	0.4819	0.4784	0.1131	1.2753
0.4	0.3549	0.2728	0.1497	0.9171
0.6	0.2985	0.2042	0.1728	0.7517
0.8	0.2655	0.1702	0.1900	0.6542
1	0.2434	0.1497	0.2037	0.5895

5.6 Concluding Remarks

In this chapter we have considered an (s,S) inventory problem with positive service time and lead time. This is the first work in inventory with service interruption---server is subject to interruption while service is in progress. No waiting space is provided for customers, other than for the one whose service gets interrupted. Hence when a service is going on an external arrival has to go to an orbit of infinite capacity. Customers do not join the orbit when inventory level is zero nor when the server is under interruption. Retrial rate is a linear function of the number of customers present in the orbit. Retrial customers, encountering the server in breakdown condition, leave the system for ever with positive probability. This leads to the system being stable always. Also a primary customer, encountering the server in breakdown condition, chooses to leave the system with positive probability or joins the orbit with complementary probability. All distributions involved are assumed to be exponential.

This system is studied by analyzing a truncated system and then we extended the results to the system with unlimited capacity for the orbit. The system is shown to be stable whenever the probability of leaving system for ever with positive probability as a consequence of the retrial customer encountering the server in break down condition. In the absence of this explicit condition for stability is derived.

A first step to extend the results here is to replace a few of the exponential distributions assumed in the chapter by more general distributions, such as the phase type. If service time duration is assumed to be phase type or at least Erlang of order two or more, then a few of the phases could be provided with protection from interruption. The cost of such protection could be incorporated to investigate the optimal number of phases to be protected. The quality of approximation of the expected waiting time will be improved in a future work.

CHAPTER 6

Production Inventory with Service Time and Interruptions

6.1 Introduction

In all the studies on inventory systems prior to Berman et al [8], it was assumed that the serving of inventory is instantaneous. However this is not the case in many practical situations. For example in a TV showroom, a customer usually spends some time with the salesperson before buying the TV or in a computer shop, after selecting the model, one might have to wait until all the required software are installed. In Berman et al [8] it has been assumed that the amount of time taken to serve an item is constant. This leads to the analysis of a queue of demands formed in an inventory system. This study was followed by numerous studies by several researchers on many kinds of inventory models with positive service time. Krishnamoorthy and Viswanath [47] introduced the idea of positive service time in to a production inventory model by considering MAP arrivals and a correlated production process. This model being a very general one as far as the modeling parameters are considered, only a numerical study of the model was carried out there. In a very recent paper by Krishnamoorthy and Viswanath [48], assuming all the underlying distributions as exponential, a product form solution for the steady state has been obtained in a production inventory model with positive service time. The above paper had been motivated by the paper by Schwarz et al. [58], where a product form solution has been obtained in an (s, S) inventory model with positive service time.

The delay in the service caused by server interruptions being a common phenomenon in almost all practical situations, White and Christie [73] was the first study to introduce this in a queueing model. Following this, there had been extensive study on these type of queueing models. We refer to the survey paper by Krishnamoorthy and Pramod [38] for more details on such studies.

Though there had been numerous studies on inventory models, where interruption occurs due to an unreliable supplier [64,63,12] and the references therein], Krishnamoorthy et al [41] can be considered as the first paper to introduce the concept of service interruption, which occurs in the middle of a service, in an inventory system. They assume that there is no bound on the number of interruptions that can occur in the middle of a single service and also that an order is instantaneously processed (zero lead-time). The steady state distribution has been obtained explicitly in product form in the above paper. In another paper [40] by the same authors, the above model has been extended by considering positive lead-time.

In an (s, S) production inventory system, once the production process is switched on (as the inventory falls from S to s), it is switched off only after the inventory level goes back to S, the maximum inventory level. This makes it distinct from an (s, S) inventory system with positive lead-time, where once the order is placed (the moment at which the inventory level hits the re-order level s), usually the ordering quantity is taken such that the inventory level goes above s as the order materializes.

In a queueing system, where the service process has certain number of phases, which are subject to interruptions, the concept of protecting certain phases of service (which may be so costly to afford an interruption) from interruption could be an important idea. Klimenok et al [25] study a multi-server queue with finite buffer and negative customers where the arrival is BMAP and service is PH-type. They assume that a negative customer can delete an ordinary customer in service if the PH-service process belongs to some given subset of the set of service phases; whereas if the service process belongs to some phase outside the above subset, the ordinary customer is protected from the effect of the negative customers. The above paper is extended by assuming an infinite buffer in Klimenok and Dudin [24]. Krishnamoorthy et al [28] introduces the idea of protection in a queueing system where the service process is subject to interruptions. They assume that the final m-n phases of the Erlang service process are protected from interruption. Whereas if the service process belongs to the first n phases, it is subject to interruption and an interrupted service is resumed/repeated after some random time. There is no reduction in the number of customers due to interruption and no bound was assumed on the number of interruptions that can possibly occur in the middle of a service. In this way, this study differs from the earlier one where at most one interruption was possible in the middle of a service and where the customer whose service got interrupted is removed from the system.

This chapter introduces the concept of service interruption to a production inventory model with positive service time. The service time and the time to produce one item are assumed to follow distinct Erlang distributions. The service process as well as the production process is subject to interruptions and certain number of phases in both these processes are protected from interruption.

This chapter introduces a production process in to the sum of the inventory models studied earlier in this thesis. This adds an item to the inventory one at a time while the production is on. We use the (s,S) policy to control the production. This policy was used in the previous chapters for replenishment (with or with out lead type). Such a situation can be regarded as production in bulk in a production cycle. We incorporate interruption in production as well as in service. Another important entity that we introduce in this chapter is 'protection' against interruption, both for production and service processes. Unlike in previous chapters, were all distributions involved were exponential, in the present chapter we go for the Erlang distribution. We introduce protection from interruption to a few among the last stages of service and production. This involves economic consideration. We look for the optimal number of stages (phases) of service/production processes to be protected. Unlike in chapter 5 here the underlying Markov chain turns out to be level – independent quasi birth and death processes.

This chapter introduces a production process in to the sum of the inventory models studied earlier in this thesis. This adds an item to the

inventory one at a time while the production is on. We use the (s,S) policy to control the production. This policy was used in the previous chapters for replenishment (with or with out lead type). Such a situation can be regarded as production in bulk in a production cycle. We incorporate interruption in production as well as in service. Another important entity that we introduce in this chapter is 'protection' against interruption, both for production and service processes. Unlike in previous chapters, were all distributions involved were exponential, in the present chapter we go for the Erlang distribution. We introduce protection from interruption to a few among the last stages of service and production. This involves economic consideration. We look for the optimal number of stages (phases) of service/production processes to be protected. Unlike in chapter 5 here the underlying Markov chain turns out to be level – independent quasi birth and death processes.

An example for the applicability of the model: In the production process, assume that less expensive components of a system are assembled first. (These could be done without any protection). Next the expensive parts are to be assembled. These need protection from negligent handling. Thus such stages of the assembly are protected which involves additional cost. Similar example could be given in the case of service in phases with last few service phases protected.

We apply a novel method, which works even if we assume general PH distributions for the production as well as the service processes, for finding an explicit expression for the stability of the system. Studies like [3, 4], have analyzed inventory system where customers are not allowed to join the system, when there is a shortage of inventory and had found that the stability of such systems is not affected by the inventory parameters. However, in the above studies, the underlying distributions were all exponential. Our proof for the stability of the system shows that the above phenomenon holds even if the underlying distributions are general PH distributions and hence it gives a characterization of the stability of inventory systems where the customers are not allowed to join the system when there is shortage of inventory.

In the section to follow, the mathematical formulation of the model is provided. Section 3 is concerned with the investigation of the stability of the system. The long run system state distribution is also given in that section. In section 4, numerous system performance measures are provided. Numerical investigation of performance measures is extensively discussed in section 5. Finally, section 6 concludes the discussion.

6.2. The Mathematical Model

The model under study is described as follows: Customers arrive to a single server counter according to a Poisson process of rate λ where inventory is served. Service time duration follows Erlang distribution, with Phase-type representation (T, β) of order m where β = (1,0,...,0) and

$$T = \begin{bmatrix} -\mu_1 & \mu_1 & 0 & . & 0 \\ 0 & -\mu_1 & \mu_1 & 0 & . \\ 0 & 0 & . & . & 0 \\ 0 & . & . & -\mu_1 & \mu_1 \\ 0 & & & 0 & -\mu_1 \end{bmatrix}.$$

Production is by one unit at a time. The production process starts whenever the inventory level falls to s and continues until the inventory level reaches S. The production process follows Erlang distribution with Phase-type representation (U, α) of order n with α = (1,0,...,0) and

$$U = \begin{bmatrix} -\mu & \mu & 0 & . & 0 \\ 0 & -\mu & \mu & 0 & . \\ 0 & 0 & . & . & 0 \\ 0 & . & & -\mu & \mu \\ 0 & & . & 0 & -\mu \end{bmatrix}.$$

Interruption to the service process occurs according to a Poisson process of rate δ_1; the server is subject to at most one interruption at a time and, further only when service is going on, the server is subject to interruption (that is an idle server is not affected by the interruption process). An exponentially distributed amount of time with parameter δ_2 is required to resume service from where it was stopped. That is the system recovers from the interruption after a repair having exponentially distributed duration with parameter δ_2. Similarly, the production process also encounters interruptions, with the interruption process following Poisson process with parameter δ_3 and recovers from it on being repaired, with repair time following an exponential distribution with parameter δ_4. In contrast to the case of service interruption, after repair an interrupted production process needs to be restarted from the beginning. In other words, we assume that an item being produced is discarded due to an interruption. For reducing the adverse effect of interruptions, we apply the concept of protection of certain phases of service as well as the production process from interruptions. Precisely, the last l_1 phases of the service process are assumed to be protected in the sense that the service will not be interrupted while being in these phases; so interruptions to the service can occur only while in service in the first $m-l_1$ phases. Similarly, the final l phases of the production process are assumed to be protected in the sense that the item being produced will not be affected and the processor will not be subject to interruption while being in these phases; so interruptions to production can occur in the first $n-l$ phases. It is important to note our assumption concerning the two interruption processes: while in interruption (service or production) another interruption cannot befall in the sense that the system behaves like a Type I counter (see Karlin and Taylor [14]). Further, a customer's service may encounter any number of interruptions. However, it may be noted that since the item being produced is discarded consequent to an interruption of the production process, there can be at most one interruption to a unit being produced. All distributions involved are assumed to be mutually independent.

For the model under discussion, we make the following assumptions:

- No inventory is lost due to a service interruption.
- The customer being served when interruption occurs, waits there until his service is completed.
- The inventory being produced is lost due to a production interruption.
- A customer, who finds no inventory on his /her arrival, leaves the system forever.
- Only when production is on, that process could get interrupted and the service gets interrupted only while the server is busy providing service.

