PERFORMANCE CHARACTERIZATION OF TWO-SERVER BATCH SERVICE QUEUE WITH SECOND OPTIONAL SERVICE

Andwilile Abrahamu George

•

Department of Mathematics and ICT, College of Business Education,
Dodoma, Tanzania.
gandwilile@gmail.com

P. Vijaya Laxmi

•

Department of Applied Mathematics, Andhra University,
Visakhapatnam, India.

Abstract

In this paper, we analyze the performance of a finite capacity two-server Markovian batch service queueing model with second optional service. The servers provide two kinds of services, the first essential service (FES), which is provided to all incoming customers and the second optional service (SOS) to those who demand it after completing FES. The service times of the two servers are identical and are exponentially distributed. Matrix-decomposition method is used to obtain the steady-state probabilities of the model. Numerical results and discussion are presented to demonstrate the impact of the model parameters on the system behavior. Furthermore, the cost model optimization is developed to determine the optimal service rates using the Quasi-Newton method to minimize the expected cost. Finally, the findings of this work show that the blocking probability is monotonically decreases as finite buffer size increases and approaches its minimum value of zero when finite buffer is sufficiently large.

Keywords: Two-server, Batch service queue, First essential service, Second optional service.

Mathematical Subject Classification: 60k25, 90B22.

I. Introduction

In everyday life, there are various queueing circumstances where all incoming customers demand the FES and only some may demand the SOS provided by the same server. Currently, this has taken a major consideration by various researchers such as Wang and Kuki [23] wherein they analysed the performance of retrial queueing system with SOS. They obtained the queue length, waiting time and busy period using the method of discrete transformation. SOS with a single server fixed batch service queueing system during repeated vacations has been studied by Ayyappan et al., [3]. They analyzed the model using the probability generating function and Rouche's theorem to obtain the probability of the number of customers present in the queue while the server is busy or on vacation.

Multi-server retrial queue with SOS has been presented by Ke et al., [13]. They derived stationary probabilities using matrix analytic approach. An $M/M/1$ queue with SOS and working breakdown has been analysed by Yang and Chen [24] who derived the stationary

probability distribution of the system size using the matrix geometric method. Other extensive studies conducted on an assortment of queueing models with SOS are found in Gupur [10]; Kalyanaraman and Murugan [12]; Ke et al., [14]; Uma and Punniyamoorthy [22], etc.
For the most part, customers get the service individually. However, this rule may not work in all circumstances, since in some places, the server provide service in batch (groups) of customers instead of serving individually. Batch services are more useful in telecommunication, where data bundle is transmitted in the accumulated large entities (batches), in the field of transportation, in a manufacturing system, in a smart city crowdsourcing application for mobile, etc.

Batch service queue has been broadly studied by numerous researchers. Goswami and Samanta [8] presented the two heterogeneous servers with a discrete-time bulk service queueing system and derived a closed-form expression of the stationary probabilities at arbitrary epoch. The manufacturing bulk service queues using Bayesian hierarchical models are presented by Armero and Conesa [2]. They developed the inferential method for the parameter of the governing model using hierarchical models. Barbhuiya and Guptal [6] analysed the $GI/M^Y/1$ queue, wherein the closed form explicit formulations of the system distribution at pre arrival and arbitrary epochs in terms of roots of the characteristic equation generated were derived using the supplementary variable approach and the difference equation method. The queue for bulk service with delayed vacation has been analyzed by Krishna Reddy and Anitha [15]. They obtained stationary probabilities and waiting time distribution of an incoming customer. Ayyappan et al., [4] studied the single server fixed batch service queueing system. The server serves exactly k items at a time. If he finds less than k items in the queue, he takes a multiple vacation of length a. Batch service with multi-server queue model has been presented by Ghimire et al., [7], wherein they computed the stationary probability, queue length size and the amount of waiting time in the queue.

A Markovian single server queueing system has been widely studied by numerous researchers for the last few decades. Due to increasing demand for services, single server operations can lead to congestion in the system such as manufacturing and production, medical clinics, in telecommunication system, etc. Multi-servers have also been studied by various researchers to reduce congestion, Hwang et al., [11]. The queueing system with two servers that are heterogeneous and a variant vacation policy has been studied by Yue and Tian [25], where the two servers take vacations together whenever the system becomes empty. They obtained the explicit expressions of the stationary distribution of the system size. Krishnamoorthy and Sreenivasan [16] investigated a queueing system using two heterogeneous servers, one of which is regularly accessible while the other goes on vacation whenever there are no customers waiting for service and service times of the two servers are exponentially distributed with different service rates. They analyzed the model using matrix geometric method to obtain the mean waiting time in the steady state regime. Kumar et al., [17] investigated the two server queue with heterogeneous servers that were vulnerable to catastrophes wherein any arriving item requires exactly one server for the service and the service rates are different. They obtained both transient and stationary probabilities of number of items in the system using probability generating function and the modified Bessel function of the first kind. Similar studies are found in Ammar [1] and Kumar and Sharma [19] with impatient behavior wherein the service times are independently and exponentially distributed with different service rate. They obtained the explicit transient and steady state probabilities by using probability generating function. Recently, Kumar et al.,[18] generalized the work of Kumar and Sharma [19] by introducing the concept of feedback and reverse balking. They used an iterative approach to obtain the probabilistic measures of the model.

Queues with limited waiting space are more realistic in real-life circumstances. For example, routers servicing incoming packets with varying speeds in a network are examples of this scenario. If the server is busy, the incoming items wait in the queue, but the incoming items are deemed lost when the waiting space is full. One of the major considerations of a system designer in this situation is to provide enough buffer space to keep the blocking probability to a minimum. The limited buffer size queues have been researched by various authors such as Sikdar and Gupta [21], who analyzed a queue for batch arrival and batch service with a limited buffer size with single and repeated vacations. They obtained stationary probability distributions of the number of items in the queue at departure and pre-arrival epochs. Moreover, they presented the average queue length, waiting time and the blocking probability as the performance measures. Gupta et al.,[9] analyzed a queue for bulk services on a single server with finite buffers and vacation. Using the supplementary variable and embedded Markov chain methods, they obtained the stationary joint distribution of the queue length and the kind of vacation that the server has taken at different epochs. Banerjee [5] presented a queue with a limited capacity and workload-dependent service. They used the embedded Markov chain technique and the supplementary variable approach to obtain stationary probabilities at a departure and arbitrary epoch.

In view of the growing demands for service in practice, our objective here is to investigate the performance of finite buffer two-server batch queue with SOS. In existing literature reviews, no work has been reported with a combination of two-server batch queue with SOS. This inspires us to investigate a two-server batch queue with SOS. Consideration of SOS with varying batch size service makes the model more versatile. The main contributions of this paper are as follows.

- First, we derive the stationary probabilities by using matrix-decomposition method.
- Second, we develop a cost function to optimize service rates for both FES and SOS using Quasi-Newton method so that the expected total cost is minimized.

The remaining part of the paper is structured as follows: Mathematical model description and its formulation are presented in Section 2. In Section 3, we discuss steady state probabilities where the servers are busy or idle in both FES and SOS. The performance measures are discussed in Section 4. The presentation of numerical analysis and discussion is in Section 5. The cost model optimization is developed in Section 6 and in Section 7, we conclude the paper.

II. Model Description and Mathematical Formulation

We consider a finite buffer $M/M^b/2/N$ queue model with SOS, where arrival occurs according to a Poisson process with parameter λ. The servers provide two kinds of service, FES, which is provided to all incoming items and SOS to those whose demand it after completing FES. The item opts the SOS with probability r $(0 \leq r \leq 1)$ or leave the system after completion FES with the probability $(1-r)$. Furthermore, It is assumed that the service times of two servers are independently, identically and exponentially distributed, each with service rate μ_1 in FES and μ_2 in SOS. The queue has limited waiting space of size N. The arriving customers are blocked from entering the queue whenever the queue size is N. The average rate of service for FES is given by

$$\mu_{i1} = \begin{cases} \mu_1, & \text{for } i = 1, \\ 2\mu_1, & \text{for } i = 2. \end{cases}$$

The average rate of service for SOS is given by

$$\mu_{j2} = \begin{cases} \mu_2, & \text{for } j = 1, \\ 2\mu_2, & \text{for } j = 2, \end{cases}$$

where i and j are the number of servers in FES and SOS, respectively.
The servers process a partial batch up to a maximum capacity size of b. If there are less than b in the queue, one of the servers begins service to those customers. If there are more than $2b$ waiting in the queue, then the servers take a batch of size b each based on the order of arrival, while others in excess of $2b$ customers, wait in the queue.

