MODELING QUEUES IN TRANSPORT USING ANALYTICAL MODELS. A CASE OF A MINING COMPANY IN ZIMBABWE

Romeo Mawonike*,
Tendai Makoni*
Musara Chipumuro*

ABSTRACT

In this paper multi-server models and some generalizations are considered. With the exception of one model, arrival time and departure time are assumed to be Markovian. Arriving units and departing units are according to the birth-and-death processes. Solutions for the models are obtained under steady state conditions. Waiting times, queue length and the number of customers in the system at any given time are discussed using analytical queuing models. Finally, sensitivity analysis is performed to reach the best optimal solution to the current problem faced by both customers and the company. The results show that two options could be pursued to reduced waiting lines and costs by customers and the company. The first one is to build more parking spaces for waiting trucks and the second option is to increase the service rate while maintaining the number of servers available.

Keywords: *Waiting time, arrival time, multi-server models, birth and death processes.*

*Department of Mathematics and Computer Science, Great Zimbabwe University, Zimbabwe.

1.0 INTRODUCTION

"Nobody likes to wait" says Deb Holler, branch Vice President at Great Western Bank in Encino, California. "We try to keep people from actually despising it". (Martinich 1997). Waiting for service is part of our daily life. We wait to eat in restaurants, we "queue up" at the checkout counters in grocery stores, and we "line up" for service in post offices. The waiting phenomenon is not an experience limited to human beings only: jobs wait to be processed on a machine, planes circle in stack before given permission to land at the airport and cars stop at traffic lights. Unfortunately, we cannot eliminate waiting without incurring inordinate expenses. In fact, all we can hope to achieve is to reduce its adverse impact to tolerable levels. It has been estimated that Americans spent 37 000 000 000 hours per year waiting in queues, USSR was notorious for the tremendously long queues that its citizens frequently had to endure just to purchase basic necessities (Hillier 2001). Queues (or waiting lines) help facilities or businesses to provide service in an orderly fashion. Forming a queue being a social phenomenon, it is beneficial to the society if it can be managed so that both the unit that waits and the one that serves get the most benefit. The analysis is based on building a mathematical model representing the process of arrival of customers who join the queue, the rules by which they are allowed into service, and the time it takes to serve the customers. Queuing theory embodies the full gamut of such models covering all perceivable systems which incorporate characteristics of a queue. In this paper, we focus on the modeling of queues of haulage trucks to find the optimum number of servers required and also to find the best structure of a queue to be adopted sensitive to time.

1.1 Queuing Behaviour

People's behaviour in queues and service encounters is often unpredictable. Reneging is the process of a customer entering the waiting line but later deciding to leave the line and server system. Balking is the process of a customer evaluating the waiting line and server system and deciding not to enter the queue. In both situations, the customer leaves the system, may not return, and a current sale or all future sales may be lost. Most analytical models assume the customer's behavior is patient and steady and they will not renege or balk.

1.2 Research Problem

Due to several trucks coming to the site for loading, waiting time and waiting costs increase**.** There are only three servers (loaders) one at each point, that is parallel servers. Space (waiting bay) for accommodating these trucks is limited only to 20 trucks. On the other hand, customer expectations, requirements and demands are not timely met. As a result, trucks are seen everywhere in the area including the residential area looking for space waiting for their turn. There have been complaints from customers due to prolonged and sluggish queues. Accidents in the residential area have become the order of the day as a result of haulage trucks. Furthermore, prostitution has also increased in the area due to truck drivers who have nothing do during their waiting period.

1.3 Main Objective

To design a Queuing system that minimizes the average waiting time, costs and length of the queue.

1.4 Assumptions of the model

- Exponential distribution of inter-arrival times and service time and Poisson distribution of arrival rate.
- Arrival and service times differ from client to client, but the average rate is known and remains constant over a particular period.
- Queue discipline is First Come First Serve (FCFS)
- The system is in the state (number of customers) N and no arrivals or service completions occurred, leaving the total to N customers
- The system is in state (number of customers) $N+1$ and a service completion occurred reducing the total to N customers
- The system is in state (number of customers) $N-1$ and a new arrival occurred, bringing the total to N customers.
- No jockeying – customers are not switched between queues if the line that there are in is slow.
- Arrivals to and departure from the system occurs completely at the random.
- Coal is always available at the service points.

