OPERATIONS RESEARCH

ANALYSIS OF SOME QUEUEING MODELS RELATED TO N-POLICY

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

BY

DEEPAK T. G.

DEPARTMENT OF MATHEMATICS
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
COCHIN – 682 022, INDIA.

MAY 2001

CERTIFICATE

This is to certify that the thesis entitled "**Analysis of some queueing models related to N-policy**" is a bonafide record of the research work carried out by Mr. Deepak T. G. under my supervision in the Department of Mathematics, Cochin University of Science and Technology. The results embodied in this 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, Department of Mathematics
Cochin University of Science & Technology
Cochin.

Cochin-682 022
24 May 2001

Contents

Chapter 1

Introduction

Over the last few years it has been increasingly realised that probability models are more realistic than deterministic models in many situations. There are many well known common and nontrivial areas of application for probability models. Queueing theory is one such area where probability models can effectively be used.

The first work on waiting line (queue) was "The theory of probabilities and telephone conversations" by A.K. Erlang [11] who published this paper in 1909. This was devoted for the study of telephone traffic congestion.

The study of queues is mainly applied in the fields of business (banks, supermarkets, booking offices etc.), industries (serving of automatic machines, production lines storage etc.), technology (telephony, communication networks, computers etc.), transportation (airports, harbours, railways, postal services etc.) and in every day life (elevators, restaurants, barber shops etc.). These are concerned with the design and planning of service facilities to meet randomly fluctuating demands for service so that congestion is minimised and the economic balance between the cost of service and the cost associated with waiting for that service is maintained.

1.1 Queueing systems and their basic characteristics

A system consisting of a servicing facility, a process of arrival of customers who wish to be served by the facility and the process of service, is called a queueing system.

The following characteristics provide an adequate description of any queueing system.

1.1.1 Arrival pattern of customers

If the arrivals and service times are strictly according to schedule, queues can be avoided, but in practice this is not the case and in most situations arrivals are controlled by factors external to the system. Therefore, the best that can be done is to represent the input process in terms of random variables. Further characterisation is required in the form of the probability distribution associated with this random process.

Arrivals may occur in batches instead of one at a time. In the event that more than one arrival can enter the system simultaneously, the input is said to occur in bulk or batch. In the bulk arrival situation not only the time between successive arrivals of the batches may be probabilistic but also the number of customers in a batch.

If the queue is too long, a customer may decide not to enter it upon arrival and he is said to have balked. On the other hand, a customer may enter the queue, but after some time he may lose his patience and may decide to leave. In this case he is said to have reneged. In the event that there are two or more parallel waiting lines, customers may switch over from one to another, jockeying for position. These three situations are examples of queues with impatient customers. If an arrival pattern does not change with time, then it is called a stationary arrival pattern; otherwise, it is called non stationary.

1.1.2 Service pattern of servers

The uncertainties involved in the service mechanism are the number of servers, the number of customers getting served at any time and the duration of service. Hence random variable representations of these characteristics seem to be essential.

Service may also be single or in batches. There are many situations where a batch of customers is served by a single server. The service rate may depend on the number of customers waiting for service. A server may work faster if he sees that the queue is building up or conversely, he may get flustered and become less efficient. Service rate can be stationary or non stationary with respect to time. In bulk service system, the batches may be of fixed size or variable size.

In some queueing systems, servers that become idle leave the system for a random period of time called vacation. These vacations may be utilised to perform additional work assigned to the servers. There are certain queueing models where the server's vacation period is the time until the accumulation of a certain number of customers (N-policy) or a certain amount of work (D-policy) or until the elapse of a certain amount of time (T-policy).

1.1.3 Queue discipline

Queue discipline is the rule according to which customers are selected for service when a queue is formed. The most common queue discipline is "first in first out" (FIFO) rule under which the customers are served in the strict order of their arrivals. Another queue discipline is "last in first out"(LIFO) rule by which the last arrival in the system is served first. Yet another queue discipline is "service in random order" (SIRO) rule according to which the arrivals are served randomly irrespective of their arrival to the system.

In some cases "priority" (PRI) discipline is followed. This discipline allows priority in service to some customers in relation to other customers waiting in the queue. Priority disciplines are classified as preemptive priority discipline and non preemptive priority discipline. According to preemptive priority discipline a customer with the highest priority is allowed to enter service immediately suspending even the service in progress to a customer with lower priority. In the non preemptive case, the highest priority customer goes to the head of the queue but gets into service only after completion of the service in progress to the customer with lower priority.

1.1.4 System capacity

The system may have either a limited or an unlimited capacity for holding customers.The source from which the customers come may be finite or infinite. In some cases it is important to limit the length of the queue to some predetermined capacity; in other cases the capacity can be considered to be infinite.

Some of the queueing processes admit the physical limitation to the amount of waiting room so that when the waiting line reaches a certain length no further customers are allowed to enter until space becomes available by a service completion. Such systems with a finite limit to the maximum queue size are called finite queueing systems. These systems can be viewed as having forced balking where a customer is forced to balk if he arrives at a time when the queue size is at its maximum limit.

In some problems service is made up of several phases and is rendered by service facilities arranged in series. Queues are allowed to build up in front of each service facility. These intermediate queues known as buffer may again have finite or infinite length.

1.1.5 Service channels

Queueing system may have several service channels to provide service. These service channels may be arranged in parallel or in series or as a more complex combination of both, depending on the design of the system's service mechanism.

In parallel channels a number of channels provide identical service facilities so that several customers may be served simultaneously. In case of series channels a customer must pass successively through the ordered channels before his service is completed. Queueing models in which there exist a series of service stations through which each calling unit must progress prior to leaving the system were studied by several researchers. Such series queueing situations are referred to as tandem queues.

A queueing system is called single server model when the system has one server only and when the system has a number of parallel servers it is known as multiserver model.

1.1.6 Notation

A queueing system is represented by the notation $A|B|C|X|Y$ where A is the interarrival time distribution, B is the service time distribution, C is the number of parallel servers, X is the system capacity and Y is the queue discipline. This is called Kendall-Lee notation. For example $M|G|10|\infty|$FIFO indicates a queueing system with exponential interarrival times, general service times, 10 parallel servers, no restriction on the maximum number allowed in the system and first-in first-out queue discipline. If a queueing system is represented by $A|B|C$, then it is understood that the system capacity is infinite and the queue discipline is FIFO.

1.2 Methods for solving queueing models

Queueing models can be broadly classified into Markovian queueing models and non-Markovian queueing models.

1.2.1 Markovian queueing models

Queueing models with inter-arrival time of customers and service time exponentially distributed are called Markovian queueing models. Markovian queueing models are analysed by

1. the difference-differential equations method or

2. the matrix-geometric algorithmic method.

Some queueing systems are studied analytically by deriving the corresponding difference - differential equations and solving them by using suitable generating functions. This method is discussed in detail by Gross and Harris [12], Kleinrock [17] and Saaty [32]. Neuts[28] developed what is called matrix-geometric algorithmic approach for studying the steady state queueing models. Matrix-geometric approach involves only real arithmetic and avoids the calculation of complex roots based on Rouche's theorem.

1.2.2 Non-Markovian queueing models

The exponential assumption on probability distribution, although certainly convenient, is not always realistic. There is a practical need for models that do not rely on strict Markov assumptions. Queueing models having the interarrival times and/or service times which are not exponentially distributed are known as non-Markovian queueing models.

The techniques generally used in studying non-Markovian queues are:

1. **Embedded Markov chain technique:** This technique, introduced by Kendall [16], is commonly used when one among the service time and interarrival time is exponentially distributed while the other is not.

2. **Supplementary variable technique:** Some non-Markovian models can be analysed by converting them into Markovian models through the introduction of one or more supplementary variables. This is known as Supplementary Variable technique. Cox[9] has analysed non-Markovian stochastic processes by the inclusion of supplementary variables.

1.3 Types of vacations

Queueing systems in which server leaves for a vacation were studied by many researchers. The non-availability of a server at the system may be termed as server's vacation. In a queueing system, if the queue is empty, then the idle time of the server can be utilised to perform additional jobs or for the preventive maintenance work which can be divided into short segments. Sometimes maintenance work may have to be done even when the queue length is empty. The following are some of the commonly applied server vacation policies.

1.3.1 Repeated vacations

In some bulk service models a server, on completion of a service, will start service again only if the system has at least a certain minimum number of customers required to start the service. Otherwise, the server will withdraw from the system for a vacation. On return after vacation period if the server finds less than the required number of customers he may immediately take another vacation. He will continue in this manner until he finds, upon returning from a vacation, the required minimum number of waiting customers.

1.3.2 Single vacation

The assumptions are same as those of repeated vacations except that, even if the server finds less than the minimum number of customers required for service when he returns from a vacation, he stays in the system waiting for the queue length to reach the minimum number for starting his next service.

1.3.3 Exceptional first vacation

In repeated vacations the duration of the first vacation and the subsequent vacations are assumed to have the same distribution. In exceptional first vacation, the duration of the first vacation is differently distributed from that of the subsequent vacations.

1.3.4 Gated vacation

In this vacation model, as soon as the server returns from a vacation, he serves only those customers who were waiting at time of his return to the system. The services of subsequent arrivals are deferred until after the next vacation. In this model when the server returns from vacation a gate closes behind the last waiting customer and the server will serve only those customers in front of the gate before leaving for another vacation.

1.3.5 Random vacation

A machine used to produce a variety of items, may breakdown randomly independent of the status of the queue. This breakdown may be regarded as server's vacation.

1.3.6 Limited service vacation

Sometimes, after producing a specific number of items a machine (server) may have to be sent for maintenance or stopped to remain idle for sometime, to make it fit for further production. In such cases, it is said that the machine (server) is allowed to take "limited service vacation".

1.4 Relevant literature survey

1.4.1 Queueing systems with vacation

Analysis of queueing models with different types of vacation were done by several researchers. Doshi [10] provides a survey of queueing systems with vacations in which he attempted to provide a methodological overview with the objective of illustrating how the seemingly diverse mix of problems are closely related in structure and can be understood in a common frame work.

Levy and Yechiali [22] considered an $M|M|S$ queueing system with servers' vacation. The distribution of the number of busy servers and the mean number of units in the system were obtained by considering repeated vacation as well as single vacation.

Takagi [36] studied exhaustively different types of vacation models. He also, analysed an $M|G|1$ queue with multiple server vacation which is very well applied to a polling model [38].

Ho Woo Lee [21] studied an $M|G|1$ queue with exceptional first vacation and obtained the transform solution of the system size distribution by defining supplementary variables.

A queueing system with single server who serves customers according to general bulk service rule and leaves the system for vacation was analysed by Nadarajan and Subramanian [24]. Both repeated vacation and single vacation of server are considered. The steady state probability vector of the number of customers in the system and the stability condition were obtained, using matrix-geometric method.

1.4.2 Control policies for a single server system

Much of the recent research in queueing theory has been concerned with optimisation. Yadin and Naor [40] obtained the optimal value the queue size has to attain in order to turn on a single server, assuming that the policy is to turn on the server when the queue size reaches a certain number, N, and turn him off when the system is empty. This is called N-policy. Heyman [13] also considered similar policies and showed the optimality of the policy under certain conditions. Balachandran [4] and Balachandran and Tijms [5] considered the D-policy, which activates the server when the cumulative service times of customers in the queue first reach (or exceed) a threshold D. Heyman [14] introduced another policy, called T-policy in which server takes a vacation of T time units after the completion of each busy period.

The above optimal policies were compared in several studies by employing different cost functions. Balachandran and Tijms [5] proved that the D-policy is superior to N-policy for exponentially distributed service times, with decreasing failure rates and for some cases with increasing failure rates, by employing a cost function based on the mean workload. Heyman [14] proved that N-policy is superior to T-policy by using a cost function based on the expected queue length. Artalejo [1] showed that the T-policy is the worst among the three under the above two cost structures and that the relation between the optimum N and D policies depends on the cost function employed.

Lee and Srinivasan [20] studied the control policies for an $M^X|G|1$ queueing system and derived the mean waiting time of an arbitrary customer. Also they presented the procedure to find the stationary optimal policy under a linear cost structure.

Takagi [37] studied time-dependent behaviour of $M|G|1$ vacation model. Also Takagi [39] considered a finite capacity $M|G|1$ queueing model with set up time under N-policy and derived certain system characteristics. A Poisson input queue, under N-policy with a general startup time was analysed by Medhi and Templeton [23].

1.4.3 Tandem queues

In queueing problems such as assembling of parts in a factory, undergoing medical check up in a clinic, or driving through several traffic intersections, service is made up of several phases and is rendered by facilities arranged in series. Queueing models in which there exist a series of service stations, through which each calling unit must progress prior to leaving the system, were studied by several researchers.

A queueing model involving tandem queues with finite waiting room in between the two servers was discussed by Neuts [26]. The study of blocking in two or more units in service with general service time distribution without intermediate buffer was considered by Avi-ltzhak and Yadin [2]. Clarke [8] investigated a tandem queueing model where in two servers are placed in series and each customer will receive service from one and only one server. Nadarajan and Audsin Mohana Dhas [25] studied a model consisting of two units 1 and 2 connected in series with a finite intermediate waiting room. The customers in the buffer are served according to a general bulk service rule with exponential times. Unit 2 is in the upstate and downstate, following exponential distribution.

1.5 Author's contribution

In this thesis the analysis of some queueing models that are related to the well known 'N-policy' has been developed and presented.

In chapter 2, an $M|M|1$ queueing model under a new operating policy, called modified N-policy, is considered as follows: The server on becoming idle waits until N units accumulate for service. ie, his vacation period ends at the arrival of N^{th} unit. These N units are served together as a batch unlike in the usual N-policy to minimise customer impatience and subsequent arrivals are served in single. Here it is assumed that arrival process is Poisson and both batch service and single service are exponentially distributed with different service rates. Steady state probabilities are obtained. Some measures of effectiveness are computed. Optimal N value is calculated. Some numerical illustrations are provided. Laplace transforms of the time dependent probabilities are obtained. Also waiting time distribution is derived.

In chapter 3, an $M|G|1$ queueing model under modified N-policy is considered. ie, the arrival process is a Poisson process and both types of services are arbitrarily distributed. By using embedded Markov chain technique, both departure time and arbitrary time probabilities are computed. Some measures of effectiveness are obtained and optimal N value is investigated. Also waiting time distribution is derived.

Chapter 4 analyses an $M|G|1$ queueing model under two different operating policies. In model 1, the operating policy is the usual N-policy, but with random N and in model 2, a system similar to the one described in chapter 3 is considered with the only difference that N is not deterministic but random. In both models the size of the queue at which initial service starts will be determined by the outcome of a random experiment. For these two models, steady state distributions are derived. Some measures of performance of the system are computed and the optimal distribution of N from a given class of distributions is investigated.

Chapter 5 is partly devoted for the transient analysis of an $M|M|1$ queue under the usual N-policy. Here transient state (time dependent) probabilities in terms of Bessel functions are obtained. Also the output distribution (distribution of time between successive departures) in the steady state under N-policy is derived.

The last chapter analyses "Tandem queue with two servers". Here we assume that the first server is a specialised one. He will be activated only after the accumulation of N units in the system. At the arrival of the N^{th} unit, he starts giving service one at a time till none is left before him. ie., N-policy is the operating policy for this server unit. After being served by this specialised server, a customer will go to the second server unit. If the server is busy at that time he will have to wait till his turn for service comes. Otherwise he can join service directly. After being served by the second unit he leaves the system. It is assumed that the second server unit is always available and there is a finite capacity waiting room between the two servers. The arrival process to the first server is assumed to be a Poisson process and service time distributions of both servers are assumed as exponential. Here the infinitesimal generator matrix has been obtained in block partitioned tridiagonal form and so the steady state probability vectors are obtained in matrix geometric form. Also the stability condition is established.

Chapter 2

Modified N-policy for the $M|M|1$ queue

N-policy for queues has been investigated by several researchers. In a queueing system, under N-policy, the server will be on vacation until N units accumulate for service for the first time after becoming idle. As soon as N units accumulate in the system, he starts service, one at a time, till the system becomes empty. The server will be turned on again when the queue size reaches the number N. The process continues in this fashion.