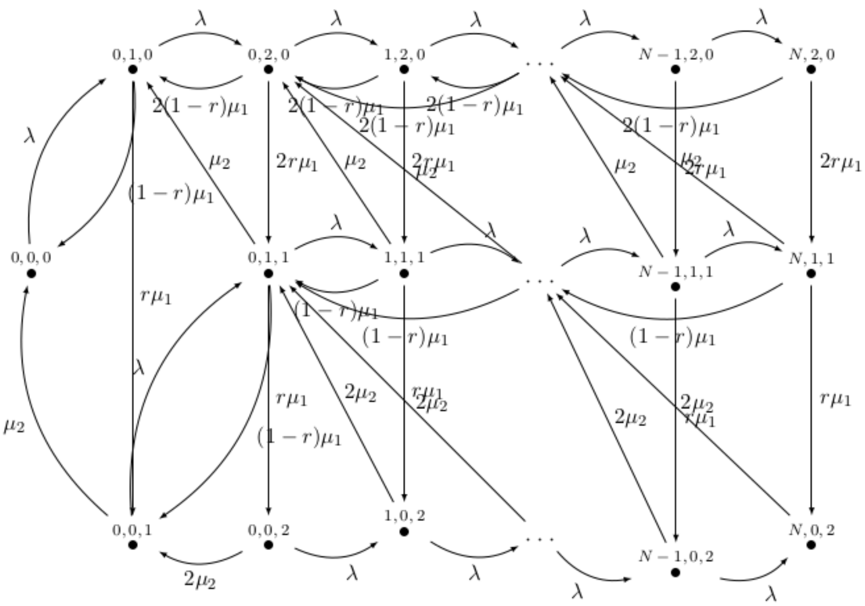

Figure 1: *Transition rate diagram for* $b = 2$

Figure 1 is the state transition diagram showing the various transition states of the queueing model.

I. Formulation of Mathematical Model

At time $t \geq 0$, let $N(t)$ be the number of items in the queue, $S(t)$ be the number of servers in FES and $J(t)$ be the number of servers in SOS, where $S(t)$ and $J(t)$ are defined as follows:

$$S(t) = \quad i, \qquad 0 \leq i \leq 2, \;\text{ and } J(t) = \quad j, \qquad 0 \leq j \leq 2.$$

We observe that, $\{(N(t), S(t), J(t)); t \geq 0\}$ defines a three dimensional Markov process in continuous time with state space

$$E = \{(n, i, j);\ 0 \leq n \leq N;\ 0 \leq i \leq 2;\ 0 \leq j \leq 2,\ i + j \leq 2\}.$$

Let the transient probabilities are defined as

$$P_{n,i,j}(t) = Pr\{N(t) = n,\ S(t) = i,\ J(t) = j\},$$
$$0 \leq n \leq N;\ 0 \leq i \leq 2;\ 0 \leq j \leq 2,\ i + j \leq 2,$$

where $P_{0,i,j}(t)$ is the transient probability that i and j servers are busy with FES or SOS with no customer waiting in the queue, where $0 \leq i \leq 2,\ 0 \leq j \leq 2,\ i + j \leq 2$.
$P_{n,i,j}(t)$ is the transient probability that all servers are busy with FES or SOS with n customer waiting in the queue, where $0 \leq i \leq 2,\ 0 \leq j \leq 2,\ i + j = 2,\ 1 \leq n \leq N$.
The above set of probabilities at steady state are denoted by $P_{0,i,j}$ and $P_{n,i,j}$, respectively. The following set of equations are obtained using the probabilistic arguments.

$$\lambda P_{0,0,0} = (1-r)\mu_1 P_{0,1,0} + \mu_2 P_{0,0,1}, \tag{1}$$
$$(\lambda+\mu_1)P_{0,1,0} = \lambda P_{0,0,0} + 2(1-r)\mu_1 P_{0,2,0} + \mu_2 P_{0,1,1}, \tag{2}$$
$$(\lambda+2\mu_1)P_{0,2,0} = \lambda P_{0,1,0} + 2(1-r)\mu_1 \sum_{i=1}^{b} P_{i,2,0} + \mu_2 \sum_{i=1}^{b} P_{i,1,1}, \tag{3}$$
$$(\lambda+2\mu_1)P_{n,2,0} = \lambda P_{n-1,2,0} + 2(1-r)\mu_1 P_{n+b,2,0} + \mu_2 P_{n+b,1,1},\ 1 \le n \le N-b, \tag{4}$$
$$(\lambda+2\mu_1)P_{n,2,0} = \lambda P_{n-1,2,0},\quad N-b+1 \le n \le N-1, \tag{5}$$
$$2\mu_1 P_{N,2,0} = \lambda P_{N-1,2,0}, \tag{6}$$
$$(\lambda+\mu_2)P_{0,0,1} = (1-r)\mu_1 P_{0,1,1} + 2\mu_2 P_{0,0,2} + r\mu_1 P_{0,1,0}, \tag{7}$$
$$(\lambda+2\mu_2)P_{0,0,2} = r\mu_1 P_{0,1,1}, \tag{8}$$
$$(\lambda+2\mu_2)P_{n,0,2} = \lambda P_{n-1,0,2} + r\mu_1 P_{n,1,1},\quad 1 \le n \le N-1, \tag{9}$$
$$2\mu_2 P_{N,0,2} = \lambda P_{N-1,0,2} + r\mu_1 P_{N,1,1}, \tag{10}$$
$$(\lambda+\mu_1+\mu_2)P_{0,1,1} = \lambda P_{0,0,1} + (1-r)\mu_1 \sum_{i=1}^{b} P_{i,1,1} + 2\mu_2 \sum_{i=1}^{b} P_{i,0,2} + 2r\mu_1 P_{0,2,0}, \tag{11}$$
$$(\lambda+\mu_1+\mu_2)P_{n,1,1} = \lambda P_{n-1,1,1} + (1-r)\mu_1 P_{n+b,1,1} + 2\mu_2 P_{n+b,0,2} + 2r\mu_1 P_{n,2,0},\quad 1 \le n \le N-b, \tag{12}$$
$$(\lambda+\mu_1+\mu_2)P_{n,1,1} = \lambda P_{n-1,1,1} + 2r\mu_1 P_{n,2,0},\quad N-b+1 \le n \le N-1, \tag{13}$$
$$(\mu_1+\mu_2)P_{N,1,1} = \lambda P_{N-1,1,1} + 2r\mu_1 P_{N,2,0}. \tag{14}$$

The steady state equations (1)-(14) can be represented as matrix-form

$$\mathbf{PQ=0}, \tag{15}$$

where $\mathbf{0}$ denotes the column vector with all elements equal to zero, and

$$\mathbf{Q} = \begin{pmatrix} -\lambda & \mathbf{M}_{12} & \mathbf{M}_{13} & 0 & \mathbf{M}_{15} & \mathbf{M}_{16} & \mathbf{M}_{17} \\ \mathbf{M}_{21} & \mathbf{M}_{22} & \mathbf{M}_{23} & \mathbf{M}_{24} & \mathbf{M}_{25} & \mathbf{M}_{26} & \mathbf{M}_{27} \\ \mathbf{M}_{31} & \mathbf{M}_{32} & \mathbf{M}_{33} & \mathbf{M}_{34} & \mathbf{M}_{35} & \mathbf{M}_{36} & \mathbf{M}_{37} \\ 0 & \mathbf{M}_{42} & \mathbf{M}_{43} & \mathbf{M}_{44} & \mathbf{M}_{45} & \mathbf{M}_{46} & \mathbf{M}_{47} \\ \mathbf{M}_{51} & \mathbf{M}_{52} & \mathbf{M}_{53} & \mathbf{M}_{54} & \mathbf{M}_{55} & \mathbf{M}_{56} & \mathbf{M}_{57} \\ \mathbf{M}_{61} & \mathbf{M}_{62} & \mathbf{M}_{63} & \mathbf{M}_{64} & \mathbf{M}_{65} & \mathbf{M}_{66} & \mathbf{M}_{67} \\ \mathbf{M}_{71} & \mathbf{M}_{72} & \mathbf{M}_{73} & \mathbf{M}_{74} & \mathbf{M}_{75} & \mathbf{M}_{76} & \mathbf{M}_{77} \end{pmatrix}$$

is a $(3N+6)\times(3N+6)$ square matrix. The entries of the matrix $\mathbf{Q}$ are listed below:

$$\mathbf{M}_{12} = \begin{pmatrix} \lambda & 0 \end{pmatrix}_{1\times 2},\ \mathbf{M}_{13} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}_{1\times N},\ \mathbf{M}_{15} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}_{1\times N},$$

$$\mathbf{M}_{27} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{2\times N},\ \mathbf{M}_{17} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}_{1\times N},\ \mathbf{M}_{21} = \begin{pmatrix} (1-r)\mu_1 \\ 0 \end{pmatrix}_{2\times 1},$$

$$\mathbf{M}_{23} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \lambda & 0 & \cdots & 0 \end{pmatrix}_{2\times N},\ \mathbf{M}_{24} = \begin{pmatrix} 0 \\ 2r\mu_1 \end{pmatrix}_{2\times 1},\ \mathbf{M}_{25} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{2\times N},$$