2.0BACKGROUND OF STUDY

The mining company was established 100 years ago and its sole business is to extract and supply coal to other manufacturing industries such as Zimbabwe Iron and Steel Company (Zisco), Sable Chemicals, Zimbabwe Alloys and so on and other countries like Zambia and Malawi. Production and Maintenance Departments have the mandate to make sure coal has been produced at optimal levels and distributed to meet customer demands. Since this company is the only company in

Zimbabwe which supplies coal, therefore it monopolizes the whole industry. Consequently, several trucks from different sectors come to this company for coal. Apparently, space to accommodate these trucks sometimes become incapacitated that is shortage of parking space.

The study of queues deals with quantifying the phenomenon of waiting in lines using representative measures of performance, such as average queue length, average waiting time in queue and average facility utilization. (Kothari, 1982). Walt Disney theme parks, a case study in New York (William 2005), realize that waiting in lines at attractions does not add to the enjoyment of their customers. They also realized that customers waiting in lines are not generating the revenue that they would if they visited restaurants and souvenir shops. Hence, they have several reasons for wanting to reduce waiting lines. Lately, they have been using reservation systems at some attractions that allow customers to reserve visit times instead of having to wait in line. As a result, customers are happier because they don't have to wait in line, and the park's potential for additional revenue increased. Bernard (2007), in a market research study, Bank of America discovered that when a customer stands in line waiting for teller, a gap develops between the actual waiting time and the customer's perceived waiting time. For example if customers have been waiting for 5 minutes, it may be seen more like 10 minutes to them. Bank of America also from prior studies that long waits have direct relationship to customer satisfaction. As a result, the bank took an experiment to see if customer's perceived waiting time could be reduced if it placed televisions above the teller in a bank branch lobby to entertain customers waiting in line and measured actual versus perceived waiting time. The results showed that the amount of waiting time customers overestimated dropped from 32% to 15% when compared to a bank without television. A. K. Erlang (1909), lays the foundation for the place of Poisson (and hence, exponential) distribution in queuing theory. His papers written in the next 20 years contain some of the most important concepts and techniques; the notion of statistical equilibrium and the method of writing down balance of state equations (later called Chapman-Kolmogorov equations) are two such examples. It should be noted that in Erlang's work, as well as the work done by others in the twenties and thirties, the motivation has been the practical problem of congestion. Pollaczek (1934) began investigations of the behavior of the system during a finite time interval. Since then and throughout his career, he did considerable work in the analytical behavioral study of queuing systems. A probabilistic approach to the analysis was initiated by Kendall (1951, 1953) when he demonstrated that imbedded Markov chains can be identified in the queue length process in systems *M/G/*1 and *GI/M/s*. Lindley (1952) derived integral equations for waiting time distributions defined at imbedded Markov points in the general queue *GI/G/*1. These investigations led to the use of renewal theory in queuing systems analysis in the 1960's. Identification of the imbedded Markov chains also facilitated the use of combinatorial methods by considering the queue length at Markov points as a random walk. Complex queuing network problems have been investigated extensively since the beginning of the 1970's. Several survey papers and books summarize the major contribution made in this area. These include Basket (1973), Kelly (1979), Whitt (1983a, b), Coffman and Hoffri (1986), describing important Computer devices and the queuing models used in analyzing their performance. Hillier's (1963) paper on economic models for industrial waiting line problems is, perhaps, the first paper to introduce standard optimization techniques to queuing problems. While Hillier considered an M/M/1 queue, Heyman (1968) derived an optimal policy for turning the server on and off in an M/G/1 queue, depending on the state of the system. Li and Liling (2000), presented a multi-objective model that includes the flexibility of jobs. They determined the number of required employees for planning period in budget, in which the objectives are in a priority level from each other. Tabari *et al* (2012), use the queuing theory to recognize the optimal number of required human resources in an educational institution carried out in Iran. The queue analysis is performed for different numbers of staff members. Finally, the result of this study shows that the staff members in this department should be reduced. McClain (1976) reviews research on models for evaluating the impact of bed assignment policies on utilization, waiting time, and the probability of turning away patients. Nosek and Wilson (2001) review the use of queuing theory in pharmacy applications with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. Green (2006a) presents the theory of queuing as applied in healthcare. She discusses the relationship amongst delays, utilization and the number of servers; the basic M/M/s model, its assumptions and extensions; and the applications of the theory to determine the required number of servers. Fomundam and Herrmann (2007), survey the contributions and applications of queuing theory in the field of healthcare. The paper summarizes a range of queuing theory results in the following areas: waiting time and utilization analysis, system design, and appointment systems and also considers results for systems at different scales, including individual departments (or units), healthcare facilities, and regional healthcare systems. The goal is to provide sufficient information to analysts who are interested in using queuing theory to model a healthcare process and want to locate the details of relevant models. Albin et al. (1990) show how one can use queuing theory to get approximate results and then use simulation models to refine them. Worthington (1987) presents an M(λq)/G/S model for service times of any fixed probability distribution and for arrival rates that decrease linearly with the queue length and the expected waiting time.