In this chapter, a modified version of the N-policy for an $M|M|1$ queueing system is considered. This modified policy is defined as follows: The server on becoming idle waits until N units accumulate for service. These N units are served in a single batch and subsequent arrivals during the busy period initiated by this batch receive single service. Service time distribution for bulk service is different from that of single service.

The above problem is motivated by some real life situations in which when a server becomes idle, he is turned off. The fixed cost of getting him back to serve may turn out to be very high. Hence service commences only after a certain number (here N) units queue up. These units are served in bulk unlike in the usual N-policy to reduce customer impatience. If N is taken too large, cost associated with waiting time of the customers, who have reached during the vacation period, will increase tremendously, whereas if N is taken very small, there will be a large number of busy cycles so that the setup cost associated with the starting of busy period will increase. Hence a trade off between these two is called for.

Let the arrival process form a Poisson process of rate λ and the service times obey the exponential distributions with parameters μ_1 or μ_2 according as the services are in single or in batch. Since the expected service time for a bulk service may be more than that of a single service, it is assumed that $\mu_2 < \mu_1$. Since the number of customers who are arriving during the batch service is the determining factor of system stability, it is also assumed that $\rho_1 = \frac{\lambda}{\mu_1} < 1$.

2.1 Notations

Let $X(t)$ be the number of units in the system at time t.

$$\text{Define } Y(t) = \begin{cases} 0 & \text{if the server is idle at } t \\ 1 & \text{if a single service is taking place at } t \\ 2 & \text{if a batch service is taking place at } t \end{cases}$$

Then $\{(X(t), Y(t)),\ t \geq 0\}$ is a continuous time Markov process with the state space

$S = \{(i,0)|0 \leq i \leq N-1\} \cup \{(i,1)|i \geq 1\} \cup \{(i,2)|i \geq N\}$

Let $P_{ij}(t)$ be the probability that the system is in state (i,j) at time t and $q_{ij} = \lim_{t\to\infty} P_{ij}(t)$.

This chapter is presented as follows. In section 2.2, the Markov chain $\{(X(t), Y(t));\ t \geq 0\}$ is analysed to get the stationary behaviour. Some measures of effectiveness are also computed in this section. In section 2.3, the optimal N value is investigated, convexity of the cost function is established and some numerical illustrations are provided. In section 2.4, waiting time distribution is derived and expected waiting time in the queue is computed. Section 2.5 is devoted to the transient or time dependent behaviour of the system.

2.2 Steady state analysis

Clearly $P_{ij}(t)$ satisfy the following system of Kolmogrov differential equations:

$$P'_{00}(t) = -\lambda P_{00}(t) + \mu_1 P_{11}(t) + \mu_2 P_{N2}(t) \tag{2.1a}$$

$$P'_{n0}(t) = -\lambda P_{n0}(t) + \lambda P_{n-1,0}(t) \quad \text{for } 1 \le n \le N-1 \tag{2.1b}$$

$$\begin{aligned} P'_{n1}(t) = &-(\lambda + \mu_1)P_{n1}(t) + \mu_1 P_{n+1,1}(t) \\ &+ \mu_2 P_{N+n,2}(t) + \lambda(1-\delta_{1n})P_{n-1,1}(t) \quad \text{for } n \ge 1 \end{aligned} \tag{2.1c}$$

$$\begin{aligned} P'_{N+n,2}(t) = &-(\lambda + \mu_2)P_{N+n,2}(t) + \lambda(1-\delta_{0n})P_{N+n-1,2}(t) \\ &+ \lambda\delta_{0n}P_{N-1,0}(t) \quad \text{for } n \ge 0 \end{aligned} \tag{2.1d}$$

where δ_{ij} is the Kronecker delta. Then the steady state probabilities q_{ij} satisfies the following system of equations.

$$0 = -\lambda q_{00} + \mu_1 q_{11} + \mu_2 q_{N2} \tag{2.2a}$$

$$0 = -\lambda q_{n0} + \lambda q_{n-1,0} \quad \text{for } 1 \le n \le N-1 \tag{2.2b}$$

$$0 = -(\lambda + \mu_1)q_{n1} + \mu_1 q_{n+1,1} + \mu_2 q_{N+n,2} + \lambda(1-\delta_{1n})q_{n-1,1} \quad \text{for } n \ge 1 \tag{2.2c}$$

$$0 = -(\lambda+\mu_2)q_{N+n,2} + \lambda(1-\delta_{0n})q_{N+n-1,2} + \lambda\delta_{0n}q_{N-1,0} \quad \text{for } n \ge 0 \tag{2.2d}$$

(2.2b) gives:

$$q_{n0} = q_{00} \quad \text{for } 1 \le n \le N-1 \tag{2.3a}$$

(2.2d) gives:

$$q_{N+n,2} = (\frac{\lambda}{\lambda+\mu_2})^{n+1} q_{00} \quad \text{for } n \geq 0 \tag{2.3b}$$

(2.2a) gives:

$$q_{11} = [\frac{\lambda^2}{\mu_1(\lambda+\mu_2)}] q_{00} \tag{2.3c}$$

Using (2.3b), (2.2c) can be written as

$$(E^2 - (1+\frac{\lambda}{\mu_1})E + \frac{\lambda}{\mu_1}) q_{n1} = \frac{-\mu_2}{\mu_1}(\frac{\lambda}{\lambda+\mu_2})^{n+2} q_{00} \quad \text{for } n \geq 2$$

which is a non-homogeneous linear differential equation of order 2 and E is the right shift operator. The general solution to this equation is

$$q_{n1} = A(\frac{\lambda}{\mu_1})^n + B - \frac{\frac{\mu_2}{\mu_1}(\frac{\lambda}{\lambda+\mu_2})^{n+2} q_{00}}{r(\frac{\lambda}{\lambda+\mu_2})}$$

where $r(E) = E^2 - (1+\frac{\lambda}{\mu 1})E + \frac{\lambda}{\mu 1}$ and A, B are arbitrary constants. Hence

$$q_{n1} = A(\frac{\lambda}{\mu_1})^n + B - \frac{\lambda q_{00}}{\lambda - \mu_1 + \mu_2}(\frac{\lambda}{\lambda+\mu_2})^n \quad \text{for } n \geq 2.$$

Since $\sum_{n=1}^{\infty} q_{n1} < 1$, $B = 0$. Therefore

$$q_{n1} = A(\frac{\lambda}{\mu_1})^n - \frac{\lambda q_{00}}{\lambda - \mu_1 + \mu_2}(\frac{\lambda}{\lambda+\mu_2})^n \quad \text{for } n \geq 2. \tag{2.3d}$$

Choose A in such a way that this result holds for $n = 1$ also. Substituting the value thus obtained for A in (2.3d), it is found that

$$q_{n1} = \frac{\lambda^{n+1}}{\lambda - \mu_1 + \mu_2}[\frac{1}{\mu_1^n} - \frac{1}{(\lambda+\mu_2)^n}] q_{00} \quad \text{for } n \geq 1 \tag{2.3e}$$

Substituting (2.3a), (2.3b), and (2.3e) in the normalising condition $\sum_{n=0}^{N-1} q_{n0} + \sum_{n=1}^{\infty} q_{n1} + \sum_{n=0}^{\infty} q_{N+n,2} = 1$, we get

$$q_{00} = \frac{\mu_2(\mu_1 - \lambda)}{\lambda\mu_1 + N\mu_2(\mu_1 - \lambda)} \tag{2.3f}$$

Lemma 2.2.1. *Average queue size when the server is busy is*

$$L = \frac{\lambda^2[\lambda\mu_2 + \mu_1(\mu_1 - \lambda)]}{\mu_2(\mu_1 - \lambda)[\lambda\mu_1 + N\mu_2(\mu_1 - \lambda)]}$$

Proof. We have

$$L = \sum_{n=1}^{\infty}(n-1)q_{n1} + \sum_{n=0}^{\infty} nq_{N+n,2} = \frac{\lambda^2[\lambda\mu_2 + \mu_1(\mu_1 - \lambda)]}{\mu_2(\mu_1 - \lambda)[\lambda\mu_1 + N\mu_2(\mu_1 - \lambda)]}$$

Lemma 2.2.2. *The expected duration l_1 of a busy period is given by*

$$l_1 = \frac{\mu_1}{\mu_2(\mu_1 - \lambda)}$$

Proof. Since the expected duration of a busy period for an ordinary $M|M|1$ queue with arrival rate λ and service rate μ is $\frac{1}{\mu - \lambda}$,

$$l_1 = \frac{1}{\mu_2} + \sum_{i=0}^{\infty} \frac{i}{\mu_1 - \lambda} e^{-\lambda/\mu_2} \frac{(\lambda/\mu_2)^i}{i!} = \frac{\mu_1}{\mu_2(\mu_1 - \lambda)}$$

□

Lemma 2.2.3. *Mean length of a busy cycle, $B = \frac{\mu_1}{\mu_2(\mu_1 - \lambda)} + \frac{N}{\lambda}$.*

Proof. Since successive busy and idle periods constitute a busy cycle,

$$B = l_1 + \frac{N}{\lambda} = \frac{\mu_1}{\mu_2(\mu_1 - \lambda)} + \frac{N}{\lambda}$$

□

2.3 Determination of optimal N

Here we investigate that value of N which minimises a suitably defined cost function. The following important costs are included in the cost function, which is denoted by F_{Mod-N}.

i) C_1: Waiting time cost per customer per unit time when the server is busy.

ii) C_2: Unit time service cost associated with batch service.

iii) C_3: Unit time service cost associated with single service.

iv) C_4: Cost towards waiting per unit time until service starts after an idle period.

v) K: Fixed cost for commencement of each busy period, that is, the set up cost.

Then the total expected cost per unit time,

$$F_{Mod-N} = C_1 L + (C_2 + K)\frac{1}{B} + C_3 \sum_{n=1}^{\infty} n q_{n1} + C_4[\frac{N-1}{\lambda} + \frac{N-2}{\lambda} + \cdots + \frac{1}{\lambda}]$$

As a particular case, choose $\mu_2 = \frac{\mu_1}{\alpha N}$ where $0.5 \leq \alpha < 1$; which means that the expected service time for a batch of N units is less than the time needed for N single services. With this modification F_{Mod-N} takes the form

$$F_{Mod-N} = C_1 \frac{\alpha\lambda^2[\lambda + \alpha N(\mu_1 - \lambda)]}{\mu_1(\mu_1 - \lambda)[\mu_1 + (\alpha - 1)\lambda]} + (C_2 + K)\frac{\lambda(\mu_1 - \lambda)}{N[\mu_1 + (\alpha - 1)\lambda]} + C_3 \frac{\alpha\lambda^2[\mu_1^2 + \alpha N\lambda(\mu_1 - \lambda)]}{\mu_1^2(\mu_1 - \lambda)[\mu_1 + (\alpha - 1)\lambda]} + C_4 \frac{N(N-1)}{2\lambda} \quad (2.4)$$

By approximating N as a continuous variable, it can be shown that the second derivative of F_{Mod-N} with respect to N is $\frac{2(C_2+K)\lambda(\mu_1-\lambda)}{N^3[\mu_1+(\alpha-1)\lambda]} + \frac{C_4}{\lambda}$ which is a positive quantity since $\rho_1 = \frac{\lambda}{\mu_1} < 1$. Hence F_{Mod-N} is convex in N. By equating the first derivative of F_{Mod-N} to zero, it can be seen that optimal N value is the root of the equation

$$2C_4\mu_1^2[\mu_1 + (\alpha-1)\lambda]N^3 + [2\alpha^2\lambda^3(C_1+C_3)(\lambda+\mu_1) - C_4\mu_1^2(\mu_1+(\alpha-1)\lambda)]N^2 + 2(C_2+K)(\lambda-\mu_1)\lambda^2\mu_1^2 = 0. \quad (2.5)$$

(2.5) is of the form $y^3 + py^2 + qy + r = 0$ and this can be reduced to the normal form $x^3 + ax + b = 0$ by the substitution $y = x - \frac{p}{3}$ where $a = \frac{1}{3}(3q - p^3)$ and $b = \frac{1}{27}(2p^3 - 9pq + 27r)$.

If p, q, r are real (hence a, b are real) and $\frac{b^2}{4} + \frac{a^3}{27} > 0$, the equation $y^3+py^2+qy+r = 0$ has exactly one real root and two conjugate imaginary roots and the real root is given by $y = A_1 + A_2 - \frac{p}{3}$ where

$$A_1 = \sqrt[3]{-b/2 + \sqrt{b^2/4 + a^3/27}} \quad \text{and} \quad A_2 = \sqrt[3]{-b/2 - \sqrt{b^2/4 + a^3/27}}.$$

In the case of (2.5) it can be proved that $b^2/4 + a^3/27 > 0$ and so it has exactly one real root, namely,

$$N^* = [A_1 + A_2 + \frac{1}{3}[\frac{1}{2} - (\frac{C_1+C_3}{C_4})\frac{\alpha^2\lambda^3(\lambda+\mu_1)}{\mu_1^2(\mu_1+(\alpha-1)\lambda)}]]^+ = [H]^+$$

where

$$A_1 = \sqrt[3]{-\frac{b}{2} + \sqrt{b^2/4 + a^3/27}}, \quad A_2 = \sqrt[3]{-b/2 - \sqrt{b^2/4 + a^3/27}},$$

$$a = -\frac{1}{3}[\frac{C_1+C_3}{C_4} \cdot \frac{\alpha^2\lambda^3(\lambda+\mu_1)}{\mu_1^2(\mu_1+(\alpha-1)\lambda)} - \frac{1}{2}]^2$$

$$b = \frac{C_2+K}{C_4}\frac{\lambda^2(\lambda-\mu_1)}{\mu_1+(\alpha-1)\lambda} + 2(\frac{-a}{3})^{3/2} \text{ and } \frac{b^2}{4} + \frac{a^3}{27} > 0.$$

Since N^* must be an integer, substitute the integers close to H in (2.4) and pick the one giving lower expected cost. The notation $[\,]^+$ refers to the integer chosen in this manner. Some numerical illustrations are shown in the following table:

λ	μ_1	α	C_1	C_2	C_3	C_4	K	N^*
5	8	0.5	25	50	40	30	200	4
10	13	0.5	30	75	45	45	250	5
15	23	0.5	30	80	55	45	260	8
22	30	0.6	35	80	55	55	210	6
15	35	0.6	40	80	55	65	200	15
20	50	0.7	45	87	62	57	220	10
20	40	0.7	45	115	70	90	350	9
25	40	0.75	50	130	75	75	300	8
30	50	0.75	50	120	75	75	280	9
40	70	0.75	60	140	80	100	400	12
30	65	0.8	55	165	90	80	500	13
45	90	0.8	55	165	90	80	500	12
50	85	0.8	60	180	90	100	500	19
55	95	0.8	60	185	90	110	400	14
60	100	0.8	70	200	95	125	500	14

2.4 Waiting time distribution

Let T represent the "time spent waiting in the queue" by an arbitrary customer and $W(\cdot)$ be the cumulative distribution function of T.
Then $W(0) = P\{T = 0\} = P\{$ the arrival finds the system in state $(N-1, 0)\} = q_{N-1,0} = q_{00}$
and $P\{0 < T \leq t\} = \sum_{(i,j)\neq(N-1,0)} P\{0 < T \leq t|$ the arrival finds the

system in state $(i,j)\}$ q_{ij}

$$
\begin{aligned}
&= \sum_{i=0}^{N-2} \int_0^t \frac{e^{-\lambda u}\lambda(\lambda u)^{N-i-2}}{(N-i-2)!} du \; q_{i0} + (1-e^{-\mu_2 t}) q_{N2} \\
&\quad + \sum_{i=N+1}^{\infty} \int_0^t \mu_2 e^{-\mu_2 u} \; * \; e^{-\mu_1 u} \frac{\mu_1(\mu_1 u)^{i-N-1}}{(i-N-1)!} du \; q_{i2} \\
&\quad + \sum_{i=1}^{\infty} \int_0^t e^{-\mu_1 u} \frac{\mu_1(\mu_1 u)^{i-1}}{(i-1)!} du \; q_{i1}
\end{aligned}
$$

where $*$ denotes convolution

$$
\begin{aligned}
&= \{(N-1)(1-e^{-\lambda t}) - e^{-\lambda t} \sum_{j=1}^{N-2} \frac{j(\lambda t)^{N-j-1}}{(N-j-1)!} + \frac{\lambda\mu_1}{\mu_2(\mu_1-\lambda)} \\
&\quad - \frac{\lambda(\mu_1-\mu_2)e^{-\mu_2 t}}{\mu_2(\mu_1-\mu_2-\lambda)} - \frac{\lambda^2 e^{-\mu_1 t}}{\mu_1-\mu_2-\lambda}[\frac{1}{\lambda+\mu_2} - \frac{e^{\lambda t}}{\mu_1-\lambda}]\} q_{00}
\end{aligned}
$$

