September 4, 2017

STATIONARY ANALYSIS OF THE M/M/1 QUEUING SYSTEM WITH TWO CUSTOMER SUBGROUPS

KHATETE MICHAEL

MSC 1153-03316-02328

A Thesis Submitted to the Department of Physical Sciences, School of Engineering and Applied Sciences, Kampala International University, Kampala-Uganda, in Partial Fulfillment of the Requirements for the Award of the Degree of Master of Science in Applied Mathematics (M.SC-AM)

2017

Declaration

I Khatete Michael, do hereby declare that this thesis entitled: "**STATIONARY ANALYSIS OF THE M/M/1 QUEUING SYSTEM WITH TWO CUSTOMER SUBGROUPS**" has never been submitted as a requirement for the award of any degree in any academic institution world wide. All citations are duly acknowledged by means of reference.

Signature and Date..

Khatete Michael

MSC: 1153-03316-02328

Certification

This is to certify that the thesis entitled "**STATIONARY ANALYSIS OF THE M/M/1 QUEUING SYSTEM WITH TWO CUSTOMER SUBGROUPS** " by Khatete Michael: MSC: 1153-03316-02328 was carried out under our supervision.

Signature and Date..

Dr. Sani Sulaiman

Supervisor I

Signature and Date..

Dr. Mbatudde Maria

Supervisor II

Approval

This dissertation has been examined and approved for the award of the Master of Science (M.Sc.) degree in Applied Mathematics.

Supervisor I.. Date........................

Supervisor II...................................... Date........................

Head of Department....................................... Date........................

Internal Examiner...................................... Date........................

External Examiner..................................... Date........................

Director; DHDR............................ Date........................

Acknowledgement

All thanks belong to God who gives me the power and energy to accomplish this research work. I wish to express my heartfelt gratitude to my main supervisor; Dr. Sani Sulaiman for his continuous guidance, counseling and support during this research work. My gratitude also goes to my co-supervisor; the current Head of Physical Sciences Department; Dr. Mbatudde Maria for her guidance. I also thank Dr. Kakuba Godwin who has been my lecturer throughout this M.Sc. period. I acknowledge the support provided to me by the staff of the School of Engineering and Applied Sciences (SEAS) of the Kampala International University especially the Dean; Dr. Jotham Sempewo, the sub-dean; Mr. Mundu Mustapha and Namaganda Shasha; the secretary of SEAS. Special thanks to the staff of the Department of Physical Sciences, SEAS; Dr. Aina Peter, Dr. Onourah Martins, Mr. Rajif Kissa and Mr. Onyesiga Living and all other staff of the Department. Finally, to my colleagues in the struggle.

Dedication

To the Almighty God, my family, relatives and friends.

Contents

Abstract

A single server Markovian queuing system called the $M/M/1$ queuing system with two distinct customer types namely; *the constraint customer type* and the *un-constraint customer type* is studied. This model is that of a production system with heterogeneous customer structures devoid of a single customer in the intersection category group wise. Initially, relevant literature covering methodologies, analysis and results derived for single server production systems are reviewed and a gap identified. The gap is that: **The application of group theory in the analysis of queuing systems is scarce in the literature. This is against the backdrop that most customers of service systems of transportation, telecommunications, computer systems and other production centers exhibit certain group characteristics**. This gap motivated us to impose a group structure on customers of the queuing system in question with the property that for the two subgroups considered in this work, their intersection is null. Using the generating function approach, we provided results on the stationary impact of one subgroup on the other. These results are fundamental basis for optimizing performance of sectors and sub-sectors of production centers relative to their distributions and expectations generally.

Chapter 1

Introduction

1.1 The Single Server Station:

In the queuing theory parlance, a single server station refers to the $M/M/1$ queuing system. By the $M/M/1$ queuing system, we mean a service station where customers arrive according to a Poisson process at a mean arrival rate say λ to receive service at a mean service rate say μ on the single server in the system. It is a Markovian service station with exponential inter-arrival and service time characteristics. The queue discipline in an $M/M/1$ queuing system may be the *first come first served* (FCFS) or the *last come first served* (LCFS). If $\frac{\lambda}{\mu} < 1$; then the queuing system is stable. Otherwise, it is considered unstable. The ratio $\frac{\lambda}{\mu}$ is called the *server utilization/occupation rate of the server* and is fundamental in defining a lot of performance measures such as the waiting time or the queue length distribution of customers in the system.

Normally, a single server Markovian station may have finite or infinite waiting space. Consequently, an $M/M/1$ queuing system where the available waiting space for customers is finite is modeled as the $M/M/1/K$. On the other hand, if the available waiting space is infinite, then the corresponding model is called an $M/M/1/\infty$. The model of a single server station that is the $M/M/1$ queuing model is memory-less because of its Markovian arrival and service characteristics. This property explains that the probability distribution of arrivals and services in an $M/M/1$queuing system are not influenced by history. Instead, the property of random variables related to the future depends only on the present. The $M/M/1$ queuing model may be discrete or continuous in time. That means the state of the $M/M/1$ queuing model is either countable or uncountable in time. This information is contained in the probability distribution of arrivals or services of the queuing model. Suppose X is a random variable whose values are defined in a finite set say $\{0, 1, 2, ..., n\}$. Under this condition, the probability distribution of X is discrete. Such a discrete probability distribution is known to be geometric. On the other hand, if the values of X are contained in the set $[0, \infty)$, then the probability distribution of X is continuous. A good example of continuous distribution is the exponential distribution.

1.2 Robustness of the $M/M/1$ Queuing Model:

The $M/M/1$ queuing model is generally robust. This is evident in the numerous analysis of the model under several descriptions in the literature. The analysis of variables such as waiting time and queue length distributions in the $M/M/1$ queuing system has been the focus of many studies and research to date. For instance, traffic intensities of roads, communication channels and telecommunications routes are analyzed with the use of the occupation rate parameter of the $M/M/1$ queuing model. Similarly, traffic congestion in service systems such as telecommunication systems, business and computer systems having direct economic bearings on nations and people are analyzed via the $M/M/1$ governing equations under added assumptions. In majority of urban areas, travel demands exceed highway capacities occasionally during peak periods. In such conditions, the standard equations of the $M/M/1$ queuing model under heavy traffic assumptions are suitable under this condition. In schools, opening times are integral times in school calendars. It should be arranged devoid of rush hour as a necessary condition. This analysis could be carried out conveniently under the governing Kolmogorov equations for the $M/M/1$ queuing model. More precisely, school pickup times and drop-off traffic times fit the $M/M/1$ equations under exponential arrival and service time distributions.

It is sequel to this robustness that this research work explores an analysis of the model under a somewhat extended structure for reasons to do with optimization and operational research. An attempt to broaden the scope of the model is carried out to accommodate realistic problems such as those problems to do with constraint growth and development as observed in many service of transportation, communication, telecommunication and business. In what follows, we provide underlying reasons why such extensions are justified.