Let $N_C(t)$ denote the number of customers in the system including the one getting service (if any) and $N_I(t)$ denote the inventory level in the system at time t. Further let

$$V(t) = \begin{cases} 0 & \text{if the server is idle} \\ 1 & \text{if the server is busy} \\ 2 & \text{if the server is on interruption} \end{cases}$$

$$P(t) = \begin{cases} 0 & \text{if the production process is in } \textit{off} \text{ mode} \\ 1 & \text{if the production process is in } \textit{on} \text{ mode} \end{cases}$$

$$V_1(t) = \begin{cases} 0 & \text{if the production process is not interrupted/off} \\ 1 & \text{if the production process is interrupted} \end{cases}$$

Finally, let $Z_1(t)$ and $Z_2(t)$ denote the phases of the service and production processes, respectively. Then $\Psi = \{(N_C(t), V(t), N_I(t), P(t), V_1(t), Z_1(t), Z_2(t));\ t \geq 0\}$ forms a continuous time Markov chain on the state space $\bigcup_{i=0}^{\infty} \tilde{L}(i)$, where $\tilde{L}(i)$'s, which are called the levels, are the collection of states defined as follows:

$$\tilde{L}(0) = \hat{L}(0,0) = \bigcup_{j=0}^{S} \bar{L}(0,0,j)$$

For $0 \leq j \leq s$, $\bar{L}(0,0,j) = \bar{\bar{L}}(0,0,j,1) = L(0,0,j,1,0) \cup L(0,0,j,1,1)$,

for $s+1 \leq j \leq S-1$, $\bar{L}(0,0,j) = L(0,0,j,0) \cup L(0,0,j,1,0) \cup L(0,0,j,1,1)$, and

$$\bar{L}(0,0,S) = L(0,0,S,0).$$

For $i \geq 1$, $\tilde{L}(i) = \hat{L}(i,0) \cup \hat{L}(i,1) \cup \hat{L}(i,2)$, where

$$\hat{L}(i,0) = \bar{L}(i,0,0) = \bar{\bar{L}}(i,0,0,1) = L(i,0,0,1,0) \cup L(i,0,0,1,1)$$

$$\hat{L}(i,l') = \bigcup_{j=1}^{S} \bar{L}(i,l',j),\ l' = 1,2, \text{ where}$$

$$\bar{L}(i,l',j) = \bar{\bar{L}}(i,l',j,1) = L(i,l',j,1,0) \cup L(i,l',j,1,1),\ 1 \leq j \leq s$$

$$\bar{L}(i,l',j) = L(i,l',j,0) \cup L(i,l',j,1,0) \cup L(i,l',j,1,1),\ s+1 \leq j \leq S-1$$

$$\bar{L}(i,l',S) = L(i,l',S,0)$$

Finally, with δ_{ij} as Kronecker delta,

$$L(i,0,0,1,j_1) = \{(i,0,0,1,j_1,j_2); 1 \leq j_2 \leq n - \delta_{1,j_1} l\},\ i \geq 0,\ j_1 = 0,1$$

$$L(0,0,j,1,j_1) = \{(0,0,j,1,j_1,j_2); 1 \leq j_2 \leq n - \delta_{1,j_1} l\},\ 1 \leq j \leq S-1,\ j_1 = 0,1$$

$$L(0,0,j,0) = L(0,0,j,0,0) = \{(0,0,j,0,0)\},\ s+1 \leq j \leq S$$

$$L(i,l',j,1,j_1)=\{(i,l',j,1,j_1,j_2,j_3);\ 1\le j_2\le m-\delta_{2,l'}l_1,\ 1\le j_3\le n-\delta_{1,j_1}l\},$$
$$i\ge 1;\ l'=1,2;\ 1\le j\le S-1;\ j_1=0,1$$

$$L(i,l',j,0)=\{(i,l',j,0,0,j_2);\ 1\le j_2\le m-\delta_{2,l'}l_1\},\ i\ge 1;\ l'=1,2;\ s+1\le j\le S.$$

It turns out that the continuous-time Markov chain Ψ is a Level Independent Quasi Birth- Death (LIQBD) process with infinitesimal generator given by

$$W=\begin{bmatrix} A_{00} & B_{00} & 0 & 0 & 0 & 0 & 0 \\ A_{10} & A_1 & A_0 & 0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & 0 \\ & & & & & & \cdot \\ & & & & & & \cdot \\ & & & & & & \cdot \end{bmatrix},$$

where the matrix A_{00} records the transition rates within the level $\tilde{L}(0)$, B_{00} those from level $\tilde{L}(0)$ to $\tilde{L}(1)$; A_{10} governs transitions from $\tilde{L}(1)$ to $\tilde{L}(0)$ and the matrices A_1 and A_0 constitute respectively, the transition rates within a level $\tilde{L}(i)$ and from level $\tilde{L}(i)$ to $\tilde{L}(i+1)$ for $i\ge 1$. Finally, the matrix A_2 governs transitions from level $\tilde{L}(i)$ to $\tilde{L}(i-1)$ for $i\ge 2$. A detailed description of the transitions that govern the generator matrix W can be found in Appendix I.

In the sequel, Q = S-s, I_n denotes the identity matrix of order n, e denotes a column matrix of 1's of appropriate order, e_n denotes a column matrix of 1's of order $n\times 1$ and 0_n denotes a zero matrix of order $n\times n$.

6.3. Analysis of the Model

6.3.1 Stability Condition

In section 2, we have assumed that the service time follows a PH distribution of order m, with representation (T,β). It follows that the service process which is subject to possible interruptions has a PH distribution with representation (T^*,β^*), where

$$\beta^* = (\beta,0) \text{ and } T^* = \begin{bmatrix} T-\delta_1 J_1 & \delta_1 J_2 \\ \delta_2 J_3 & \delta_2 I_{(m-l_1)} \end{bmatrix}, \text{ with } J_1 = \begin{bmatrix} I_{(m-l_1)} & 0 \\ 0 & 0 \end{bmatrix}_{(m\times m)},$$

$$J_2 = \begin{bmatrix} I_{(m-l_1)} \\ 0 \end{bmatrix}_{(m\times m-l_1)}, J_3 = \begin{bmatrix} I_{(m-l_1)} & 0 \end{bmatrix}_{(m-l_1\times m)}.$$

Let T^{0^*} be the column matrix such that $T^*e+T^{0^*}=0$ and let $2m-l_1=m'$.

Similarly the production process subject to possible interruptions also has a PH distribution with representation (U^*,α^*) where $\alpha^*=(\alpha,0)$ and

$$U^* = \begin{bmatrix} U-\delta_3 J_4 & \delta_3 J_5 \\ \delta_4 J_6 & -\delta_4 I_{(n-l)} \end{bmatrix}, J_4 = \begin{bmatrix} I_{n-l} & 0 \\ 0 & 0 \end{bmatrix}_{(n\times n)},$$

$$J_5 = \begin{bmatrix} I_{n-l} \\ 0 \end{bmatrix}_{(n\times n-l)}, J_6 = \begin{bmatrix} e_{n-l} & 0 \end{bmatrix}_{(n-l\times n)}.$$

Let U^{0^*} be the column matrix such that $U^*e+U^{0^*}=0$ and let $2n-l=n'$.

Then $\Psi^* = (N(t),L(t),P(t),Z_1^*(t),Z_2^*(t))$, where $Z_1^*(t)$ and $Z_2^*(t)$ denote the phases of the above service and production process, models the same Markov process as Ψ. The infinitesimal generator matrix of the process Ψ^* is given by

$$W^* = \begin{bmatrix} A^*_{00} & B^*_{00} & 0 & 0 & 0 & 0 & 0 \\ A^*_{10} & A^*_1 & A^*_0 & 0 & 0 & 0 & 0 \\ 0 & A^*_2 & A^*_1 & A^*_0 & 0 & 0 & 0 \\ 0 & 0 & A^*_2 & A^*_1 & A^*_0 & 0 & 0 \\ & & & & & \cdot & \\ & & & & \cdot & \cdot & \\ & & & & \cdot & & \cdot \end{bmatrix}.$$

Since the form of the matrices A^*_{00}, B^*_{00} and A^*_{10} will not affect the stability of the Markov chain Ψ^*, we do not give their detailed description here; the other block matrices are given as follows:

$$A^*_0 = \begin{bmatrix} 0 & 0 \\ 0 & \lambda I \end{bmatrix};$$

$$A^*_2 = \begin{bmatrix} 0 & & & & & & & & & \\ A_2^{*(0,0)}(1) & 0 & & & & & & & & \\ & A_2^{*(0,1)}(1) & 0 & & & & & & & \\ & & \cdot & & & & & & & \\ & & & A_2^{*(0,1)}(1) & & & & & & \\ & & & & A_2^{*(0,2)}(1) & 0 & & & & \\ & & & & & A_2^{*(0,3)}(1) & 0 & & & \\ & & & & & & \cdot & & & \\ & & & & & & A_2^{*(0,3)}(1) & 0 & & \\ & & & & & & & A_2^{*(0,4)}(1) & 0 & \end{bmatrix}$$

,

where $A_2^{*(0,0)}(1) = T^{0^*} \otimes I_{n'}$, $A_2^{*(0,1)}(1) = T^{0^*}\beta^* \otimes I_{n'}$, $A_2^{*(0,4)}(1) = \begin{bmatrix} T^{0^*}\beta^* & 0 \end{bmatrix}$,

$A_2^{*(0,2)}(1) = \begin{bmatrix} T^{0^*}\beta^* \otimes \alpha^* \\ T^{0^*}\beta^* \otimes I_{n'} \end{bmatrix}$, $A_2^{*(0,3)}(1) = \begin{bmatrix} T^{0^*}\beta^* & 0 \\ 0 & T^{0^*}\beta^* \otimes I_{n'} \end{bmatrix}$ and

$$A^*_1 = \begin{bmatrix} A_1^{*(0,0)}(1) & A_1^{*(1,0)}(1) & & & & & & \\ & A_1^{*(0,1)}(1) & A_1^{*(1,1)}(1) & & & & & \\ & & \ddots & & & & & \\ & & & A_1^{*(0,1)}(1) & A_1^{*(1,2)}(1) & & & \\ & & & & A_1^{*(0,2)}(1) & A_1^{*(1,3)}(1) & & \\ & & & & & \ddots & \ddots & \\ & & & & & & A_1^{*(0,2)}(1) & A_1^{*(1,4)}(1) \\ & & & & & & & A_1^{*(0,3)}(1) \end{bmatrix}$$

with $A_1^{*(0,2)}(1) = \begin{bmatrix} T^* - \lambda I_{m'} & 0 \\ 0 & T^* \oplus U^* - \lambda I_{m'n'} \end{bmatrix}$; $A_1^{*(0,0)}(1) = U^*$; $A_1^{*(0,1)}(1) = T^* \oplus U^*$;

$A_1^{*(0,3)}(1) = T^* - \lambda I_{m'}$; $A^{*}_1{}^{(1,0)}(1) = \beta^* \otimes \left(U^{0^*} \alpha^*\right)$; $A^{*}_1{}^{(1,1)}(1) = I_{m'} \otimes \left(U^{0^*} \alpha^*\right)$;

$A^{*}_1{}^{(1,2)}(1) = \begin{bmatrix} 0 & I_{m'} \otimes \left(U^{0^*} \alpha^*\right) \end{bmatrix}$; $A^{*}_1{}^{(1,3)}(1) = \begin{bmatrix} 0 & 0 \\ 0 & I_{m'} \otimes \left(U^{0^*} \alpha^*\right) \end{bmatrix}$;

$A^{*}_1{}^{(1,4)}(1) = \begin{bmatrix} 0 \\ I_{m'} \otimes \left(U^{0^*}\right) \end{bmatrix}$.