$$\mathbf{M}_{16} = \begin{pmatrix} 0 & 0 \end{pmatrix}_{1\times 2},\ \mathbf{M}_{26} = \begin{pmatrix} r\mu_1 & 0 \\ 0 & 0 \end{pmatrix}_{2\times 2},\ \mathbf{M}_{31} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{N\times 1},\mathbf{M}_{32} = \begin{pmatrix} 0 & r_{i2} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{pmatrix}_{N\times 2},$$

$$\mathbf{M}_{51} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{N\times 1},\ \mathbf{M}_{33} = \begin{pmatrix} \alpha & \lambda & 0 & \cdots & 0 & 0 \\ A_{10} & \alpha & \lambda & \cdots & 0 & 0 \\ A_{20} & A_{21} & \alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ A_{N-20} & A_{N-21} & A_{N-22} & \cdots & \alpha & \lambda \\ A_{N-10} & A_{N-11} & A_{N-12} & \cdots & A_{N-1N-2} & -2\mu_1 \end{pmatrix}_{N\times N},$$

where

$$\alpha = -(\lambda + 2\mu_1),\ r_{i2} = 2(1-r)\mu_1,\ \text{for } i = 1,\ 2,\ \cdots, b,\ \text{ and } b \le N,$$

$$A_{ij} = \begin{cases} 2(1-r)\mu_1, & \text{if } \frac{i-j}{b} = 1 \text{ for } 1 \le i \le N-1,\ 0 \le j \le N-2,\ b \le N, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{M}_{34} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{N\times 1},\ \mathbf{M}_{35} = \begin{pmatrix} 2r\mu_1 & 0 & \cdots & 0 \\ 0 & 2r\mu_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2r\mu_1 \end{pmatrix}_{N\times N},\ \mathbf{M}_{44} = -(\lambda + \mu_1 + \mu_2),$$

$$\mathbf{M}_{37} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{N\times N},\ \mathbf{M}_{42} = \begin{pmatrix} \mu_2 & 0 \end{pmatrix}_{1\times 2},\ \mathbf{M}_{43} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}_{1\times N},$$

$$\mathbf{M}_{45} = \begin{pmatrix} \lambda & 0 & \cdots & 0 \end{pmatrix}_{1\times N},\ \mathbf{M}_{46} = \begin{pmatrix} (1-r)\mu_1 & r\mu_1 \end{pmatrix}_{1\times 2},\ \mathbf{M}_{47} = \begin{pmatrix} 0 & 0 & \cdots & 0 \end{pmatrix}_{1\times N},$$

$$\mathbf{M}_{52} = \begin{pmatrix} 0 & s_{i2} \\ \vdots & \vdots \\ 0 & 0 \\ 0 & 0 \end{pmatrix}_{N\times 2},\ \mathbf{M}_{53} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ B_{10} & 0 & 0 & \cdots & 0 & 0 \\ B_{20} & B_{21} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ B_{N-20} & B_{N-21} & B_{N-22} & \cdots & 0 & 0 \\ B_{N-10} & B_{N-11} & B_{N-12} & \cdots & B_{N-1N-2} & 0 \end{pmatrix}_{N\times N},$$

where

$$s_{i1} = (1-r)\mu_1 \text{ and } s_{i2} = \mu_2,\ \text{for } i = 1,\ 2,\ \cdots, b,\ \text{ and } b \le N,$$

$$B_{ij} = \begin{cases} \mu_2, & \text{if } \frac{i-j}{b} = 1 \text{ for } 1 \le i \le N-1,\ 0 \le j \le N-2,\ b \le N, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{M}_{55} = \begin{pmatrix} \lambda' & \lambda & 0 & \cdots & 0 & 0 \\ C_{10} & \lambda' & \lambda & \cdots & 0 & 0 \\ C_{20} & C_{21} & \lambda' & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{N-20} & C_{N-21} & C_{N-22} & \cdots & \lambda' & \lambda \\ C_{N-10} & C_{N-11} & C_{N-12} & \cdots & C_{N-1N-2} & -(\mu_1+\mu_2) \end{pmatrix}_{N\times N},\ \mathbf{M}_{61} = \begin{pmatrix} \mu_2 \\ 0 \end{pmatrix}_{2\times 1},$$

where $\lambda' = -(\lambda + \mu_1 + \mu_2)$.

$$C_{ij} = \begin{cases} (1-r)\mu_1, & \text{if } \frac{i-j}{b} = 1 \text{ for } 1 \le i \le N-1,\ 0 \le j \le N-2,\ b \le N, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{M}_{22} = \begin{pmatrix} -(\lambda+\mu_1) & \lambda \\ 2(1-r)\mu_1 & -(\lambda+2\mu_1) \end{pmatrix}_{2\times 2},\ \mathbf{M}_{57} = \begin{pmatrix} r\mu_1 & 0 & \cdots & 0 \\ 0 & r\mu_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r\mu_1 \end{pmatrix}_{N\times N},$$

$$\mathbf{M}_{62} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}_{2\times 2},\ \mathbf{M}_{63} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{2\times N},\ \mathbf{M}_{67} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \lambda & 0 & \cdots & 0 \end{pmatrix}_{2\times N},$$

$$\mathbf{M}_{65} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{2\times N}, \mathbf{M}_{66} = \begin{pmatrix} -(\lambda+\mu_2) & 0 \\ 2\mu_2 & -(\lambda+2\mu_2) \end{pmatrix}_{2\times 2}, \mathbf{M}_{64} = \begin{pmatrix} \lambda \\ 0 \end{pmatrix}_{2\times 1},$$

$$\mathbf{M}_{71} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{N\times 1}, \mathbf{M}_{72} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{N\times 2}, \mathbf{M}_{73} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{N\times N},$$

$$\mathbf{M}_{74} = \begin{pmatrix} w_{i1} \\ \vdots \\ 0 \\ 0 \end{pmatrix}_{N\times 1}, \mathbf{M}_{75} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ D_{10} & 0 & 0 & \cdots & 0 & 0 \\ D_{20} & D_{21} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ D_{N-20} & D_{N-21} & D_{N-22} & \cdots & 0 & 0 \\ D_{N-10} & D_{N-11} & D_{N-12} & \cdots & D_{N-1N-2} & 0 \end{pmatrix}_{N\times N},$$

where

$$w_{i1} = 2\mu_2, \text{ for } i = 1,\ 2,\ \cdots, b, \text{ and } b \le N,$$

$$D_{ij} = \begin{cases} 2\mu_2, & \text{if } \frac{i-j}{b} = 1 \text{ for } 1 \le i \le N-1,\ 0 \le j \le N-2,\ b \le N, \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathbf{M}_{54} = \begin{pmatrix} s_{i1} \\ \vdots \\ 0 \\ 0 \end{pmatrix}_{N\times 1}, \mathbf{M}_{36} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{N\times 2}, \mathbf{M}_{76} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{N\times 2}, \mathbf{M}_{56} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}_{N\times 2},$$

$$\mathbf{M}_{77} = \begin{pmatrix} -(\lambda+2\mu_2) & \lambda & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda+2\mu_2) & \lambda & \cdots & 0 & 0 \\ 0 & 0 & (-\lambda+2\mu_2) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda+2\mu_2) & \lambda \\ 0 & 0 & 0 & \cdots & 0 & -2\mu_2 \end{pmatrix}_{N\times N},$$

In the following sequel, we use a matrix decomposition approach to obtain the stationary probabilities in a recursive manner. Let $P_{0,0,0}$, P_1, P_2, $P_{0,1,1}$, P_3, P_4, P_5 be the vectors of stationary probabilities where $P_1 = (P_{0,1,0},\ P_{0,2,0})$, $P_2 = (P_{1,2,0},\ P_{2,2,0}, \cdots,\ P_{N,2,0})$, $P_3 = (P_{1,1,1},\ P_{2,1,1}, \cdots,\ P_{N,1,1})$, $P_4 = (P_{0,0,1},\ P_{0,0,2})$, $P_5 = (P_{1,0,2},\ P_{2,0,2}, \cdots,\ P_{N,0,2})$.