1.3 Statement of the Research Problem:

A look at the literature on queuing systems reveals that the use of group theory in the analysis of queuing systems is not exploited generally. Intuitively, it can be envisaged that this stand is connected with the fact that in most queuing analysis, the assumption of homogeneity is a strong assumption. There are two reasons why this no exploit is weak. First, homogeneity of objects is unrealistic; Krishnamoorthy [3]. Secondly, the homogenous assumption employed in most queuing analysis limits the scope of application of models such as the $M/M/1$ queuing model in question. Suppose one is faced with a heterogeneous customer problem as we see everyday in banks where counters have distinct customer subgroups such as withdrawal counter for cashing customers, forex counter for currency exchange customers, etc. Since the assumption of homogeneity in most queuing models is strong, one cannot use these equations to capture this kind of problem. Generally, other basic assumptions governing most queuing models suffice for problems such as the one exemplified above. The absence of heterogeneity assumption makes adoption impossible and unrealistic. As in Sulaiman et al. [37], suppose one entrepreneur wishes to create jobs in a production system where there are two distinct types of jobs in the system with no relationship of any sort. Because of the near imperfect job structure in this case, homogenous job assumption embedded in the classical $M/M/1$ model will be unrealistic a.s. More precisely, it can make attainment of investment goals difficult. Such difficult-to-attain goals arising from homogenous counts in a heterogeneous situation may have direct consequences on entrepreneurship and business. Without doubt, it could lead to loss of jobs due to poor instrumentation and analysis. Unfortunately, existing $M/M/1$ global balanced equations are not free from homogenous assumption problem. Thus, these equations realistically cannot analyzed such production centers effectively since they cannot guarantee near exact results under imperfect considerations. More so, adopting such equations in their classical form without taking into account customer sub-groupings for quantification will impact negatively on the performance of production centers because of embedded homogeneity therein. Thus, there is need to extend the geometry of the said equations from their classical form to a group theory form so that their usage transcend originally designed homogenous problems with extension to heterogeneous problems.

1.4 Aim and Objectives of the Research:

The aim of this research is to extend the basic equations of the classical $M/M/1$ queuing model to the heterogeneous case via group theory analysis. This is carried out by assuming that in an arbitrary production center whose queuing feature is that of the $M/M/1$ queuing system has two distinct customer subgroups and a null intersection, with a view to arriving at an extended $M/M/1$ model equations capable of solving even optimization problems of heterogeneous-customer-base-production centers.

To achieve this aim, the following objectives are drawn

1. To construct an extended model of the $M/M/1$ queuing system with two heterogeneous customer subgroups.
2. To prove the stationary probability distribution and expectations of the model constructed in (1) above.
3. To analytically compare results proved for the model constructed in (1) above with results for similar models in the literature.
4. To provide remarks on the performance of the model in (1) above relative to those models compared.

1.5 Significance of the Research:

This research is significant to the following category of persons:

1. Government willing to understand the impact of sectors and subsectors on the entire service system for improving performance.
2. Academics as a source of reference.
3. Business owners and entrepreneurs whose businesses have heterogeneous customer structures for optimization and better service delivery.

1.6 Scope and Limitations of the Research:

The scope of this research covers all areas where existing equations of the classical $M/M/1$ queuing model can be applied. In addition, it encompasses additional constraint problems true for centers with heterogeneous customer structures. However, results of this research depend to a large extent on the rate equality principle which drives Poisson arrival processes to steady state.

1.7 Definition of Terms

Definition 1 (Random Variable) *: A random variable can be defined as a function acting on a given event space in terms of a numerical value, about which a probability statement can be made.*

If X is a random variable then, the above definition can be expressed mathematically as

$$X : \Omega \longrightarrow [0, 1],$$

where Ω is a subset of R^n.

Definition 2 (Probability Generating Function) *: A probability generating function (PGF) of a discrete random variable X with probability P_x and intensity z is defined as;*

$$P(z) = E[z^X] = \sum_{x=0}^{\infty} P_x z^x. \tag{1.1}$$

P(z) is convergent on the unit disc $|z| \leq 1$.

Definition 3 (Laplace Transform of a Probability Function) *: The Laplace transform of a probability function $f^*(z)$ for a random variable X is defined as*

$$f^*(z) = E[e^{-zX}] = \sum_{x=0}^{\infty} P_x e^{-zx}. \tag{1.2}$$

Definition 4 (Geometric Distribution) *A random variable X with parameter p denoting the probability that the xth trial out of x-trials is a success is said to be geometric if*

$$P(X = x) = (1 - p)^{x-1} p, \quad x = 0,\ 1\ ,\ 2... \tag{1.3}$$

Definition 5 (Poisson Distribution) *A random variable X with parameter λ is said to have a Poisson distribution if its probability density function is given by*

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0,\ 1\ ,\ 2, .. \tag{1.4}$$

where x is a non negative integer representing the probability that, there are exactly x-arrivals and e=2.71828.

Definition 6 (Exponential Distribution) *A random variable X with parameter μ is said to have an exponential distribution if its probability density function is given by*

$$P(X = x) = \mu e^{-\mu x}; \quad x > 0, \quad \mu > 0. \tag{1.5}$$

Definition 7 (Service Facility) *: A service facility is a point in a service system where customers take service.*

A service facility may be an ATM post, a production center, a parking space, a petroleum station, a mosque, a church or even a mall.

Definition 8 (Arrival Rate) *: An arrival rate is the intensity of arriving customers in a system.*

Definition 9 (Service Rate) *: The service rate is the average time customers spent on the service facility.*

It is the expected value of all individual service times for a given system.

Definition 10 (Steady State) *: Steady state means limiting state for a queuing system.*

In steady state, the rate at which customers arrive a system equals to the rate at which they depart away from the system.

Definition 11 (The Server Utilization) *: The server utilization is the fraction of time the server in a system is busy.*

Definition 12 (Sojourn Time) *: The sojourn time of a customer is the amount of time the customer stays in a system until he receives service.*

It is denoted by T and is simply given by;

$$T = W + S, \tag{1.6}$$

where W is the waiting time and S is the service time.

Definition 13 (Waiting Time) *: The waiting time of a customer in a system is the amount of time the customer stays in a wait (queue) before given the opportunity to enter the service facility.*

Definition 14 (Service Time) *: The service time of a customer is the time a customer spends on a service facility measured from the instant he starts receiving service, until the end of his service period.*

Definition 15 (Queuing Discipline) *: The queuing discipline is a rule describing the manner to which customers access the service facility to receive service.*

It defines the order to which service is to be received. A queueing discipline may be First-Come-First-Served ($FCFS$), a Last-Come-First-Served ($LCFS$), Processor Sharing (PS) or Priority Discipline (PD).

Definition 16 (The M/M/C Queue) *: By the M/M/C queue, we mean a queuing system where by, customers enter the system according to a Poisson process* ($first\ M\ means\ Memoryless$) *independent of any arrival with an expected value of* λ, *then receive service in a time that is exponentially distributed* ($M\ means\ Memoryless$) *with expected value* μ, *on any of the available* C *servers* ($C\ means\ servers$) *in the system.*

The number of servers in a system specifies the individual $M/M/C$ queue. For Instance, if $C = 1$ *or* $C = \infty$ then,the $M/M/C$ queue becomes, an $M/M/1$ or an $M/M/\infty$ queue.

Definition 17 (The M/G/C Queue) *: By the M/G/C queue we mean a queuing system with a Poisson process arrival (M means $Memoryless$) independently at an expected value of λ, to receive service at a service time that is not specified (G means $General$) on any of the C servers in the system (C is the number of servers).*

Definition 18 (Markov Process) *: A Markov process is a process whose values at different time instants are independent at a distance of one step.*

The Markov process that marks the feature of queuing systems are those referred to as embedded Markov chains since, we are looking at embedded points on the time axis at a specific chosen instant such as the departure instant and so on. The residual service time denoted by R_{res}, is the remaining service time for the current customer on service as seen by an arriving customer.

Definition 19 (Accessibility and Communication) *: A state j in a Markov chain is said to be accessible from another state j^*, if there exist a positive probability P such that, starting from j^*, the Markov chain will visit state j after a finite number of steps. This statement could be expressed for some n number of steps as;*

$$j^* \to j \;\; if \;\; for \;\; some \;\; n \geq 0 : P^n_{j^*,j} > 0. \tag{1.7}$$

If states j and j^* are accessible from each other then, the two states communicate and is written as, $j \leftrightarrow j^*$. A Markov chain having all states communicating is called an irreducible Markov chain.