Let $A = A^*_2 + A^*_1 + A^*_0$ and $x = (x_0, x_1, \ldots, x_s, x_{s+1,0}, x_{s+1,1}, \ldots, x_{S-1,0}, x_{S-1,1}, x_S)$ be the steady state vector of A, where $x_{i,0}$ and $x_{i,1}$, $s+1 \le i \le S-1$, represent the probabilities corresponding to the production *off* and *on* states respectively. Then $xA = 0$ gives us the following system of equations.

$$x_0 U^* + x_1 T^{0^*} \otimes I_{n'} = 0 \tag{3.1.1}$$

$$x_0 \beta^* \otimes U^{0^*} \alpha^* + x_1 (T^* \oplus U^*) + x_2 T^{0^*} \beta^* \otimes I_{n'} = 0 \tag{3.1.2}$$

$$x_{i-1} I_{m'} \otimes U^{0^*} \alpha^* + x_i (T^* \oplus U^*) + x_{i+1} T^{0^*} \beta^* \otimes I_{n'} = 0, \; 2 \le i \le s-1 \tag{3.1.3}$$

$$x_{s-1} I_{m'} \otimes U^{0^*} \alpha^* + x_s (T^* \oplus U^*) + x_{s+1,0} T^{0^*} \beta^* \otimes \alpha^* + x_{s+1,1} T^{0^*} \beta^* \otimes I_{n'} = 0 \tag{3.1.4}$$

$$x_s I_{m'} \otimes U^{0^*}\alpha^* + x_{s+1,1}(T^* \oplus U^*) + x_{s+2,1} T^{0^*}\beta^* \otimes I_{n'} = 0 \quad (3.1.5)$$

$$x_{i-1,1} I_{m'} \otimes U^{0^*}\alpha^* + x_{i,1}(T^* \oplus U^*) + x_{i+1,1} T^{0^*}\beta^* \otimes I_{n'} = 0, \ s+2 \le i \le S-2 \quad (3.1.6)$$

$$x_{S-2,1} I_{m'} \otimes U^{0^*}\alpha^* + x_{S-1,1}(T^* \oplus U^*) = 0 \quad (3.1.7)$$

$$x_{S-1,1} I_{m'} \otimes U^{0^*} + x_S T^* = 0 \quad (3.1.8)$$

$$x_{i,0} T^* + x_{i+1,0} T^{0^*}\beta^* = 0, \ s+1 \le i \le S-2 \quad (3.1.9)$$

$$x_{S-1,0} T^* + x_S T^{0^*}\beta^* = 0. \quad (3.1.10)$$

Noting that $T^* e = -T^{0^*}$ and that $\beta^* e = 1$, right multiply each equation in the system of equations (3.1.9) and (3.1.10) by the column vector e to obtain the system of equations:

$$x_{i,0} T^{0^*} = x_{i+1,0} T^{0^*}, \ s+1 \le i \le S-2 \quad (3.1.11)$$

$$x_{S-1,0} T^{0^*} = x_S T^{0^*} \quad (3.1.12)$$

Now, (3.1.2) and (3.1.3) together result in:

$$x_{i,0}\left(T^* + T^{0^*}\beta^*\right) = 0, \ s+1 \le i \le S-1. \quad (3.1.13)$$

From (3.1.11) and (3.1.13), we get that $x_{s+1,0} = x_{s+2,0} = ... = x_{S-1,0}$, which is the left eigen vector of the irreducible generator matrix $(T^* + T^{0^*}\beta^*)$.

Right multiplying equation (3.1.1) by $e_{n'}$, we get the equation

$$-x_0 U^{0^*} + x_1 T^{0^*} \otimes e_{n'} = 0$$

Noticing that $\beta^* \otimes x_0 U^{0^*} = x_0 \beta^* \otimes U^{0^*}$ and that $\beta^* \otimes T^{0^*} = T^{0^*}\beta^*$, we take the left Kronecker product in the above equation with β^* to obtain the equation:

$$-x_0 \beta^* \otimes U^{0^*} + x_1 T^{0^*}\beta^* \otimes e_{n'} = 0$$

(3.1.14)

Right multiplying each equation (3.1.2) to (3.1.7) by $I_{m'} \otimes e_{n'}$, we get the following equations

$$x_0 \beta^* \otimes U^{0^*} + x_1(T^* \otimes e_{n'}) - x_1(I_{m'} \otimes U^{0^*}) + x_2 T^{0^*} \beta^* \otimes e_{n'} = 0$$

(3.1.15)

$$x_{i-1} I_{m'} \otimes U^{0^*} + x_i(T^* \otimes e_{n'}) - x_i(I_{m'} \otimes U^{0^*}) + x_{i+1} T^{0^*} \beta^* \otimes e_{n'} = 0, \; 2 \le i \le s-1$$

(3.1.16)

$$x_{s-1} I_{m'} \otimes U^{0^*} + x_s(T^* \otimes e_{n'}) - x_s(I_{m'} \otimes U^{0^*}) + x_{s+1,0} T^{0^*} \beta^* + x_{s+1,1} T^{0^*} \beta^* \otimes e_{n'} = 0$$

(3.1.17 A)

$$x_s I_{m'} \otimes U^{0^*} + x_{s+1,1}(T^* \otimes e_{n'}) - x_{s+1,1}(I_{m'} \otimes U^{0^*}) + x_{s+2,1} T^{0^*} \beta^* \otimes e_{n'} = 0$$

(3.1.18)

$$x_{i-1,1} I_{m'} \otimes U^{0^*} + x_{i,1}(T^* \otimes e_{n'}) - x_{i,1}(I_{m'} \otimes U^{0^*}) + x_{i+1,1} T^{0^*} \beta^* \otimes e_{n'} = 0, \; s+2 \le i \le S-2$$

(3.1.19)

$$x_{S-2,1} I_{m'} \otimes U^{0^*} + x_{S-1,1}(T^* \otimes e_{n'}) - x_{S-1,1}(I_{m'} \otimes U^{0^*}) = 0$$

(3.1.20)

From equations (3.1.11) and (3.1.12), we note that $x_{s+1,0} T^{0^*} = x_{S-1,0} T^{0^*} = x_S T^{0^*}$; using which, we re-write the equation (3.1.17 A) as

$$x_{s-1} I_{m'} \otimes U^{0^*} + x_s(T^* \otimes e_{n'}) - x_s(I_{m'} \otimes U^{0^*}) + x_S T^{0^*} \beta^* + x_{s+1,1} T^{0^*} \beta^* \otimes e_{n'} = 0$$

(3.1.17)

Now, adding equations (3.1.8) and (3.1.13) to (3.1.20), we get the equation

$$(x_1 + x_2 + \ldots + x_s + x_{s+1,1} + \ldots + x_{S-1,1})((T^* + T^{0^*} \beta^*) \otimes e_{n'}) +$$
$$(x_{s+1,0} + \ldots + x_{S-1,0} + x_S)(T^* + T^{0^*} \beta^*) = 0$$

Noticing that, for any row vector ξ of dimension $1 \times m'n'$, we have

$\xi((T^* + T^{0^*}\beta^*) \otimes e_{n'}) = (\xi(I_{m'} \otimes e_{n'}))(T^* + T^{0^*}\beta^*)$, we write the above equation as,

$$((x_1 + x_2 + ... + x_s + x_{s+1,1} + ... + x_{S-1,1})(I_{m'} \otimes e_{n'}))(T^* + T^{0^*}\beta^*) +$$
$$(x_{s+1,0} + ... + x_{S-1,0} + x_S)(T^* + T^{0^*}\beta^*) = 0.$$

(3.1.21)

Since the generator matrix $T^* + T^{0^*}\beta^*$ is irreducible, equation 3.1.21 implies that

$$(x_1 + x_2 + ... + x_s + x_{s+1,1} + ... + x_{S-1,1})(I_{m'} \otimes e_{n'}) + (x_{s+1,0} + ... + x_{S-1,0} + x_S) = a\rho,$$

where ρ is the steady state vector of the generator matrix $T^* + T^{0^*}\beta^*$ and a is a scalar. Now, since $xe = 1$, it follows that $a = 1 - x_0 e_{n'}$.

The Markov chain Ψ^* and hence the Markov chain Ψ are stable if and only if $xA^*_0 e < xA^*_2 e$ (see Neuts [15]). From the structure of the matrices A^*_0 and A^*_2, it follows that,

$xA^*_0 e = (1 - x_0 e_{n'})\lambda$ and

$$xA^*_2 e = ((x_1 + x_2 + ... + x_s + x_{s+1,1} + ... + x_{S-1,1})(I_{m'} \otimes e_{n'}) + (x_{s+1,0} + ... + x_{S-1,0} + x_S))T^{0^*}$$
$$= (1 - x_0 e_{n'})\rho T^{0^*}$$

Hence the inequality $xA^*_0 e < xA^*_2 e$ reduces to

$$(1 - x_0 e_{n'})\lambda < (1 - x_0 e_{n'})\rho T^{0^*}.$$

That is, the Markov chain Ψ is stable if and only if $\lambda < \rho T^{0^*}$. We summarize the above as:

Theorem 6.3.1

A necessary and sufficient condition for the stability of the Markov chain Ψ is $\lambda < \rho T^{0^*}$.

Note:

We notice that the above stability condition is the same as that for an $M/PH/1$ queueing system, where the arrival process is Poisson with parameter λ and the service time follows a Phase-type distribution with representation (T^*, β^*) of order m'. Also, notice that the production process as well as the re-order and maximum inventory levels have no influence on the stability of the model studied. The proof of Theorem 3.1 reveals that barring customers from joining the system when there is shortage of inventory is the reason behind this phenomenon and hence characterizes the stability of such inventory systems.

6.3.2 Computation of the Steady State Vector

Next we compute the steady state vector of Ψ numerically. Let $\pi = (\pi_0, \pi_1, \pi_2,$ ...), be the steady state vector of Ψ, where $\pi_0 = \pi_{00} = (\pi_{00}(1), \pi_{00}(2), \pi_{00}(3))$;

$\pi_{00}(1) = (\pi_{00}(0,1,0,*), \pi_{00}(0,1,1,\bullet), \pi_{00}(1,1,0,*), \pi_{00}(1,1,1,\bullet),....., \pi_{00}(s,1,0,*), \pi_{00}(s,1,1,\bullet))$;

$\pi_{00}(2) = (\pi_{00}(s+1,0,0), \pi_{00}(s+1,1,0,*), \pi_{00}(s+1,1,1,\bullet), \pi_{00}(s+2,0,0), \pi_{00}(s+2,1,0,*),$ $\pi_{00}(s+2,1,1,\bullet),....., \pi_{00}(S-1,0,0), \pi_{00}(S-1,1,0,*), \pi_{00}(S-1,1,0,\bullet))$; $\pi_{00}(3) = \pi_{00}(S,0,0)$. Here $\pi_{00}(i,1,0,*)$ is a row vector of dimension n and that of $\pi_{00}(i,1,1,\bullet)$ is n-l. For $r \geq 1$,

$\pi_{r0}(0) = (\pi_{r0}(0,1,0,*), \pi_{r0}(0,1,0,\bullet))$;

$\pi_{r0}(1) = (\pi_{r0}(1,1,0,\circ,*), \pi_{r0}(1,1,1,\circ,\bullet), \pi_{r0}(2,1,0,\circ,*), \pi_{r0}(2,1,1,\circ,\bullet),......., \pi_{r0}(s,1,0,\circ,*), \pi_{r0}(s,1,1,\circ,\bullet))$; $\pi_{r0}(2) = (\pi_{r0}(s+1,0,0,\circ), \pi_{r0}(s+1,1,0,\circ,*), \pi_{r0}(s+1,1,1,\circ,\bullet),$

$\pi_{r0}(s+2,0,0,\circ), \pi_{r0}(s+2,1,0,\circ,*), \pi_{r0}(s+2,1,1,\circ,\bullet),.......,$ $\pi_{r0}(S-1,0,0,\circ), \pi_{r0}(S-1,1,0,\circ,*), \pi_{r0}(S-1,1,1,\circ,\bullet))$; $\pi_{r0}(3) = (\pi_{r0}(S,0,0,\circ)$.

Here $\pi_{r0}(0,1,0,*)$ is a row vector of length n, $\pi_{r0}(0,1,0,\bullet)$ is a row vector of length n-l, $\pi_{r0}(i,1,0,\circ,*)$ is a row vector of length mn, $\pi_{r0}(i,1,1,\circ,\bullet)$ is a row vector of length $m(n-l)$, $\pi_{r0}(i,0,0,\circ)$ is a row vector of length m , The description of π_{r1} is identical to that of π_{r0} except for $\pi_{r1} = (\pi_{r1}(1), \pi_{r1}(2), \pi_{r1}(3))$ and $m - l_1$ coming in place of m.