From equation (15) it follows that

$$-\lambda P_{0,0,0} + P_1 M_{21} + P_4 M_{61} = 0, \tag{16}$$
$$P_{0,0,0} M_{12} + P_1 M_{22} + P_2 M_{32} + P_{0,1,1} M_{42} + P_3 M_{52} = 0, \tag{17}$$
$$P_1 M_{23} + P_2 M_{33} + P_3 M_{53} = 0, \tag{18}$$
$$P_1 M_{24} - P_{0,1,1} M_{44} + P_3 M_{54} + P_4 M_{64} + P_5 M_{74} = 0, \tag{19}$$
$$P_2 M_{35} + P_{0,1,1} M_{45} + P_3 M_{55} + P_5 M_{75} = 0, \tag{20}$$
$$P_1 M_{26} + P_{0,1,1} M_{46} + P_4 M_{66} = 0, \tag{21}$$
$$P_3 M_{57} + P_4 M_{67} + P_5 M_{77} = 0. \tag{22}$$

Equation (21) yields

$$P_4 = P_1 \Phi_1 + P_{0,1,1} \Phi_2, \tag{23}$$

where $\Phi_1 = -\mathbf{M}_{26}\mathbf{M}_{66}^{-1}$ and $\Phi_2 = -\mathbf{M}_{46}\mathbf{M}_{66}^{-1}$.
Using equation (23) into equation (16), we obtain

$$P_{0,0,0} = \frac{1}{\lambda}\left[P_1\Phi_3 + P_{0,1,1}\Phi_4\right], \tag{24}$$

where $\Phi_3 = \mathbf{M}_{21} + \Phi_1\mathbf{M}_{61}$ and $\Phi_4 = \Phi_2\mathbf{M}_{61}$.
From equation (22), we obtain

$$P_5 = P_3\Phi_5 + P_1\Phi_6 + P_{0,1,1}\Phi_7, \tag{25}$$

where $\Phi_5 = -\mathbf{M}_{57}\mathbf{M}_{77}^{-1}$, $\Phi_6 = -\Phi_1\mathbf{M}_{67}\mathbf{M}_{77}^{-1}$ and $\Phi_7 = -\Phi_2\mathbf{M}_{67}\mathbf{M}_{77}^{-1}$.
Substituting equation (25) into equation (20), we get

$$P_3 = (P_2\Phi_8 + P_{0,1,1}\Phi_9 + P_1\Phi_{10})\Phi_{11}^{-1}, \tag{26}$$

where $\Phi_8 = -\mathbf{M}_{35}$, $\Phi_9 = -(\mathbf{M}_{45} + \Phi_7\mathbf{M}_{75})$, $\Phi_{10} = -\Phi_6\mathbf{M}_{75}$
and $\Phi_{11} = \mathbf{M}_{55} + \Phi_5\mathbf{M}_{75}$.
Using equation (26) into equation (18), we get

$$P_2 = (P_1\Phi_{12} + P_{0,1,1}\Phi_{13})\Phi_{14}^{-1}, \tag{27}$$

where $\Phi_{12} = -(\mathbf{M}_{23} + \Phi_{10}\mathbf{M}_{53}\Phi_{11}^{-1})$, $\Phi_{13} = -\Phi_9\mathbf{M}_{53}\Phi_{11}^{-1}$
and $\Phi_{14} = \mathbf{M}_{33} + \Phi_8\mathbf{M}_{53}\Phi_{11}^{-1}$.
Using equations (27) into equation (26), we have

$$P_3 = P_1\Phi_{15} + P_{0,1,1}\Phi_{16}, \tag{28}$$

where $\Phi_{15} = \Phi_{12}\Phi_{14}^{-1}\Phi_8\Phi_{11}^{-1} + \Phi_{10}\Phi_{11}^{-1}$ and $\Phi_{16} = \Phi_{13}\Phi_{14}^{-1}\Phi_8\Phi_{11}^{-1} + \Phi_9\Phi_{11}^{-1}$.
Substituting equations (24), (27) and (28) into equation (17), we have

$$\begin{aligned} P_1 &= P_{0,1,1}\Phi_{17}, \qquad (29)\\ \text{and } \Phi_{17} &\text{ is given by}\\ \Phi_{17} &= -AB^{-1},\\ \text{where } A &= \Phi_4\mathbf{M}_{12} + \lambda\Phi_{13}\Phi_{14}^{-1}\mathbf{M}_{32} + \lambda\mathbf{M}_{42} + \lambda\Phi_{16}\mathbf{M}_{52},\\ B &= \Phi_3\mathbf{M}_{12} + \lambda\mathbf{M}_{22} + \lambda\Phi_{12}\Phi_{14}^{-1}\mathbf{M}_{32} + \lambda\Phi_{15}\mathbf{M}_{52}. \end{aligned}$$

Substituting the value of P_1 into equations (23), (24), (25), (27) and (28), respectively, we obtain

$$\begin{aligned} P_4 &= P_{0,1,1}(\Phi_{17}\Phi_1 + \Phi_2), && (30)\\ P_{0,0,0} &= P_{0,1,1}\frac{1}{\lambda}(\Phi_{17}\Phi_3 + \Phi_4), && (31)\\ P_5 &= P_{0,1,1}(\Phi_{17}\Phi_{15}\Phi_5 + \Phi_{16}\Phi_5 + \Phi_{17}\Phi_6 + \Phi_7), && (32)\\ P_2 &= P_{0,1,1}(\Phi_{17}\Phi_{12}\Phi_{14}^{-1} + \Phi_{13}\Phi_{14}^{-1}), && (33)\\ P_3 &= P_{0,1,1}(\Phi_{17}\Phi_{15} + \Phi_{16}). && (34) \end{aligned}$$

Now all probabilities have been expressed as a function of $P_{0,1,1}$. The normalization condition is

$$P_{0,0,0} + P_{0,1,1} + P_1e_1 + P_2e_2 + P_3e_2 + P_4e_1 + P_5e_2 = 1, \tag{35}$$

where e_1 and e_2 are vectors with all of the entries equal to one of dimensions (2×1) and $(N \times 1)$, respectively.

Substituting equations (29), (30), (31), (32), (33) and (34) into (35), we get

$$P_{0,1,1} = \frac{\lambda}{\Phi_{18}}, \tag{36}$$

where

$$\begin{aligned}\Phi_{18} &= \Phi_{17}\Phi_3 + \Phi_4 + \lambda(1 + \Phi_{17}e_1 + \Phi_{17}\Phi_{12}\Phi_{14}^{-1}e_2 + \Phi_{13}\Phi_{14}^{-1}e_2 \\ &+ \Phi_{17}\Phi_{15}e_2 + \Phi_{16}e_2 + \Phi_{17}\Phi_1e_1 + \Phi_2e_2 + \Phi_{17}\Phi_{15}\Phi_5e_2 \\ &+ \Phi_{16}\Phi_5e_2 + \Phi_{17}\Phi_6e_2 + \Phi_7e_2).\end{aligned}$$

The derivation is complete for all stationary probabilities, which can be used to find the measures of performance of the model.

III. Measures of Performance

In this section, we list out some key measures of performance that reflect the bahaviour of the model.

- Let L_s denote the mean number of customers in the system when the servers are busy

$$L_s = \sum_{l=1}^{b}\sum_{m=1}^{b}\sum_{n=0}^{N}(n+l+m)P_{n,i,j},\ 0 \le i \le 2,\ 0 \le j \le 2,\ i+j=2.$$

- Let L_q denote the mean number of customers in the queue when the servers are busy

$$L_q = \sum_{n=1}^{N} nP_{n,i,j},\ 0 \le i \le 2,\ 0 \le j \le 2,\ i+j=2. \tag{37}$$

- The arriving customers are blocked from entering the queue whenever the queue size is N. In this case the blocking probability (P_{block}) is given by

$$P_{block} = P_{N,2,0} + P_{N,1,1} + P_{N,0,2}. \tag{38}$$

- The effective arrival rate is given by

$$\lambda_{eff} = \lambda(1 - P_{block}) = \lambda\left(\sum_{n=0}^{N-1} P_{n,i,j}\right),\ 0 \le i \le 2,\ 0 \le j \le 2,\ i+j=2.$$

  The expected waiting time in the system using Little's law, we get

$$W_s = \frac{L_s}{\lambda_{eff}}.$$

- The expected waiting time in the queue using Little's law, we get

$$W_q = \frac{L_q}{\lambda_{eff}}. \tag{39}$$

- Percentage Variation (P.V.) in waiting time (W_q) is defined as

$$P.V. = \frac{|(W_q)_{b2} - (W_q)_{b1}|}{(W_q)_{b1}} \times 100\%,$$

  where $(W_q)_{b1}$ and $(W_q)_{b2}$ are the waiting time of two values of b.

I. Practical Application of the Model

The queueing model described in this paper has an application in hospital management systems. The model can be applied to situations where the outpatients request an appointment for HIV testing in the clinic centre. The clinic officer monitors the length of booking windows for appointments of the outpatients, but since, in most cases, there is a limited number of doctors, it leads to an unbalanced ratio between the number of outpatients and the doctors. This situation leads to an increase in the length of the booking window and brings the long waiting for appointments. Therefore, to shorten outpatients waiting time, we limit the size of the booking window and assume the clinic centre has a limited slots capacity during a limited length of the booking window. The clinic centre capacity can be divided into sessions/periods of an equal amount of slots. The slots are termed as the maximum size of the outpatient appointments per session. In this scenario, we consider the pooled testing being used to screen the blood of outpatients to detect HIV infections. The slots are taken into service as the primary test for blood sample testing, which is tested in groups. The slots that test positive results opt for a secondary test for further testing. In contrast, all outpatients in the slots are cleared and exempted from further testing if slots test negative. In this situation, appointment, doctors, pooled test (testing in group), limited length of the booking window, primary test and secondary test correspond to the arrivals, servers, batch testing, limited queue capacity, FES and SOS, respectively, in queueing terminology. In the next section, this model application in practice is analyzed with a numerical investigation.