2.1 Notation

Commonly used notation in queuing theory can be defined as follows:

n – The number of customers or units in the system.

$p_n(t)$ – The time dependent probability of exactly n customers being in the system at time t, given that the system started at time zero.

p_n – The time independent or steady-state probability of exactly n customers being in the system.

λ – The number of customers or units arriving per unit time, that is the mean arrival rate.

µ – The number of customers or units served per unit time, that is the mean service rate per busy server.

m – The number of parallel service channel providing identical service.

ρ – Utilization capacity of the server.

W_s – The mean waiting time per customer in the system.

W_q – The mean waiting time per customer in the queue.

L_s – The mean number of customers in the system.

L_q – The mean number of customers in the queue.

3.0METHODOLOGY

3.1 Basic structure of a Queuing Model

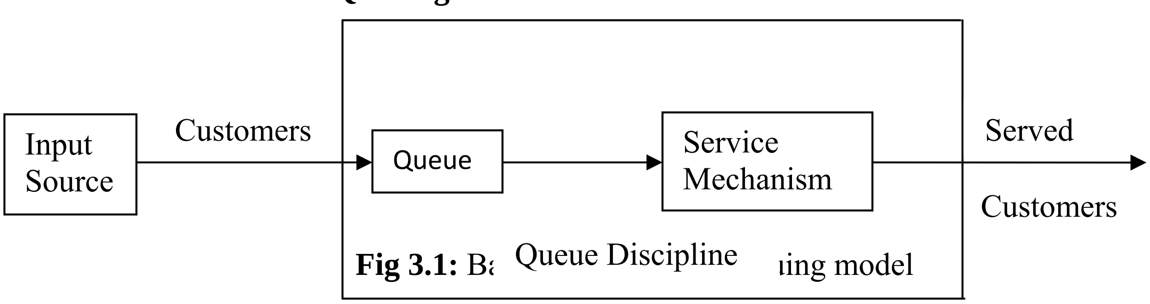

Fig 3.1: Ba[illegible]ing model

Where:

- Input source is the calling population characterized by the total number of potential customers.

- Customers are trucks waiting to be loaded.
- Queue is where customers (trucks) wait before being served.
- Queue Discipline refers to the order in which members of the queue are selected for service (in this case First In First Served, FIFO is the order).
- Service mechanism consists of one or more service facilities, each of which contains one or more parallel service channels called servers. (In this paper three servers are considered).

3.2 Data Collection

Hourly data has been collected from the site where trucks are being loaded. Data is for 20 days that is 480 hours since there is break due to some work shifts by workers. The loading time to each truck has been noted as well as the arrival times and inter-arrival times of trucks coming to the parking bay have been also noted. Currently, there are three (3) identical loaders (servers) stationed at the site but there is only one queue. Space prepared for accommodating these trucks while waiting for service is limited. Therefore, truck drivers out of the parking bay will find other means provided they maintain their queue position since they would have come from other cities of the country and even out of the country as far as Zambia and Malawi. Survey has been carried out and it was discovered that each driver spends \$5 per hour on average while in the queue.

3.3 Birth and Death Process.

In this paper, birth-and-death models are used since there are both arrival and departure. The birth-and-death process is a special case of continuous time Markov process, where the states (for example) represent a current size of a population and the transitions are limited to birth and death. When a birth occurs, the process goes from state i to state $i+1$. Similarly, when death occurs, the process goes from state i to state $i-1$. It is assumed that the birth and death events are independent of each other. The birth-and-death process is characterized by the birth rate $\{\lambda_i\}_{i=0,1,2\ldots,}$ and death rate $\{\mu_i\}_{i=0,1,2\ldots,}$, which vary according to state i of the system. We can define a Pure Birth Process as a birth-death process with $\lambda_i = 0$ for all i. Similarly, a Pure Death Process corresponds to a birth-death process with $\mu_i = 0$ for all i. The general description of the Birth-and-Death process can be as follows: after the process enters state i, it holds (sojourns) in the given state for some random length of time, exponentially distributed with parameter $(\lambda_i + \mu_i)$. When leaving$i$, the process enters either $i+1$ with probability,

$$\frac{\lambda_i}{\lambda_i + \mu_i}$$

Or $i - 1$ with probability,

$$\frac{\mu_i}{\lambda_i + \mu_i}$$

If the next state chosen is $i + 1$, then the process sojourns in this state according to the exponential distribution with parameter $\lambda_{i+1} + \mu_{i+1}$ and then chooses the next state etc. The number of visits back to the same state is ignored since in a continuous time process transitions from state i back to i would not be identifiable.