Definition 20 (Aperiodicity) *: A state j in a Markov chain is said to be aperiodic if, the set of all transitions to state j from another state j^* has a greatest common divisor (gcd) of one.*

It is interesting to know that, if the states in a Markov chain form a communicating class then, each state in the chain is aperiodic.

Definition 21 (An Ergodic Markov Chain) *: A Markov chain is called ergodic if it is irreducible and aperiodic.*

Definition 22 (An Optimization Problem) *: An optimization problem can be defined as a problem to do with selecting the best solution out of many feasible solutions.*

Definition 23 (A Constraint Customer) *: A customer is called constraint if there exists certain functions that deny the normal functioning of the customer in a service system.*

Definition 24 (A Normal Customer) *: A customer is called normal if all his service functions are equivalent to that provided by a service system.*

Definition 25 (The rate equality principle) *: The rate equality principle states that in steady state, the rate in which customers enter a service system is equal to the rate at which they leave the system.*

Chapter 2

Literature Review

2.1 Introduction

Generally, whenever the demand for a given service exceeds the capacity to provide it, there will always exist congestion; Medhi [16]. Congestion, delays and queuing problems are most common features not only in our daily-life situations,[1] but also in more technical environments such as manufacturing, computer networking and telecommunications. For example, in the United States of America (USA), Hillier & Lieberman [10] estimated that Americans spend 37,000,000,000 patience-hours per year waiting in various queues. If this time could be spent productively instead, it would amount to nearly 20 million persons-years of useful work each year. The mathematical study of congestion, delays and queues is called queuing theory. More precisely, queuing theory is the mathematical study of congestion and waiting lines. It utilizes mathematical models and performance measures to assess and improve the flow of customers in a queuing system. Queuing theory has many applications and has been used extensively by service industries. Additionally, queuing theory is used to assess staff schedules, working environment, productivity sectors, patient's waiting

[1] Such as at a bank or postal office, at a ticketing office, in public transportation or in a traffic jam

time, etc.

Queuing theory itself does not directly find solutions to delays, congestion and queue formations, it contributes vital information required for such decisions by predicting various characteristics of the queues such as the average waiting time, the average number of customers in the queue; (Gross & Harris [7]; Hillier & Lieberman [10]). Similarly, Willig [2] described a queuing system as a service center together with a population of customers that may enter the service center at various points of time in order to get service. In many cases, the servers in a given service center can only serve a limited number of customers at a time. If a new customer arrives and there is no free server among the servers in a system, the customer enters a waiting line and waits until a server becomes available. The term customer is used in a general sense and doesn't imply necessarily a human customer. For example, Gross & Harris [7] stated that the word *customer* could be used for a ball bearing waiting to be polished, an airplane waiting in line to take off, or a computer command waiting to be performed.

As customers do not like to wait in queues, queue managers of establishments do not also like customers to wait. This is because customer impatience has a negative impact on business of firms as it leads to loss of potential customers. Thus, various queuing models have been studied to provide solutions to congestion and delay problems experienced in service systems. For instance, Kumar [28], Kapodistria [34], Fuhrmann & Cooper [38]. The most common service station is that station with a single server present in the system. The literature covering this kind of service stations are dominant generally.

2.2 The Single Server Queue:

By a single server queue, we mean a waiting line generated in a uni-server station under a given arrival and service distributions. Models under single server system category may include; the M/M/1 queuing model [Def. 16], the M/D/1 queuing model, the M/G/1 queuing model, the G/G/1 queuing model and etc.

Over the years, a lot of research has been carried out in the analysis of single server queuing systems. For instance, Abdelkader & Maryam [13] studied a single-server Markovian queuing system called the M/M/1 queue. Using the method of order statistics, Abdelkader & Maryam [13] computed some performance measures for the M/M/1 queue. The result shows that there is an inverse relationship between the traffic intensity of customers and idle service intervals. Mohammad [36] studied the total minimum expectation cost of a bank in solving the waiting line problems. Using the M/M/1 queuing model and linear programming technique, the performance of the system was estimated. Poongodi & Muthulakshmi [39], studied the construction of control charts for systems involving congestion and traffic problems. Using the M/M/1 queuing model, the performance of the system was improved. Modares & Fakher [26] studied material flow under heavy traffic conditions. Using the M/M/1 queuing model, an analysis of the in-site traffic was obtained and applied in determining optimum stock levels of materials at destination shops. The least stock levels at destination shops for a predetermined production interruption rate was obtained. Boer et al. [29] studied various properties of the M/M/1 queuing system. Boer et al. [29] obtained analytically the properties of flexibility and robustness of the M/M/1 queuing model. In addition, the work provided a strong evidence for the cross entropy optimality.

Tsitsiklis & Xu [17] studied queuing system topologies with limited flexibility using the M/M/n and M/M/1 queuing models. Comparing the performance of the two models, the M/M/1 queuing model obtained the fully flexible system with a much larger capacity region. Hence the M/M/1 queuing model is robust to uncertainties or changes in the arrival rates. Girish & Hu [25] developed some higher order approximations for estimating the performance measures in a GI/G/1 queue with splitting, merging and feedback phenomena. Approximations for the moments and the lag-1 auto correlations of splits from a renewal process or from a departure process of the model in question was derived. The moments of the split process were obtained explicitly from that of the original process. Similarly, the auto correlation was expressed in terms of the moments and auto correlations of the system times. Girish & Hu [25] obtained that the superposition of two renewal processes is a Markov renewal process.

Whitt [43] approximated the performance of time- varying $G_t/GI/1$ queue using the diffusion approximations and heavy-traffic limits. It was established that a heavy traffic limit theorem for the model in question exists. Adan & Kulkarni [14] studied a single-server MAP/G/1 queuing system where the inter-arrival times of customers and their service times [Def. 14] depend on a common discrete time Markov chain. The service times of customers are assumed to be independent and identically distributed random variables. Using Lindley's integral equation, Adan & Kulkarni [14] derived the steady-state waiting time and queue length distributions leading to a recursive equations for the calculation of moments of the function model. Natalia [41] reviewed the stochastic decomposition for the number of customers in an M/G/1 [Def. 17] retrial queues under reliable server regime and when the server is subjected to breakdowns. Under exponential assumption for retrial times, an approximation in the non-exponential case was obtained. Similarly, an approximated solution for the steady-state queue size distribution was derived. Natalia [41] proved that increasing the traffic intensity and the coefficient of variation of service times and that of retrial times have an adverse effect on the performance of the approximation.Song et al. [46] studied the optimal service policies in an M/G/1 queuing system with consecutive vacations. Using a finite state Semi-Markov decision model, the optimal service policy to minimize the long-term average cost for the vacation system was obtained. Haviv [22] studied the M/M/1 queuing system with a particular interest on customer behavior.[2] Haviv [22] obtained the appropriate non-cooperative games and their Nash equilibria. Khew et al. [35] provided a new approach for finding basic per-

[2]With respect to their cost/reward parameters

formance measures of the GI/G/1 queuing system. Given that the inter arrival and service times are discrete random variables, the steady-state waiting time distribution for the continuous time GI/G/1 queue model was obtained. Thangaraj & Vanitha [42] studied the M/G/1 queuing system using Bernoulli probability distribution and uni vacation policy. Given that the server takes a vacation at an exponentially distributed period, the time dependent probability generating function was derived. A steady state results for the mean queue length and mean waiting time [Def. 13] were computed.