The sub vectors π_i of the vector π are given by $\pi_i = -\pi_0 B_{00}\left(A_1 + RA_2\right)^{-1} R^{i-1},\ i \geq 1$, where R is the minimal non-negative solution of the matrix-quadratic equation $R^2 A_2 + RA_1 + A_0 = 0$ (see Neuts [15]). The sub vector π_0 is first found as the steady state probability vector of the generator matrix $A_{00} - B_{00}\left(A_1 + RA_2\right)^{-1} A_{10}$ and then normalized using the condition $\pi_0 e + \sum_{i \geq 1} \pi_i e = \pi_0 e + \pi_1 \left(I - R\right)^{-1} e = 1$. For computing the R matrix, we applied the logarithmic reduction algorithm by Latouche and Ramaswami [17].

6.4. System Performance Measures

1. The probability that server is busy

$$PSB = \sum_{r \geq 0} \pi_{r0}\, e$$

2. The probability that server is on interruption

$$PSI = \sum_{r \geq 1} \pi_{r1}\, e$$

3. The expected inventory level when there are no customers in the system

$$EIL\,(0) = \sum_{i=1}^{S-1}\sum_{k=1}^{n} i\pi_{00}(i,1,0,k) + \sum_{i=1}^{S-1}\sum_{k=1}^{n-l} i\pi_{00}(i,1,1,k) + \sum_{i=s+1}^{S-1} i\pi_{00}(i,0,0) +$$

$$S\pi_{00}(S,0,0)$$

4. The expected inventory level when there is at least one customer in the system and server is busy

$$EIL\,(1) = \sum_{r\geq 1}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n} i\pi_{r0}(i,1,0,k_1,k_2) + \sum_{r\geq 1}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n-l} i\pi_{r0}(i,1,1,k_1,k_2) +$$

$$\sum_{r\geq 1}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m} i\pi_{r0}(i,0,0,k_1) + \sum_{r\geq 1}\sum_{k_1=1}^{m} i\pi_{r0}(S,0,0,k_1)$$

5. The expected inventory level when there is at least one customer in the system and

server is on interruption

$$EIL\,(2) = \sum_{r\geq 1}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n} i\pi_{r1}(i,1,0,k_1,k_2) + \sum_{r\geq 1}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n-l} i\pi_{r1}(i,1,1,k_1,k_2) +$$

$$\sum_{r\geq 1}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m-l_1} i\pi_{r1}(i,0,0,k_1) + \sum_{r\geq 1}\sum_{k_1=1}^{m-l_1} S\pi_{r1}(S,0,0,k_1)$$

6. The expected inventory level in the system

$EIL= EIL\,(0) + EIL\,(1) + EIL\,(2)$

7. The expected number of customers in the system

$$ENCS = \sum_{r\geq 0} r\pi_{r0}e + \sum_{r\geq 0} r\pi_{r1}e$$

8. The expected production switch *off* rate

$$PSWOF = \pi_{00}(S-1,1,0,\bullet)U^0 + \sum_{r=1}^{\infty}\pi_{r0}(S-1,1,0,\bullet,\bullet)e_m \otimes U^0$$

$$+\sum_{r=1}^{\infty}\pi_{r1}(S-1,1,0,\bullet,\bullet)e_{m-l_1} \otimes U^0\,.$$

9. The expected production commencement rate

$$PCOM = \sum_{r=1}^{\infty}\pi_{r0}(s+1,0,0,k_1)T^0$$

10. The expected interruption rate of production

$$EIRP = \delta_3 \{ \sum_{i=0}^{S-1}\sum_{k=1}^{n-l}\pi_{00}(i,1,1,k) + \sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n-l}\pi_{r0}(i,1,1,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n-l}\pi_{r1}(i,1,1,k_1,k_2) \}$$

11. The expected rate of interruption of service

$$EIRS = \delta_1 \sum_{r\geq 1} \pi_{r1} e$$

12. Fraction of time service is in protected stage

$$FSP = \sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=m-l_1+1}^{m}\sum_{k_2=1}^{n}\pi_{r0}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=m-l_1+1}^{m}\pi_{r0}(i,0,0,k_1) +$$

$$\sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=m-l_1+1}^{m}\sum_{k_2=1}^{n-l}\pi_{r0}(i,1,1,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{k_1=m-l_1+1}^{m}\pi_{r0}(S,0,0,k_1)$$

13. Fraction of time production is in protected stage

$$FPP = \sum_{i=0}^{S-1}\sum_{k=n-l+1}^{n}\pi_{00}(i,1,0,k) + \sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=n-l+1}^{n}\pi_{r0}(i,1,0,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{i=0}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=n-l+1}^{n}\pi_{r1}(i,1,0,k_1,k_2)$$

14. Loss rate of customers due to zero inventory

$$LZI = \lambda\{ \sum_{k=1}^{n}\pi_{00}(0,1,0,k) + \sum_{k=1}^{n-l}\pi_{00}(0,1,1,k) + \sum_{r=1}^{\infty}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n}\pi_{r0}(0,1,0,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n-l}\pi_{r0}(0,1,1,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n}\pi_{r1}(0,1,0,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n-l}\pi_{r1}(0,1,1,k_1,k_2) \}.$$

15. Probability that inventory level is greater than s and production is in *off* mode

$$P_{prod\,off}\,(I>s) = \sum_{i=s+1}^{S-1} \pi_{00}(i,0,0) + \pi_{00}(S,0,0) + \sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m}\pi_{r0}(i,0,0,k_1) +$$

$$\sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m-l_1}\pi_{r1}(i,0,0,k_1) + \sum_{r=1}^{\infty}\sum_{k_1=1}^{m}\pi_{r0}(S,0,0,k_1)$$

16. Probability that inventory level is greater than s and production is in *on* mode

$$Pprod_{on}\,(I>s) = \sum_{i=s+1}^{S-1}\sum_{k=1}^{n}\pi_{00}(i,1,0,k) + \sum_{i=s+1}^{S-1}\sum_{k=1}^{n-l}\pi_{00}(i,1,1,k) +$$

$$\sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n}\pi_{r0}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2=1}^{n-l}\pi_{r0}(i,1,1,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n}\pi_{r1}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=s+1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n-l}\pi_{r1}(i,1,1,k_1,k_2)$$

17. Probability that inventory level is greater than s,

$P\,(I>s) = P_{prod\,off}\,(I>s) + P_{prod\,on}\,(I>s)$.

18. Fraction of time production is in *on* mode

$$FPon = \sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2}^{n}\pi_{r0}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m}\sum_{k_2}^{n-l}\pi_{r0}(i,1,1,k_1,k_2) +$$

$$\sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2}^{n}\pi_{r1}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2=1}^{n-l}\pi_{r1}(i,1,1,k_1,k_2) +$$

$$\sum_{i=0}^{S-1}\sum_{k=1}^{n}\pi_{00}(i,1,0,k) + \sum_{i=0}^{S-1}\sum_{k=1}^{n-l}\pi_{00}(i,1,1,k)$$

19. Fraction of time server is interrupted and production is in *on* mode

$$FPon/SI = \sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2}^{n}\pi_{r1}(i,1,0,k_1,k_2) + \sum_{r=1}^{\infty}\sum_{i=1}^{S-1}\sum_{k_1=1}^{m-l_1}\sum_{k_2}^{n-l}\pi_{r1}(i,1,1,k_1,k_2)$$

20. Fraction of time production is in *on* mode with no customers in the system

$$FPon/0C = \sum_{i=0}^{S-1}\sum_{k=1}^{n}\pi_{00}(i,1,0,k) + \sum_{i=0}^{S-1}\sum_{k=1}^{n-l}\pi_{00}(i,1,1,k)$$

21. Fraction of time production is in *on* mode with server active

FPon / SA= FPon - (FPon/0C + FPon/SI)

6.5. Numerical Illustration

In this section, we provide the results of the numerical experiments that has been carried out for studying the impact of different parameters on various system performance measures.

6.5.1 Effect of the Service Interruption Rate δ_1

Intuitively, as the interruption rate δ_1 increases, the length of a service, which is subject to interruptions, increases, which in turn leads to lesser number of service completions and to increased queue length and loss rate. The increase in the measures *PSI*, *ENCS* and *LZI* in Table 1 supports the above intuition. From Table 1, one can observe that the sum of the two fractions *PSB* and *PSI* is increasing, whereas the fraction *PSB*, which is the fraction of time, the server remains active, is decreasing. This shows that the decrease in the server idle probability is not in favor of the system. We think the possible decrease in the number of service completions with increase in δ_1 could be the reason behind the decrease in the server active probability *PSB*. The same reasoning can be attributed to the decrease in the fraction of time the service process remains in protected phases *FSP*. The decrease in the service completion rate as δ_1 increases, leading to a slow depletion rate of inventory in the system, can be pointed out as the reason for the increase in the expected inventory level *EIL*. In the Table, one can see that *EIL* is above the production switch on level *s* and is increasing; this can be pointed out as the reason behind the increase in the production switch *off* rate *PSWOF* as well as the production commencement rate *PCOM*.

Table 1: Effect of the Service Interruption Rate δ_1 on Various Performance

Measures

$l = 3,\ l_1 = 3, n = 7, m = 7,\ \mu = 25,\ \mu_1 = 35, s = 4, S = 10, \lambda = 2,\ \delta_2 = 2,$
$\delta_3 = 2,\ \delta_4 = 3$

δ_1	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPon
1.4	0.378467	0.15138	5.5648	1.33858	0.1622	0.06733	0.067331	0.107662	0.795554
1.6	0.378193	0.172888	5.58341	1.49551	0.162083	0.06745	0.067451	0.10903	0.794978
1.8	0.377923	0.19436	5.6026	1.669662	0.161967	0.06755	0.067545	0.11038	0.794411
2	0.377658	0.215804	5.622427	1.86421	0.191854	0.06762	0.067616	0.111701	0.793854
2.2	0.377399	0.237222	5.6429	2.08313	0.161743	0.06766	0.067648	0.112988	0.79331
2.4	0.377147	0.258615	5.66404	2.331542	0.161635	0.06769	0.067692	0.114237	0.792782

6.5.2 Effect of the Service Repair Rate δ_2

When the repair rate increases, one would expect faster service completions, which leads to decreased queue length and loss rate. The decrease in the measures *PSI*, *ENCS* and *LZI* in Table 2 supports the above intuition. In contrast to the case of increase in the interruption rate δ_1, here the server active probability *PSB* and the server idle probability are increasing. For a moment one may suspect that the increase in the server idle probability is not in favor of the system; however, a closer look at Table 2 reveals that the increase in the server idle probability = 1-*PSB*-*PSI* is due to the high decrease in the fraction *PSI* as compared to the increase in the fraction *PSB*. The increase in the server active probability suggests increasing the repair rate as a remedy to interruption; however to what extent one can achieve this may depend on the particular situation at hand. Faster recovery from interruption can be thought of as the reason behind the slight increase in the fraction of time the service process is in protected phases *FSP*.

The decrease in the expected number of customers as well as in the loss rate of customers points to a faster depletion of inventory in the system. However, the Table shows a narrow increase in the expected inventory level *EIL*. This can be thought of as due to an increase in the production by a narrow margin as indicated by the narrow increase in the fraction *FPon*.

Table 2: Effect of the Service Repair Rate δ_2 on Various Performance Measures

$l = 3$, $l_1 = 3$, $n = 7$, $m = 7$, $\mu = 25$, $\mu_1 = 35$, $s = 3$, $S = 8$, $\lambda = 2$, $\delta_1 = 2$, $\delta_3 = 2$, $\delta_4 = 3$

δ_2	*PSB*	*PSI*	*EIL*	*ENCS*	*FSP*	*PSWOF*	*PCOM*	*LZI*	*FPon*
1.4	0.363462	0.2967	4.269289	3.57877	0.15577	0.08748	0.087481	0.182685	0.764012
1.6	0.365842	0.261316	4.28675	2.70606	0.15679	0.08778	0.087776	0.170786	0.769015
1.8	0.367506	0.233337	4.298384	2.196706	0.157502	0.08777	0.087774	0.16245	0.772513
2	0.368702	0.210686	4.305717	1.862118	0.158015	0.08762	0.08762	0.156484	0.775026
2.2	0.369575	0.191987	4.310007	1.641239	0.15839	0.0874	0.087398	0.152119	0.776863
2.4	0.370225	0.176297	4.312084	1.476432	0.158668	0.08715	0.087152	0.148873	0.778227

6.5.3 Effect of the Production Interruption Rate δ_3

As the production becomes slower due to an increase in the interruption rate, we expect a decrease in the measures like expected inventory level *EIL*, production switch *off* rate *PSWOF* and the production commencement rate *PCOM*. Table 3 supports these intuitions. The increase in the fraction *FPon* can be seen to be due to an increase in the length of the production process brought by the increase in the interruption rate. The decrease in the expected inventory level leads to an increase in the loss rate *LZI*. The possible decrease in the number of service completions due to an increased loss rate can be seen to be the reason behind the decrease in the server active probability *PSB*. The decrease in *PSB* can be thought of as the reason for the slight decrease in the server interruption probability *PSI* and hence this decrease in *PSI* is not in favor of the system under study. Again the decrease in *PSB* can be thought of as the reason behind the narrow increase in the expected number of customers in the system.