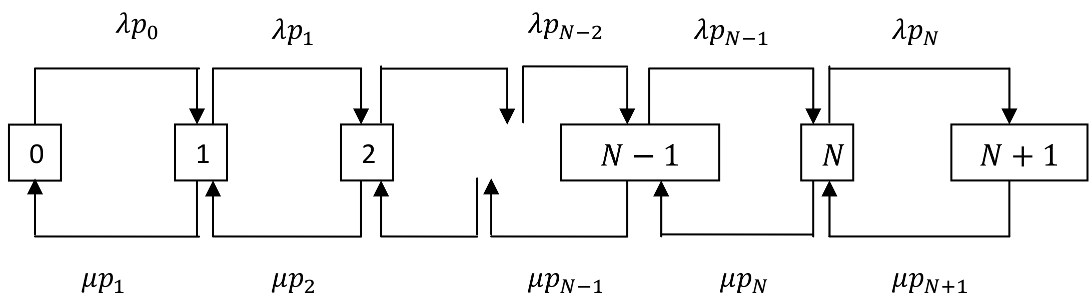

Fig 3.2: The state transition diagram illustrating the birth and death process for $M/M/1$ queue.

If we write the global balance equations for this system, we can recognize a regular pattern emerging:

$$\lambda p_0 = \mu p_1 \quad (1)$$

$$(\lambda + \mu)p_1 = \lambda p_0 + \mu p_2 \quad (2)$$

$$(\lambda + \mu)p_2 = \lambda p_1 + \mu p_3 \quad (3)$$

.

.

.

Using simple algebra we can re-write these equations as:

$$p_1 = \frac{\lambda}{\mu} p_0 \quad (4)$$

$$p_2 = \frac{\lambda}{\mu} p_1 = \left(\frac{\lambda}{\mu}\right)^2 p_0 \tag{5}$$

$$p_3 = \frac{\lambda}{\mu} p_2 = \left(\frac{\lambda}{\mu}\right)^3 p_0 \tag{6}$$

.

.

.

$$p_n = \frac{\lambda}{\mu} p_{n-1} = \left(\frac{\lambda}{\mu}\right)^n p_0 \tag{7}$$

Taking $\frac{\lambda}{\mu}$ as ρ we have,

$$p_n = \rho^n p_0 \tag{8}$$

3.4 M/M/m Model.

In this model, it is assumed that m parallel servers provide the service simultenously (where $m > 1$). All service channels are identical and the service time distribution is exponential with mean μ per unit time. A customer is attended to by only one server, even if some of the other service channels are free. We have the following probability equations:

$$p_0(t + \Delta t) = p_0(t)(1 - \lambda\Delta t) + p_1(t)\mu\Delta t \tag{9}$$

$$p_n(t + \Delta t) = p_n(t)(1 - \lambda\Delta t)(1 - n\mu\Delta t) + p_{n-1}(t)\lambda\Delta t + p_{n+1}(t)(n+1)\mu\Delta t, n < m \tag{10}$$

and

$$p_n(t + \Delta t) = p_n(t)(1 - \lambda\Delta t)(1 - m\mu\Delta t) + p_{n-1}(t)\lambda\Delta t + p_{n+1}(t)\mu\Delta t, \qquad n \geq m \tag{11}$$

Letting $\Delta t \to 0$ in (9),(10) and (11)

$$p_0'(t) = -\lambda p_0(t) + \mu p_1'(t) \tag{12}$$

$$p_n'(t) = -(\lambda + n\mu)p_n(t) + \lambda p_{n-1}(t) + (n+1)\mu p_{n+1}(t), \ \ 0 < n < m \tag{13}$$

and

$$p_n'(t) = -(\lambda + m\mu)p_n(t) + \lambda p_{n-1}(t) + m\mu p_{n+1}(t), \quad n \geq m \tag{14}$$