IV. Numerical Investigation

In this section, we present the applicability of the solutions obtained using matrix-decomposition method. We compute the model numerically by taking the model parameters that have close incidence with the practical situations as $N = 14,\ b = 3, \lambda = 0.6,\ \mu_1 = 0.4,\ \mu_2 = 0.3,\ r = 0.4$, with the assumption that $b \leq N$, where

- λ = Appointment rate of outpatients to the clinic centre,
- μ_1 = Service rate during primary test,
- μ_2 = Service rate during secondary test,
- r = Probability of opting secondary test,
- b = Maximum number of appointment of outpatients per session,
- N = Maximum capacity of the outpatients at the clinic centre.

Table 1: *Variation in different measures of performance with the change in service rate (μ_1 and μ_2)*

μ_1	L_q	W_q	P_{block}	μ_2	L_q	W_q	P_{block}
0.1	3.745370	6.480820	0.038017	0.1	6.193580	10.32330	0.105915
0.2	1.727790	2.884150	0.001948	0.2	1.423120	2.371980	0.001781
0.3	1.116660	1.861360	0.000308	0.3	0.836311	1.393880	0.000123
0.4	0.836311	1.393880	0.000123	0.4	0.655543	1.092590	0.000029
0.5	0.678974	1.131630	0.000078	0.5	0.569728	0.949556	0.000014
0.6	0.580261	0.967104	0.000061	0.6	0.520022	0.866710	0.000009
0.7	0.513680	0.856134	0.000053	0.7	0.487796	0.812998	0.000008
0.8	0.466371	0.777284	0.000048	0.8	0.465319	0.775536	0.000007
0.9	0.431387	0.718978	0.000045	0.9	0.448812	0.748024	0.000006
1.0	0.404684	0.674474	0.000042	1.0	0.436214	0.727027	0.000005

Table 1 shows the effect of the service rate of primary test (μ_1) and secondary test (μ_2) on some performance measures. We observe that increasing $\mu_1(\mu_2)$ decreases L_q, W_q and P_{block}. This is because as $\mu_1(\mu_2)$ increases, outpatients are served faster so that the queue length and the waiting time decrease. Moreover, P_{block} tends to zero due to faster services.

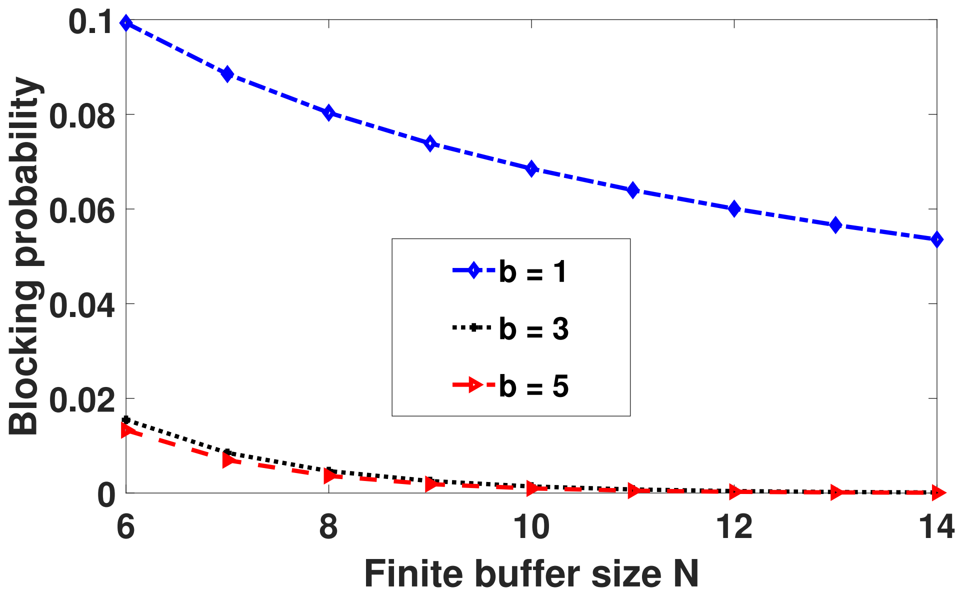

Figure 2: *Effect of buffer size on the blocking probability*

Figure 2 shows the impact of buffer size (N) on the blocking probability (P_{block}) for different batch sizes b. We observe that P_{block} monotonically decreases as N increases and reaches its minimum value zero as N is sufficiently larger. Moreover, when N is kept fixed, P_{block} decreases with the increase of b, as we expect. This is because as b incrases, more number of outpatients are taken for service in a batch, which leads to a decrease in size of the queue. Hence P_{block} decreases.

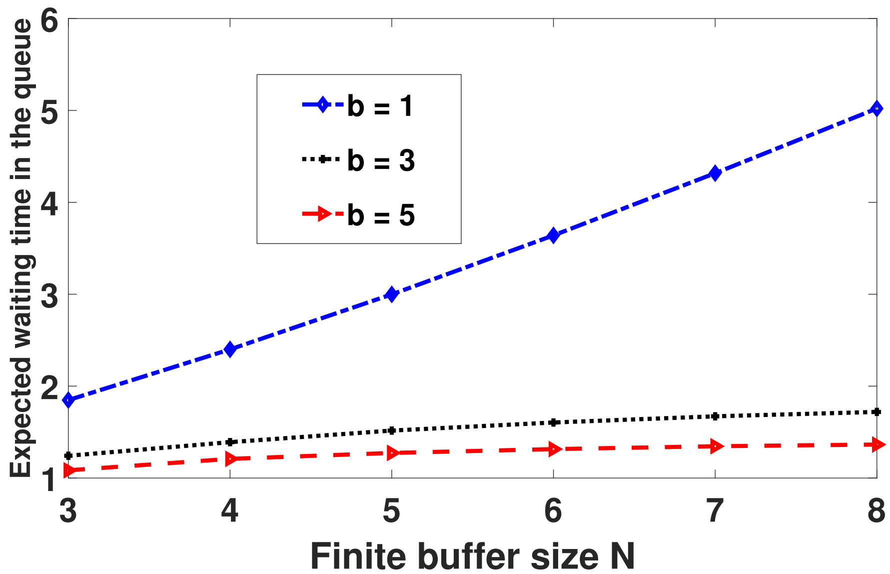

Figure 3: *Effect of finite buffer size on the expected waiting time in the queue* (W_q)

In Figure 3, we observe that for fixed N, W_q decreases as b increases. This is because, as b increases, more outpatients are served in a batch at a time, as a result, W_q decreases. Further, for fixed b, except $b = 1$, the waiting time is more prominently increasing when N lies in $[3, 6]$. However, when the buffer size increases beyond 6, the waiting time is not much affected by increasing N, since the arrival rate is constant and the doctors serve the outpatients in batches $b > 1$. When $b = 1$, the waiting time increases monotonically as N increases. The reason is that by increasing the buffer size, more outpatients accumulate in the queue, and the doctor serves one outpatient at a time, this leads to an increase in the queue length. Hence, the waiting time increases compared to $b = 3$ ($b = 5$).

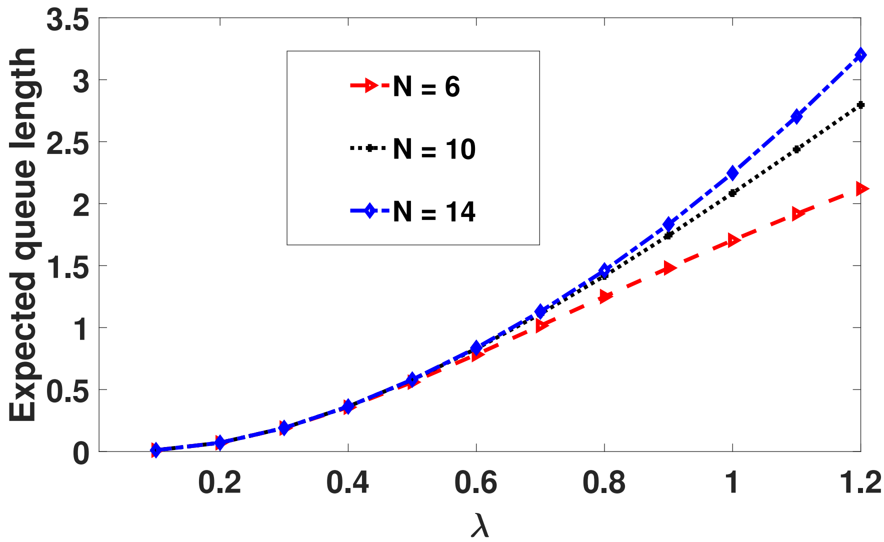

Figure 4: *Impact of the arrival rate on the expected queue length*

Figure 4 shows the impact of λ on the expected queue length (L_q) for different values of N. It is clear that as λ increases, L_q increases for all values of N, as it should be. Moreover, for a fixed λ, as N increases, more outpatients accumulate in the queue thereby an increasing trend can be seen in L_q.