Table 3: Effect of the Production Interruption Rate δ_3 on Various Performance Measures $l = 3, l_1 = 3, n = 7, m = 7, \mu = 25, \mu_1 = 35, s = 3, S = 8, \lambda = 2, \delta_1 = 2, \delta_2 = 3, \delta_4 = 2$

δ_3	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPON
1.4	0.36876	0.14048	4.319611	1.177874	0.15804	0.09197	0.091972	0.156195	0.766493
1.6	0.362298	0.138018	4.1462.5	1.1784	0.155271	0.08233	0.082324	0.188505	0.791344
1.8	0.355355	0.135373	3.96877	1.17886	0.152295	0.07313	0.073134	0.223219	0.814541
2	0.347983	0.132565	3.788857	1.179261	0.149135	0.06471	0.064706	0.26008	0.836012
2.2	0.34024	0.129615	3.608022	1.179601	0.145817	0.05696	0.056963	0.298799	0.855723
2.4	0.332188	0.126547	3.42779	1.179879	0.142366	0.04991	0.049906	0.339057	0.873673

6.5.4 Effect of the Production Repair Rate δ_4

As the production repair rate δ_4 increases, the average span of production process being interrupted decreases and hence the production rate increases. The increase in the production switch *off* rate *PSWOF* as well as the expected inventory level *EIL* and the decrease in the fraction of time that the production process is in *on* mode *FPon* can be seen to be due to an increase in the overall production rate. The same reasoning can be applied to explain the increase in the production commencement rate *PCOM*. Since the expected inventory level increases, the expected loss rate *LZI* of customers decreases. This decrease in *LZI* can be thought of as the reason for the slight increase in the expected number of customers in the system *ENCS*. Note that in the case of increase in δ_2, which was discussed earlier, the decrease in the loss rate does not lead to an increase in

ENCS. Here, one can notice that in the case of increase in δ_2, it is the faster service completion, that leads to a decrease in the number of customers and an increase in the service completion rate can't be expected in the case of increase in δ_4. The increase in the fraction of time *PSB* as well as *PSI* may be attributed to the increase in *ENCS*.

Table 4: Effect of the Production Repair Rate δ_4 on Various Performance Measures $l = 3$, $l_1 = 3$, $n = 7$, $m = 7$, $\mu = 25$, $\mu_1 = 35$, $s = 3$, $S = 8$, $\lambda = 2$, $\delta_1 = 2$, $\delta_2 = 3$, $\delta_3 = 2$

δ_4	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPON
1.4	0.314368	0.179639	3.259093	1.85844	0.134729	0.04832	0.048318	0.428149	0.876677
1.6	0.327259	0.187006	3.476104	1.860745	0.140254	0.05482	0.05482	0.363696	0.859958
1.8	0.337441	0.192824	3.658436	1.862656	0.144618	0.06083	0.060828	0.312789	0.844474
2	0.345574	0.197471	3.812373	1.864207	0.148103	0.06635	0.066347	0.272124	0.830223
2.2	0.352143	0.201224	3.943064	1.865446	0.150918	0.0714	0.0714	0.239281	0.817153
2.4	0.357505	0.204288	4.054683	1.866421	0.153216	0.07602	0.076017	0.21247	0.805187

6.5.5 Effect of the Maximum Inventory Level *S*

When the maximum inventory level *S* increases, it takes more time for switching *off* the production process, once it is switched *on*. This is reflected as the decrease in the production switch *off* rate in Table 5. The decrease in the production commencement rate *PCOM* and the increase in the fraction *FPon* have the same reasoning. Since the production remains in *on* mode for a longer time, the expected inventory level *EIL* increases with increase in *S*. The presence of sufficient inventory in the system leads to a decrease in the loss rate *LZI* and also to an increase in the server active probability *PSB*. The increase in the server interruption probability *PSI* is due an increase in *PSB*. In Table 5, it can be seen that the expected number of customers is decreasing in spite of a decrease in the loss rate. This may seem contradictory with what we said in the case of increase in the production repair rate, δ_4. We point out the significant increase in the expected inventory level with an increase in the maximum inventory level, compared to the increase when δ_4 increases as the reason for this contradiction.

Table 5: Effect of the Maximum Inventory Level S on Various Performance Measures

$l = 3, l_1 = 3, n = 7, m = 7, \mu = 25, \mu_1 = 35, s = 3, \lambda = 2, \delta_1 = 2, \delta_2 = 3, \delta_3 = 2, \delta_4 = 3$

S	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPon
9	0.374236	0.142566	4.861727	1.177649	0.160387	0.070329	0.070329	0.128815	0.78666
10	0.37667	0.143493	5.4196	1.176987	0.161429	0.05898	0.05898	0.11665	0.791774
11	0.378755	0.144288	5.97946	1.176425	0.162323	0.050608	0.050608	0.106222	0.79616
12	0.380555	0.144973	6.53631	1.175947	0.163091	0.044205	0.044205	0.09722	0.799944
13	0.38212	0.145569	7.08564	1.175538	5.163765	0.039168	0.039167	0.089399	0.803232
14	0.383487	0.14609	7.629048	1.175189	0.164351	0.035112	0.035112	0.082562	0.806107

6.5.6 Effect of the Re-Order Level *s*

As the re-order level *s* increases, production switch *on*'s and switch *off* 's occur more frequently and hence the increase in the measures *PCOM*, *PSWOF*, *FPon* and *EIL* are expected. Increase in the expected inventory level can be thought of as the reason for decrease in the loss rate *LZI* as well as the increase in the server busy probability *PSB*. As in the case of the increase in the maximum inventory level *S*, here also the expected number of customers is decreasing despite a decrease in the loss rate. The same reasoning as in the case of *S* can be given in this case also.

Table 6: Effect of the Re-Order Level *s* on Various Performance Measures

$l = 3, l_1 = 3, n = 7, m = 7, \mu = 25, \mu_1 = 35, S = 15, \lambda = 2, \delta_1 = 2, \delta_2 = 3, \delta_3 = 2, \delta_4 = 3$

s	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPon
3	0.384688	0.146548	8.16732	1.174887	0.164866	0.031784	0.031784	0.076552	0.808633
4	0.386787	0.147349	8.432331	1.1744407	0.165765	0.033856	0.033856	0.066061	0.813045
5	0.388528	0.14801	8.712697	1.174057	0.166512	0.036487	0.036486	0.057358	0.816704
6	0.389985	0.148565	8.999142	1.173733	0.167136	0.039826	0.039826	0.050075	0.819766
7	0.391212	0.149033	9.280952	1.173458	0.167662	0.044108	0.044108	0.043936	0.822347

6.5.7 Effect of the Number l of Protected Phases of the Production Process

As the number l of protected phases in the production process increases, since the harm due to interruptions is reduced, production becomes faster. Recall that the production repair rate δ_4, when increased, accelerates the production process by reducing the adverse effect of interruption. Hence, as l increases, we expect a similar impact on the performance measures as in the case of increase in the production repair rate δ_4. Comparison of Tables 4 and 7 shows that, the expected number of customers, *ENCS* is the only measure, which shows an opposite behavior in the two tables. Precisely as δ_4 increases, *ENCS* increases and when l increases, *ENCS* shows a narrow decreasing nature. As we pointed out in the case of the parameters S and δ_4, here also the comparatively high inventory level in the case of increase in l as compared to the increase in δ_4 could be the reason for the above phenomenon.

Table 7: Effect of the Number l of Protected Phases of the Production Process on Various Performance Measures

$l_1 = 3, n = 8, m = 7, \mu = 45, \mu_1 = 45, s = 3, S = 8, \lambda = 2, \delta_1 = 2, \delta_2 = 3, \delta_3 = 2, \delta_4 = 2$

l	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPON
2	0.291561	0.11107	4.61743	0.755968	0.124955	0.141702	0.141701	0.125668	0.642062
3	0.297146	0.113198	4.864789	0.755864	0.127348	0.16162	0.161619	0.089762	0.591293
4	0.303594	0.114893	4.86764	0.755677	0.12925	0.181432	0.181432	0.061168	0.540687
5	0.305052	0.116211	5.23527	0.755422	0.130736	0.200782	0.200782	0.038935	0.491164
6	0.307681	0.117212	5.39456	0.755122	0.131863	0.219385	0.219385	0.022041	0.443453
7	0.309633	0.117955	5.50981	0.754797	0.132699	0.237032	0.237032	0.009491	0.398101

6.5.8 Effect of the Number l_1 of Protected Phases of the Service Process

As the number of protected phases of the service process increases, intuitively, service becomes faster, which leads to a decrease in the expected number of customers *ENCS* as well as in the server interruption probability *PSI.* Table 8 supports the above intuition. Notice that in Table 8, the expected inventory level is close to 5.5, whereas the re-order level is 3. Since there is sufficient inventory, the customer loss rate is not significant. This could be thought of as the reason behind the almost unchanged values for the performance measures *PSB, PCOM, PSWOF* and *FPon.*

Table 8: Effect of the Number l_1 of Protected Phases of the Service Process on Various Performance Measures

$l = 3, n = 7, m = 11, \mu = 45, \mu_1 = 45, s = 3, S = 8, \lambda = 2, \delta_1 = 2, \delta_2 = 3, \delta_3 = 2, \delta_4 = 3$

l_1	PSB	PSI	EIL	ENCS	FSP	PSWOF	PCOM	LZI	FPON
3	0.484998	0.23515	5.536363	2.3564	0.132272	0.220386	0.220385	0.015909	0.446394
4	0.484974	0.205746	5.56041	2.02771	0.176354	0.220061	0.22006	0.016011	0.446371
5	0.484954	0.176347	5.54865	1.7548	0.220433	0.219788	0.219788	0.016093	0.446353
6	0.48494	0.146951	5.5554	1.52402	0.264513	0.219569	0.219569	0.016152	0.44634
7	0.484932	0.117559	5.532303	1.325833	0.308593	0.219406	0.219406	0.016184	0.446333
8	0.484932	0.088169	5.546095	1.153372	0.352677	0.219299	0.219299	0.016188	0.446332

6.5.9 Cost Function

For studying the optimality of the parameters like the re-order level s, the maximum inventory level S, the number of protected phases of the service as well as the production processes, we construct the following cost function:

Cost = CI*EIL + CN*ENCS + CIP*EIRP + CIS*EIRS + CPON*FPON + CZ*LZI + CPPR*FPP + CSPR*FSP + CPRR*FPRS + CSRR*FSRS + CPCOM*PCOM,

where ***CI*** is the inventory holding cost per item per unit time, ***CN*** is the holding cost per unit time per customer, ***CIP*** is the cost per unit time per interruption, ***CIS*** is the unit time cost incurred due to server interruptions, ***CPON*** is the unit time cost incurred for running the production process, ***CZ*** is the cost incurred for the loss of customers due to shortage of inventory, ***CPPR*** is the unit time cost incurred due to protection of the production process, ***CSPR*** is the unit time cost incurred due to the protection of the service process, ***CPRR*** is the unit time cost incurred due to repair of the production process, ***CSRR*** is the unit time cost incurred due to the repair of the service process and ***CPCOM*** is the unit time cost for switching *on* the production process. Using this cost function, the optimality of the parameters s, S, l and l_1 has been studied. Following are a few illustrations.