Letting $t \to \infty$ in (12), (13) and (14) it follows that,

$$-\lambda p_0 + \mu p_1 = 0 \tag{15}$$

$$-(\lambda + n\mu)p_n + \lambda p_{n-1} + (n+1)\mu p_{n+1} = 0 \tag{16}$$

and

$$-(\lambda + m\mu)p_n + \lambda p_{n-1} + m\mu p_{n+1} = 0 \tag{17}$$

Solving (15), (16) and (17) together with,

$$\sum_{n=1}^{\infty} p_n = 1 \tag{18}$$

We obtain the values of p_0 and p_n as follows:

$$p_0 = \left[1 + \frac{(m\rho)^m}{m!\,(1-\rho)} + \sum_{n=1}^{m-1} \frac{(m\rho)^n}{n!}\right]^{-1} \tag{19}$$

$$p_n = \begin{cases} \dfrac{\lambda^n}{n!\,\mu^n} p_{0,} & n=1,2,3,\ldots,m-1 \\ \dfrac{\lambda^n}{m!\,m^{n-m}\mu^n} p_{0,} & n=m,m+1,\ldots,\infty \end{cases} \tag{20}$$

Or equivalently (with $\rho = \frac{\lambda}{m\mu}$)

$$p_n = \begin{cases} \dfrac{(m\rho)^n}{n!} p_{0,} & n=1,2,3,\ldots,m-1 \\ \dfrac{\rho^n m^m}{m!} p_{0,} & n=m,m+1,\ldots,\infty \end{cases} \tag{21}$$

3.5 Measures of performance of the M/M/m model.

$$l_q = \sum_{n=m+1}^{\infty} (n-m)p_n \tag{22}$$

$$= \frac{\rho^{m+1}}{(m-1)!\,(m-\rho)^2} p_0 \tag{23}$$

$$l_s = \sum_{n=0}^{\infty} np_n \tag{24}$$

$$= l_q + \rho$$

$$W_q = \frac{l_q}{\lambda} \tag{25}$$

$$W_s = \frac{l_s}{\lambda} \tag{26}$$

$$= W_q + \frac{1}{\mu}$$

3.6 Research Model

Only three (3) front loaders are available to give service to the trucks and there is no restriction to the number of these trucks to come for service. According to the Kendal Lee notation, the model is as follows:

$$(M/M/3) : (FIFO/\infty/\infty)$$

Where:

- First M is the Markovian arrival rate (assuming Poisson arrival rate, λ).
- Second M is also the Markovian service rate (assuming Exponential service rate, μ).
- The number of servers is 3.
- General discipline is First In First Out.
- Number of customers allowed for service is unlimited.
- Calling population from which customers come from is unlimited.

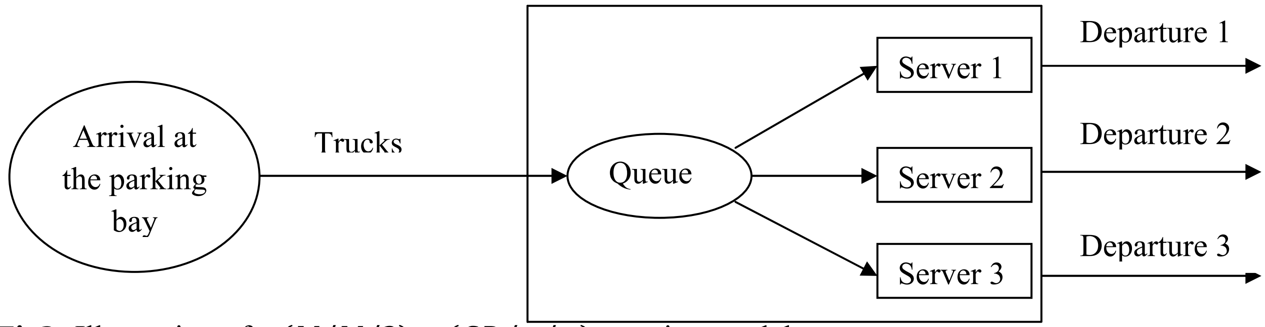

Fig2: Illustration of a $(M/M/3)\ :\ (GD/\infty/\infty)$ queuing model.

4.0RESULTS AND DISCUSSION

Table 1: Data related to arrivals and service of customers.