Hence
$$
\begin{aligned}
\frac{d}{dt}W(t) &= \frac{d}{dt}P\{0 < T \le t\} + W(0) \\
&= \{\lambda e^{-\lambda t} \sum_{j=0}^{N-2} \frac{(\lambda t)^j}{j!} + \frac{\lambda(\mu_1-\mu_2)}{\mu_1-\mu_2-\lambda} e^{-\mu_2 t} \\
&\quad + \frac{\lambda^2 \mu_1 e^{-\mu_1 t}}{(\lambda+\mu_2)(\mu_1-\mu_2-\lambda)} - \frac{\lambda^2 e^{-(\mu_1-\lambda)t}}{\mu_1-\mu_2-\lambda} + 1\} q_{00}
\end{aligned}
$$

It can be verified that $\int_0^\infty \frac{d}{dt}P\{0 < T \le t\}dt + W(0) = 1$. Now the expected waiting time in the queue,

$$
\begin{aligned}
E(T) = 0 \cdot q_{00} + q_{00} \int_0^\infty & t[\lambda e^{-\lambda t} \sum_{j=0}^{N-2} \frac{(\lambda t)^j}{j!} + \frac{\lambda(\mu_1-\mu_2)}{\mu_1-\mu_2-\lambda} e^{-\mu_2 t} \\
&+ \frac{\lambda^2 \mu_1 e^{-\mu_1 t}}{(\lambda+\mu_2)(\mu_1-\mu_2-\lambda)} - \frac{\lambda^2 e^{-(\mu_1-\lambda)t}}{\mu_1-\mu_2-\lambda}] dt
\end{aligned}
$$

$$= [\frac{N(N-1)}{2\lambda} + \frac{\lambda(\mu_1-\mu_2)}{\mu_2^2(\mu_1-\mu_2-\lambda)} + \frac{\lambda^2}{\mu_1(\lambda+\mu_2)(\mu_1-\mu_2-\lambda)} - \frac{\lambda^2}{(\mu_1-\lambda)^2(\mu_1-\mu_2-\lambda)}]q_{00}$$

2.5 Transient behaviour

Assume that the system size at time 0 is i where $i < N$ and the server is not activated. ie., the initial service will start only after the accumulation of $N - i$ more units. Thus $P_{i0}(0) = 1$ and $P_{jk}(0) = 0$ for $(j,k) \neq (i,0)$ Define the probability generating functions

$$G_1(z,t) = \sum_{n=1}^{\infty} z^n P_{n0}(t) \text{ (where } P_{n0}(t) = 0 \text{ for } n \geq N),$$

$$G_2(z,t) = \sum_{n=1}^{\infty} z^n P_{n1}(t) \quad \text{and} \quad G_3(z,t) = \sum_{n=0}^{\infty} z^n P_{n2}(t)$$

(where $P_{n2}(t) = 0$ for $n < N$) such that all these series are convergent for $|z| \leq 1$. Multiplying equation (2.1b) throughout by z^n and summing over n from 1 to $N - 1$, it is seen that

$$\frac{\partial}{\partial t}G_1(z,t) + (\lambda - \lambda z)G_1(z,t) = \lambda z P_{00}(t) - \lambda z^N P_{N-1,0}(t).$$

Taking Laplace transform of both sides and rearranging, we get

$$\bar{G}_1(z,s) = \frac{z}{s+\lambda-\lambda z}\{z^{i-1}[1-(\frac{\lambda z}{s+\lambda})^{N-i}] + \lambda\bar{P}_{00}(s)[1-(\frac{\lambda z}{s+\lambda})^{N-1}]\} \tag{2.6}$$

where $\bar{G}_1(z,s)$ and $\bar{P}_{00}(s)$ are the Laplace transforms of $G_1(z,t)$ and $P_{00}(t)$, respectively. Multiplying (2.1d) throughout by $z^{N+n}(n \geq 0)$ and summing over n from 0 to ∞ gives

$$\frac{\partial}{\partial t}G_3(z,t) + (\lambda + \mu_2)G_3(z,t) = \lambda z G_3(z,t) + \lambda z^N P_{N-1,0}(t)$$

Taking Laplace transform of both sides and substituting the value of $\bar{P}_{N-1,0}(s)$ obtained from (2.6), it is found that

$$\bar{G}_3(z,s) = \frac{\lambda^{N-i} z^N}{(s+\lambda+\mu_2-\lambda z)(s+\lambda)^{N-i}}\left[1 + \frac{\lambda^i}{(s+\lambda)^{i-1}}\bar{P}_{00}(s)\right] \quad (2.7)$$

Multiplying (2.1c) throughout by $z^n (n \geq 1)$, summing with respect to n from 1 to ∞, adding the resulting equation and (2.1a) results in

$$P'_{00}(t) + \lambda P_{00}(t) + \frac{\partial}{\partial t}G_2(z,t) = (\lambda z - \lambda - \mu_1 + \frac{\mu_1}{z})G_2(z,t) + \frac{\mu_2}{z^N}G_3(z,t)$$

Taking Laplace transform of both sides and rearranging, it is found that

$$\bar{G}_2(z,s) = \frac{\mu_2 \bar{G}_3(z,s) - (s+\lambda)z^N \bar{P}_{00}(s)}{z^{N-1}[(s+\lambda+\mu_1)z - \lambda z^2 - \mu_1]} \quad (2.8)$$

where $\bar{G}_3(z,s)$ is given by (2.7).

Since the Laplace transform $\bar{G}_2(z,s)$ converges in the region $|z| \leq 1, Re(s) > 0$ wherever the denominator of the quotient in (2.8) has zeroes in that region, so must the numerator. The zeroes of the denominator are

$$z_1 = \frac{(s+\lambda+\mu_1) - \sqrt{(s+\lambda+\mu_1)^2 - 4\lambda\mu_1}}{2\lambda} \text{ and}$$
$$z_2 = \frac{(s+\lambda+\mu_1) + \sqrt{(s+\lambda+\mu_1)^2 - 4\lambda\mu_1}}{2\lambda}$$

Using Rouche's theorem, it can be proved that z_1 is the only zero of the denominator in $|z| \leq 1$. Therefore $(s+\lambda)z_1^N \bar{P}_{00}(s) = \mu_2 \bar{G}_3(z_1,s)$ so that

$$\bar{P}_{00}(s) = \frac{\mu_2 \bar{G}_3(z_1,s)}{(s+\lambda)z_1^N}.$$

Then (2.7) yields :

$$\bar{P}_{00}(s) = \frac{\mu_2 \lambda^{N-i}(s+\lambda)^{i-1}}{(s+\lambda)^N(s+\lambda+\mu_2-\lambda z_1) - \mu_2\lambda^N} \tag{2.9}$$

Thus in (2.6), (2.7) and (2.8), each of $\bar{G}_i(z,s)$, is expressed in terms of $\bar{P}_{00}(s)$ and $\bar{P}_{00}(s)$ is given by (2.9).

Chapter 3

Modified N-policy for the $M|G|1$ queue

The queueing model that is being discussed in this chapter is the one similar to that was discussed in the previous chapter with the only exception that the service times are arbitrarily distributed.

Let the arrivals form a Poisson process of rate λ and both single service times and batch service times are independent sequences of independent and identically distributed random variables having arbitrary distribution functions $B_1(\cdot)$ and $B_2(\cdot)$ with service rates μ_1 and μ_2, respectively. It is assumed that both service times have finite second moments.

3.1 Notations

Let $X(t)$ be the number of units in the system at time t.

$$\text{Define } Y(t) = \begin{cases} 0 & \text{if the server is idle at } t. \\ 1 & \text{if the forthcoming service is a single service or} \\ & \text{a single service is taking place at } t \text{ according as } t \text{ is} \\ & \text{a departure epoch or arbitrary epoch, respectively.} \\ 2 & \text{if a batch service is taking place at } t. \end{cases}$$

Then $\{(X(t), Y(t)) \ : \ t \geq 0\}$ is a continuous time stochastic process with the state space $S = \{(i,0)|0 \leq i \leq N-1\} \cup \{(i,1)|i \geq 1\} \cup \{(i,2)|i \geq N\}$.

Material of this chapter will appear in *Computers and Operations Research*

Let q_{ij} and π_{ij} be the steady-state probabilities that the system is in state (i,j) at an arbitrary epoch and at a departure epoch, respectively.

3.2 Analysis

The embedded stochastic process $\{(X(t_i), Y(t_i))\}$ where $t_0 = 0, t_1, t_2, t_3, \ldots$ are successive times of completion of service, is a Markov chain with state space, $\{(0,0), (1,1), (2,1), (3,1), \ldots\}$. For the time being the existence of a steady-state solution is assumed. Then the arbitrary time probabilities q_{ij} and departure point probabilities π_{ij} are connected as follows.

For $0 \leq i \leq N-1$, $q_{i0} = \pi_{00} \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^i}{i!} dt$

$$\text{ie.} \quad q_{i0} = \frac{\pi_{00}}{\lambda} \qquad \text{for } 0 \leq i \leq N-1$$

For $i \geq N$,

$$q_{i2} = \pi_{00} \int_0^\infty \int_0^t \frac{\lambda^N u^{N-1} e^{-\lambda u}}{(N-1)!} [1 - B_2(t-u)] e^{-\lambda(t-u)} \frac{[\lambda(t-u)]^{i-N}}{(i-N)!} du\, dt$$

$$\text{and} \quad q_{i1} = \sum_{j=1}^{i} \pi_{j1} \int_0^\infty \frac{e^{-\lambda u} (\lambda u)^{i-j}}{(i-j)!} [1 - B_1(u)] du \qquad \text{for } i \geq 1$$

Let Q be the transition probability matrix of the embedded Markov chain $\{(X(t_i), Y(t_i))\}$. Then

$$Q = \begin{array}{c} \\ (0,0) \\ (1,1) \\ (2,1) \\ (3,1) \\ \vdots \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} (0,0) & (1,1) & (2,1) & \ldots \end{array} \\ \begin{pmatrix} c_0 & c_1 & c_2 & \ldots \\ k_0 & k_1 & k_2 & \ldots \\ 0 & k_0 & k_1 & \ldots \\ 0 & 0 & k_0 & \ldots \\ \vdots & \vdots & \vdots & \ldots \\ \vdots & \vdots & \vdots & \ldots \end{pmatrix} \end{array}$$

where

$$\begin{aligned} c_n &= \Pr\{n \text{ arrivals during a batch service }\} \\ &= \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\,dB_2(t) \quad \text{and} \\ k_n &= \Pr\{n \text{ arrivals during a single service }\} \\ &= \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}\,dB_1(t) \end{aligned}$$

$\Pi = \{\pi_{ij}\}$ can be found as the solution to the stationary equation $\Pi Q = \Pi$. This yields:

$$\pi_{00} = \pi_{00}c_0 + \pi_{11}k_0 \tag{3.1}$$

$$\pi_{i1} = \pi_{00}\,c_i + \sum_{j=1}^{i+1}\pi_{j1}k_{i-j+1} \text{ for } i \geq 1 \tag{3.2}$$

Define the probability generating functions

$$\Pi(z) = \pi_{00} + \sum_{i=1}^{\infty}\pi_{i1}z^i, \qquad K(z) = \sum_{i=0}^{\infty}k_i z^i$$

and $C(z) = \sum_{i=0}^{\infty} c_i z^i$ such that all these series converge for $|z| \leq 1$. Multiplying (3.2) by z^i, summing over i from 1 to ∞, adding the resulting equation to (3.1), it is found that

$$\Pi(z) = \pi_{00}\frac{[zC(z) - K(z)]}{z - K(z)}$$

Since $\Pi(1) = 1$, we have $\pi_{00} = \dfrac{1-\rho_1}{1-\rho_1+\rho_2}$ where $\rho_1 = \lambda E(X_1)$, $\rho_2 = \lambda E(X_2)$ and $E(X_1)$, $E(X_2)$ are the expected durations of single service and batch service, respectively. It is assumed that $E(X_1) < E(X_2)$. ie., $\mu_2 < \mu_1$.

Now the expected system size at a departure point,

$$L' = \Pi'(1) = [\frac{d}{dz}\Pi(z)]_{z=1} = \frac{(1-\rho_1)[\lambda^2\sigma^2_{X_2} + \rho_2^2 + 2\rho_2] + \rho_2[\lambda^2\sigma^2_{X_1} + \rho_1^2]}{2(1-\rho_1)(1-\rho_1+\rho_2)} \tag{3.3}$$

where $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$ are the variances of the single service and batch service, respectively.

The application of Foster's theorem in a fashion similar to that of section 5.1.4 of [12] shows that the embedded Markov chain is ergodic and hence possesses stationary distribution when $\rho_1 = \lambda E(X_1) = \lambda/\mu_1 < 1$ provided $E(X_1) < E(X_2)$. Now consider the following lemmas.

Lemma 3.2.1. *Average queue size when the server is busy is given by*

$$L = (L' - 1 + \pi_{00})E(X_1) + \frac{\lambda(1-\pi_{00})}{2}E(X_1^2) + \frac{\lambda\pi_{00}}{2}E(X_2^2)$$

where L' is given by (3.3) *and $E(X_i^2)$ is the second row moment of X_i.*

Proof.

$$\begin{aligned} L &= \sum_{i=1}^{\infty}(i-1)q_{i1} + \sum_{i=N}^{\infty}(i-N)q_{i2} \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{i}(i-1)\pi_{j1}\int_0^{\infty}\frac{e^{-\lambda u}(\lambda u)^{i-j}}{(i-j)!}[1-B_1(u)]\,du \\ &\quad + \pi_{00}\int_0^{\infty}\int_0^{t}\frac{e^{-\lambda u}\lambda(\lambda u)^{N-1}}{(N-1)!}\lambda(t-u)(1-B_2(t-u))\,dudt \end{aligned}$$

$$\begin{aligned} &= \sum_{j=1}^{\infty}\pi_{j1}\int_0^{\infty}[(j-1)+\lambda u](1-B_1(u))\,du \\ &\quad + \pi_{00}\int_0^{\infty}\int_u^{\infty}\frac{\lambda^2 e^{-\lambda u}(\lambda u)^{N-1}}{(N-1)!}(t-u)(1-B_2(t-u))\,dtdu \end{aligned}$$

$$= \sum_{j=1}^{\infty} (j-1)\pi_{j1} \int_0^{\infty} (1 - B_1(u))du + \lambda \sum_{j=1}^{\infty} \pi_{j1} \int_0^{\infty} u(1 - B_1(u))\, du$$
$$+ \lambda \pi_{00} \int_0^{\infty} u(1 - B_2(u))\, du$$
$$= (L' - 1 + \pi_{00}) E(X_1) + \frac{\lambda(1-\pi_{00})}{2} \int_0^{\infty} u^2\, dB_1(u) + \lambda \frac{\pi_{00}}{2} \int_0^{\infty} u^2 dB_2(u)$$

since $\int_0^{\infty} u^2\, dB(u) = 2 \int_0^{\infty} u(1 - B(u))\, du.$

Thus

$$L = (L' - 1 + \pi_{00}) E(X_1) + \frac{\lambda(1-\pi_{00})}{2} E(X_1^2) + \lambda \frac{\pi_{00}}{2} E(X_2^2).$$

Hence the proof. □

Lemma 3.2.2. *Expected duration of a busy period,*

$$l_1 = \frac{E(X_2)}{1 - \rho_1}.$$

Proof. Let $H_1(\cdot)$ and $H_2(\cdot)$ be the CDFs of the busy periods generated by a single customer and a batch of N customers, respectively. Then

$$H_2(x) = \int_0^{x} \text{Pr(given first service time} = t\text{, busy period generated by all arrivals occurring during the time} \le x - t)\; dB_2(t)$$

ie.
$$H_2(x) = \int_0^{x} \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad H_1^{*n}(x - t) \quad d\,B_2(t) \tag{3.4}$$

where $H_1^{*n}(x)$ is the n-fold convolution of $H_1(x)$. Let $\overline{H_i}(s)$ and $\overline{B_i}(s)$ be the Laplace-Stieltjes transformation (LSTs) of $H_i(t)$ and $B_i(t)$, respectively.