6.5.9.1 Optimality of the Re-Order Level *s*

Table 6 shows that loss rate *LZI* and the expected number of customers, *ENCS* are the only measures that shows a decreasing nature, which are involved in the cost function, as the re-order level increases. Therefore unless we select the cost *CZ* and *CN* so as to capture this decrease, the cost function will be linearly increasing. Figure1(a) shows that by taking a comparatively high cost *CZ*, optimal value of *s* is 5 and in Figure 1(b) the cost curve shows a linearly increasing nature suggesting *s*=2 as the optimal value.

Figure 1(a): Effect of the Re-Order Level *s* on the Cost Function

$l = 3$, $n = 7$, l_1=3,$m = 7$, $k = 25$, $k_1 = 35$, $S = 17$, $\lambda = 1.2$, $\delta_1 = 2$, $\delta_2 = 1$, $\delta_3 = 2.5$,

$\delta_4 = 1.5$

CN=1000, CI=100, CPPR=750, CPRR=750, CSPR=750, CSRR=750,
CPON=500, CPCOM=500, CIP=500, CIS=500, CZ=17000

s	COST
2	3504.491
3	3429.888
4	3389.797
5	**3374.475**
6	3377.203
7	3393.082
8	3418.647
9	3451.228
10	3489.021
11	3531.185
12	3576.578

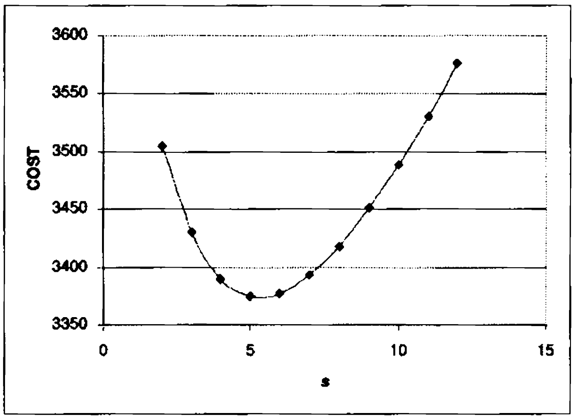


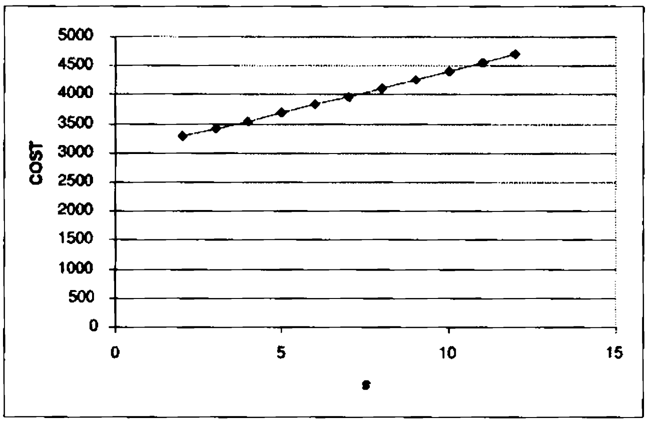
Figure 1(b): Effect of the Re-Order Level *s* on the Cost Function

$l = 3$, $n = 7$, l_1=3,$m = 7$, $k = 25$, $k_1 = 35$, $S = 17$, $\lambda = 1.2$, $\delta_1 = 2$, $\delta_2 = 1$, $\delta_3 = 2.5$,

$\delta_4 = 1.5$

CN=200, CI=300, CPPR=200, CPRR=400, CSPR=100, CSRR=200, CPON=250, CPCOM=100, CIP=250, CIS=300, CZ=1350

s	COST
2	3287.48
3	3413.14
4	3545.43
5	3683.73
6	3826.6
7	3972.22
8	4119.32
9	4266.22
10	4411.81
11	4555.92
12	4695.78



6.5.9.2 Optimality of the Maximum Inventory Level *S*

Table 5 shows that as the maximum inventory level *S* increases, the measures which show a decreasing nature are *ENCS, PSWOF, PCOM* and *LZI*. But the magnitude of the above decrease is small as compared to the increase in the measures like *EIL*. Figure 2(a) shows that particular choice of high costs corresponding to the measures which show a decreasing nature, an optimal value for the maximum inventory level is attainable. However, if we use the same costs as for Figure 1(b), we get a linearly increasing cost function as in Figure 2(b).

Figure 2(a): Effect of the Maximum Inventory Level *S* on the Cost Function

$l = 3$, $n = 7$, $l_1=3$,$m = 7$, $k = 25$, $k_1 = 35$, $s = 3$, $\lambda = 1.2$, $\delta_1 = 2$, $\delta_2 = 1$, $\delta_3 = 2.5$, $\delta_4 = 1.5$

CN=1000, CI=200, CPPR=500, CPRR=1000, CSPR=500, CSRR=500, CPON=5000, CPCOM=5000, CIP=500, CIS=1000, CZ=10000

S	COST
7	6702.57
8	6622.267
9	6590.386
10	**6590.016**
11	6611.220
12	6647.06
13	6693.465
14	6747.94
15	6808.712
16	6842.117

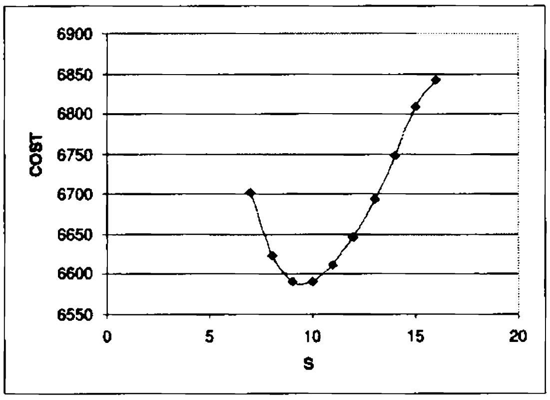
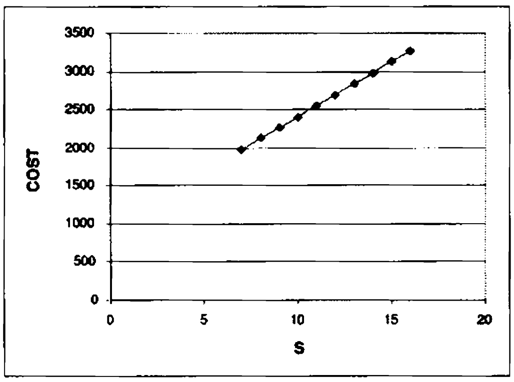


Figure 2(b): Effect of the Maximum Inventory Level *S* on the Cost Function $l = 3, n = 7, l_1=3, m = 7, k = 25, k_1 = 35, s = 3, \lambda = 1.2,$ $\delta_1 = 2,\ \delta_2 = 1,\ \delta_3 = 2.5,\ \delta_4 = 1.5$

CN=200, CI=300, CPPR=200, CPRR=400, CSPR=100, CSRR=200, CPON=250, CPCOM=100, CIP=250, CIS=300, CZ=1350

S	COST
7	**1982.47**
8	2123.93
9	2265.64
10	2408.41
11	2552.34
12	2696.15
13	2839.48
14	2982.56
15	3125.48
16	3264.49



6.5.9.3 Optimality of the Number l_1 of Protected Phases of the Service Process

Figure 3(a) shows an optimal value for the number of protected phases of the service process for the selected costs, whereas if we use the costs as for Figure

1(b), the cost curve shows a decreasing nature suggesting that to attain the optimal cost, we have to protect all the service phases.

Figure 3(a): Effect of l_1 on the Cost Function

$l = 3$, $n = 6$, $m = 14$, $K = 35$, $K_l = 55$, $s = 3$, $S = 8$, $\lambda = 1.2$, $\delta_1 = 2$, $\delta_2 = 3$, $\delta_3 = 2.5$, $\delta_4 = 1.5$;

CN=200,*CI*=100,*CPPR*=700,CPRR=5000,CSPR=5000,CSRR=5000, CPON=1000,CPOCOM=500,CIP=500,CIS=800,CZ=200

l_1	COST
2	4024.602
3	3913.678
4	3866.162
5	3841.629
6	**3839.682**
7	3854.263
8	3875.564
9	3902.444
10	3934.289

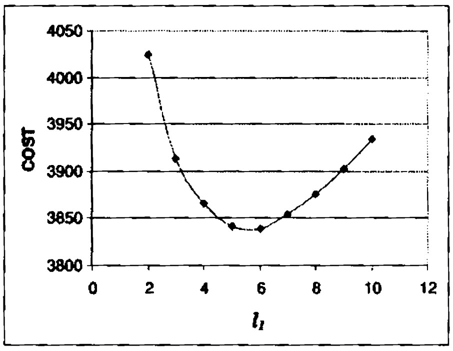
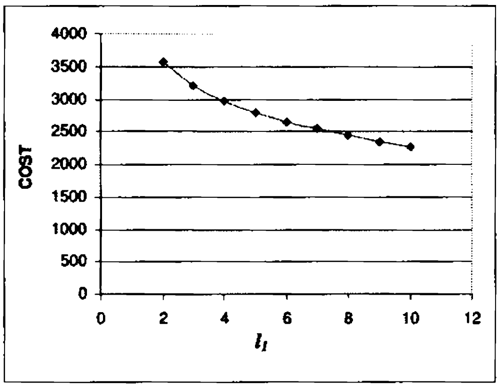


Figure 3(b): Effect of l_1 on the Cost Function

$l = 3$, $n = 6$, $m = 14$, $K = 35$, $K_l = 55$, $s = 3$, $S = 8$, $\lambda = 1.2$, $\delta_1 = 2$, $\delta_2 = 3$, $\delta_3 = 2.5$, $\delta_4 = 1.5$

CN=200, CI=300, CPPR=200, CPRR=400, CSPR=100, CSRR=200, CPON=250, CPCOM=100, CIP=250, CIS=300, CZ=1350

l_1	COST
2	3569.37
3	3206.14
4	2977.84
5	2784.9
6	2643.32
7	2535.52
8	2438.76
9	2350.96
10	2272.2



5.9.4 Optimality of the Number *l* of Protected Phases of the Production Process

Figure 4(a) shows an optimal value, for the costs selected, for the number of phases to be protected for the production process; however, if we take the costs as for Figure 2(a), the cost function decreases, which suggests that protecting all the production phases leads to the optimal value of the cost function.