	Number of trucks Served	Number of Days	Trucks served per day	Trucks served per hour	Time spend from bay to server
Server 1	1501	20	75.1	3.1	10 minutes
Server 2	1421	20	71.1	3.0	10 minutes
Server 3	1550	20	77.5	3.2	10 minutes
	4472				

Number of trucks arrived in 20 days is 4320

The table above shows the total number of trucks arrived in 20 days, the number of trucks served at each server in 20 days and the average number of trucks which could be served per day as well as per hour. Calculating the average of the averages of trucks served per hour, we see that the system has the following parameters:

Average number of trucks served per hour per server (μ) is 3.1.

Average number of trucks arriving at the site per hour (λ) is 9.

Therefore, our queuing model for this system becomes,

$$(9/2.3/3)\ :\ (FIFO/\infty/\infty).$$

Table 2 : Costs to maintain each server during service per hour.

	Cost per hour
Server 1	$50
Server 2	$50
Server 3	$50
Total	**$150**

Table 3: Measures of performance of the system.

Lambda	Mu	Servers	Rho	Ls	Lq	Ws	Wq
9	3.1	3	2.903	31.1	28.2	3.45	3.13

Table 3 above shows that in the system we can expect 32 trucks both at the server and in the queue at any given time, 29 trucks waiting for service (in queue) at any given time which means that our measures are stationary (independent of time). The expected time a truck can be in the system is 3.45 hours that is both the time it can wait for service and the time of service. Furthermore, we expect a truck to wait for at least 3.13 hours in the queue before it gets the service. The inter-arrival time for each truck is expected to be 19.2 minutes, which is the time between being called for service and the actual time of receiving that service. The capacity utilization of each server is 0.968 thus is each server will be busy 96.8% of the time (464.6 hours in 20 days). Each server lie idle only 3.2% of the time (15.4 hours in 20 days). The fact that the expected number of trucks in the queue at any time is 29, it means that all arriving trucks will not be accommodated at the space provided. At least 9 trucks will be out of the parking bay since only 20 trucks get the parking provided.

Table 4: Costs incurred per hour.

	Costs/hour
Customers	$155.47
Servers	$150.00
Total	**$305.47**

Table 4 above shows that the total costs per time period (per hour) incurred by both the servers and customers is $305.47.

Table 5: Values of p(n) for n=0 to 247, else p(n) < .00001

0	0.00743	21	0.01680	41	0.00872	61	0.00453	81	0.00235	101	0.00122
1	0.02158	22	0.01626	42	0.00844	62	0.00438	82	0.00227	102	0.00118
2	0.03132	23	0.01573	43	0.00817	63	0.00424	83	0.00220	103	0.00114
3	0.03031	24	0.01523	44	0.00790	64	0.00410	84	0.00213	104	0.00110
4	0.02933	25	0.01473	45	0.00765	65	0.00397	85	0.00206	105	0.00107
5	0.02839	26	0.01426	46	0.00740	66	0.00384	86	0.00199	106	0.00103
6	0.02747	27	0.01380	47	0.00716	67	0.00372	87	0.00193	107	0.00100
7	0.02659	28	0.01335	48	0.00693	68	0.00360	88	0.00187	108	0.00097
8	0.02573	29	0.01292	49	0.00671	69	0.00348	89	0.00181	109	0.00094
9	0.02490	30	0.01251	50	0.00649	70	0.00337	90	0.00175	110	0.00091
10	0.02410	31	0.01210	51	0.00628	71	0.00326	91	0.00169	111	0.00088
11	0.02332	32	0.01171	52	0.00608	72	0.00316	92	0.00164	112	0.00085
12	0.02257	33	0.01133	53	0.00588	73	0.00305	93	0.00158	113	0.00082
13	0.02184	34	0.01097	54	0.00569	74	0.00295	94	0.00153	114	0.00080
14	0.02113	35	0.01061	55	0.00551	75	0.00286	95	0.00148	115	0.00077
15	0.02045	36	0.01027	56	0.00533	76	0.00277	96	0.00144	116	0.00075
16	0.01979	37	0.00994	57	0.00516	77	0.00268	97	0.00139	117	0.00072
17	0.01915	38	0.00962	58	0.00499	78	0.00259	98	0.00135	118	0.00070
18	0.01854	39	0.00931	59	0.00483	79	0.00251	99	0.00130	119	0.00068
19	0.01794	40	0.00901	60	0.00468	80	0.00243	100	0.00126	120	0.00065
20	0.01736									121+	0.01898