Taking LSTs of both sides of (3.4), it is found that

$$\overline{H_2}(s) = \overline{B_2}(s + \lambda - \lambda \overline{H_1}(s))$$

Hence the mean length of the busy period generated by a batch of N units $= -\frac{d}{ds}\overline{H}_2(s)\mid_{s=0} = \frac{E(X_2)}{1-\rho_1}$, since $[\frac{d}{ds}\bar{H}_1(s)]_{s=0} = -\frac{E(X_1)}{1-\lambda E(X_1)}$ from section 5.1.7 of [12].

Thus $l_1 = \frac{E(X_2)}{1-\rho_1}$. Hence the lemma. □

Mean length of the busy cycle is given by

$$B = l_1 + \frac{N}{\lambda} = \frac{E(X_2)}{1-\rho_1} + \frac{N}{\lambda} = \frac{\lambda\mu_1 + N\mu_2(\mu_1 - \lambda)}{\lambda\mu_2(\mu_1 - \lambda)}$$

If the costs C_1, C_4 and K, which are stated in the previous chapter are the only costs considered, the unit time cost function F_{Mod-N} assumes the form

$$\begin{aligned} F_{Mod-N} &= C_1 L + K\frac{1}{B} + C_4[\frac{N-1}{\lambda} + \frac{N-2}{\lambda} + \ldots + \frac{1}{\lambda}] \\ &= C_1\Bigg[(L' - \frac{\rho_2}{1-\rho_1+\rho_2})E(X_1) + \frac{\lambda\rho_2}{2(1-\rho_1+\rho_2)}E(X_1^2) \\ &\quad + \frac{\lambda(1-\rho_1)}{2(1-\rho_1+\rho_2)}E(X_2^2)\Bigg] \\ &\quad + K\frac{\lambda\mu_2(\mu_1-\lambda)}{\lambda\mu_1 + N\mu_2(\mu_1-\lambda)} + C_4\frac{N(N-1)}{2\lambda} \end{aligned} \quad (3.5)$$

By treating N as continuous, it can be shown that the second derivative of F_{Mod-N} with respect to N is

$$K\frac{2\lambda\mu_2^3(\mu_1-\lambda)^3}{[\lambda\mu_1 + N\mu_2(\mu_1-\lambda)]^3} + \frac{C_4}{\lambda}$$

which is greater than zero since $\rho_1 = \frac{\lambda}{\mu_1} < 1$. Hence F_{Mod-N} is convex in N.

Equating the first derivative of F_{Mode-N} to zero, the following cubic equation is obtained;

$$2C_4\mu_2^2(\mu_1-\lambda)^2N^3+[4C_4\lambda\mu_1\mu_2(\mu_1-\lambda)-C_4\mu_2^2(\mu_1-\lambda)^2]N^2$$
$$+[2C_4\lambda^2\mu_1^2-2C_4\lambda\mu_1\mu_2(\mu_1-\lambda)]N$$
$$=2K\lambda^2\mu_2^2(\mu_1-\lambda)^2+C_4\lambda^2\mu_1^2$$

By a procedure similar to the one used in the case of (2.5) in chapter 2, here also it can be proved that $\frac{b^2}{4}+\frac{a^3}{27}>0$ and hence this equation has exactly one real root, namely,

$$N^*=[A_1+A_2+\frac{1}{6}[1-\frac{4\lambda\mu_1}{\mu_2(\mu_1-\lambda)}]]^+=[H]^+$$

where

$$A_1=\sqrt[3]{-b/2+\sqrt{\frac{-K\lambda^2}{4C_4}(2b+\frac{K\lambda^2}{C_4})}}$$

$$A_2=\sqrt[3]{-b/2-\sqrt{\frac{-K\lambda^2}{4C_4}(2b+\frac{K\lambda^2}{C_4})}}$$

$$\text{and}\quad b=-\frac{1}{108}[1+\frac{2\lambda\mu_1}{\mu_2(\mu_1-\lambda)}]^3-\frac{K\lambda^2}{C_4}$$

Since N^* must be an integer, substitute the integers close to H in (3.5) and choose the one giving the lower expected cost. Some numerical illustrations are given in the following table. It provides optimal N values corresponding to various input parameter values.

λ	μ_1	μ_2	C_4	K	N^*
5	20	15	20	100	5
10	25	20	25	150	8
12	25	23	30	125	8
15	35	30	40	200	10

λ	μ_1	μ_2	C_4	K	N^*
15	50	40	50	225	10
10	50	45	30	150	8
5	60	30	50	300	5
30	60	50	50	250	16
20	80	55	60	350	13
25	60	40	70	400	15
30	100	50	70	400	17
40	100	80	80	500	21
45	50	40	60	450	17
50	60	40	70	250	16
50	75	70	65	200	18

3.3 Waiting time distribution

Let T be the random variable representing the time spent waiting in the queue by an arbitrary customer and $W(t)$ be its cumulative distribution function. Then

$$W(0) = \Pr\{T = 0\} = \Pr\{\text{the arrival finds the system in state}(N-1,0)\}$$
$$= q_{N-1,0} = q_{00}$$

and $\Pr\{0 < T \leq t\}$

$$= \sum_{(i,j)\neq(N-1,0)} \Pr\{0 < T \leq t | \text{ the arrival finds the system in state } (i,j)\} q_{i,j}$$

$$
\begin{aligned}
&= \sum_{i=0}^{N-1} \int_0^t \frac{e^{-\lambda u}\lambda(\lambda u)^{N-i-2}}{(N-i-2)!} du \, q_{i0} \\
&\quad + \sum_{i=N}^{\infty} \int_0^t \left(b_1^{*(i-N)}(t-y) * \int_0^{\infty} \frac{b_2(u+y)}{1-B_2(u)} du \right) dy \, q_{i2} \\
&\quad + \sum_{i=1}^{\infty} \int_0^t \left(b_1^{*(i-1)}(t-y) * \int_0^{\infty} \frac{b_1(u+y)}{1-B_1(u)} du \right) dy \, q_{i1}
\end{aligned}
$$

where $b_i(t) = \frac{d}{dt}B_i(t)$ for $i = 1, 2$ and $b_i^{*j}(\cdot)$ is the j-fold convolution of $b_i(\cdot)$. Then with itself $E(T) = 0 \cdot q_{00} + \int_0^{\infty} t \frac{d}{dt} \Pr\{0 < T \leq t\}$.

Chapter 4

Random N-policies for the $M|G|1$ queue

In a queueing model under N-policy, the server takes a vacation until a fixed number, N of customers accumulate for service since the completion of the last busy period. But there arise some practical situations where, upon emptying the queue, the server decides on a random number N of customers to accumulate before he is activated. Hence, the number N may vary with different cycles. This policy is called the random N-policy.

In this chapter, an $M|G|1$ queue under two types of random N-policies are considered. In Model 1 an $M|G|1$ queue, under the random N-policy described above, is analyzed. This model was earlier studied by Chatschik Bisdikian [7], who obtained the Z-transform of the queue size and Laplace-Stieltjes transform (LST) of the waiting time of a customer under both FIFO and LIFO service disciplines. For the same model that is being discussed in this chapter, we have computed the average queue size and mean length of a busy period. Also the optimal distribution of N from a given class of distributions is investigated. In Model 2, the operating policy is the modified N-policy considered in the previous chapters with the only difference that N is not deterministic, but random. Here steady state distribution is derived. Some measures of performance of the system are computed and the optimal distribution of N from a given set of distributions is enquired.

4.1 Model 1

Let the arrival process be Poisson of rate λ and service times of customers be independent and identically distributed (iid) random variables following an arbitrary distribution with mean service time $\frac{1}{\mu}$ and variance σ^2. Also it is assumed that $\rho = \frac{\lambda}{\mu} < 1$.

Here, following a busy period, the server remains idle until N customers accumulate in the queue. Once N units are accumulated, service starts one at a time till the system becomes empty. Contrary to the classical N-policy, here N is assumed to be a random variable taking a finite number of values, say $1, 2, \cdots m$, with probabilities $p_1, p_2, \ldots, p_m$, respectively. Let B_k be the expected length of a busy period that starts with k customers and N_k be the expected number of customers served in such a busy period. In [13], it is shown that

$$B_k = \frac{k}{\mu(1-\rho)}, \quad k = 0, 1, \cdots \text{ and } N_k = \frac{k}{1-\rho}, \quad k = 0, 1, \cdots .$$

Hence for this model, mean length of a busy period

$$= \sum_{k=1}^{m} B_k p_k = \frac{1}{\mu(1-\rho)} \sum_{k=1}^{m} k p_k$$

Since the idle time of the server is k/λ, where the initial service starts only on arrival of k units, the mean length of an idle period is $\sum_{k=1}^{m} \frac{k}{\lambda} p_k$.
Therefore mean length B of a busy cycle is

$$B = \frac{1}{\mu(1-\rho)} \sum_{k=1}^{m} k p_k + \frac{1}{\lambda} \sum_{k=1}^{m} k p_k = \frac{\mu}{\lambda(\mu-\lambda)} \sum_{k=1}^{m} k p_k$$

Let W and L be the mean wait in the system and mean number in the system, respectively. To obtain L, W will be found first and then use Little's theorem. Let W_k and L_k be the mean wait in the system and mean number in the system for an $M|G|1$ queue that starts each busy period with $k \geq 1$ customers. In [13], it is shown that $L_k = L_{M|G|1} + \frac{k-1}{2}$ where $L_{M|G|1}$ is given by the Pollaczek-Khinchine equation

$$L_{M|G|1} = \rho + (\rho^2(\sigma^2\mu^2 + 1))/(2(1-\rho)).$$

By Little's theorem, $W_k = \frac{L_k}{\lambda}$. Now W is the average taken over the waiting times of all customers. If this average is formed by combining the waiting times of those customers that were served in a busy period that starts with k customers, noting the contribution to the sum of all the waiting times from this subset is proportional to both the average number of customers served in these busy periods and how often these busy periods occured, it is found that

$$W = \frac{\sum_{k=1}^{m} W_k N_k p_k}{\sum_{k=1}^{m} N_k p_k} = \frac{\sum_{k=1}^{m}(W_{M|G|1} + (k-1)/2\lambda)k/(1-\rho)p_k}{\sum_{k=1}^{m} k/(1-\rho)p_k}.$$

where $W_{M|G|1} = L_{M|G|1}/\lambda$ is the average waiting time of a customer in the system for an ordinary $M|G|1$ queue. ie,

$$W = W_{M|G|1} + \frac{1}{2\lambda}\left(\frac{\sum_{k=1}^{m} k^2 p_k}{\sum_{k=1}^{m} k p_k} - 1\right) \tag{4.1}$$

Applying Little's theorem to (4.1) yields

$$L = L_{M|G|1} + \frac{1}{2}\left(\frac{\sum_{k=1}^{m} k^2 p_k}{\sum_{k=1}^{m} k p_k} - 1\right)$$

Let C_1 be the holding cost per customer per unit time and K be the fixed cost for activating the server. By considering these two costs only, the per unit time average cost of running the system, denoted by F_N, assumes the form

$$F_N = C_1 \times L + K \times \frac{1}{B}$$
$$= C_1\big[L_{M|G|1} + \frac{1}{2}(\frac{\sum_{k=1}^{m} k^2 p_k}{\sum_{k=1}^{m} k p_k} - 1)\big]$$
$$+ K \times \frac{\lambda(\mu - \lambda)}{\mu}(\sum_{k=1}^{m} k p_k)^{-1} \quad (4.2)$$

Now we are investigating the optimal distribution of N (distribution which minimizes F_N) from a given set of distributions. For this purpose the following three cases are being considered.

Case 1: N *uniformly distributed.*

Here $p_k = \frac{1}{m}$ for $1 \leq k \leq m$. Then

$F_N = C_1\big[L_{M|G|1} + \frac{m-1}{3}\big] + \frac{2K\lambda(\mu - \lambda)}{(m+1)\mu}$. Clearly F_N is convex in m.

Hence if m^* is the optimal value for m, m^* satisfies the relations

$F_N(m^*) \leq F_N(m^*+1)$ and $F_N(m^*) \leq F_N(m^*-1)$

The first relation yields:

$$\frac{6K\lambda(\mu - \lambda)}{C_1\mu} \leq (m^*+1)(m^*+2) \quad (4.3)$$

The second relation yields:

$$m^*(m^*+1) \leq \frac{6K\lambda(\mu - \lambda)}{C_1\mu} \quad (4.4)$$

Combining (4.3) and (4.4), we get

$$m^*(m^*+1) \leq \frac{6K\lambda(\mu - \lambda)}{C_1\mu} \leq (m^*+1)(m^*+2) \quad (4.5)$$

Some numerical illustrations are provided below.

m^*	1	2	3	4	5	6	7	8	9	10
$m^*(m^*+1)$	2	6	12	20	30	42	56	72	90	110
$(m^*+1)(m^*+2)$	6	12	20	30	42	56	72	90	110	132

From the above table, for a given value of $\frac{6K\lambda(\mu-\lambda)}{C_1\mu}$, the corresponding value for m^* will be obtained. For example, if

$$\frac{6K\lambda(\mu-\lambda)}{C_1\mu} = 25, m^* = 4$$

and if $\dfrac{6K\lambda(\mu-\lambda)}{C_1\mu} = 56$, m^* has two values, namely 6 and 7.
However, for large m^*, (4.5) can be approximated as

$$(m^*+1)^2 \approx \frac{6K\lambda(\mu-\lambda)}{C_1\mu}$$

so that

$$m^* \approx \sqrt{\frac{6K\lambda(\mu-\lambda)}{C_1\mu}} - 1$$

Case 2: *Distribution of N unimodal and symmetric with respect to a maximum.*

Here we assume that $m = 2n+1$, an odd number. Then

$$p_k = p_{2n-k+2} = \frac{k}{(n+1)^2} \quad \text{for } k = 1, 2, \ldots n+1$$

$$\text{and} \qquad F_N = C_1\left[L_{M|G|1} + \frac{n(7n+8)}{12(n+1)}\right] + \frac{K\lambda(\mu-\lambda)}{(n+1)\mu}$$

Clearly F_N is convex in n. If n^* is the optimal value for n, then

$$F_N(n^*) \leq F_N(n^*+1) \text{ and } F_N(n^*) \leq F_N(n^*-1)$$

These two relations yield

$$7n^{*^2} + 7n^* + 1 \leq \frac{12K\lambda(\mu - \lambda)}{C_1\mu} \leq 7(n^* + 1)^2 + 7(n^* + 1) + 1 \quad (4.6)$$

Some numerical illustrations are shown below.

n^*	1	2	3	4	5	6	7	8	9	10
$7n^{*^2} + 7n^* + 1$	15	43	85	141	211	295	393	505	631	771
$7(n^* + 1)^2 + 7(n^* + 1) + 1$	43	85	141	211	295	393	505	631	771	925

By using the above table, for a given value of $\frac{12K\lambda(\mu-\lambda)}{C_1\mu}$, the corresponding value for n^* can be computed. For example, if $\frac{12K\lambda(\mu-\lambda)}{C_1\mu} = 375$, $n^* = 6$ so that $m^* = 13$.

However, for large n^* (4.6) gives an approximate value for n^*, namely

$$n^* \approx \frac{\sqrt{21 + \frac{336K\lambda(\mu-\lambda)}{C_1\mu}}}{14} - \frac{1}{2}$$

Case 3:*Distribution of N symmetric with respect to a minimum.*

As in case 2, here also we assume that $m = 2n + 1$, an odd number.