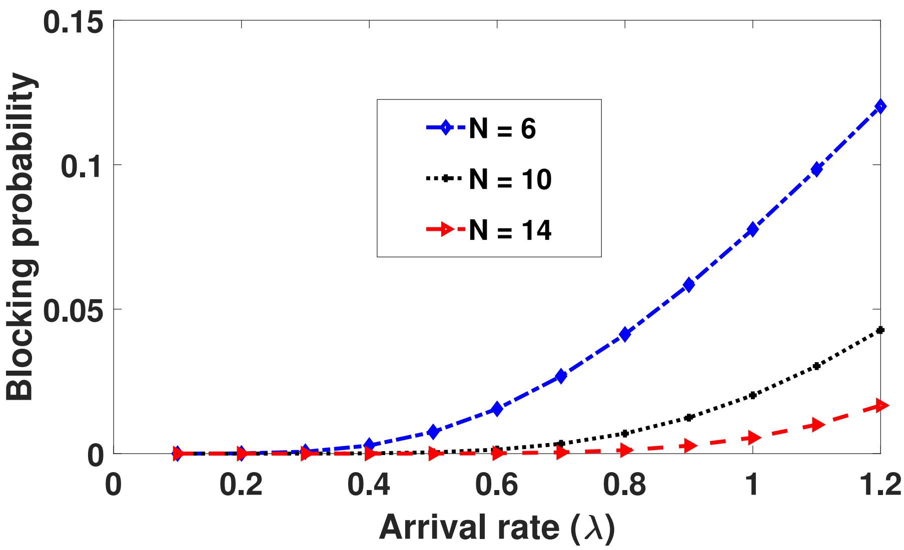

Figure 5: *Impact of arrival rate on the blocking probability*

Figure 5 shows the impact of λ on P_{block}. As intuitively expected, P_{block} increases with the increase of λ. Furthermore, for a fixed λ, P_{block} is high for smaller values of N, which is true.

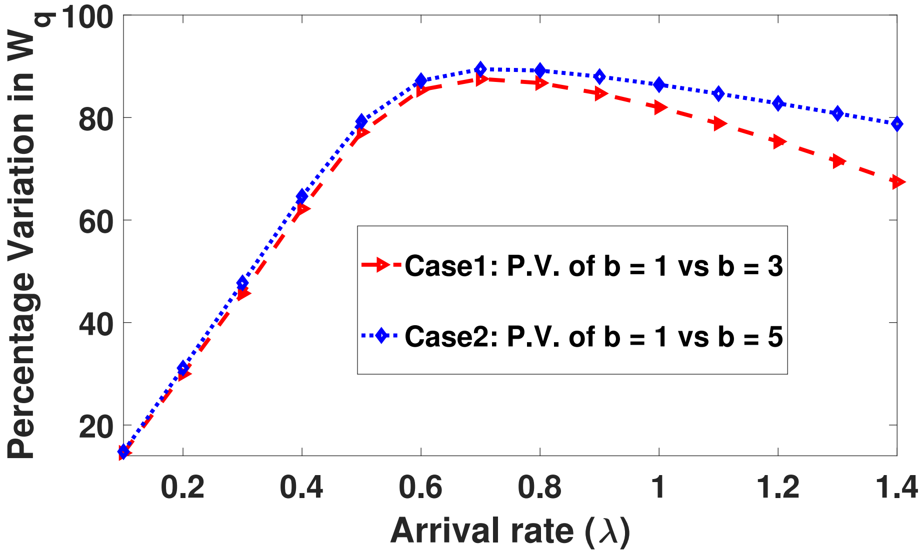

Figure 6: *Impact of arrival rate on the percentage variation in expected waiting time*

Figure 6 shows the impact of arrival rate λ on the P.V. in W_q in two cases, case 1: P.V. of $b = 1$ vs $b = 3$ and case 2: $b = 1$ vs $b = 5$. As λ increases, the P.V. in W_q for case 1 and case 2 initially increases up to $\lambda = 0.7$ and for $\lambda > 0.7$ it slightly decreases. Moreover, the P.V. in W_q varies widely as λ increases. This means that as arrival rate increases, there is a high probability of blocking outpatients to enter the clinic centre.

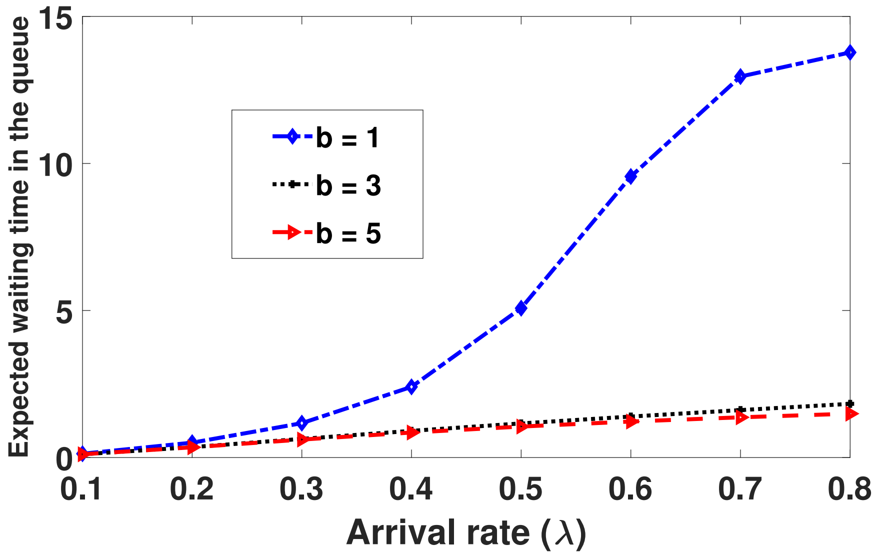

Figure 7: *Impact of arrival rate on the anticipated waiting time*

For different values of b, we show the impact of λ on the anticipated waiting time (W_q) in Figure 7. As λ increases, the inflow of outpatients to the clinic centre increases, which tends to a longer queue. Hence, W_q increases. Further, W_q shows an opposite trend with the increase of batch size taken for the service, as intuitively expected.

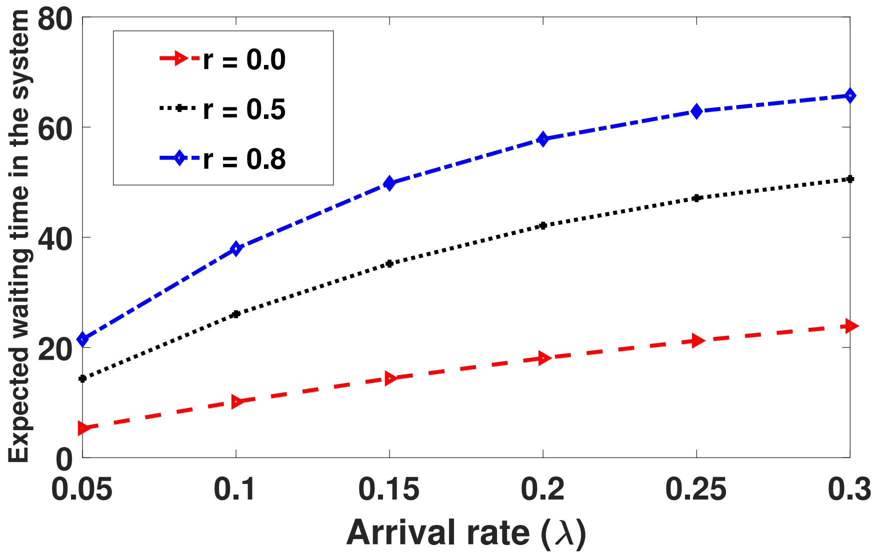

Figure 8: *Impact of arrival rate on the anticipated waiting time in the system*

Figure 8 demonstrates the impact of λ on the anticipated waiting time in the system (W_s) for different values of r. For a fixed r, it is observed that W_s increases with the increase of λ, as we expect. Furthermore, for a fixed λ, W_s is smaller in the absence of secondary test ($r = 0.0$), and as r increases, more number of outpatients are tend to secondary test (SOS), which leads to an increase in W_s.