Overall, the literature on single server queue is enormous. Additionally, the numerous results proved for this queuing system especially the $M/M/1$ model emphasize the continuous interests of scholars and experts on the model.

2.3 The Single Server Queues under Constraints:

Maglaras et al. [5] studied a multi-product M/G/1 queuing system with the intention of controlling the lead time for service. A tractable approach to obtain the lead time constraints through the admission and sequencing control was provided using a deterministic fluid-model. Finally, a general and a relatively simpler constraint approach more simpler than a heavy-traffic approach with specific admission policies and consistent with heavy traffic analysis was developed. Cheng et al. [6] studied a two stage M/M/1 tandem queuing system under process queue time (PQT) constraints and proposed a batch process admission control (BPAC) model. The queuing system studied consists of an upstream batch process machine and a downstream single process machine where the waiting time of each job in the downstream queue is constrained by an upper limit. Violation of this limit causes scrap of the job. Simulation result shows that the proposed BPAC model outperforms other methods in every key system performance indices.

Noah & Zhou [12] investigated an inbound telephone call center queuing system that processes two types of work under service level constraints. Using dynamic programming methods, an optimal policy normalization was derived. The results show that when the expected service times of the two classes differ, the policies are optimal within the class of priority policies. The determination of optimal policy parameters can be obtained through the solution of a linear program with $O(c^3)$ variables and $O(c^2)$ constraints. Sundaresan [33] studied the capacity constraint problem in an exponential server timing channel ./M/1 queuing system via point process channels. The point-process approach enables the timing channels that arose in both single and multi server queues to be studied. Sundaresan [33] provided an analytical bounds for the queuing system and highlighted a method to obtain achievable rates using simulation technique.

Abdolghani [9] studied the Markovian queuing system with Poisson inter arrival rate and exponential service times for customers. A queuing system having finite storage capacity with additional customers get rejected whenever the buffer is full was assumed. Using Bellman equations for bounded and unbounded action space and Lagrangian analysis, an idea of constrained Markov decision process is obtained. Xiaofei & Feinberg [44] studied the problem of optimal admission of arriving customers in a Markovian finite-capacity queuing system with several customer types. The study considered that the system managers be rewarded for serving customers and penalized for rejecting customers. The rewards and penalties depend on customer types. Xiaofei & Feinberg [44] aimed at maximizing the average rewards per unit time subject to the constraint on the average penalties per unit time. Using a Linear Programming transformation method and optimal policies based on Lagrangian optimization, it was shown that the existence of a 1-randomized trunk reservation optimal policy with an acceptance thresholds for different customer types exists.

Sumitha & Chandrika [8] studied a batch arrival retrial $M^x/G/1$ queuing system with two phases of heterogeneous service and controllable arrivals. The study assumed that when a server is idle, batch customers are admitted into the system for service.

In addition, upon completion of service, the server may go for a single vacation or remain idle in the system. Under steady state analysis, the solution for a retrial queue with admission control and vacation was obtained. Ayyappan & Shyamala [11] studied a single server batch arrival non-Markovian retrial queuing system with non persistent customers. Using probability generating function methods, the steady state solution with performance measures of the system and the reliability indices were obtained. Jeeva1 & Rathnakumari [24] studied a retrial M/G/1 queuing system with modified vacation policy, random server breakdown, balking and optional re-service. Using supplementary variable technique, the probability generating functions of the number of customers in the system when it is idle, busy, on vacation and under repair are obtained. In addition, performance measures of this queuing system was derived. Artelajo & Falin [15] compared a single server M/G/1 and the multi server M/M/c models of retrial queues with emphasis on similarities and differences between the retrial queues and their counterparts. The study demonstrated that although retrial queues are closely connected with these standard queuing models, they however possess unique distinguished features. For instance, the standard queuing models do not take into account the phenomenon of retrials and therefore cannot be applied in solving a number of practically important problems. Similarly, early works on retrial theory have shown that retrial queues are suitable mathematical models for the modeling of subscribers behavior in telephone networks. Laghaie et al. [20] studied a single server queuing system subject to two different deteriorating conditions. Using the conditions of a planning horizon on the D/M/1 queuing system in accordance with certain maintenance policy, a new model was developed. The model controlled sojourn times for optimal values of arrival rates. Similarly, the model controlled maintenance policy and cost for reasons to do with minimization. Ayestaa et al. [40] studied a single server M/G/1 processor sharing (PS) queuing system with multiple vacations. The server only takes a vacation when it is empty. If the system is empty upon return after vacation of the server, the server takes another vacation, and so on. Under the following constraints, the service discipline satisfies a branching property. Similarly, the arrival processes at various queues are independent Poisson processes. Ayestaa et al. [40] determined the sojourn time distribution of an arbitrary customer and sojourn time distribution in the M/M/1-PS queuing system of a polling model. Chih-ping et al. [21] studied two convex optimization problems in a multi-class M/G/1 queuing system with a control service rates.The minimizing convex functions of the average delay vector and average service cost were both subject to per-class delay constraints. Using virtual queue techniques, the delay and rate-optimal control in a multi-class priority queue with adjustable service rates was obtained . The model was analyzed and validated through simulations.

Pant & Ghimire [1] studied the M (t)/M/1 queuing system with customer arrival rate following a sinusoidal time process and the server's rate is an exponential time process. The aim is to derive a model for this system. For this model, Pant & Ghimire [1] obtained the expected number of customers in the system, the expected number of customers in the queue, and etc. In addition, numerical results for various parameter change were obtained. Krenzler & Daduna [30] studied the M/D/1 queuing system with infinite waiting room in a random environment where the service system and the environment interact in both directions. It is supposed that whenever the environment enters a pre specified subset of its state space, the service process is completely blocked, interrupted and newly arriving customers are lost. Using queuing-inventory and reliability theory, Krenzler & Daduna [30] obtained a product form equilibrium of the embedded Markov chain under general conditions. In addition, numerical results for various parameters were obtained.

Kumar et al. [31] studied the M/M/1 single server queuing system with retention of reneged customers. Using economic analysis of an optimum strategy for a firm to maximize its profit under the constraints, Kumar et al. [31] obtained a recursive formula for the steady-state solution. In addition, the total expected cost, total expected revenue and total expected profit functions were derived. Similarly, the optimum system capacity and optimum service rate was obtained. Kumar et al. [32] studied a Markov modulated Poisson process of the single server queuing system called the MMPP/M/1. The study aimed at minimizing a combination of effort cost and holding cost incurred per unit time using the discounted and average cost optimality criterion. Kumar et al. [32] characterized the structure of an optimal service rate as being monotone in the queue length for each arrival rate. In particular, the manner in which the process switches between the arrival rates plays an important role in determining the structure of the optimal policy. When the transition matrix governing the MMPP is stochastically monotone, then optimal control arrival rates are obtained.