Then

$$p_k = p_{2n-k+2} = \frac{n - k + 2}{(n + 1)^2 + n} \text{ for } k = 1, 2, \cdots n + 1.$$

and

$$F_N = C_1 \left(L_{M|G|1} + \frac{n(9n^2 + 25n + 8)}{12(n^2 + 3n + 1)} \right) + \frac{K\lambda(\mu - \lambda)}{(n + 1)\mu}.$$

Obviously F_N is convex in n. If n^* is the optimal value of n, then

$$F_N(n^*) \leq F_N(n^* + 1) \text{ and } F_N(n^*) \leq F_N(n^* - 1)$$

From the above two relations, it is obtained that

$$\frac{n^*(n^*+1)[9n^{*^4}+36n^{*^3}+22n^{*^2}-17n^*-8]}{(n^{*^2}+3n^*+1)((n^*-1)^2+3(n^*-1)+1)} \leq \frac{12K\lambda(\mu-\lambda)}{C_1\mu}$$

$$\leq \frac{(n^*+1)(n^*+2)[9(n^*+1)^4+36(n^*+1)^3+22(n^*+1)^2-17(n^*+1)-8]}{[(n^*+1)^2+3(n^*+1)+1](n^{*^2}+3n^*+1)}$$

$$\text{ie.} \qquad A_{n^*} \leq \frac{12K\lambda(\mu-\lambda)}{C_1\mu} \leq A_{n^*+1} \tag{4.7}$$

Some numerical illustrations are shown in the following table.

n^*	1	2	3	4	5	6	7	8	9	10
A_n^*	10.5	52.14	105.65	177.28	267	374.78	500.6	644.5	806.34	986.23
A_{n^*+1}	52.14	105.65	177.28	267	374.78	500.6	644.5	806.34	986.23	1184.14

From the above table, for a given value of $\frac{12K\lambda(\mu-\lambda)}{C_1\mu}$, the corresponding value for n^* will be obtained.

4.2 Model 2

Here an $M|G|1$ queue under modified N-policy, with random N, is considered. Thus following a busy period, the server remains idle until N units accumulate in the queue and these N units are served together as a batch and the subsequent arrivals are taken into service one by one. Here, unlike in chapter 3, it is assumed that N is not deterministic, but a non-degenerate random variable.

Let the arrivals form a Poisson process of rate λ and service in single have independent and identically distributed random duration following an arbitrary distribution with mean service time $\frac{1}{\mu}$ and variance σ_X^2. Let the random variable N take the values $1, 2, \cdots, m$ with probabilities $p_1, p_2, \cdots p_m$, respectively. Assume that the service time of a batch of k units $(1 \leq k \leq m)$ is also arbitrarily distributed with mean service time $\frac{1}{\mu_k}$ and variance $\sigma_{X_k}^2$ for

$1 \leq k \leq m$ and $\mu_1 > \mu_2 > \mu_3 \cdots > \mu_m$.
Following notations are used in this model.

$X(t)$ – Number of units in the system at time t.

$$Y(t) = \begin{cases} 0 & \text{if the server is idle at } t. \\ 1 & \text{if the forthcoming service is a single service or a single} \\ & \text{service is taking place at t, according as } t \text{ is a departure} \\ & \text{epoch or an arbitrary epoch, respectively.} \\ 2 & \text{if a batch service is taking place at } t. \end{cases}$$

If $Y(t) = 2$, define $Z(t) =$ Number of units that are being served at t.
$\{(X(t), Y(t), Z(t)) | t \geq 0\}$ is a stochastic process with the state space

$$\{(i,0)|0 \leq i \leq m-1\} \cup \{(i,1)|i \geq 1\} \cup \{(i,2,j)|i \geq 1, j \leq i\}$$

In particular, the embedded stochastic process $\{(X(t_i), Y(t_i))\}$ where $t_0 = 0, t_1, t_2, t_3, \cdots$ are the successive times of completions of service is a Markov chain with the state space $S = \{(0,0)\} \cup \{(i,1)|i \geq 1\}$.
Let π_{ij} and $q_{ij}(q_{i2j})$ be the steady state probabilities that the system is in state $(i,j)((i,2,j))$ at a departure epoch and an arbitrary epoch, respectively. Assume that the steady state solution exists. Then the departure point probabilities π_{ij} and general time probabilities $q_{ij}(q_{i2j})$ are related as follows:
For $0 \leq i \leq m-1$,

$$q_{i0} = \pi_{00} \prod_{k=1}^{i} (1 - p_k) \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^i}{i!} \, dt,$$

$$\text{ie,} \quad q_{i0} = \frac{\pi_{00}}{\lambda} \prod_{k=1}^{i} (1 - p_k) \quad \text{for } 0 \leq i \leq m-1.$$

For $i \geq 1$ and $j \leq i$,

$$q_{i2j} = \pi_{00}p_j \prod_{k=1}^{j-1}(1-p_k) \int_0^\infty \int_0^t \frac{e^{-\lambda u}\lambda(\lambda u)^{j-1}}{(j-i)!}[1 - B_j(t-u)]e^{-\lambda(t-u)} \frac{[\lambda(t-u)]^{i-j}}{(i-j)!}\, du\, dt$$

and

$$q_{i1} = \sum_{k=1}^{i} \pi_{k1} \int_0^\infty \frac{e^{-\lambda u}(\lambda u)^{i-k}}{(i-k)!}[1 - B(u)]\, du.$$

where $B_j(\cdot)$ and $B(\cdot)$ are the distribution functions of the service times of a batch of j units and a single unit, respectively. Let Q be the transition probability matrix of the embedded Markov chain $\{(X(t_i), Y(t_i))\}$.
Then

$$Q = \begin{array}{c} \\ (0,0) \\ (1,1) \\ (2,1) \\ (3,1) \\ \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccc} (0,0) & (1,1) & (2,1) & (3,1) & \cdots \end{array} \\ \begin{pmatrix} c_0 & c_1 & c_2 & c_3 & \cdots \\ k_0 & k_1 & k_2 & k_3 & \cdots \\ 0 & k_0 & k_1 & k_2 & \cdots \\ 0 & 0 & k_0 & k_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \end{array}$$

where

$$\begin{aligned} c_n &= Pr\{n \text{ arrivals during a batch service}\} \\ &= \sum_{k=1}^{m} \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!}\, dB_k(t) p_k \end{aligned}$$

and $k_n = Pr\{n \text{ arrivals during a single service}\} = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} dB(t)$

$\Pi = \{\pi_{ij}\}$ can be found as the solution to the stationary equation $\Pi Q = \Pi$.

This yields :

$$\pi_{00} = \pi_{00}c_0 + \pi_{11}k_0 \tag{4.8}$$

$$\pi_{i1} = \pi_{00}c_i + \sum_{j=1}^{i+1} \pi_{j1}k_{i-j+1} \quad \text{for } i \geq 1. \tag{4.9}$$

Define the probability generating functions $\Pi(z) = \pi_{00} + \sum_{i=1}^{\infty} \pi_{i1}z^i$, $C(z) = \sum_{i=0}^{\infty} c_i z^i$ and $K(z) = \sum_{i=0}^{\infty} k_i z^i$ such that all these series converge for $|z| \leq 1$.

Multiplying (4.9) by z^i, taking summation over i from 1 to ∞, adding the resulting equation and (4.8), it can be found that

$$\Pi(z) = \pi_{00}\left[\frac{zC(z) - K(z)}{z - K(z)}\right].$$

Then $\Pi(1) = 1$ gives $\pi_{00} = \frac{1-\rho}{1-\rho+\sum_{k=1}^{m}\rho_k p_k}$ where $\rho = \frac{\lambda}{\mu}$ and $\rho_k = \lambda/\mu_k$

Now the expected system size at a departure epoch,

$$L' = \Pi'(1) = \left[\frac{d}{dz}\Pi(z)\right]_{z=1}$$
$$= \frac{(1-\rho)\sum_{k=1}^{m}(\lambda^2\sigma_{X_k}^2 + \rho_k^2 + 2\rho_k)p_k + (\lambda^2\sigma_X^2 + \rho^2)\sum_{k=1}^{m}\rho_k p_k}{2(1-\rho)(1-\rho+\sum_{k=1}^{m}\rho_k p_k)}$$

Expected queue size at an arbitrary time point when the server is busy is given by

$$L = \sum_{i=1}^{\infty}(i-1)q_{i1} + \sum_{j=1}^{m}\sum_{i=j}^{\infty}(i-j)q_{i2j}$$
$$= \sum_{i=1}^{\infty}\sum_{k=1}^{i}(i-1)\pi_{k1}\int_0^{\infty}\frac{e^{-\lambda u}(\lambda u)^{i-k}}{(i-k)!}(1-B(u))\,du + \pi_{00}\sum_{j=1}^{m}\prod_{k=1}^{j-1}p_j(1-p_k)$$
$$\int_0^{\infty}\int_0^{t} e^{-\lambda u}\frac{\lambda(\lambda u)^{j-1}}{(j-1)!}(\lambda(t-u))[1-B_j(t-u)]\,dudt$$
$$= \frac{L'-1+\pi_{00}}{\mu} + \frac{\lambda(1-\pi_{00})}{2}E(X^2) + \frac{\lambda\pi_{00}}{2}\sum_{j=1}^{m}\prod_{k=1}^{j-1}p_j(1-p_k)E(X_j^2)$$

where $E(X^2)$ and $E(X_j^2)$ are the second raw moments of single service time and batch service times containing j units, respectively. The application of Foster's theorem in a fashion similar to that of section 5.1.4 of [12] shows that the embedded Markov chain is ergodic and hence possesses stationary distribution when $\rho = \lambda/\mu < 1$ provided $\mu_k < \mu$ for $k > 1$.

Since by Lemma 3.2.2 of chapter 3, the expected duration of a busy period that starts with k customers is $\frac{E(X_k)}{1-\rho}$, the expected duration of a busy period for this model is $\sum_{k=1}^{m} \frac{E(X_k)p_k}{1-\rho}$. Thus we have

Lemma 4.2.1. *The expected duration of a busy period is* $\sum_{k=1}^{m} \frac{E(X_k)p_k}{1-\rho} = \sum_{k=1}^{m} = \frac{P_k/\mu_k}{1-\rho}$.

Since the idle time of the server in a model where the initial service starts only with the arrival of the k^{th} unit is k/λ, the expected idle time of the server for this model is $\frac{1}{\lambda}\sum_{k=1}^{m} kp_k$. Hence we have

Lemma 4.2.2. *Mean length of a busy cycle,*

$$B = \sum_{k=1}^{m} \frac{p_k/\mu_k}{1-\rho} + \frac{1}{\lambda}\sum_{k=1}^{m} kp_k.$$

Now, W, the expected total waiting time of all those customers who have reached the system during an idle period

$$= \sum_{k=1}^{m} E\,(\text{Waiting time of all those customers who have reached}$$

during an idle period given that the idle period ends at

the arrival of the k^{th} unit)p_k

$$= \sum_{k=1}^{m}\sum_{j=1}^{k-1} \frac{j}{\lambda} p_k = \sum_{k=2}^{m} \frac{k(k-1)}{2\lambda} p_k$$

If the costs K and C_4, that have been already stated in the previous chapters, are the only costs considered, the per unit time cost function for this model, denoted by F_{Mod-N} assumes the form

$$F_{Mod-N} = K\frac{1}{B} + C_4\, W$$
$$= K\left[\frac{\sum_{k=1}^{m} p_k/\mu_k}{1-\rho} + \frac{1}{\lambda}\sum_{k=1}^{m} kp_k\right]^{-1} + \frac{C_4}{2\lambda}\sum_{k=2}^{m} k(k-1)p_k \quad (4.10)$$

Now we investigate the optimal form of the distribution of N from the same three types of distributions that have been considered in Model 1. Also as a particular case, it is assumed that $\mu_k = \frac{\mu}{\alpha k}$; $0.5 < \alpha < 1$ for $1 \le k \le m$. This means that expected service time of a batch consisting of k units is less than the time required for k single services.

Case 1 : N *uniformly distributed.*

Since N is discrete uniform, $p_k = \frac{1}{m}$ for $1 \le k \le m$.

Then

$$F_{Mod-N} = K\frac{2\lambda(\mu-\lambda)}{(m+1)[\mu+(\alpha-1)\lambda]} + C_4\frac{m^2-1}{6\lambda}$$

Clearly F_{Mod-N} is convex in m. Hence if m^* is the optimal value for m,

$F_{Mod-N}(m^*) \le F_{Mod-N}(m^*+1)$ and $F_{Mod-N}(m^*) \le F_{Mod-N}(m^*-1)$

These two relations yield :

$$m^*(m^*+1)(2m^*-1) \le 12\frac{K\,\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]} \le (m^*+1)(m^*+2)(2m^*+1) \quad (4.11)$$

Some numerical illustrations are provided in the following table.

m^*	1	2	3	4	5	6	7	8	9	10
$m^*(m^*+1)(2m^*-1)$	2	18	60	140	270	462	728	1080	1530	2090
$(m^*+1)(m^*+2)(2m^*+1)$	18	60	140	270	462	728	1080	1530	2090	2772

From the above table, we see that for given values of C_4, K, α, λ and μ the corresponding value for m^* can be obtained. For example, if

$$12\frac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}=175, \qquad m^*=4.$$

However, for large value of m^*, from (4.11) we get

$$m^*(m^*+1)(2m^*-1)\approx 12\frac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}.$$

so that $m^* \approx A_1 + A_2 - \frac{1}{6}$

where $A_1 = \sqrt[3]{-b/2+\sqrt{b^2/4+a^3/27}}$,

$A_2 = \sqrt[3]{-b/2-\sqrt{b^2/4+a^3/27}}$, $a = -7/12$ and

$b = \frac{1}{27}(\frac{5}{2} - \frac{162K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]})$

Case 2 : *Distribution of N unimodal and symmetric with respect to a maximum.*

Under this assumption let $m = 2n+1$, odd. Then $p_k = p_{2n-k+2} = \frac{k}{(n+1)^2}$ for $k = 1, 2, \ldots, n+1$ and

$$F_{Mod-N} = K\frac{\lambda(\mu-\lambda)}{(n+1)[\mu+(\alpha-1)\lambda]}+\frac{C_4 n(7n+8)}{12\lambda}$$

F_{Mod-N} is convex in n. If n^* is the value of n that minimises F_{Mod-N},

$F_{Mod-N}(n^*) \leq F_{Mod-N}(n^*+1)$ and $F_{Mod-N}(n^*) \leq F_{Mod-N}(n^*-1)$

These two relations yield :

$$n^*(n^*+1)(14n^*+1) \leq 12\frac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]} \leq (n^*+1)(n^*+2)(14n^*+15) \qquad (4.12)$$

Some numerical illustrations are shown below.

n^*	1	2	3	4	5	6	7	8	9	10
$n^*(n^*+1)(14n^*+1)$	30	174	516	1140	2130	3570	5544	8136	11430	15510
$(n^*+1)(n^*+2)(14n^*+15)$	174	516	1140	2130	3570	5544	8136	11430	15510	20460

From the above table, for a given value of $12\dfrac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}$, ie, given values of C_4, K, α, λ and μ, the corresponding value for n^* will be obtained.

For example, if

$12\dfrac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]} = 4000$, $n^* = 6$ so that $m^* = 2n^* + 1 = 13$.

However, for large value of n^*, (4.12) gives

$$n^*(n^*+1)(14n^*+1) \approx 12\frac{K\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}$$

so that $n^* \approx A_1 + A_2 - 5/14$

where $A_1 = \sqrt[3]{-b/2+\sqrt{b^2/4+a^3/27}}$,

$A_2 = \sqrt[3]{-b/2-\sqrt{b^2/4+a^3/27}}$, $a = -61/196$ and

$b = \dfrac{45}{686} - \dfrac{6K\lambda^2(\mu-\lambda)}{7C_4[\mu+(\alpha-1)\lambda]}$

Case 3 : *Distribution of N symmetric with respect to a minimum.*

Just like in case 2, assume that $m = 2n+1$, odd.