Figure 4(a): Effect of *l* on the Cost Function

$l_1=3$, $n=14$, $m=6$, $K=25$, $K_1=30$, $s=3$, $S=8$, $\lambda=1.2$, $\delta_1=2$, $\delta_2=1$, $\delta_3=2.5$, $\delta_4=1.5$

CN =200, *CI* =300, *CPPR*=200, *CPRR=400*, *CSPR=100*, *CSRR=200*, *CPON* =250, *CPCOM*=100, *CIP* =250, *CIS* =300, CZ=1350

l	COST
2	2240.677
3	2240.808
4	2215.624
5	2192.397
6	**2089.139**
7	2160.83
8	2161.505
9	2164.486
10	2171.003

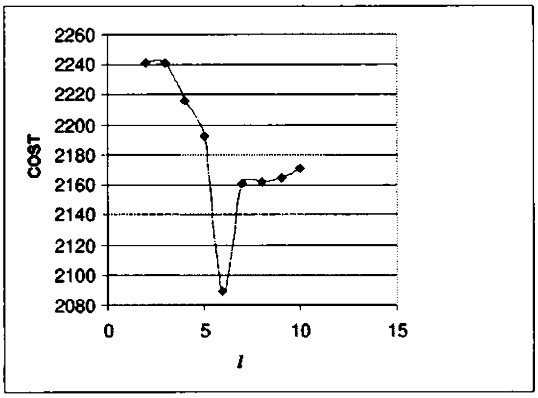
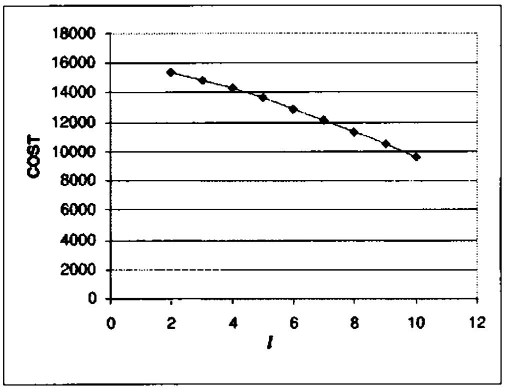


Figure 4(b): Effect of *l* on the Cost Function

l_1=3, n=14, m=6, K=25, K_1=30, s=3, S=8, λ=1.2, δ_1=2, δ_2=1, δ_3=2.5, δ_4=1.5

CN=1000, CI=200, CPPR=500, CPRR=1000, CSPR=500, CSRR=500, CPON=5000, CPCOM=5000, CIP=500, CIS=1000, CZ=10000

l	COST
2	15351.15
3	14868.72
4	14305.12
5	13672.22
6	12911.49
7	12194.89
8	11370.66
9	10514.08
10	**9665.41**



6.6 Conclusion

We analyzed an (s, S) production inventory system, where the processing of inventory requires some random amount time. Erlang distributions are used to model the service as well as the production processes, which are subject to multiple interruptions. For reducing the adverse effect of the interruptions, the concept of protecting certain phases of the service as well as the production process from interruption has been introduced. We further assumed that no customers would be allowed to join the system if there is no inventory in the system. This assumption lead us to derive an explicit expression for stability condition, which even holds if one assumes general PH in place of the assumed Erlang distributions.

In Studies like [3, 4] on inventory systems where customers are barred to join the system when there is shortage of inventory, the authors were able to show that the stability of such inventory systems is not affected by the inventory parameters and also that their steady state distributions can be obtained in product form. However, in these studies, the underlying distributions were all exponential. The proof, which we have given for the stability of the system, can be used to characterize the stability of the above said inventory systems with more general underlying distributions. However, we have not yet able to check whether some kind of product form expression for the steady state distribution is possible, which will be really interesting if one can do that.

6.7 Appendix I

I. Transitions leading to an increase in the level:

Transitions due to arrival of customers

From $L(0,0,j,1,j_1) \to L(1,1,j,1,j_1)$ is governed by

$\beta \otimes \lambda I_{n-\delta_{1,j_1}l}$; $1 \le j \le S-1$; $j_1 = 0,1$

From $L(0,0,j,0) \to L(1,1,j,0)$ is governed by $\lambda\beta$; $s+1 \le j \le S$

From $L(i,1) \cup L(i,2) \to L(i+1,1) \cup L(i+1,2)$ is governed by λI (note that these are diagonal transitions)

II. Transitions leading to a decrease in the level:

Transitions due to service completion

$L(1,1,j,1,j_1) \to L(0,0,j-1,1,j_1)$ is governed by $T^0 \otimes I_{n-\delta_{1,j_1}l}$; $1 \le j \le S-1$; $j_1 = 0,1$

$L(1,1,s+1,0,0) \to L(0,0,s,1,0)$ is governed by $T^0 \otimes \alpha$ (note that the production process needs to be switched to *on* mode the moment such a transition occurs)

$L(1,1,j,0,0) \to L(0,0,j-1,0,0)$ is governed by T^0; $s+2 \le j \le S$

For $i \ge 2$,

$L(i,1,1,1,j_1) \to L(i-1,0,0,1,j_1)$ is governed by $T^0 \otimes I_{n-\delta_{1,j_1}l}$; $j_1 = 0,1$

$L(i,1,j,1,j_1) \to L(i-1,1,j-1,1,j_1)$ is governed by

$T^0 \otimes \beta \otimes I_{n-\delta_{1,j_1}l}$; $2 \le j \le S-1$; $j_1 = 0,1$

$L(i,1,s+1,0,0) \to L(i-1,1,s,1,0)$ is governed by $T^0 \otimes \beta \otimes \alpha$

$L(i,1,j,0,0) \rightarrow L(i-1,1,j-1,0,0)$ is governed by $T^0 \otimes \beta$; $s+2 \leq j \leq S$

III. Transitions where no level change occurs:

III(a) Transitions due to a production completion

$L(0,0,j,1,0) \rightarrow L(0,0,j+1,1,0)$ is governed by $U^0 \otimes \alpha$; $0 \leq j \leq S-2$

$L(0,0,S-1,1,0) \rightarrow L(0,0,S,0,0)$ is governed by U^0

$L(0,0,j,1,0) \rightarrow L(0,0,j+1,1,0)$ is governed by $U^0 \otimes \alpha$; $0 \leq j \leq S-2$

$L(i,0,0,1,0) \rightarrow L(i,1,1,1,0)$ is governed by $\beta \otimes U^0 \otimes \alpha$; $i \geq 1$

$L(i,l',j,1,0) \rightarrow L(i,l',j+1,1,0)$ is governed by

$I_{m-\delta_{2,l'}l_1} \otimes U^0 \otimes \alpha$; $i \geq 1$; $l'=1,2$; $1 \leq j \leq S-2$

$L(i,l',S-1,1,0) \rightarrow L(i,l',S,0,0)$ is governed by $I_{m-\delta_{2,l'}l_1} \otimes U^0$; $i \geq 1$; $l'=1,2$

III(b) Transitions due to a production interruption

$L(0,0,j,1,0) \rightarrow L(0,0,j,1,1)$ is governed by $\delta_3 E$; $0 \leq j \leq S-1$, where $E = \begin{bmatrix} I_{n-l} \\ 0 \end{bmatrix}$

$L(i,0,0,1,0) \rightarrow L(i,0,0,1,1)$ is governed by $\delta_3 E$; $i \geq 1$

$L(i,l',j,1,0) \rightarrow L(i,l',j,1,1)$ is governed by

$I_{m-\delta_{2,l'}l_1} \otimes \delta_3 E$; $i \geq 1$; $l'=1,2$; $1 \leq j \leq S-1$

III(c) Transitions due to completion of repair of an interrupted production process

$L(0,0,j,1,1) \rightarrow L(0,0,j,1,0)$ is governed by $\delta_4 e \otimes \alpha$; $0 \leq j \leq S-1$

$L(i,0,0,1,1) \rightarrow L(i,0,0,1,0)$ is governed by $\delta_4 e \otimes \alpha;\ i \geq 1$

$L(i,l',j,1,1) \rightarrow L(i,l',j,1,0)$ is governed by

$I_{m-\delta_{2,j'} l_1} \otimes \delta_4 e \otimes \alpha;\ i \geq 1;\ l' = 1,2;\ 1 \leq j \leq S-1$

III(d) Transitions due to a service interruption

$L(i,1,j,1,j_1) \rightarrow L(i,2,j,1,j_1)$ is governed by $\delta_1 \tilde{E}_{j_1}^T;\ i \geq 1;\ 1 \leq j \leq S-1;\ j_1 = 0,1$,

where $\tilde{E}_{j_1} = \begin{bmatrix} I_{(m-l_1)(n-\delta_{1,j_1} l)} & 0 \end{bmatrix}$

$L(i,1,j,0) \rightarrow L(i,2,j,0)$ is governed by $\delta_2 \left(\tilde{E}^0\right)^T;\ i \geq 1;\ s+1 \leq j \leq S$, where

$\tilde{E}^0 = \begin{bmatrix} I_{(m-l_1)} & 0 \end{bmatrix}$

III(e) Transitions due to completion of repair of an interrupted service

$L(i,2,j,1,j_1) \rightarrow L(i,1,j,1,j_1)$ is governed by $\delta_2 \tilde{E}_{j_1};\ i \geq 1;\ 1 \leq j \leq S-1;\ j_1 = 0,1$

$L(i,2,j,0) \rightarrow L(i,1,j,0)$ is governed by $\delta_2 \tilde{E}^0;\ i \geq 1;\ s+1 \leq j \leq S$

CONCLUSION

In this thesis we have developed a few inventory models in which items are served to the customers after a processing time. This leads to a queue of demand even when items are available. In chapter two we have discussed a problem involving search of orbital customers for providing inventory. Retrial of orbital customers was also considered in that chapter; in chapter 5 also we discussed retrial inventory model which is sans orbital search of customers. In the remaining chapters (3, 4 and 6) we did not consider retrial of customers, rather we assumed the waiting room capacity of the system to be arbitrarily large. Though the models in chapters 3 and 4 differ only in that in the former we consider positive lead time for replenishment of inventory and in the latter the same is assumed to be negligible, we arrived at sharper results in chapter 4. In chapter 6 we considered a production inventory model with production time distribution for a single item and that of service time of a customer following distinct Erlang distributions. We also introduced protection of production and service stages and investigated the optimal values of the number of stages to be protected. In all problems investigated closed form expressions for the system stability were derived. Our conclusions in each chapter are provided with numerous illustrations.

Bibliography

[1] Aissani and J R. Artalejo, On the single server retrial queue subject to breakdowns; *Queueing Systems* Vol. 30 (3-4), 309-321, 1998.

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

[3] J R .Artalejo , A. G´omez-Corral, Unreliable retrial queues due to service interruptions arising from facsimile networks. *Belg. J. Oper. Res. Statist. Comput. Sci.* Vol. 38 (1), 31-41, 1998.

[4] J R Artalejo and A Gómez-Corral; Waiting time in the M/M/c queue with finite retrial group. *Bulletin of Kerala Mathematics Association* Vol. 2, pp. 1-17, 2005.

[5] J R Artalejo, V C Joshua and A. Krishnamoorthy; An M/G/1 retrial queue with orbital search by the server. In Advances in Stochastic Modelling, Eds. J R Artalejo and A Krishnamoorthy; pages 41–54. *Notable Publications Inc.*, New Jersey, 2002.

[6] J R Artalejo, A Krishnamoorthy and M J Lopez-Herrero; Numerical analysis of (s, S) inventory systems with repeated attempts; *Ann Oper Res*; (2006) 141: 67-83 .

[7] B Avi-Itzhak, and P Naor,; Some Queueing Problems with the Service Station subject to breakdown, *Operations Research*, Vol. 11, Issue 3, 303-320, 1963 .

[8] O. Berman, E. Kim, and D. G. Shimshack, Deterministic approximations for inventory management at service facilities, *IIE Transactions 25* (1993), no. 5, 98–104 .

[9] O. Berman and E. Kim, Stochastic models for inventory management at service facilities, Communications in Statistics. *Stochastic Models* 15 (1999), no. 4, 695–718.

[10] O. Berman and K. P. Sapna, Inventory management at service facilities for systems with arbitrarily distributed service times, Communications in Statistics. *Stochastic models* 16 (2000), no. 3-4, 343–360.

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

[12] Y Chen and X Li; The effect of customer segmentation on an inventory system in the presence of supply disruptions; *Proceedings of the 2009 Winter Simulation Conference*; Eds. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, eds. pp 2343-2352 .

[13] T G Deepak, A Krishnamoorthy, C.N.Viswanath and K Vineetha; Inventory with service time and transfer of customers and/ inventory; *Annals of Operations Research*, PP.191-213, Volume 160, Number 1 / April, 2008 .

[14] T G Deepak, C.N.Viswanath and V C Joshua; On an Inventory with positive service time having optional processing time; *Bulletin of Kerala Mathematics Association*, Vol. 4, No.2, 2007, December, 75-86 .

[15] A.N Dudin, A Krishnamoorthy, V C Joshua and G V Tsarenkov; Analysis of the BMAP/G/1 retrial system with search of customers from the orbit; *European Journal of Operational Research* 157 (2004) 169-179.