Maglaras [4] studied the problem of dynamic pricing for a multi product make-to-order system using a multi class single server queuing system Mn/M/1 model with controllable arrival rates, general demand curves, and linear holding costs. The problem of maximizing the expected revenues minus holding costs by selecting a pair of dynamic pricing and sequencing policies was studied. Using a deterministic and continuous (fluid model) relaxation the solution for the optimal greedy sequencing and the optimal pricing and sequencing decisions decouple in finite time was obtained. Artalejo & Lopez [18] provided a survey on information theoretic technique for estimating the performance characteristics of the M/G/1 retrial single server queuing system. Basically, the survey focused on the limiting distribution of the system state, the length of a busy period and the waiting time. Using the principle of maximum entropy (PME) to estimate probability distributions under given constraints, the solutions were obtained. Similarly, the numerical experiments showed the goodness of the maximum entropy solutions. Wanga et al. [19] studied an M/G/1 single removeable queuing system operating under the N policy in steady-state where a server is turned on at arrival epochs or off at departure epochs. Using the maximum entropy principle with several well-known constraints, the approximate formula for the probability distributions of the number of customers and the expected waiting time in the queue was developed. Similarly, a comparative analysis between the approximate results with exact analytic results for three different service time distributions, exponential, 2-stage Erlang, and 2-stage hyper-exponential was obtained. The maximum entropy approximation approach is accurate enough for practical purposes. On illustration, the maximum entropy approximation approach proved accurate enough for practical purpose.

Guo et al. [45] studied optimal probability routing in distributed parallel queuing system. Guo et al. [45] established the stochastic process limit for the GI/GI/1 queuing system, analyzed diffusion process to obtain an approximation for the corresponding equilibrium sojourn time. Using this approximation process, the optimal routing probability was calculated. Herlicha et al. [23] studied the energy efficient queuing system with delayed activation and deactivation using an M/G/1 single server. The study analytically determined the steady state distribution and derived a closed form formula for both power consumption and latency depending on the rate of arrival, processing, activation, deactivation, activation delay, and deactivation delay. Using simulation, the results obtained demonstrated that it also holds for other random distributions.

2.4 Literature Gap:

From the literature reviewed, it can be seen that a lot of works on the single server queuing system has been carried out. Similarly, single server queuing systems with constraints have equally been studied. The most striking feature of the literature covering queuing system analysis is combined customer analysis (homogenous considerations). Very little exists on queuing analysis under distinct customer group consideration. Even though, it is evidently clear that customers of service systems where queuing results are required are mostly organized in groups and subgroups. A problem area depicting this type of systems is an imperfect production center whose imperfection is sequel to certain latency problems in the system. Analyzing this type of system is necessary for performance evaluation relative to cost of service, quality of service and optimization purposes. Though in Sulaiman et al. [37], a similar problem with arithmetic properties is studied, the model presented cannot capture beyond 1-cycle of latency in a system. There are numerous production centers with more than 1-cycle of latency as a result of customer behaviors, service provider's behavior and etc. Additionally, there are several service systems for instance, ATM systems, banking systems, computer systems and etc with constraint customers whose sizes are known to be monotonically increasing.

To the best of our knowledge, models covering this kind of industry problems are not fully discussed in the literature. Most importantly, this kind of industry problem requires group analysis for better posing and desired results. The next chapter provides a methodology on the modeling of a case problem for a single server Markovian queuing system.

Chapter 3

Modeling

3.1 Basic Assumptions:

Consider an optimization problem $O(.)$ [Def. 22] to do with the number of customers N in an $M/M/1$ queuing system. Let $\{N\} = \{c\} \otimes \{n\}$; $\{c\} \cap \{n\} = \{\}$. Where $\{c\}$ is a constraint subgroup of $\{N\}$ [Def. 23] and $\{n\}$ is an un-constraint subgroup of $\{N\}$ [Def. 24]. Suppose that customers arrive the queuing system according to a Poisson process [Def. 5] at a rate of λ to receive [1] service on the single server in the system. The service time of customers is assumed to follow the exponential distribution [Def. 6] with service parameter μ. For reasons to do with stability, we suppose that the occupation rate [Def. 11] $\rho = \frac{\lambda}{\mu} < 1$.

3.2 Model Evolution

Generally, the $M/M/1$ queuing system will evolve in time; see Medhi [16]. That means, as $t \to \infty$, the time dependent group of customers $\{N(t)\} \to \{N\}$. Consequently, an analysis of the $M/M/1$ queuing system under this condition is a stationary

[1] That means there are two types of customers in the system with distinct characteristics

analysis. Let j denote the present state of the stationary group process $\{N\}$. Then the associated Markov chain for this group process is two-state since a transition from j resides only in $(j-1)$. In addition, $j \leftrightarrow (j-1)$. Hence, the Markov chain is irreducible, [Def. 19]. Let k denote the number of steps to reach $(j-1)$ given that the chain is initially in j. Then the greatest common divisor (gcd) of the set containing the number of times the chain goes to j given that it was in $(j-1)$ is unity. That is, the $gcd(k) = 1$. Hence, the Markov chain is aperiodic, [Def. 20]. One can conclude that the Markov chain for the customer process N is an ergodic Markov chain, [Def. 21]. Now, suppose that the rate equality principle for Poisson arrival processes holds. Let P be a probability measure on the random number [Def. 1] $N \in \{N\}$ such that $P[N=j] = P_j$ gives the probability that there are j customers in $\{N\}$. Then from the known queuing results below

$$\lambda P_0 = \mu P_1; \qquad j = 0 \tag{3.1}$$

$$(\lambda + \mu) P_N = \lambda P_{N-1} + \mu P_{N+1}; \qquad j > 0 \tag{3.2}$$

$$P_N = \left[1 - \left(\frac{\lambda}{\mu}\right)\right] \left(\frac{\lambda}{\mu}\right)^N \qquad j \geq 0, \tag{3.3}$$

given that $\{N\} = \{c\} \otimes \{n\}$; $\{c\} \cap \{n\} = \{\}$ such that

$$N = nc = c, 2c, 3c, ..., rc, ... \qquad c \in \mathbf{N}, \quad n = 0, 1, 2, 3, ... \tag{3.4}$$

Then the size of the group $\{N\}$ of stationary customers in the system can be computed under multiplier effect of the size of $\{c\}$ on the subgroup $\{n\}$. The direct interpretation of this kind of the $M/M/1$ queuing system is that of a uni-server system with two distinct customer subgroups where the effect of one subgroup is dominant on the other subgroup.

Now, denote by $V(z)$ the probability generating function (PGF) for the stationary number of customers in $\{N\}$. In view of [Def. 2] and the definition of N in (3.4) above, one can write

$$V(z) = \sum_{N=0}^{\infty} P_N z^N = \sum_{cn|n=0}^{\infty} P_{cn} z^{cn}, \quad c \in \mathbf{N} \tag{3.5}$$

Equation (3.5) contains all vital information on the stationary performance characteristics of the optimization problem $O(.)$ described above. More precisely, if $V(z)$ is known then, one can compute all vital characteristics of $O(.)$ such as the waiting time and queue length distributions and expectations for one subgroup under the effect of the other subgroup for optimization purposes and better service delivery.

3.3 Model validation

The extended model of the $M/M/1$ queuing system designed in this work will be validated using the classical $M/M/1$ model. For varying sizes of $\{n\}$ and a fixed size of $\{c\}$ as in (3.4), we proposed that the extended model is valid if there exists an $n \in \aleph$ and a $\lambda \in \Re$ such that

$$P_{\{c\}\otimes\{n\}}(extended) = P_N(classical) \tag{3.6}$$

3.4 Discussion of Results

Analytic results derived in this work will be compared with those of two models namely; the classical $M/M/1$ model and the additive model presented in Sulaiman et. al. [37]. This comparison will be carried out as a discussion on areas of gains of the model presented in this work on the impact of subgroup $\{c\}$ on the entire distributions of customers in the system.