Then $p_k = p_{2n-k+2} = \frac{n-k+2}{(n+1)^2+n}$ for $k = 1, 2, \ldots, n+1$ and

$$F_{Mod-N} = K\frac{\lambda(\mu-\lambda)}{(n+1)[\mu+(\alpha-1)\lambda]} + \frac{C_4 n(n+1)(9n^2+25n+8)}{12\lambda((n+1)^2+n)}$$

Here again F_{Mod-N} is convex in n. If n^* is the optimal value for n, then we have

$F_{Mod-N}(n^*) \leq F_{Mod-N}(n^*+1)$ and $F_{Mod-N}(n^*) \leq F_{Mod-N}(n^*-1)$

These two relations yield :

$$\frac{n^{*^2}(n^*+1)(9n^{*^4}+35n^{*^3}+23n^{*^2}-17n^*-8)}{(n^{*2}+n^*-1)((n^*+1)^2+n^*)} \leq \frac{6R\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}$$
$$\leq \frac{(n^*+1)^2(n^*+2)(9n^{*^4}+71n^{*^3}+182n^{*^2}+170n^*+42)}{((n^*+1)^2+n^*)((n^*+2)^2+n^*+1)}$$

$$\text{ie.} \qquad A_{n^*} \leq \frac{6R\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]} \leq A_{n^*+1} \tag{4.13}$$

Some numerical illustrations are shown below.

n^*	1	2	3	4	5	6	7	8	9	10
A_{n^*}	16.8	103.42	313.84	482.72	1322.37	2228.55	3474.7	5114.8	7202.9	9793
A_{n^*+1}	103.42	313.84	482.72	1322.37	2228.55	3474.7	5114.8	7202.9	9793	12939

From the above table, for a given value of $\frac{6R\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]}$, the corresponding value for n^* will be obtained. For example, if $\frac{6R\lambda^2(\mu-\lambda)}{C_4[\mu+(\alpha-1)\lambda]} = 2000$, then $n^* = 5$ and so $m^* = 2n^* + 1 = 11$.

Chapter 5

Transient analysis of the $M|M|1$ queue under N-policy

The transient (time dependent) analysis of queueing models are quite complicated procedures. The derivation of the transient behaviour of even the $M|M|1$ model, which is the simplest among all queueing models, becomes quite messy. The transient solution of the $M|M|1$ model postdated that of the basic Erlang work by nearly half a century, with the first published solution due to Lederman and Reuter [19], in which they used spectral analysis for the general birth-death process. In the same year, another paper by Bailey [3], on the solution of the problem, appeared. This was followed by the one by Champernowne [6]. Bailey's approach to the time-dependent problem was via generating functions for the partial differential equation and Champernowne's was via advanced combinatorial methods. It is Bailey's approach that has been the most popular over the years and the same approach is being followed in this chapter also. Recently, Parthasarathy [29] has provided a simple approach to the transient analysis. Syski [35] also has made some contribution to this. Parthasarathy and Sharafali [30] have studied transient behaviour of $M|M|C$ queue. A different approach to transient analysis was developed by Sharma and Shobha [33].

Steady state analysis of both $M|M|1$ and $M|G|1$ models, under N-policy have been carried out by several researchers. In this chapter, the transient behaviour of an $M|M|1$ queue, under N-policy is studied. Here the system state probabilities in finite time are obtained in terms of Bessel function. Also, for the same model, the output distribution (distribution of time between successive departures) in the steady state is derived.

Let the arrivals follow Poisson process of rate λ and service times be exponentially distributed with rate μ.

5.1 Transient analysis

Let $X(t)$ be the number of units in the system at time t and

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

Then $\{(X(t), Y(t)) : t \geq 0\}$ is a continuous time Markov process with the state space $S = \{(0,0)(1,0)\cdots,(N-1,0)(1,1)(2,1)\cdots\}$. Let $P_{ij}(t)$ be the probability that the system is in state (i,j) at time t. Then the differential-difference equations satisfied by $P_{ij}(t)$ are:

$$P'_{00}(t) = -\lambda P_{00}(t) + \mu P_{11}(t) \tag{5.1}$$

$$P'_{n0}(t) = -\lambda P_{n0}(t) + \lambda P_{n-1,0}(t) \quad \text{for } 1 \leq n \leq N-1 \tag{5.2}$$

$$P'_{11}(t) = -(\lambda+\mu)P_{11}(t) + \mu P_{21}(t) \tag{5.3}$$

$$\begin{aligned} P'_{N1}(t) = -(\lambda+\mu)P_{N1}(t) + \lambda P_{N-1,1}(t) + \mu P_{N+1,1}(t) \\ + \lambda P_{N-1,0}(t) \end{aligned} \tag{5.4}$$

$$P'_{n1}(t) = -(\lambda+\mu)P_{n1}(t) + \lambda P_{n-1,1}(t) + \mu P_{n+1,1}(t) \quad \text{for } 2 \leq n \leq N-1,\ n \geq N+1 \tag{5.5}$$

Let it be assumed that the initial system size at time 0 is i where $i < N$. ie, the initial service will start only after the accumulation of $N - i$ more units.

Then $P_{i0}(0) = 1$ and $P_{jk}(0) = 0$ for all other j and k.

Define $G_1(z,t) = \sum_{n=1}^{N-1} P_{n0}(t)z^n$ and $G_2(z,t) = \sum_{n=1}^{\infty} P_{n1}(t)z^n$ such that the infinite series is convergent for $|z| \leq 1$. When (5.2) is multiplied throughout by z^n and then summed with respect to n from 1 to $N-1$, we get

$$\frac{\partial}{\partial t}G_1(z,t) = -\lambda G_1(z,t) + \lambda z P_{00}(t) + \lambda z G_1(z,t) - \lambda z^N P_{N-1,0}(t)$$

Let $\bar{G}_i(z,s)$ and $\bar{P}_{ij}(s)$ be the Laplace transforms of $G_i(z,t)$ and $P_{ij}(t)$ respectively. Then by taking Laplace transform of both sides of the above equation and rearranging, it is found that

$$\bar{G}_1(z,s) = \frac{z}{s+\lambda-\lambda z}[z^{i-1} + \lambda\bar{P}_{00}(s) - \lambda z^{N-1}\bar{P}_{N-1,0}(s)] \tag{5.6}$$

Since $\bar{G}_1(z,s) = \sum_{n=1}^{N-1} \bar{P}_{n0}(s)z^n$, from (5.6) we get

$$\bar{P}_{n0}(s) = \begin{cases} (\frac{\lambda}{s+\lambda})^n \bar{P}_{00}(s) & \text{for } n < i \\ \frac{1}{\lambda}(\frac{\lambda}{s+\lambda})^{n-i+1} + (\frac{\lambda}{s+\lambda})^n \bar{P}_{00}(s) & \text{for } i \leq n \leq N-1 \end{cases} \tag{5.7}$$

Multiplying (5.3) by z, (5.4) by z^N and each of the equations in (5.5) by z^n, adding all these equations and adding the resulting equation to (5.1), we get

$$\frac{\partial}{\partial t}G_2(z,t) + P'_{00}(t) = (\lambda z - \lambda - \mu + \frac{\mu}{z})G_2(z,t) - \lambda P_{00}(t) + \lambda P_{N-1,0}(t)z^N$$

Taking Laplace transform of both sides and substituting the value of $\bar{P}_{N-1,0}(s)$ obtained from (5.7), it is found that

$$\bar{G}_2(z,s) = z\frac{\{z^N(\frac{\lambda}{s+\lambda})^{N-i} + [\lambda z^N(\frac{\lambda}{s+\lambda})^{N-1} - (s+\lambda)]\bar{P}_{00}(s)\}}{-\lambda z^2 + (s+\lambda+\mu)z - \mu} \tag{5.8}$$

Since the Laplace transform $\bar{G}_2(z,s)$ converges in the region $|z| \leq 1$, $Re(s) > 0$, wherever the denominator has zeroes in that region, so must the numerator. The zeroes of the denominator are

$$z_1 = \frac{(s+\lambda+\mu) - \sqrt{(s+\lambda+\mu)^2 - 4\lambda\mu}}{2\lambda} \text{ and}$$
$$z_2 = \frac{(s+\lambda+\mu) + \sqrt{(s+\lambda+\mu)^2 - 4\lambda\mu}}{2\lambda}$$

Using Rouche's theorem, we can easily prove that z_1 is the only zero of the denominator in $|z| \leq 1$. Hence z_1 is a zero of the numerator also.

Thus we get

$$z_1^N(\frac{\lambda}{s+\lambda})^{N-i} + \left[\lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1} - (s+\lambda)\right]\bar{P}_{00}(s) = 0,$$

since $z_1 \neq 0$. From this we get

$$\bar{P}_{00}(s) = \frac{z_1^N(\frac{\lambda}{s+\lambda})^{N-i}}{(s+\lambda) - \lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}} \tag{5.9}$$

Substituting the value of $\bar{P}_{00}(s)$ in (5.8), it is found that

$$\bar{G}_2(z,s) = \frac{z(\frac{\lambda}{s+\lambda})^{N-i}}{-\lambda(z-z_1)(z-z_2)}\left\{z^N + \frac{z_1^N[\lambda(\frac{\lambda}{s+\lambda})^{N-1}z^N - s - \lambda]}{(s+\lambda) - \lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}}\right\}$$

Since $\bar{G}_2(z,s) = \sum_{n=1}^{\infty} \bar{P}_{n1}(s)z^n$, from the above expression for $\bar{G}_2(z,s)$ we get

$$\bar{P}_{n1}(s) = (\frac{\lambda}{s+\lambda})^{N-i}\frac{1}{\lambda z_2}z_1^{N-n}\frac{1}{(1-\frac{z_1}{z_2})}$$
$$+\frac{z_1^N(\frac{\lambda}{s+\lambda})^{N-i}}{\lambda z_2[(s+\lambda) - \lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}]}\frac{1}{(1-\frac{z_1}{z_2})}\left\{\lambda(\frac{\lambda}{s+\lambda})^{N-1}z_1^{N-n} - (s+\lambda)\frac{1}{z_2^n}\right\}$$
$$\text{for } 1 \leq n \leq N, \tag{5.10}$$

and

$$\bar{P}_{n1}(s) = (\frac{\lambda}{s+\lambda})^{N-i}\frac{1}{\lambda z_2}(\frac{1}{z_2})^{n-N}\frac{1}{(1-\frac{z_1}{z_2})}$$
$$+\frac{z_1^N(\frac{\lambda}{s+\lambda})^{N-i}}{\lambda z_2[(s+\lambda)-\lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}]}\frac{1}{(1-\frac{z_1}{z_2})}$$
$$\left\{\lambda(\frac{\lambda}{s+\lambda})^{N-1}\frac{1}{z_2^{n-N}}-(s+\lambda)\frac{1}{z_2^n}\right\} \quad (5.11)$$

for $n \geq N$, since $|\frac{z_1}{z_2}| < 1$.

Now we find the inverse Laplace transform of each of these terms. Consider

$$\begin{aligned}\frac{1}{\lambda z_2}z_1^{N-n}(1-\frac{z_1}{z_2})^{-1} &= \frac{1}{\lambda z_2}z_1^{N-n}(1+\frac{z_1}{z_2}+(\frac{z_1}{z_2})^2+\cdots)\\ &= \frac{1}{\lambda}(\frac{z_1}{z_2})^{N-n}\left[z_2^{N-n-1}+z_1z_2^{N-n-2}+z_1^2z_2^{N-n-3}+\cdots\right]\\ &= \frac{1}{\lambda}\frac{(z_1z_2)^{N-n}}{z_2^{2(N-n)}}\left[z_2^{N-n-1}+z_1z_2^{N-n-2}+z_1^2z_2^{N-n-3}+\cdots\right]\\ &= \frac{1}{\lambda}(\frac{\mu}{\lambda})^{N-n}\left[\frac{1}{z_2^{N-n+1}}+\frac{z_1}{z_2^{N-n+2}}+\frac{z_1^2}{z_2^{N-n+3}}+\cdots\right]\end{aligned}$$

since $z_1z_2 = \frac{\mu}{\lambda}$.

Now we use the fact that

$$L^{-1}(\frac{1}{z_2^m}) = e^{-(\lambda+\mu)t}m(\frac{\lambda}{\mu})^{m/2}t^{-1}I_m(2\sqrt{\lambda\mu}t)$$

where $I_m(y) = \sum_{k=0}^{\infty}\frac{(y/2)^{m+2k}}{k!(m+k)!}$, $(m > -1)$, is the modified Bessel function of the first kind. Using this result in the above relation, we get

$$L^{-1}\{\frac{1}{\lambda z_2}z_1^{N-n}(1-\frac{z_1}{z_2})^{-1}\}$$
$$= \frac{e^{-(\lambda+\mu)t}}{\lambda t}(\frac{\mu}{\lambda})^{\frac{N-n-1}{2}}\sum_{j=0}^{\infty}(N-n+2j+1)I_{N-n+2j+1}(2\sqrt{\lambda\mu}t)$$

and $L^{-1}\left((\frac{\lambda}{s+\lambda})^{N-i}\frac{z_1^{N-n}}{\lambda z_2}(1-\frac{z_1}{z_2})^{-1}\right)$

$$= \int_{u=0}^{t} e^{-\lambda(t-u)}\frac{[\lambda(t-u)]^{N-i-1}}{(N-i-1)!}\frac{e^{-(\lambda+\mu)u}}{u}(\frac{\mu}{\lambda})^{\frac{N-n-1}{2}} \sum_{j=0}^{\infty}(N-n+2j+1)I_{N-n+2j+1}(2\sqrt{\lambda\mu u})du$$

$$= (\frac{\mu}{\lambda})^{\frac{N-n-1}{2}} e^{-\lambda t}\sum_{j=0}^{\infty}\int_{u=0}^{t}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{N-i-1}}{(N-i-1)!} (N-n+2j+1)I_{N-n+2j+1}(2\sqrt{\lambda\mu u})du \quad (5.12)$$

In the above we used the fact that

$$L^{-1}(\bar{f}(s)\bar{g}(s)) = \int_{u=0}^{t} f(u)g(t-u)du.$$

Now consider

$$\frac{(\frac{\lambda}{s+\lambda})^{2N-i-1}z_1^{2N-n}}{z_2[(s+\lambda)-\lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}]}(1-\frac{z_1}{z_2})^{-1}$$
$$= (\frac{\lambda}{s+\lambda})^{2N-i-1}\frac{z_1^{2N-n}}{(s+\lambda)z_2}(1-(\frac{\lambda z_1}{s+\lambda})^N)^{-1}(1-\frac{z_1}{z_2})^{-1}$$
$$= \frac{1}{\lambda}(\frac{\lambda}{s+\lambda})^{2N-i}\frac{z_1^{2N-n}}{z_2}\left\{\sum_{j=0}^{N-1}(\frac{z_1}{z_2})^j + \sum_{k=1}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}(\frac{\lambda}{s+\lambda})^{lN}\frac{z_1^{kN+j}}{z_2^{(k-l)N+j}}\right\}$$

since $|\lambda z_1| < \lambda < |s+\lambda|$ because $Re(s) > 0$ and $|z_1| < 1$.