[16] G I Falin, and J G C Templeton ; Retrial Queues, Chapman and Hall (1997).

[17] A Federgruen and L Green; Queueing systems with service interruptions; *Operations Research*,vol. 34(5):752–768, 1986.

[18] D Gaver, Jr. A waiting line with interrupted service, including priorities. *Journal of the Royal Statistical Society*, vol. B24:73–90, 1962.

[19] U Ibe and K.S Trivedi :Two queues with alternating service and server breakdowns, *Queueing Systems*, 253-268,1990.

[20] N K Jaiswal; Preemptive resume priority queue; *Operations Res.*, Vol. 9, No. 5, 732-742, 1961.

[21] N K Jaiswal ; Time dependant solution of the head of the line priority queue; *J. Roy. Stat. Soc.* B24, 91-101, 1962.

[22] S. Karlin and H.M Taylor (1975); A First Course in Stochastic Process, 2nd Edition, academic Press, Newyork.

[23] J Keilson, Queues subject to service interruptions, *Ann. Math. Statistics*, 33, 1314-1322, 1962.

[24] V. Klimenok and A.N. Dudin :A BMAP/PH/N Queue with negative customers and partial protection of service, To appear in *Journal of statistical Planning & inference* 2011

[25] V Klimenok, C S Kim and A Kuznetsov; A Multi-Server Queue with Negative Customers and Partial Protection of Service; *Proceedings of the 13th International Conference on Analytical and Stochastic Modelling Techniques and Applications* (ASMTA06), 28-31 May 2006, Bonn, Germany; Eds. K Al-Begain; pp 143-148.

[26] A Krishnamoorthy, T G Deepak and V C Joshua; An M/G/1 retrial queue with non-persistant customers and orbital search; *Stochastic Analysis and Applications*; Vol. 23, No. 5, 2005, pp. 975-997.

[27] A Krishnamoorthy, T G Deepak, C.N.Viswanath and K Vineetha; Effective utilization of idle time in an (s, S) inventory with positive service time, *Journal of Applied Mathematics and Stochastic Analysis* (JAMSA), vol. 2006, Article ID 69068, 13 pages, 2006. doi:10.1155/JAMSA/2006/69068 .

[28] A Krishnamoorthy, B Gopakumar and C N Viswanath; An M/Em/1 Queue with Protected and Unprotected Phases from Interruption; 5th International Conference on Queueing Theory and Network Applications (QTNA 2010), July 24-26, 2010, Beijing, China.

[29] A Krishnamoorthy, B Gopakumar and C.N.Viswanath ; A Queueing model with Interruption Resumption/Restart and Reneging; *Bulletin of Kerala Mathematics Association*, Special Issue. 29-45, October 2009; Guest Editor: S. R. S. Varadhan FRS.

[30] A Krishnamoorthy and M E Islam; Production inventory with retrials of customers in an (s,S) Policy; *Stochastic Modelling and Applications*; Vol. 6, No. 2, pp.1-11, 2003.

[31] A Krishnamoorthy and M E Islam; Production inventory with random life time and retrial of customers; In *Proceedings of the Second National Conference on Mathematical and Computational Models*, NCMCM 2003; Eds. G Arulmozhi and R Nadarajan; Coimbatore, India, pp. 89-110.

[32] A Krishnamoorthy and M E Islam; (s, S) inventory system with postponed demands; *Stochastic Analysis and Applications*; Volume 22(3), 2004, 827-842.

[33] A Krishnamoorthy, M E Islam and C.N.Viswanath; Retrial Inventory with Batch Markovian Arrival and Positive Service Time, *Stochastic Modelling and Applications* Vol. 9, Number 2, pp. 38-53, 2006.

[34] A Krishnamoorthy and K P Jose; Comparison of inventory systems with service, positive lead-time, loss, and retrial of customers; *Journal of Applied Mathematics and Stochastic Analysis*; Volume 2007, Article ID 37848, 23 pages, 2007. doi:10.1155/2007/37848.

[35] A Krishnamoorthy, K P Jose and C.N.Viswanath ; Numerical Investigation of a PH/PH/1 Inventory System with Positive Service Time and Shortage; *Neural, Parallel & Scientific Computations*; Volume 16, Number 4, December 2008, Dynamic Publishers, USA .

[36] A.Krishnamoorthy, B.Lakshmy and R.Manikadan: A survey on Inventory models with positive service time, *OPSEARCH*, 2011, DOI 10.1007/s12597-010-0032-z.

[37] A Krishnamoorthy, P K Pramod and T G Deepak; On a queue with interruptions and repeat or resumption of service, *Nonlinear Analysis: Theory, Methods & Applications*, Volume 71, Issue 12, 15 December 2009, Pages e1673-e1683.

[38] A Krishnamoorthy, Chakravathy, P K Pramod; Queues With Interruption and Repeat/Resumption of Service- A Survey and Further Results, Fourth International Conference on Neural, Parallel & Scientific Computations; August 11-14, 2010, MorehouseCollege, Atlanta, Georgia, USA.

[39] A Krishnamoorthy, S.N. Sajeev and C.N.Viswanath; An Inventory Model with Retrial and Orbital Search ; *Bulletin of Kerala Mathematics Association*, Special Issue. 47-65, October 2009; Guest Editor: S. R. S. Varadhan FRS .

[40] A Krishnamoorthy, S. N. Sajeev and C N Viswanath; An Inventory model with positive service time and server interruptions; [unpublished].

[41] A Krishnamoorthy, S.N. Sajeev and C.N.Viswanath; An Inventory Model with Server Interruptions; Presented at 5th International Conference on Queueing Theory and Network Applications; July 24-26, 2010, Beijing, China .

[42] A Krishnamoorthy and P V Ushakumari. On an M/G/1 retrial queue with disaster to the customer in service; In International Workshop on Retrial Queues, Madrid, September 1998.

[43] A Krishnamoorthy, C.N.Viswanath, T G Deepak and K Vineetha; Control Policies for Inventory with Service Time; *Stochastic Analysis and Applications*; Volume 24, Number 4/2006, PP.889-899 .

[44] A Krishnamoorthy, C.N.Viswanath and M E Islam; On a Production Inventory with Service Time and Retrial of Customers; *Proceedings of the 11th International Conference on Analytical and Stochastic Modelling Techniques and Analysis*; ASMTA 2004 pp. 238-247, SCS- Publishing House- Magdeburg.

[45] A Krishnamoorthy, C.N.Viswanath and M E Islam; Retrial Production Inventory with MAP and Service Time; Queues, Flows, Systems, Networks. *Proceedings of the International Conference "Modern Mathematical Methods of Analysis and Optimization of Tele - communication Networks"*, Gomel, 2003, pp.148-156.

[46] A Krishnamoorthy, C N Viswanath, V C Joshua and T G Deepak; An MAP/PH/1 retrial queue with service interruption and orbital search; Accepted for publication in *Int. Journal of Applied Mathematics.*

[47] A Krishnamoorthy and C N Viswanath; Production inventory with service time and vacation to the server; *IMA Journal of Management Mathematics*; doi: 10.1093/imaman/dpp025.

[48] A Krishnamoorthy and C N Viswanath; Stochastic Decomposition in Production Inventory with Service Time; [Communicated].

[49] V G Kulkarni and B D Choi; Retrial queues with server subject to breakdowns and repairs; *Queueing Systems*; Vol. 7, No.2, 191-208, 1990.

[50] V G Kulkarni ,V.F. Nicola and K.S Trivedi ;The completion time of a job on multi node -systems, *Advances in Applied Probability*, Vol 19,No:4,pp.932-954,1987.

[51] V.G. Kulkarni, V.F. Nicola, K.S. Trivedi, "Queueing Analysis of Fault-Tolerant Computer Systems," *IEEE Transactions on Software Engineering*, vol. 13, no. 3, pp. 363-375, March 1987

[52] G Latouche. and V. Ramaswami (1999), Introduction to Matrix Analytic Methods in Stochastic Modeling, SIAM.

[53] H Masuyama and T Takine. Stationary queue length in a FIFO single server queue with service interruptions and multiple batch Markovian arrival streams; *Journal of the Operations Research Society of Japan* vol. 46(3), 319–341, 2003.

[54] J.Medhi; Stochastic processes (Second edition). New age International (P), LTD Publishers, New Delhi.

[55] R.Marie and K.Trivedi ,A note on the Effect of Preemptive Policies on the Stability of a Priority Queue , *Information Processing Letters*, Vol.24 (No 6) ,pp.397-401,1987

[56] M F Neuts; Matrix-geometric solutions in stochastic models-An algorithmic approach. John-Hopkins Univ. Press, 1981. Also published by Dover, 2002.

[57] M F Neuts and B M Rao; Numerical investigation of a multiserver retrial model; *Queueing Systems*; Volume 7, Number 2, 169-189, June 1990.

[58] M Schwarz, C Sover, H Daduna, R Kulik and R Szekli: M/M/1Queueing systems with inventory, *Queueing Systems*, 54,55-78, 2006.

[59] N.P Sherman,. and J.P. Kharoufeh; An M/M/1 retrial queue with unreliable server. *Operations Research Letters*, 34 (6), 697-705, 2006 .

[60] N.P Sherman, J.P Kharoufeh and M.A. Abramson; An M/G/1 retrial queue with unreliable server for streaming multimedia applications. *Probability in the Engineering and Informational Sciences*, 23, 281-304, 2009 .

[61] B Sivakumar; A perishable inventory system with retrial demands and a finite population; *Journal of Computational and Applied Mathematics*; Volume 224(1), February, 2009, 29 – 38.

[62] B Sivakumar; Two-commodity inventory system with retrial demand; *European Journal of Operational Research*; Volume 187(1), May 2008; 70 - 83.

[63] L.E Stanfel and B.D Sivazlian; *Analysis of Systems in Operations Research*, Prentice Hall, 1975.

[64] T Takine and B Sengupta; A single server queue with service interruptions; *Queueing Systems*, vol. 26:285–300, 1997.

[65] K Thiruvengadam. Queuing with breakdowns. *Operations Research*, vol. 411(1):62–71, 1963.

[66] B. Tomlin; Impact of supply learning when suppliers are unreliable; *Manufacturing & Service Operations Management*; Vol. 11, No. 2, pp 192-209, 2009.

[67] B. Tomlin; On the value of mitigation and contingency strategies for managing supply chain disruption risks; *Management Science*, Vol. 52, No. 5, pp 639-657, May 2006.

[68] P V Ushakumari; On (s, S) inventory system with random lead time and repeated demands; *Journal of Applied Mathematics and Stochastic Analysis*,Volume 2006 (2006), Article ID 81508,22pages; doi:10.1155/JAMSA/2006/81508.

[69] C.N.Viswanath, T G Deepak, A Krishnamoorthy and B Krishnakumar; On an (s,S) Inventory system with service time vacation to the server and correlated lead time; *Quality Technology and Quantitative Management* (QTQM), PP.129-143, Volume 5, Number 2 June 2008 .

[70] N Van Dijk; Simple bounds for queueing systems with breakdowns; *Performance Evaluation*, vol. 8(2):117–128, 1988.

[71] J Wang, J Cao and Q L Li; Reliability analysis of the retrial queue with server breakdowns and repairs. *Queueing Systems*, 38 (4), 363-380, 2001 .

[72] J Wang, B Liu and J Li; Transient analysis of an M/G/1 retrial queue subject to disasters and server failures; *European Journal of Operational Research*; Vol. 189, Issue 3, 1118-1132, 2008.

[73] H White and L Christie; Queuing with preemptive priorities or with breakdown, *Operations Research*, vol. 6 (1), 79–95, 1958.

[74] P Wuchner, J Sztrik, and H de Meer. Homogeneous finite-source retrial queues with search of customers from the orbit, In GI/ITG Konferenz Messung, *Modellierung und Bewertung von Rechen- und Kommunikationssystemen* (MMB 2008), Dortmund, Germany, March 2008.