Chapter 4

Results and Discussions

Lemma 4.0.1 *Given the customer group* $\{N\} = \{c\} \otimes \{n\} : \{c\} \cap \{n\} = \{\}$*; the stationary probability* P_N *that there are* N *customers in the group* $\{N\}$ *is given by*

$$P_N = \rho_c^c(1-\rho_c)\rho_n^n(1-\rho_n) \tag{4.1}$$

Proof Suppose an ordered paired process $\{N(t) = c(t)n(t), \zeta(t)\}_{t\geq 0}$ is given. Where $N(t)$ denotes the number of customers in the system at time t and $\zeta(t)$ is the past service time of an arbitrary customer in $\{N\}$. Looking at the system at departure instants upon completion of service of an arbitrary customer, the bi-variate process $\{N(t), \zeta(t)\}_{t\geq 0}$ is a Markov process; see [Def. 18]. Assuming that the service time of customers is continuous and that the system is empty at time $t = 0$. Then one can apply the supplementary variable technique on $\{N(t), \zeta(t)\}_{t\geq 0}$; Boxma et al. [27]. Let P be a probability measure on $\{N(t), \zeta(t)\}_{t\geq 0}$ such that[1]

$$P[\{N(t), \zeta(t)\} = 0] = P[cad\{N(t), \zeta(t)\} = 0]; cad\{c\} = 0, cad\{n\} = 0. \tag{4.2}$$

[1]The cardinality of a group denoted by (cad) in this work is the number of customers, elements or objects in the group

$$P[\{N(t), \zeta(t)\} = c] = P[cad\{N(t), \zeta(t)\} = c]; cad\{c\} = c, cad\{n\} = 1. \tag{4.3}$$

$$P[\{N(t), \zeta(t)\} = cn] = P[cad\{N(t), \zeta(t)\} = cn]; cad\{c\} = c, cad\{n\} = n. \tag{4.4}$$

Now, given that $\lambda < \mu$ for all arrivals within $\{N\}$, then as $t \to \infty$, the time dependent probabilities $P[cad\{N(t), \zeta(t)\} = 0]$, $P[cad\{N(t), \zeta(t)\} = c]$ and finally, $P[cad\{N(t), \zeta(t)\} = cn]$ converge to $P[cad\{N, \zeta\} = 0] = P_0$, $P[cad\{N, \zeta\} = c] = P_c$ and $P[cad\{N, \zeta\} = cn] = P_{cn}$ respectively, [Def. 10]. Similarly, the time dependent process $\{N(t), \zeta(t)\}_{t \geq 0} \to \{N, \zeta\}$; see Medhi [16]. Let $P[N = n|c]$ denote the probability that there $n \in \{N\}$ customers in the system given that there were $c \in \{N\}$ fixed customers. Suppose that every arrival $n \in \{N\}$ is followed by a departure (rate-equality principle)with the condition that $\{c\} \cap \{n\} = \{\}$. Then by the ergodic theorem, the following difference-differential equations are satisfied by the stationary process $\{N, \zeta\}$.

$$\lambda_c P_0 = \mu_c P_1 : c = 0 \tag{4.5}$$

$$(\lambda_c + \mu_c) P_c = \lambda_c P_{c-1} + \mu_{c+1} : c > 0 \tag{4.6}$$

$$\lambda_n P(n = 0|c) = \mu_n P(n = 1|c) : n = 0 \tag{4.7}$$

$$(\lambda_n + \mu_n) P(n = n|c) = \lambda_n P(n = |n - 1|c) + \mu_n P(n = n + 1|c) : n > 0 \tag{4.8}$$

By the Markov property of the system as claimed and combining (4.5) and (4.6) for $c = 0, 1, 2, 3, ...$, we have

$$P_c = \left(\frac{\lambda_c}{\mu_c}\right)^c P_0 \tag{4.9}$$

The zero state probability P_0 that clears $c \in \{c\}$ customers is obtained from the normalization condition

$$\sum_{c=0}^{\infty} P_c = P_0 + P_1 + P_2 + ... = 1 \tag{4.10}$$

Consequently,

$$P_c = \left(\frac{\lambda_c}{\mu_c}\right)^c \left(1 - \frac{\lambda_c}{\mu_c}\right) \tag{4.11}$$

Similarly, combining (4.7) and (4.8) for $n = 0, 1, 2, 3, ...$ when there are $c \geq 0$ customers, we have

$$P(n = n|c) = \left(\frac{\lambda_n}{\mu_n}\right)^n P(n = 0|c) \tag{4.12}$$

$$P(n = n|c) = \left(\frac{\lambda_n}{\mu_n}\right)^n \left[1 - \left(\frac{\lambda_n}{\mu_n}\right)^n\right] \tag{4.13}$$

The lemma holds good upon combining (4.11) and (4.13) for $\frac{\lambda_i}{\mu_i} = \rho_i$.

Lemma 4.0.2 *The stationary expected number of customers in the subgroup* $\{n\}$ *when there are* $c \in \{c\}$ *customers is given by*

$$E[n|c] = cP_c. \left[E[N]_{M/M/1}\right] \tag{4.14}$$

Proof Denote by $V(z)$ the PGF [Def. 3] of the number of customers in the group $\{N\}$ generally. In view of (3.5) when c is fixed and $n = 0, 1, 2, 3, ...$, we have

$$V(z) = P_0 + P_{1.c}z^c + P_{2.c}z^{2c} + P_{3.c}z^{3c} + ... \tag{4.15}$$

So that (4.15) becomes

$$V(z) = P_0 + P_c z^c + P_{2c} z^{2c} + P_{3c} z^{3c} + ... \tag{4.16}$$

In view of (4.1), one can write (4.16) as

$$V(z) = \rho_c^c (1-\rho_c)(1-\rho_n)\{1 + \rho_n z^c + \rho_n^2 z^{2c} + \rho_n^3 z^{3c} + ...\} \tag{4.17}$$

Which gives in view of (4.1) again that

$$V(z) = P_0(1 + \rho_n z^c + (\rho_n z^c)^2 + (\rho_n z^c)^3 + ...) \tag{4.18}$$

(4.18) is an infinite geometric series with common ration $\rho_n z^c$. Hence, we have

$$V(z) = \frac{P_0}{1-\rho_n z^c} \tag{4.19}$$

Differentiating (4.19) with respect to z yields

$$V^{'}(z) = \frac{cP_0 \rho_n z^{(c-1)}}{(1-\rho_n z^c)^2} \tag{4.20}$$

Where $V'(z)$ gives the expectation of customers in the subgroup $\{n\}$ when there are $c \in \{c\}$ customers in the system. The lemma follows upon evaluating (4.20) at $z = 1$ with simple re-arrangement and simplification.

Lemma 4.0.3 *Suppose $N < \infty$ such that $N \to K$ for some $K \in \Re^+$. Then the stationary expected number of customers $n \in \{n\}$ given $c \in \{c\}$ is given by*

$$E[n|c] = \frac{c\rho_n {\rho_c}^c (1-\rho_c)[(1-\rho_n^K) - K\rho_n^{K-1}(1-\rho_n)]}{(1-\rho_n)} \tag{4.21}$$

Proof Denote by $U(z)$ the PGF of the number of customers $n \in \{n\}$ when there are $c \in \{c\}$ customers in the system. By the definition of PGF [Def. 3] for $c \geq 0$ and $n = 0, 1, 2, 3, ...$, we have

$$U(z) = P_0 + P_c z^c + P_{2c} z^{2c} + P_{3c} z^{3c} + ... + P_{cK} z^{cK} \tag{4.22}$$

In view of (4.1) with some rearrangement, we have

$$U(z) = {\rho_c}^c(1-\rho_c)(1-\rho_n)[1+\rho_n z^c + (\rho_n z^c)^2 + (\rho_n z^c)^3 + ... + (\rho_n z^c)^K] \quad (4.23)$$

Upon further rearrangement and application of certain basic properties of geometric series, we have

$$U(z) = {\rho_c}^c(1-\rho_c)(1-\rho_n)\left[\frac{1-(\rho_n z^c)^K}{1-\rho_n z^c}\right] = P_0\left[\frac{1-(\rho_n z^c)^K}{1-\rho_n z^c}\right] \quad (4.24)$$

Letting the numerator of (4.24) as ψ and the denominator as ϕ and then differentiating the said equation with respect to z yields

$$U'(z) = \frac{-cKP_0\phi(z)\rho_n^K z^{cK-1} + \psi(z)c\rho_n z^{c-1}}{(1-\rho_n z^c)^2} \quad (4.25)$$

The lemma follows after evaluating (4.25) at $z=1$ upon further simplification.