$$= \frac{1}{\lambda}\left\{\sum_{j=0}^{N-1}(\frac{\lambda}{s+\lambda})^{2N-i}\frac{z_1^{2N-n+j}}{z_2^j}\right.$$

$$\left.+\sum_{k=1}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}(\frac{\lambda}{s+\lambda})^{(l+2)N-i}\frac{z_1^{(k+2)N-n+j}}{z_2^{(k-l)N+j+1}}\right\}$$

$$= \frac{1}{\lambda}\sum_{j=0}^{N-1}(\frac{\lambda}{s+\lambda})^{2N-i}\frac{(\mu/\lambda)^{2N-n+j}}{z_2^{2N-n+2j+1}} + \sum_{k=1}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}(\frac{\lambda}{s+\lambda})^{(l+2)N-i}$$

$$\frac{(\mu/\lambda)^{(k+2)N-n+j}}{z_2^{(2k-l+2)N-n+2j+1}}$$

$$L^{-1}\left\{\frac{(\frac{\lambda}{s+\lambda})^{2N-i-1}z_1^{2N-n}}{z_2[(s+\lambda)-\lambda z_1^N(\frac{\lambda}{s+\lambda})^{N-1}]}(1-\frac{z_1}{z_2})^{-1}\right\}$$

$$= \sum_{j=0}^{N-1}\int_{u=0}^{t}\frac{e^{-\lambda(t-u)}[\lambda(t-u)]^{2N-i-1}}{(2N-i-1)!}(\frac{\mu}{\lambda})^{\frac{2N-n-1}{2}}\frac{e^{-(\lambda+\mu)u}}{u}$$

$$(2N-n+2j+1)I_{2N-n+2j+1}(2\sqrt{\lambda\mu}u)du$$

$$+\sum_{k=1}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}\int_{u=0}^{t}\frac{e^{-\lambda(t-u)}[\lambda(t-u)]^{(l+2)N-i-1}}{((l+2)N-i-1)!}(\frac{\mu}{\lambda})^{\frac{(l+2)N-n-1}{2}}\frac{e^{-(\lambda+\mu)u}}{u}$$

$$[(2k-l+2)N-n+2j+1]I_{(2k-l+2)N-n+2j+1}(2\sqrt{\lambda\mu}u)du$$

$$= (\frac{\mu}{\lambda})^{\frac{2N-n-1}{2}}e^{-\lambda t}\sum_{k=0}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}\int_{u=0}^{t}(\frac{\mu}{\lambda})^{\frac{lN}{2}}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{(l+2)N-i-1}}{((l+2)N-i-1)!}$$

$$[(2k-l+2)N-n+2j+1]I_{(2k-l+2)N-n+2j+1}(2\sqrt{\lambda\mu}u)du \quad (5.13)$$

Similarly we get

$$L^{-1}\left\{(\frac{\lambda}{s+\lambda})^{N-i}\frac{z_1^N}{\lambda z_2^{n+1}}(1-(\frac{\lambda z_1}{s+\lambda})^N)^{-1}(1-\frac{z_1}{z_2})^{-1}\right\}$$

$$= (\frac{\mu}{\lambda})^{\frac{N-n-1}{2}}e^{-\lambda t}\sum_{k=0}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}\int_{u=0}^{t}(\frac{\mu}{\lambda})^{lN/2}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{(l+1)N-i-1}}{((l+1)N-i-1)!}$$
$$[(2k-l+1)N+2j+n+1]I_{(2k-l+1)N+2j+n+1}(2\sqrt{\lambda\mu})du. \quad (5.14)$$

Inverting (5.10) and using the results (5.12),(5.13) and (5.14), we get

$$P_{n1}(t) = (\frac{\mu}{\lambda})^{\frac{N-n-1}{2}}e^{-\lambda t}\sum_{j=0}^{\infty}\int_{u=0}^{t}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{N-i-1}}{(N-i-1)!}$$
$$(N-n+2j+1)I_{N-n+2j+1}(2\sqrt{\lambda\mu}u)du$$
$$+(\frac{\mu}{\lambda})^{\frac{2N-n-1}{2}}e^{-\lambda t}\sum_{k=0}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}\int_{u=0}^{t}(\frac{\mu}{\lambda})^{lN/2}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{(l+2)N-i-1}}{((l+2)N-i-1)!}$$
$$((2k-l+2)N-n+2j+1)I_{(2k-l+2)N-n+2j+1}(2\sqrt{\lambda\mu}u)du$$
$$-(\frac{\mu}{\lambda})^{\frac{N-n-1}{2}}e^{-\lambda t}\sum_{k=0}^{\infty}\sum_{j=0}^{N-1}\sum_{l=0}^{k}\int_{u=0}^{t}(\frac{\mu}{\lambda})^{lN/2}\frac{e^{-\mu u}}{u}\frac{[\lambda(t-u)]^{(l+1)N-i-1}}{((l+1)N-i-1)!}$$
$$[(2k-l+1)N+2j+n+1]I_{(2k-l+1)N+2j+n+1}(2\sqrt{\lambda\mu}u)du$$
$$\text{for } i \le n \le N \quad (5.15)$$

In a similar manner, by inverting (5.11) we get

$$P_{n1}(t) = (\frac{\lambda}{\mu})^{\frac{n-N+1}{2}} e^{-\lambda t} \sum_{j=0}^{\infty} \int_{u=0}^{t} \frac{e^{-\mu u}}{u} \frac{[\lambda(t-u)]^{N-i-1}}{(N-i-1)!} (n-N+2j+1)$$

$$I_{n-N+2j+1}(2\sqrt{\lambda\mu u})du$$

$$+ (\frac{\lambda}{\mu})^{\frac{n-2N+1}{2}} e^{-\lambda t} \sum_{k=0}^{\infty} \sum_{j=0}^{N-1} \sum_{l=0}^{k} \int_{u=0}^{t} \frac{e^{-\mu u}}{u} \frac{[\lambda(t-u)]^{(l+2)N-i-1}}{((l+2)N-i-1)!}$$

$$(\frac{\mu}{\lambda})^{lN/2}((2k-l)N+2j+n+1)I_{(2k-l)N+2j+n+1}(2\sqrt{\lambda\mu u})du$$

$$- (\frac{\lambda}{\mu})^{\frac{n-N+1}{2}} e^{-\lambda t} \sum_{k=0}^{\infty} \sum_{j=0}^{N-1} \sum_{l=0}^{k} \int_{u=0}^{t} \frac{e^{-\mu u}}{u} \frac{(\lambda(t-u))^{(l+1)N-i-1}}{((l+1)N-i-1)!} (\frac{\mu}{\lambda})^{lN/2}$$

$$((2k-l+1)N+2j+n+1)I_{(2k-l+1)N+2j+n+1}(2\sqrt{\lambda\mu u})du \text{ for } n \geq N. \tag{5.16}$$

From (5.9) we have

$$\bar{P}_{00}(s) = \frac{\lambda^{N-i} z_1^N}{(s+\lambda)^{N-i+1}} \sum_{j=0}^{\infty} (\frac{\lambda z_1}{s+\lambda})^{jN} = \sum_{j=0}^{\infty} \frac{\lambda^{(j+1)N-i}}{(s+\lambda)^{(j+1)N-i+1}} z_1^{(j+1)N}$$

$$= \sum_{j=0}^{\infty} \frac{\lambda^{(j+1)N-i}}{(s+\lambda)^{(j+1)N-i+1}} \frac{(\mu/\lambda)^{(j+1)N}}{z_2^{(j+1)N}}$$

Taking inverse Laplace transforms of both sides we get

$$P_{00}(t) = e^{-\lambda t} \sum_{j=0}^{\infty} \int_{u=0}^{t} (\frac{\mu}{\lambda})^{\frac{(j+1)N}{2}} \frac{e^{-\mu u}}{u} (j+1)NI_{(j+1)N}(2\sqrt{\lambda\mu u})du \tag{5.17}$$

Inverting both equations in (5.7) we get

$$P_{n0}(t) = \begin{cases} \int_0^t e^{-\lambda(t-u)} \frac{\lambda(\lambda(t-u))^{n-1}}{(n-1)!} P_{00}(u)du & \text{for } 0 < n < i \\ \frac{e^{-\lambda t}(\lambda t)^{n-i}}{(n-i)!} + \int_0^t e^{-\lambda(t-u)} \frac{\lambda(\lambda(t-u))^{n-1}}{(n-1)!} P_{00}(u)du & \\ \quad \text{for } i \le n \le N-1. & \end{cases} \tag{5.18}$$

Thus the equations (5.15), (5.16), (5.17) and (5.18) determine all time dependent probabilities.

5.2 Output distribution

The steady state probabilities q_{ij} are obtained as $q_{i0} = q_{00}$ for $0 \le i \le N-1$,

$$q_{i1} = \begin{cases} \frac{\rho}{1-\rho}(1-\rho^i)q_{00} & \text{for } 1 \le i \le N-1 \\ \frac{1-\rho^N}{1-\rho}\rho^{i-N+1}q_{00} & \text{for } i \ge N. \end{cases}$$

and $q_{00} = \frac{1}{N}(1-\rho)$ where $\rho = \frac{\lambda}{\mu}$.

Let T represent "time between successive departures" and $\pi_{ij}(t) = \Pr\{$ the system is in state (i,j) at a time t since the last departure and $T > t\}$. Then $\pi_{i0}(t) = q_{00}\dfrac{e^{-\lambda t}(\lambda t)^i}{i!}$ for $0 \le i \le N-1$ and

$$\pi_{i1}(t) = \begin{cases} \sum_{j=1}^{i} q_{j1} \frac{e^{-\lambda t}(\lambda t)^{i-j}}{(i-j)!} e^{-\mu t} & \text{for } 1 \le i \le N-1 \\ q_{00} \int_{u=0}^{t} \frac{e^{-\lambda u}\lambda(\lambda u)^{N-1}}{(N-1)!} e^{-\mu(t-u)} e^{-\lambda(t-u)} \frac{[\lambda(t-u)]^{i-N}}{(i-N)!} du & \\ \quad + \sum_{j=1}^{i} q_{j1} \frac{e^{-\lambda t}(\lambda t)^{i-j}}{(i-j)!} e^{-\mu t} & \text{for } i \ge N \end{cases}$$

Now
$$Pr(T > t) = \sum_{i=0}^{N-1} \pi_{i0}(t) + \sum_{i=1}^{\infty} \pi_{i1}(t)$$
$$= q_{00} \left\{ e^{-\lambda t} \sum_{i=0}^{N-1} \frac{(\lambda t)^i}{i!} + \frac{e^{-(\lambda+\mu)t} \rho}{1-\rho} \sum_{i=1}^{N-1} \sum_{j=1}^{i} (1-\rho^j) \frac{(\lambda t)^{i-j}}{(i-j)!} \right.$$
$$+ e^{-(\lambda+\mu)t} \sum_{i=N}^{\infty} \int_0^t \frac{\lambda^i e^{\mu u} u^{N-1}}{(N-1)!} \frac{(t-u)^{i-N}}{(i-N)!} du$$
$$+ \frac{e^{-(\lambda+\mu)t} \rho}{1-\rho} \sum_{i=N}^{\infty} \sum_{j=1}^{N-1} (1-\rho^j) \frac{(\lambda t)^{i-j}}{(i-j)!}$$
$$\left. + e^{-(\lambda+\mu)t} \frac{1-\rho^N}{1-\rho} \sum_{i=N}^{\infty} \sum_{j=N}^{i} \rho^{j-N+1} \frac{(\lambda t)^{i-j}}{(i-j)!} \right\}$$

ie.
$$Pr(T > t) = q_{00} \left\{ e^{-\lambda t} \sum_{i=0}^{N-1} \frac{(\lambda t)^i}{i!} + \frac{e^{-(\lambda+\mu)t} \rho}{1-\rho} \sum_{i=1}^{N-1} \sum_{j=1}^{i} (1-\rho^j) \frac{(\lambda t)^{i-j}}{(i-j)!} \right.$$
$$+ \left(\frac{\rho}{1-\rho}\right)^N e^{-\mu t} \left[1 - e^{-(\lambda-\mu)t} \sum_{i=0}^{N-1} \frac{((\lambda-\mu)t)^i}{i!}\right]$$
$$\left. + \frac{\rho(1-\rho^N)}{(1-\rho)^2} e^{-\mu t} + \frac{e^{-(\lambda+\mu)t} \rho}{1-\rho} \sum_{i=N}^{\infty} \sum_{j=1}^{N-1} (1-\rho^j) \frac{(\lambda t)^{i-j}}{(i-j)!} \right\} \quad (5.19)$$

and $Pr(T \leq t) = 1 - Pr(T > t)$.

Chapter 6

Queues with two servers in series – A specialised server and a regular server

Queueing models consisting of two units in series/tandem with an intermediate waiting room of finite capacity were studied by several researchers. A model with a finite waiting room in between the two service stations was discussed by Neuts[26]. Service station I of this model contains one server with general service time distribution and second service station consists of 'c' parallel exponential servers.

The study of blocking in two or more service stations in series with general service time distribution, without intermediate buffer was considered by Avi-ltzhak and Yadin[2]. Clarke[8] investigated a tandem queueing model wherein two servers are placed in series and each customer will receive service from one and only one server. The novel feature of this model is that a busy service unit prevents the access of new customers to servers further down the line. A departing customer may also be temporarily prevented from leaving by occupied service units down line. Kandasamy[15] analysed tandem queue with general service rule and server's vacation. Prabhu[31] studied transient analysis in a tandem queue. Models of related type with finite total number of customers were treated by Sharma[34].

In this chapter, we deal with a queueing model consisting of two servers connected in series with a finite intermediate waiting room of capacity k.

Here we assume that server I is a specialised server. He will activate only on accumulation of N units in front of him once he becomes idle due to absence of customers at that counter. At the arrival of the N^{th} unit, he starts service till none is left before him; ie, the service rule of server I is governed by N-policy. After being served by server I, the customer goes to server II and joins service directly if server II is idle at that time. Otherwise he waits in the queue till his turn for service occurs. When the number of customers in the intermediate waiting room becomes k, the service given by server I will be blocked and it will restart only after one departure from server II. Arrivals to server I occur according to a Poisson process of rate λ and service provided by both servers have exponentially distributed duration with rates μ_1 and μ_2 respectively. For this model, the steady state probability vector and the stability condition are obtained using matrix- geometric method.

6.1 Steady state analysis

Let $X(t)$ and $Y(t)$ be the number of customers queued up at time t in front of server I and server II, respectively. Define

$$Z(t) = \begin{cases} 0 & \text{if server I is idle at } t \\ 1 & \text{if server I is busy at } t \\ 2 & \text{if server I is available, but service is blocked.} \end{cases}$$

Then $\{(X(t), Y(t), Z(t)) \mid t \geq 0\}$ is a continuous time Markov chain with state space, $S = \{(i, j, 0) | 0 \leq i \leq N - 1, 0 \leq j \leq k\}$

$\cup \{(i, j, 1) | i \geq 1, 0 \leq j \leq k - 1\} \cup \{(i, k, 2) | i \geq 1\}$

To facilitate the representation of the infinitesimal generator Q of the continuous time Markov chain with the above state space, we define first the sub-matrices $A_0, A_1, A_2, B_i (1 \leq i \leq N - 1), C_i (0 \leq i \leq 8)$ and

$D_i (0 \leq i \leq 7)$.