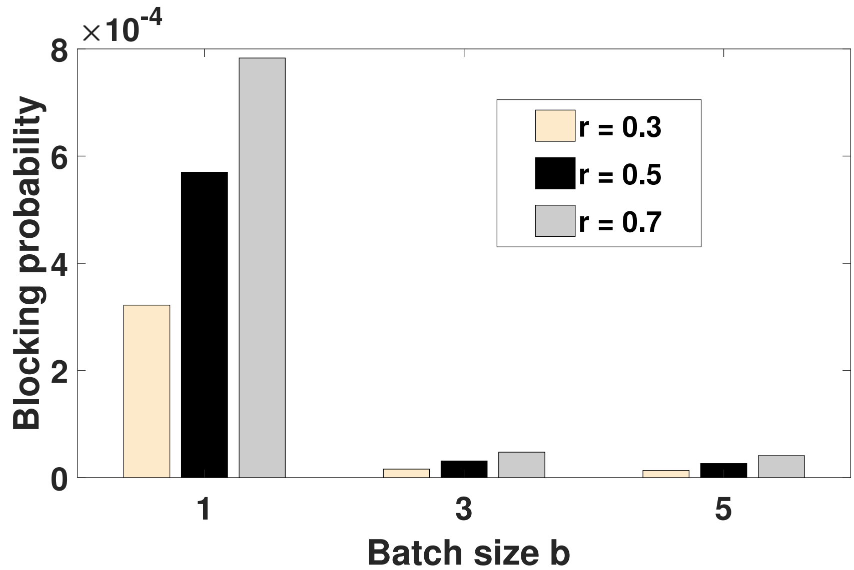

Figure 9: *Effect of batch size on blocking probability*

Figure 9 shows the effect of b on the blocking probability (P_{block}) for different r. As b increases, more outpatients are served in batch and leave the queue, which results in a decrease of P_{block}. Furthermore, for fixed b, P_{block} increases as r increases. The reason is that the outpatients opting for secondary test increase the expected waiting time of other outpatients to be served.

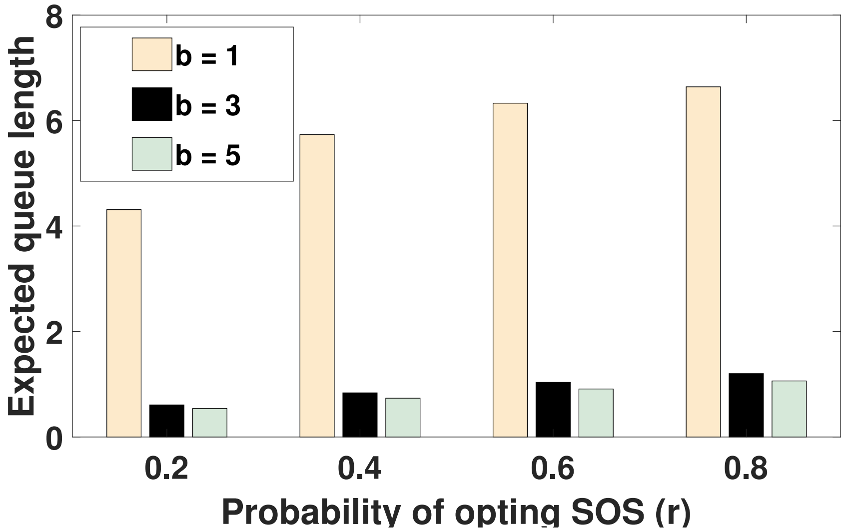

Figure 10: *Effect of probability r on the expected queue length*

Figure 10 depicts the effect of r on the anticipated queue length (L_q). We observe that as r increases, outpatients opt for secondary test, thereby increasing the expected queue length. Also, for a fixed r, as batch size increases, more outpatients are served at a time, resulting in a decrease in L_q.

V. Cost Model Optimization

In this section, we present the total anticipated cost function per item per unit time. The main goal is to figure out the optimal value of service rates μ_1 and μ_2 for primary and secondary tests, respectively, so that the cost is minimized. We apply the Quasi-Newton technique to derive the optimal values of the service rates.

I. Quasi-Newton Method

Quasi-Newton method is reliable and efficient for finding a minimizer of a nonlinear function by estimating the curvature along a sequence of search directions with some fixed tolerance (say ϵ). Let μ^i denote the current point at iteration $i = 0,\ 1,\ 2, \cdots$ The gradient of f at μ^i is denoted $\overrightarrow{\nabla} f(\mu^i)$ and abbreviated to $\overrightarrow{\nabla} f(\mu)$. We use $H(\mu^i)$ to denote a positive definite matrix which is an estimate of the inverse Hessian $\overrightarrow{\nabla}^2 f(\mu^i)^{-1}$. It is important to note that if f is not differentiable at μ^{i+1} we say that the algorithm breaks down (in theory) and if $\overrightarrow{\nabla} f(\mu) = 0$ we say it terminates at a smooth stationary point (for more details the reader may refer [**?**]).

II. Algorithm for Quasi-Newton method

The steps of Quasi-Newton method can be described as follows:

Step 1: Pick any starting trial solution for $\mu^0 = (\mu_1^0, \mu_2^0)$ and compute $f(\mu_1^0, \mu_2^0)$.

Step 2: While $|\frac{\partial f}{\partial \mu_1}| > \epsilon$ or $|\frac{\partial f}{\partial \mu_2}| > \epsilon$; do step $3-5$.

Step 3: Compute the cost gradient $\overrightarrow{\nabla} f(\mu) = [\frac{\partial f}{\partial \mu_1}, \frac{\partial f}{\partial \mu_2}]^T$ and the cost Hessian matrix $H(\mu) = \begin{bmatrix} \frac{\partial^2 f}{\partial \mu_1^2} & \frac{\partial^2 f}{\partial \mu_1 \partial \mu_2} \\ \frac{\partial^2 f}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 f}{\partial \mu_2^2} \end{bmatrix}$ at point $\overrightarrow{\mu}^{(i)}$, provided that $\frac{\partial^2 f}{\partial \mu_1^2}(\frac{\partial^2 f}{\partial \mu_2^2}) - (\frac{\partial^2 f}{\partial \mu_1 \partial \mu_2})^2 \neq 0$, which enables the existence of inverse of $H(\mu)$.

Step 4: Update the new trial solution $\mu^{(i+1)} = \mu^{(i)} - [H(\mu^{(i)})]^{-1}\vec{\nabla} f(\mu^{(i)})$.

Step 5: Set $i = i + 1$ and return to step 2. If the gradient is sufficiently smaller than ϵ, then stop.

To implement the above algorithm, we propose the cost function $f(\mu_1, \mu_2)$ per item per unit time as

$$f(\mu_1, \mu_2) = c_q L_q + c_1\mu_1 + c_2\mu_2 + 2c_3. \tag{40}$$

Let us consider the following optimization problem:

$$f(\mu_1^*, \mu_2^*) = \underset{s.t \quad \mu_1>\mu_2}{\text{Minimize}} f(\mu_1, \mu_2),$$

where the various cost components are defined as follows:
c_q = cost per unit per item present in the queue,
c_1 = cost per unit time when the servers are in FES,
c_2 = cost per unit time when the servers are in SOS,
c_3 = fixed cost for purchase of one server.
The goal of using this cost model is to give emphasis on the service rates in order to have a cost benefit and less congestion at the queueing system.
We apply the Quasi-Newton technique to obtain the optimal value of service rates (μ_1, μ_2). Assuming the values to the coefficients of the expected total cost function in (40) as c_q = \$600, c_1 = \$300, c_2 = \$400, c_3 = \$200, $\epsilon = \epsilon_1 = \epsilon_2 = 10^{-7}$ and $\lambda = 0.6$, $b = 3$, $r = 0.4$ and $N = 14$. Starting with the initial values $(\mu_1^0, \mu_2^0) = (0.4, 0.3)$, the following Table values have been calculated.

Table 2: *Quasi-Newton Method.*

Iteration	0	1	2	3
μ_1	0.40000	0.53331	0.64672	0.69485
μ_2	0.30000	0.37052	0.43434	0.46377
$f(\mu_1, \mu_2)$	1141.79	1010.26	974.409	971.328
L_q	0.83631	0.50344	0.34443	0.29561
$\frac{\partial f}{\partial \mu_1}$	-914.739	-308.569	-68.0794	-2.25317
$\frac{\partial f}{\partial \mu_2}$	-1275.38	-437.473	-99.3126	-3.79388
Hessian	$\begin{bmatrix} 6861.83 & 2.13400 \\ 2.13400 & 18086.7 \end{bmatrix}$	$\begin{bmatrix} 2720.75 & 2.82643 \\ 2.82643 & 6854.12 \end{bmatrix}$	$\begin{bmatrix} 1414.49 & 2.10218 \\ 2.10218 & 3374.53 \end{bmatrix}$	$\begin{bmatrix} 1098.71 & 1.79815 \\ 1.79815 & 2530.27 \end{bmatrix}$
Iteration	4	5	6	7
μ_1	0.69690	0.69667	0.69669	**0.69669**
μ_2	0.46530	0.46513	0.46515	**0.46515**
$f(\mu_1, \mu_2)$	971.324	971.324	971.324	**971.324**
L_q	0.29358	0.29378	0.29376	0.29376
$\frac{\partial f}{\partial \mu_1}$	0.25677	-0.02454	0.00303	-0.00029
$\frac{\partial f}{\partial \mu_2}$	0.34083	-0.04239	0.00402	-0.00050
Hessian	$\begin{bmatrix} 1087.19 & 1.78470 \\ 1.78470 & 2494.97 \end{bmatrix}$	$\begin{bmatrix} 1088.50 & 1.78605 \\ 1.78605 & 2498.18 \end{bmatrix}$	$\begin{bmatrix} 1088.38 & 1.78590 \\ 1.78590 & 2497.79 \end{bmatrix}$	$\begin{bmatrix} 1088.39 & 1.78592 \\ 1.78592 & 2497.83 \end{bmatrix}$

From Table 2, we find that the minimum cost per unit time is $f(\mu_1^*, \mu_2^*)$ = \$971.324 at $(\mu_1^*, \mu_2^*) = (0.69669, 0.46515)$ achieved at seventh iteration.