4.1 Discussions and Remarks

In this section, we compare results derived in this work with those of the model presented in Sulaiman et al. [37]. We wish to study the impact of certain considerations adopted here and how such considerations aid the understanding of imperfect production centers for better strategies and operations.

Remark 4.1.1 *Suppose the subgroup $\{n\} = \{\}$. Under the condition of no arrival into the subgroup $\{n\}$, then the stationary probability P_N in (4.1) is that of the classical $M/M/1$ queuing system.*

Intuitively, if $\{n\} = \{\}$, then the stability condition of the system is saddled on the subgroup $\{c\}$. Consequently, the total number of customers in the entire customer group $\{N\} = \{c\} \cup \{n\} = \{c\} \cup \{\} = \{c\}$. Putting $n = 0$ in (4.1) yields

$$P_N = \rho_c^c(1-\rho_c)(1-\rho_0) \quad (4.26)$$

Where $\rho_{n=0} = \rho_0$ is the probability that no arrival occurring at the rate λ_n in the system takes place. Finally, the remark holds if one lets $\rho_0 = 0$ in (4.26) trivially. Under this condition, there is an equivalence relation between (4.26) at $\rho_0 = 0$ and (5.1) in the appendix-1. This equivalence relation proves the validity of the model presented in this work.

Remark 4.1.2 *There exists a direct relationship between the size of the subgroup* $\{n\}$ *and the stationary probability* P_c.

This is in view of Lemma 4.02. More precisely, the lemma implies a direct relationship between the size of the elements in the subgroup $\{n\}$ and the stationary probability P_c given that c and $E[M/M/1]$ are constants. Thus, the maximizer of the elements of $\{n\}$ is any real valued function that increases the tendency P_c in $(0, 1)$.

Remark 4.1.3 *Under similar arrival and service conditions of the extended model presented in this work with that of Sulaiman et al. [37], it canbe concluded that* $E[N](extended) < E[N](additive)$ *a.s.*

This is in view of comparative analysis of (4.21) and (6.3). Additionally, by the comparative analysis of (4.1) and (6.1) with that of (6.2) that shows the slow nature of reversal of (4.1) to (6.1) more slower than that of (6.2) to (6.1).

Chapter 5

Summary, Recommendations and Conclusion

5.1 Summary

In this research work, an extended model of the classical $M/M/1$ queuing system with two customer subgroups is developed. Using group theory and analysis, stationary probability distribution P_N that there are $n \in \{n\}$ customers when there are already $c \in \{c\}$ customers is obtained. Similarly, the expected number $E[n|c]$ for $n \in \{n\}$ in the system given $c \in \{c\}$ customers are derived. The derivations considered both finite and infinite buffer assumptions. Other areas covered include a comparison on the performance of the extended model relative to the classical $M/M/1$ model and the additive subgroup model presented in Sulaiman et al. [37] leading to some remarks. The remarks show a direct relationship between the size of the subgroup $\{n\}$ and the stationary probability P_c. Additionally, the work proves a unique condition under which the probability distribution P_N of the extended model goes to that of the classical $M/M/1$ queuing model. Finally, it is valid to propose group theory in queuing system performance analysis especially where heterogeneity is part and parcel of the system.

5.2 Recommendations:

For the purpose of extending this research work, it is recommended that:

1. Another class of constraint be studied via the group theory methodology presented in this work.
2. Performance analysis and derivations covering multi server queuing systems via group theory analysis will be an interesting research.
3. Extension of the group theory analysis to data modeling will be an interesting research either.

5.3 Conclusion:

Result obtained in this research work can be applied in service systems with homogeneous or heterogeneous customer structures for reasons to do with optimality.

Chapter 6

Appendices

6.1 Appendix-1

In the queuing literature, it is well known for N-homogenous jobs that the stationary probability of maintaining these jobs in a uni-server production center is given by

$$P_N = (1-\rho)\rho^N \tag{6.1}$$

The parameter ρ is the occupation rate of the center; Medhi [16]

6.2 Appendix-2

Lemma 6.2.1 *Given* $\{N\} = \{c\} \oplus \{n\}; \{c\} \cap \{n\} = \{\}$; *then*

$$E[N] = \frac{\rho^c[\rho + c(1-\rho)]}{1-\rho}. \tag{6.2}$$

6.3 Appendix-3

Lemma 6.3.1 *For a finite capacity imperfect production center with two distinct job groups* $\{n\}$ *and* $\{c\}$*, we have*

$$E[N] = \frac{\rho^c\left[(1-\rho^K)(c(1-\rho)+\rho) - K(1-\rho)\rho^K\right]}{1-\rho} \tag{6.3}$$

Bibliography

[1] A. P. Pant & R. P. Ghimire (2015). M(t)/M/1 Queueing System with Sinusoidal Arrival Rate. *Journal of the Institute of Engineering,* 11(1), 120-127.

[2] A. Willig (2005). Performance evaluation techniques - summer 2004. *http://www.tkn.tu-berlin.de /-awillig/user*$_i$*ncludes/Pet*$_s$*s*2005/*skript.pdf*, $20-04-2017$ *at* $22:15$ *Hrs*.

[3] B. Krishnarmoothi (1963). On Poisson queue with Two Herterogenuos service. *Operation Research,* 11(3), 321-330.

[4] C. Maglaras (2006). Revenue Management for a Multiclass Single-Server Queue via a Fluid Model Analysis *OPERATIONS RESEARCH,* 54(5), 914-932.

[5] C. Maglaras & J. A. V. Mieghem (2004). Queuing systems with lead-time constraints: A fluid-model approach for admission and sequencing control. *European Journal of Operational,* 167, 179-207.

[6] C. Wu & Y. Cheng & P. Tang & J. Y. Yu (2012). OPTIMAL BATCH PROCESS ADMISSION CONTROL IN TANDEM QUEUEING SYSTEMS WITH QUEUE TIME CONSTRAINT CONSIDERATIONS. *Proceedings of the 2012 Winter Simulation Conference,* 1-6. 9-12-2012.

[7] D. Gross & C. M. Harris (1998). Fundamentals of queuing theory. *Third edition*, New York: John Wiley Sons.

[8] D. Sumitha & K. U. Chandrika (2014). $\{M^X/G/1$ Retrial queue with second optional service, feedback, admission control and vacation. *International Journal of Innovative Research in Science, Engineering and Technology,* 3(1), 8320-8329.

[9] E.Abdolghani (2015). Rate control of a queue with quality-of-service constraint under bounded and unbounded action spaces. *A Master of science thesis, Iowa State University, Ames, Iowa,* 1-39.

[10] F. G. Hillier & G. J. Lieberman (2005). Introduction to operations research. *Eighth edition*, New York: McGraw-Hill.

[11] G. Ayyappan & S. Shyamala (2014). Transient behavior of $\{M^X/G/1$ Retrial Queuing Model with Non Persistent Customers, Random break down, Delaying Repair and Bernoulli Vacation. *International Journals of Mathematical Trends and Technology,* 7(2), 126-135.