The matrices A_0, A_1 and A_2 are square matrices of order $k+1$, defined by $A_0 = \lambda I$ where I is the identity matrix of order $k+1$,

$$A_1 = \begin{bmatrix} -(\lambda+\mu_1) & 0 & 0 & \cdots & 0 & 0 \\ \mu_2 & -(\lambda+\mu_1+\mu_2) & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & -(\lambda+\mu_1+\mu_2) & 0 \\ 0 & 0 & 0 & \cdots & \mu_2 & -(\lambda+\mu_2) \end{bmatrix}$$

and

$$A_2 = \begin{bmatrix} 0 & \mu_1 & 0 & \cdots & 0 \\ 0 & 0 & \mu_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \mu_1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

The matrices B_i $(1 \leq i \leq N-1)$ are of order $(2k+1) \times (N+k)$ such that in B_i, $(1, i+1)^{th}$ element is μ_2 and rest of the elements are zeroes.
The matrices C_i $(0 \leq i \leq 8)$ are defined by

$$C_0 = \begin{bmatrix} D_0 & 0 \\ 0 & D_1 \end{bmatrix}_{(N+k)\times(N+k)}, C_1 = \begin{bmatrix} D_2 & 0 \\ 0 & D_3 \end{bmatrix}_{(2k+1)\times(N+k)},$$

$$C_2 = \begin{bmatrix} 0 & 0 \\ \lambda I & 0 \end{bmatrix}_{(N+k)\times(2k+1)}$$, where I is the identity matrix of order k,

$$C_3 = \begin{bmatrix} D_4 \\ 0 \end{bmatrix}_{(N+k)\times(k+1)}, C_4 = \begin{bmatrix} D_5 \\ \lambda I \end{bmatrix}_{(2k+1)\times(k+1)}$$

where I is the identity matrix of order $k+1$,

$$C_5 = \begin{bmatrix} D_1 & 0 \\ 0 & D_6 \end{bmatrix}_{(2k+1)\times(2k+1)},$$

$C_6 = \lambda I$ where I is the identity matrix of order $2k+1$,

$$C_7 = \begin{bmatrix} 0 & 0 \\ 0 & D_7 \end{bmatrix}_{(2k+1)\times(2k+1)} \quad \text{and } C_8 = \begin{bmatrix} 0 & D_7 \end{bmatrix}_{(k+1)\times(2k+1)}.$$

Here '0' represents zero matrix of appropriate order. The matrices D_i $(0 \leq i \leq 7)$ are defined as follows:

$$D_0 = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -\lambda \end{bmatrix}_{N\times N}$$

$$D_1 = \begin{bmatrix} -(\lambda+\mu_2) & 0 & 0 & \cdots & 0 & 0 \\ \mu_2 & -(\lambda+\mu_2) & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \mu_2 & -(\lambda+\mu_2) \end{bmatrix}_{k\times k}$$

D_2 is a matrix of order $k \times N$ in which the $(1,2)^{th}$ element, μ_2 is the only non-zero element.

$$D_3 = \begin{bmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mu_1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}_{(k+1)\times k}$$

D_4 is a matrix of order $N \times (k+1)$ in which the $(N,1)^{th}$ element, λ is the only non-zero element.

$$D_5 = \begin{bmatrix} 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}_{k\times(k+1)}$$

D_6 and D_7 are square matrices of order $k+1$ defined by

$$D_6 = \begin{bmatrix} -(\lambda+\mu_1) & 0 & \cdots & 0 & 0 \\ \mu_2 & (-\lambda+\mu_1+\mu_2) & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mu_2 & -(\lambda+\mu_2) \end{bmatrix}$$

and

$$D_7 = \begin{bmatrix} 0 & \mu_1 & 0 & \cdots & 0 \\ 0 & 0 & \mu_1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \mu_1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

Now, the infinitesimal generator Q of the continuous time Markov chain, with state space defined in the beginning of this section, has the block partitioned structure as shown below:

$$Q = \begin{array}{c} \\ \overline{0} \\ \overline{1} \\ \overline{2} \\ \overline{3} \\ \vdots \\ \vdots \\ \overline{(N-1)} \\ \overline{N} \\ \overline{(N+1)} \\ \overline{(N+2)} \\ \vdots \\ \vdots \end{array} \begin{array}{cccccccccc} \overline{0} & \overline{1} & \overline{2} & \overline{3} & & \overline{(N-1)} & \overline{N} & \overline{(N+1)} & \overline{(N+2)} & \\ C_0 & C_2 & 0 & 0 & \cdots & 0 & C_3 & 0 & 0 & \\ B_1 & C_5 & C_6 & 0 & & 0 & 0 & 0 & 0 & \cdots \\ B_2 & C_7 & C_5 & C_6 & & 0 & 0 & 0 & 0 & \cdots \\ B_3 & 0 & C_7 & C_5 & & 0 & 0 & 0 & 0 & \\ & & & & & & & & & \\ \cdots & & & & & \cdots & \cdots & & & \\ B_{N-1} & 0 & 0 & 0 & & C_5 & C_4 & 0 & 0 & \\ 0 & 0 & 0 & 0 & & C_8 & A_1 & A_0 & 0 & \\ 0 & 0 & 0 & 0 & & 0 & A_2 & A_1 & A_0 & \\ 0 & 0 & 0 & 0 & & 0 & 0 & A_2 & A_1 & \\ & & & & & & & & & \\ \cdots & & & & & & & & & \end{array}$$

where

$$\overline{0} = ((i,0,0),(0,j,0)), \quad 0 \le i \le N-1, 1 \le j \le k$$

$$\overline{i} = ((i,j,0),(i,j,1),(i,k,2)), 0 \le j \le k-1 \quad \text{for } 1 \le i \le N-1$$

and $\overline{i} = ((i,j,1),(i,k,2)), \quad 0 \le j \le k-1$ for $i \ge N$

Let us denote by $\underline{X}$ the vector of steady state probabilities associated with

Q, such that

$$\underline{X}Q = 0, \qquad \underline{X}\,\underline{e} = 1 \tag{6.1}$$

where $\underline{e} = (1, 1, 1 \cdots 1)^T$. Let us partition $\underline{X}$ as $\underline{X} = (\underline{X}_0, \underline{X}_1, \underline{X}_2, \cdots)$ where $\underline{X}_0$ is a $1 \times (N + k)$ vector, $\underline{X}_i$ for $1 \leq i \leq N - 1$ are $1 \times (2k + 1)$ vectors and $\underline{X}_i$ for $i \geq N$ are $1 \times (k + 1)$ vectors

In the stable case, following Neuts[27], we examine the existence of a solution of the form

$$\underline{X}_i = \underline{X}_{N+1} R^{i-(N+1)}, \qquad i \geq N + 1 \tag{6.2}$$

For this we find from (6.1)

$$A_0 + RA_1 + R^2 A_2 = 0 \tag{6.3}$$

The matrix R is the minimal solution to (6.3). ie, R is an irreducible non-negative matrix of spectral radius, less than one. Latouche and Neuts [18] have proposed an iterative approach for finding the matrix R as follows:

$$\begin{aligned} R(0) &= 0 \\ R(n+1) &= -A_0 A_1^{-1} - R^2(n) A_2 A_1^{-1}, \quad n \geq 0 \end{aligned} \tag{6.4}$$

For Markov process with this type of generator, Neuts [27] obtained the stability condition

$$\underline{\pi} A_0 \underline{e} < \underline{\pi} A_2 \underline{e} \tag{6.5}$$

where the row vector $\underline{\pi}$ is defined as below:

Consider the infinitesimal generator $A = A_0 + A_1 + A_2$. A is irreducible.

Then there is a unique row vector $\underline{\pi} \geq 0$ such that

$$\underline{\pi} A = 0 \text{ and } \underline{\pi}\,\underline{e} = 1. \tag{6.6}$$

In this case $\underline{\pi} = (\pi_0, \pi_1, \pi_2 \cdots \pi_k)$ Hence (6.6) yields:

$$\pi_i = (\frac{\mu_1}{\mu_2})^i \pi_0 \text{ for } 0 \le i \le k$$
$$\text{and } \pi_0 = \frac{1 - \mu_1/\mu_2}{1 - (\mu_1/\mu_2)^{k+1}}.$$

Now $\underline{\pi} A_0 \underline{e} = \lambda$ and

$$\underline{\pi} A_2 \underline{e} = \mu_1(1 - \pi_k) = \mu_1(1 - (\frac{\mu_1}{\mu_2})^k \pi_0)$$

So (6.5) gives:

$$\frac{\lambda}{\mu_1} < \frac{1 - (\frac{\mu_1}{\mu_2})^k}{1 - (\frac{\mu_1}{\mu_2})^{k+1}} \tag{6.7}$$

which is the required stability condition.

We are now left with finding $(\underline{X}_0, \underline{X}_1 \cdots \underline{X}_{N+1})$. We define Q^* by

$$Q^* = \begin{array}{c} \\ \overline{0} \\ \overline{1} \\ \overline{2} \\ \overline{3} \\ \vdots \\ \vdots \\ \overline{(N-1)} \\ \overline{N} \\ \overline{(N+1)} \end{array} \begin{array}{c} \begin{array}{cccccccc} \overline{0} & \overline{1} & \overline{2} & \overline{3} & \cdots & \overline{(N-1)} & \overline{N} & \overline{(N+1)} \end{array} \\ \left(\begin{array}{cccccccc} C_0 & C_2 & 0 & 0 & & 0 & C_3 & 0 \\ B_1 & C_5 & C_6 & 0 & & 0 & 0 & 0 \\ B_2 & C_7 & C_5 & C_6 & & 0 & 0 & 0 \\ B_3 & 0 & C_7 & C_5 & & 0 & 0 & 0 \\ \cdots & & \cdots & \cdots & \cdots & & & \\ \cdots & & \cdots & \cdots & \cdots & \cdots & \cdots & \\ B_{N-1} & 0 & 0 & 0 & & C_5 & C_4 & 0 \\ 0 & 0 & 0 & 0 & & C_8 & A_1 & A_0 \\ 0 & 0 & 0 & 0 & & 0 & A_2 & A_1 + RA_2 \end{array} \right) \end{array}$$

We shall prove that $Q^* \underline{e} = 0$. Since the first N rows of Q^* are identical to the first N rows of Q, we need only prove that (last row of Q^*) $\underline{e} = 0$, Now (last row of Q^*) $\underline{e}$

$$
\begin{aligned}
&= (A_2 + A_1 + RA_2)\underline{e} = (I + R)A_2\underline{e} + A_1\underline{e} \\
&= (I - R)^{-1}[(I - R^2)A_2\underline{e} + (I - R)A_1\underline{e}] \\
&= (I - R)^{-1}[A_2\underline{e} + A_1\underline{e} - RA_1\underline{e} - R^2A_2\underline{e}] \\
&= (I - R)^{-1}[-A_0\underline{e} - RA_1\underline{e} - R^2A_2\underline{e}] = 0 \quad \text{using (6.3)}
\end{aligned}
$$

This implies $Q^*\underline{e} = 0$ and so Q^* is an infinitesimal generator. Also, the matrix Q^* is irreducible.

Let $\underline{X}^* = (\underline{X}_0, \underline{X}_1, \cdots \underline{X}_{N+1})$ be a solution of $\underline{X}^*Q^* = 0$. Then

$$
\begin{aligned}
&\underline{X}_0C_0 + \underline{X}_1B_1 + \underline{X}_2B_2 + \cdots + \underline{X}_{N-1}B_{N-1} = 0 \\
&\underline{X}_0C_2 + \underline{X}_1C_5 + \underline{X}_2C_7 = 0 \\
&\underline{X}_1C_6 + \underline{X}_2C_5 + \underline{X}_3C_7 = 0 \\
&\underline{X}_2C_6 + \underline{X}_3C_5 + \underline{X}_4C_7 = 0 \\
&\cdots \\
&\underline{X}_{N-2}C_6 + \underline{X}_{N-1}C_5 + \underline{X}_NC_8 = 0 \\
&\underline{X}_0C_3 + \underline{X}_{N-1}C_4 + \underline{X}_NA_1 + \underline{X}_{N+1}A_2 = 0 \\
&\underline{X}_NA_0 + \underline{X}_{N+1}(A_1 + RA_2) = 0
\end{aligned}
$$

The vectors $\underline{X}_i (0 \le i \le N)$ can be expressed in terms of $\underline{X}_{N+1}$ using the above set of equations and $\underline{X}_{N+1}$ may be normalised by using $\sum_{i=0}^{N} \underline{X}_i\underline{e} + \underline{X}_{N+1}(I - R)^{-1}\underline{e} = 1$.

References

[1] Artalejo, J. R. (1992): A unified cost function for $M|G|1$ queueing systems with removable server, *Trabajos De Investigacion Operativa*, 7:95–104.

[2] Avi-ltzhak, B. and Yadin, M. (1965): A sequence of two queues with no intermediate queue, *Management Science*, 11:553–564.

[3] Bailey, N. T. J. (1954): A continuous time treatment of a single queue using generating functions, *J. Roy. Statist. Soc. Ser.*, 16:288–291.

[4] Balachandran, K. R. (1973): Control policies for a single server system, *Management Science*, 19:1013–1018.

[5] Balachandran, K. R. and Tijms, H. (1975): On the D-policy for the $M|G|1$ queue, *Management Science*, 21:1073–1076.

[6] Champernowne, D. G. (1956): An elementary method of solution of the queueing problem with a single server and a constant parameter, *J. Roy. Statist. Soc. Ser.*, 18:125–128.

[7] Chatschik Bisdikian (1994): The random N-policy, *Research report*, IBM research division, New York.

[8] Clarke, A. B. (1977): A two server tandem queueing system with storage between servers, *Mathematics report*, 50, W. M. University, Kalamazoo.

[9] Cox, D. R. (1955): The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables, *Proc. Camb. Phil. Soc.*, 51:433–441.

[10] Doshi, B. T. (1986): Queueing systems with vacation - A survey, *Queueing systems*, 1:29–66.

[11] Erlang, A. K. (1909): The theory of probabilities and telephone conversations, *Nyt Tidsskrift Matematik*, B.20, 33–39.

[12] Gross, D. and Harris, C. M. (1985): *Fundamentals of queueing theory*, John Wiley, New York.

[13] Heyman, D. P. (1968): Optimal operating policies for $M|G|1$ queueing systems, *Operations Research*, 16:362–382.

[14] Heyman, D. P. (1977): The T-policy for the $M|G|1$ queue, *Management Science*, 23:775–778.

[15] Kandasamy, P. R. (1990): Matirx-geometric algorithmic approach to some Markovian queueing and inventory models, Ph.D. thesis, PSG college of Technology, Coimbatore, India.

[16] Kendall, D. G. (1953): Stochastic processes occuring in the theory of queues and their analysis by the method of imbedded Markov chains, *Ann. Math. Statist.*, 24:338–354.

[17] Kleinrock, L. (1975): *Queueing systems*, Volume 1, John Wiley, New York.

[18] Latouche, G. and Neuts, M. F. (1980): Efficient algorithmic solutions to exponential tandem queues with blocking, *SIAM J. Algebraic and Discrete Method*, 1:93–106.

[19] Lederman, W. and Reuter, G. E. (1954): Spectral theory for the differential equations of simple birth and death process, *Phil. Trans. Roy. Soc. London Ser.*, 246:321–369.

[20] Lee, H. S. and Srinivasan, M. M. (1989): Control policies for the $M^X|G|1$ queueing systems, *Management Science*, 35:708–721.

[21] Lee, H. W. (1988): $M|G|1$ queue with exceptional first vacation, *Computers & Operations Research*, 15:441–445.

[22] Levy, Y. and Yechiali, U. (1976): An $M|M|S$ queue with server's vacations, *INFOR*, 14:153–163.

[23] Medhi, J. and Templeton, J. G. C. (1992): A Poisson input queue under N-policy and with a general startup time, *Computers & Operations Research*, 19:35–41.

[24] Nadarajan, R. and Subramanian, A. (1984): A general bulk service queue with server's vacation, *Operational research in management system*, Academic Pub., 127–135.

[25] Nadarajan, R. and Audsin Mohana Dhas, D. (1989): Two units connected in series with general bulk service and random breakdown in unit II, *Microelectron. Reliab.*, 29:761–763.

[26] Neuts, M. F. (1968): Two queues in series with a finite intermediate waiting room, *J. Appl. Prob.*, 5:123–142.

[27] Neuts, M. F. (1978): Markov chains with applications in queueing theory which have matrix-geometric invariant probability vector, *Advances in Appl. Prob.*, 10:185–212.

[28] Neuts, M. F. (1981): *Matrix-geomtric solutions in stochastic models - an algorithmic approach*, The Johns Hopkins University Press, Baltimore.

[29] Parthasarathy, P. R. (1987): Transient solution to $M|M|1$ queue: a simple approach, *Adv. Appl. Prob.*, 19:997–998.

[30] Parthasarathy, P. R. and Sharafali, M. (1989): Transient solution for many-server Poisson queue - a simple approach, *J. Appl. Prob.*, 26:584–594.

[31] Prabhu, N. U. (1966): Transient behaviour of a tandem queue, *Management Science*, 13:631–639.

[32] Saaty, T. L. (1961): *Elements of queueing theory with applications*, McGraw Hill Book Co., New York.

[33] Sharma, O. P. and Shobha, B. (1984): A new approach to the $M|M|1$ queue, *J. Engg. Prod.*, 7:70–79.

[34] Sharma, S. D. (1974): On a continuous/discrete time queueing system with arrivals in batches of variable size and correlated departures, *J. Appl. Prob.*, 12:115–129.

[35] Syski, R. (1988): Further comments on the solution of $M|M|1$ queue, *Adv. Appl. Prob.*, 20:693.

[36] Takagi, H. (1987): Queueing analysis of vacation models, Part I : $M|G|1$ and Part II : $M|G|1$ with vacations, *Tech. report TR 87-0032*, IBM Tokyo Research Laboratory, Tokyo.

[37] Takagi, H. (1990): Time dependent analysis of $M|G|1$ vacation models with exhaustive service, *Queueing Systems*, 6:369–389.

[38] Takagi, H. (1992): Analysis of an $M|G|1$ queue with multiple server vacations and its application to a polling model, *J. Op. Res. Soc. of Japan*, 35:300–315.

[39] Takagi, H. (1993): $M|G|1|K$ queues with N-policy and setup times, *Queueing Systems*, 14:79–98.

[40] Yadin, M. and Naor, P. (1963): Queueing systems with a removable service station, *Operations Research Quarterly*, 14:393–405.