Table 6: Cumulative values of p(n) for n=0 to 247

0	0.00743	21	0.49603	41	0.73842	61	0.86423	81	0.92953	101	0.96343
1	0.02901	22	0.51229	42	0.74686	62	0.86861	82	0.93181	102	0.96461
2	0.06033	23	0.52802	43	0.75503	63	0.87285	83	0.93401	103	0.96575
3	0.09064	24	0.54324	44	0.76293	64	0.87695	84	0.93614	104	0.96685
4	0.11998	25	0.55798	45	0.77058	65	0.88092	85	0.93820	105	0.96792
5	0.14837	26	0.57224	46	0.77798	66	0.88476	86	0.94019	106	0.96896

6 0.17584	27 0.58603	47 0.78514	67 0.88848	87 0.94212	107 0.96996
7 0.20242	28 0.59939	48 0.79207	68 0.89208	88 0.94399	108 0.97093
8 0.22815	29 0.61231	49 0.79878	69 0.89556	89 0.94579	109 0.97186
9 0.25305	30 0.62482	50 0.80527	70 0.89893	90 0.94754	110 0.97277
10 0.27715	31 0.63692	51 0.81155	71 0.90219	91 0.94923	111 0.97365
11 0.30046	32 0.64863	52 0.81763	72 0.90534	92 0.95087	112 0.97450
12 0.32303	33 0.65997	53 0.82351	73 0.90840	93 0.95246	113 0.97532
13 0.34487	34 0.67094	54 0.82921	74 0.91135	94 0.95399	114 0.97612
14 0.36600	35 0.68155	55 0.83472	75 0.91421	95 0.95547	115 0.97689
15 0.38645	36 0.69182	56 0.84005	76 0.91698	96 0.95691	116 0.97764
16 0.40624	37 0.70176	57 0.84521	77 0.91966	97 0.95830	117 0.97836
17 0.42540	38 0.71138	58 0.85020	78 0.92225	98 0.95965	118 0.97905
18 0.44393	39 0.72070	59 0.85503	79 0.92476	99 0.96095	119 0.97973
19 0.46187	40 0.72970	60 0.85971	80 0.92718	100 0.96221	120 0.98038
20 0.47923					121+ 1.0000

Table 5 above shows the probabilities of having n (where n=0 – 247) trucks in the system at any given time. It shows that the probability of having no units (zero units) in the system at any given time is 0.00743 that is 0.74% of the time the servers will be idle. Ideally this means that the chances of having an empty queue are very slim. The probabilities increase as the expected number trucks in the system increase to 3 then falls back but not as much as having no trucks in the system. The system has chances of having as many as 247 trucks at any given time since the probabilities are there but small. Table 6 shows the cumulative probabilities of having n (where n=0 - 247) trucks is the system at any given time. It shows that the probability of having 0-20 trucks in the system is 0.47923 which is even less than 0.5. This tells us that we expect more than 20 trucks in the system at any given time. The probability of having between 32 and 100 trucks in the system at any given time is 0.305 = ($P(32 < X < 100)$, where X stands for trucks. We expect more than 32 trucks in the system at any time since the probability is high. The cumulative probabilities show that more than 20 trucks are in the system at any given time. Therefore, some of the trucks are not accommodated in the space provided by the company. The probability that an arriving truck will have to wait outside the parking bay is 0.521 and the probability that an arriving truck will get a normal parking space is 0.479.

Table 7: Expected costs incurred by servers and customers.

	Expected number of hours in 20 days	**Costs**
Server 1	464.6	$23230.00
Server 2	464.6	$23230.00
Server 3	464.6	$23230.00
customers	13521.6	$67608.00
Total		**$137298.00**

Table 7 shows the total costs incurred by servers during the service and customers while waiting for the service. On average each customer can spend $15.65 while in queue. The total cost for all severs within 20 days is $69690.00.

4.1 Sensitivity Analysis

We considered two scenarios, (a) is to place an additional server in the system, (b) to find an optimal service rate for the current servers so as to accommodate all waiting trucks in the space provided.

(a)

Table 8: Measures of performance of a new system with an addition of one server.

Lambda	Mu	Servers	Rho	Ls	Lq	Ws	Wq
9	3.1	4	2.903	4.15	1.24	0.46	0.14

Table 8 shows the measures of performance of the new system after an additional of one server identical to other three. The results show a decrease in the waiting time and queue length. The length of the queue and the number of trucks in the system has been reduced to 1.24 trucks and 4.15 respectively which is a 96% decrease. On the other hand, the waiting time in the queue and the waiting time in the system have also been reduced to 0.14 hours (8 mins) and 0.46 hours (27.6 mins) respectively. As a result, the space for accommodating trucks in the queue becomes more than sufficient. Each server is busy only 72.6% that is 27.4% of the time it is lying idle.