Table 3: *The optimal service rates* $(\mu_1,\ \mu_2)$ *and cost function* $f(\mu_1, \mu_2)$ *obtained in variation of b and r.*

		μ_1^*	μ_2^*	L_q	$f(\mu_1^*, \mu_2^*)$
$b = 1$	$r = 0.2$	0.80571	0.572062	0.234602	1011.3
	$r = 0.5$	0.728169	-	-	-
	$r = 0.8$	0.691397	-	-	-
$b = 3$	$r = 0.2$	0.687462	0.357727	0.243295	895.306
	$r = 0.5$	0.698797	0.505399	0.314466	1000.48
	$r = 0.8$	0.700593	0.598757	0.367520	1070.19
$b = 5$	$r = 0.2$	0.670994	0.327201	0.250341	882.383
	$r = 0.5$	0.687876	0.475721	0.321209	989.377
	$r = 0.8$	0.697033	0.570477	0.373146	1061.19
$b = 7$	$r = 0.2$	0.667993	0.319684	0.252881	880.000
	$r = 0.5$	0.685428	0.469603	0.323526	987.585
	$r = 0.8$	0.695602	0.565179	0.375082	1059.80
$b = 9$	$r = 0.2$	0.667434	0.317632	0.25369	879.497
	$r = 0.5$	0.684913	0.468318	0.32414	987.285
	$r = 0.8$	0.695243	0.564205	0.375558	1059.59

In Table 3, we observe that

- For fixed b, as r increases, we observe that μ_1^* shows an increasing trend. However, for $b = 1$, as r increases, the reverse trend is observed in μ_1^*. This is necessary in order to balance the system in the profitable manner. On the other hand, it is obvious that μ_2^*, L_q and $f(\mu_1^*, \mu_2^*)$ increase with the increase of r.
- Similarly, for a fixed r, as b increases, μ_1^*, μ_2^* and $f(\mu_1^*, \mu_2^*)$ all decreases except L_q which increases as b increase. L_q increase because the service rates decrease in order to balance the system profitably.
- For $b = 1$ and $r = (0.5)0.8$, the service rate in secondary test (SOS) does not satisfy the condition $\mu_1 > \mu_2$. Hence, we exclude those values.

In general, we observe the following features of the queueing system:

- As service rates increases, blocking probability, expected waiting time, and queue size decrease.
- The blocking probability is higher for smaller values of buffer size.
- An increase in arrival rate increases the blocking probability, expected waiting time, and queue size.
- An increase in batch size decreases the blocking probability, expected waiting time, queue size, and optimum cost.
- An increase of r probability increases the optimum cost.

VI. Conclusion

A two-server batch service queueing model with SOS is studied using matrix-decomposition method. The cost model optimization was also developed to determine the optimal service rates to minimize the cost. Performance measures and numerical illustrations discussed in this paper provide valuable insights about the functioning of clinic centre in providing the services of primary and secondary test. The clinic offices will benefit from the simplified numerical simulations, which will help them increase system efficiency.

In future work, we will incorporate a two-server batch service queue with SOS, adding the concepts of working vacations and vacation interruption. Also, we will consider the transient state in the current model with batch arrival.

References

[1] Ammar, S. (2014). Transient analysis of a two-heterogeneous servers queue with impatient behavior. *Egyptian Mathematical Society*, 22(1), 90–95.

[2] Armero, C. and Conesa, D. (2006). Bayesian hierarchical models in manufacturing bulk service queues. *Journal of Statistical Planning and Inference*, 136(2), 335–354.

[3] Ayyappan, G., Devipriya, G., and Subramanian, A. (2014). Analysis of single server batch service queueing system under multiple vacations with second optional service. *International Journal of Mathematical Archive*, 5(2), 126–144.

[4] Ayyappan, G., Devipriya, G., and Subramanian, A. (2014). Analysis of single server fixed batch service queueing system under multiple vacations with gated service. *International Journal of Computer Applications*, 89(5),15–19.

[5] Banerjee, A. (2016). Analysis of finite buffer queue with state dependent service and correlated customer arrivals. *Egyptian Mathematical Society*, 24(2), 295–302.

[6] Barbhuiya, F. and Guptal, U. (2019). A difference equation approach for analysing a batch service queue with the batch renewal arrival process. *Journal of Industrial and Production Engineering*, 25(2), 233–242.

[7] Ghimire, S., Ghimire, R., Thapa, G. B., and Fernandes, S. (2017). Multi-server batch service queueing model with variable service rates. *International Journal of Applied Mathematics & Statistical Sciences*, 6(4), 43–54.

[8] Goswami, V. and Samanta, S. (2009). Discrete time bulk service queue with two heterogeneous servers. *Computers and Industrial Engineering*, 56(4), 1348–1356.

[9] Gupta, G. K., Banerjee, A., and Gupta, U. C. (2019). On finite-buffer batch-size-dependent bulk service queue with queue-length dependent vacation. *Quality Technology and Quantitative Management*, 17(6), 501–527.

[10] Gupur, G. (2014). On the eigenvalues of the $m/m/1$ queueing model with second optional service. *Journal of Pseudo-Differential Operators and Applications*, 5(3), 359–410.

[11] Hwang, J., Gao, L., and Jang, W. (2010). Joint demand and capacity management in a restaurant system. *European journal of operational research*, 207(1), 465–472.

[12] Kalyanaraman, R. and Murugan, S. (2008). A single server queue with additional optional service in batches and server vacation. *Applied mathematical sciences*, 2(56), 2765–2776.

[13] Ke, J. C., Wu, C. H., and Pearn, W. L. (2011). Multi-server retrial queue with second optional service: Algorithmic computation and optimisation. *International journal of systems science*, 42(10), 1755–1769.

[14] Ke, J. C., Wu, C. H., and Pearn, W. L. (2013). Infinite capacity multi-server queue with second optional service channel. *Journal of physics. Conference series*, 410, 1–4.

[15] Krishna Reddy, G. and Anitha, R. (1998). Markovian bulk service queue with delayed vacations. *Computers and Operations Research*, 25(12), 1159–1166.

[16] Krishnamoorthy, A. and Sreenivasan, C. (2012). An $m/m/2$ queueing system with heterogeneous servers including one with working vacation. *International Journal of Stochastic Analysis*, 2012, 1–16.

[17] Kumar, B. S., Madheswari, S., and Venkatakrishnan, K. S. (2007). Transient solution of an $m/m/2$ queue with heterogeneous servers subject to catastrophes. *International Journal of Information and Management Sciences*, 18(1), 63–80.

[18] Kumar, B. S., Sharma, V., and Seth, A. (2020). An $m/m/2$ heterogeneous service markovian feedback queuing model with reverse balking, reneging and retention of reneged customers. *Advances in Computing and Intelligent Systems, Algorithms for Intelligent System*, pages 291–296.

[19] Kumar, R. and Sharma, S. (2017). Transient solution of a two-heterogeneous server's queuing system with retention of reneging customers. *Bulletin of the Malaysian Mathematical Sciences Society*, 42(4), 1–18.

[20] Lewis, A. S. and Overton, M. L. (2013). Nonsmooth optimization via quasi-newton methods. *Mathematical Programming, Series B*, 141, 135–166.

[21] Sikdar, K. and Gupta, U. (2008). On the batch arrivals batch service queue with finite buffer under server's vacation: $m^X/g^Y/1/n$. *Computers and Mathematics with Applications*, 56(11), 2861–2873.

[22] Uma, S. and Punniyamoorthy, K. (2016). Single server bulk queue with second optional service, balking and compulsory vacation. *International journal of engineering research and applications*, 6(10), 15–20.

[23] Wang, J., W. F. S. J. and Kuki, A. (2017). Finite source retrial queues with two phase service. *International Journal of Operational Research*, 30(4), 421–440.

[24] Yang, D. Y. and Chen, Y. H. (2018). Computational and optimization of a working breakdown queue with second optimal service. *Journal of Industrial and Production Engineering*, 35(3), 81–88.

[25] Yue, D., Y. W. and Tian, R. (2010). Analysis of two-server queues with a variant vacation policy. *JThe Ninth International Symposium on Operations Research and Its Applications (ISORA), Chengdu-Jiuzhaigou, China, August 19-23*, pages 483–491.