[12] G. Noah Y. Zhou (2003). CALL-ROUTING PROBLEM WITH SERVICE-LEVEL CONSTRAINTS. *Operations Research,* 51(2), 255-271.

[13] H. M. Barakat & Y. H. Abdelkader (2004). Computing the moments of order statistics from non identical random variables. Statistical Methods Applications. *Statistical Methods& Applications,* 13, 15-26 . block On the Single Server Retrial Queue with Batch Arrivals.

[14] I. J. B. F. Adan & V. G. Kulkarni (2003). Single-server queue with Markov dependent inter-arrival and service times. *Queuing Systems,* 45(2), 113-134.

[15] J. ARTALEJO & G. FALIN (2002). STANDARD AND RETRIAL QUEUING SYSTEMS: A COMPARATIVE ANALYSIS. *REVISTA MATEMATICA COMPLUTENSE,* 15(1), 101-129.

[16] J. Medhi (2003). Stochastic models in queuing theory. *Second edition,* Carlifonia press Inc.

[17] J. N. Tsitsiklis & K. Xu (2013). Queueing System Topologies with Limited Flexibility. *Sigmetrics, Pittsburgh, Performance Evaluation Reveiw,* 41(1), 167-178.

[18] J. R. Artalejo & M. J. L. Herrero (2004). The M/G/1 retrial queue: An information theoretic approach. *SORT, http://hdl.handle.net/2099/3755,* 29(1), 119-138.

[19] K. H. Wang & S. L. Chuang & W. L. Pearnba (2001). Maximum entropy analysis to the N policy M/G/1 queuing system with a removable server. *Applied Mathematical Modeling,* 26, 1151-1162.

[20] K. S. Laghaie & M. S. Mehrabad & A. M. Larijani (2011). Optimization Models for a Deteriorating Single Server Queuing Production System. *International Journal of Industrial Engineering Production Research,* 22, 251-260.

[21] L. Chih-ping & M. J. Neely (2012). Delay and Rate-Optimal Control in a Multi-Class Priority Queue with Adjustable Service Rates. *Proceedings IEEE INFOCOM-12,* 2976-2980.

[22] M. Haviv (2009). Customer behavior in a single server queue.

[23] M. Herlicha & N. Bredenbalsa & H. Karla (2013). Average and Competitive Analysis of Latency and Power consumption of a Queuing System with a Sleep Mode. *In Proceedings of the International Conference on Future Energy Systems,* 14, 110.

[24] M. Jeeva & E. Rathnakumari (2014). Steady State Analysis of Repairable M/G/1 Retrial Queuing Model with Modified Server Vacation Balking and Optional Re-service. Journal of Mathematics, 10(3), 12-20.

[25] M. K. Girish & J. Hu (1995). Higher order approximations for tandem queuing networks. *Questa,Technical Report, Dept. of Manufacturing Engineering, Boston University (April 1995),* 22(3), 249-276.

[26] M. Mohammad & H. K. Fakher (2009). Traffic Flow Analysis Based on Queuing Models. *Journal of Industrial Engineering,* 2, 19-23.

[27] O.J Boxma, Q. Deng and A.P Zwart (2002). Waiting time asymtotics for the M/G/2 queue with heterogenous servers. *Queueing Systems*, 40:5–31.

[28] P. B. KUMAR (2013). QUEUING THEORY AND CUSTOMER SATISFACTION: A REVIEW OF TERMINOLOGY, TRENDS, AND APPLICATIONS TO HOSPITAL PRACTICE. *Asia Pacific Journal of Marketing Management Review,* 2(6), 2319-2336.

[29] P. T. de-Boer & D. P. Kroese & R. Y. Rubinstein (2004). A Fast Cross-Entropy Method for Estimating Buffer Overflows in Queuing Networks. *MANAGEMENT SCIENCE,* 50(7), 883-895.

[30] R. Krenzler & H. Daduna (2013). Loss systems in a random environment. *http://preprint.math.uni-hamburg.de/public/papers/prst/prst2013/-02.pdf,* May 2013.

[31] R. Kumar & N. J. Kumar & B. S. Kumar (2014). OPTIMIZATION OF AN M/M/1/N FEEDBACK QUEUE WITH RETENTION OF RENEGED CUSTOMERS. *Advanced Modeling and Optimization-AMO,* 14(3), 681-668. .

[32] R. Kumar & M. E. Lewis & H. Topaloglu (2011). Dynamic Service Rate Control for a Single Server Queue with Markov Modulated Arrivals. *Naval research logistics,* 60(8), 661-677.

[33] R. Sundaresan (2006). Capacity of Queues Via Point-Process Channels. *IEEE TRANSACTIONS ON INFORMATION THEORY,* 52(6), 1-13.

[34] S. Kapodistria (2011). The M/M/1 Queue with Synchronized Abandonments. *Queuing Systems,* 68(1), 79-109.

[35] S. Khew Koh & A. H. Pooi & Y. F. Tan (2014). New Approach for Finding Basic Performance Measures of Single Server Queue. *Hindawi Publishing Corporation Journal of Probability and Statistics, http://dx.doi.org/10.1155/2014/851738,* 1-11.

[36] S. R. C. Mohammad (2013). QUEUING THEORY MODEL USED TO SOLVE THE WAITING LINE OF A BANK -A STUDY ON ISLAMI BANK BANGLADESH LIMITED,CHAWKBAZAR BRANCH, CHITTAGONG. *ASIAN JOURNAL OF SOCIAL SCIENCES* & *HUMANITIES,* 2(3), 468-478.

[37] S. Sani, H. Musa and M.L. Malumfashi (2017). Optimal Job Creation Strategy In the Presence of $A-$-Type of Constraints. *Applied Mathematics, To Appear.*

[38] . S. W. Fuhrmann & R. B. Cooper (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. *Operations Research,* 33, 1117-1129.

[39] T. Poongodi & S. Muthulakshmi (2013). Control Chart for Waiting Time in System of (M/M/1): $(\infty/FCFS)$ Queuing Model. Journal of Computer Applications 63(4), 1-6.

[40] U. Ayestaa & O. J. Boxmac & I.M. Verloopb. Sojourn times in a processor sharing queue with multiple vacations. *Queuing Systems,* 71(1), 53-78.

[41] V. D. Natalia (2006). ON THE SINGLE-SERVER RETRIAL QUEUE. *Yugoslav Journal of Operations Research,* 1, 45-53.

[42] V. Thangaraj & S. Vanitha (2010). A Single Server M/G/1 Feedback Queue with Two Types of Service Having General Distribution. *International Mathematical Forum,* 1, 15-33.

[43] W. Whitt (2016). Recent Papers on the Time-Varying Single-Server Queue. *http://www.columbia.edu/ ww2040/allpapers.html,* 18-September-2016, 1-4.

[44] X. Fan-Orzechowski & E. A. Feinberg (2005). Optimal Admission Control for a Markovian Queue Under the Quality of Service Constraint. *Proceedings of the* 44^{th} *IEEE Conference on Decision and Control, and the European Control Conference 2005 Seville,* 12-15.

[45] X. Guo, Y. Lu & M. S. Squillate (2003). Optimal stochastic routing in distributed parallel queues. Technical report, IBM Research Division, November 2013, Revised, April 2014, 53-54.

[46] Y. Song & Member & IACSIT (2014). The Optimal Service Policies in an M/G/1 Queue with Consecutive Vacations. *International Journal of Modeling and Optimization,* 4(2), 100-103.

100