Table 9: Expected costs incurred by servers and customers of a new system with one server more.

	Expected number of hours in 20 days	**Costs**
Server 1	348.36	$17418.00
Server 2	348.36	$17418.00
Server 3	348.36	$17418.00
Server 4	348.36	$17418.00
customers	604.8	$3024.00
Total		**$72696.00**

In table 9, costs for each incurred by each customer has been reduced drastically $0.70 while waiting and the total costs for servers have decreased from $69690.00 to $69672.00.

(b)

Table 10: Measures of performance of a new system with an adjusted service rate.

Lambda	Mu	Servers	Rho	Ls	Lq	Ws	Wq
9	3.14	3	2.87	22.5	19.6	2.50	2.18

Table 10 shows the measures of performance of the system with an adjusted average service rate for each server. Results show that if could have an average service rate of 3.14 trucks per hour, then the waiting line could be reduced to 20 trucks and the expected number of trucks in the system could be reduced to 23 at any given time. Furthermore, the waiting time in queue and in the system could be reduced to 2.18 hours and 2.5 hours respectively. To come to this service rate, more 17 trucks should be served in 20 days by each server. Each server is 95.7% busy.

Table 11: Expected costs incurred by servers and customers of a new system with an adjusted service rate.

	Expected number of hours in 20 days	**Costs**
Server 1	459.2	$22960.00
Server 2	459.2	$22960.00

Server 3	459.2	$22960.00
customers	9417.6	$47088.00
Total		**$115968.00**

Table 11 shows the total costs incurred by servers and customers after an adjustment to the service rate for each server to 3.14 trucks per hour. The total cost for servers has decreased from $69690.00 to $68880.00. Each customer can spend $10.90 while waiting in the queue.

5.0 CONCLUSION

In this research, the use of queuing models was very useful in achieving the main objective. The results show that currently 29 trucks will be in queue waiting for service and 32 trucks will be in the system at any given time. 9 trucks at any given time will be outside the parking bay, which means no space for parking is provided to them. The chance that an arriving truck gets the parking is 0.479 which is less than a half. The current situation at the company needs to be addressed in at least two options. One is to add an extra server to the loading site to minimize queue length of customers by 96%. Total costs of the servers would be reduced by $22.00 provided an additional server is not bought but donated to the company. If the company is going to buy an additional server, then costs will be higher than expected. Total costs for customers would be reduced to only $0.70 while in queue.

The second option is to increase the service rate to 3.14 trucks per hour. An additional of 17 trucks could be required to be served within 20 days by each server. Using his system, trucks in the queue would be reduced to 20 and 23 in the system. The system will break even as it expects to accommodate all arriving trucks in the parking bay since only 20 trucks will be accommodated in the parking lot. Waiting time by customers in the queue and in the whole system would be reduced by 30%. Total costs incurred by customers while in the queue are considerable since they would wait for at least 2 hours 11 minutes as compared to the current scenario where they spend at least 3 hours 8 minutes.

Given these options we would advise the company either to go for the second option which is to increase the service rate to 3.14 or to increase the parking space for trucks to at least 30 spaces but less than 50 spaces since ground for doing that is limited.

REFERENCE

[1] Albin, S.L., Barrett, J., Ito, D. and Mueller, J.E. (1990) A queueing network analysis of a health center. Queueing Systems, 7, 51-61.

[2] Bernard W. Taylor III (2007), Introduction To Management Science, 9th Edition.

[3] C. R. Kothari (1982), An Introduction to Operations Research, 1st Edition, Oxford University.

[4] Fomundam. S. and Herrmann. F, (2007). A Survey of Queuing Theory Applications in Healthcare. The Institute for Systems Research.

[5] Frederick S. Hillier and Gerald J. Lieberman (2001), Introduction To Operations Research, 7th Edition, Stanford University – India.

[6] Green, L. (2006a) Queueing analysis in healthcare, in Patient Flow: Reducing Delay in Healthcare Delivery, Hall, R.W., ed., Springer, New York, 281-308.

[7] Heyman, D. P. (1968). "Optimal operating policies for M/G/1 queueing systems", Operations Res.16, 362-382.

[8] Hillier, F. S. (1963). "Economic models for industrial waiting line problems", Management Sci. 10, 119-130.

[9] Joseph S. Martinich (1997), Introduction To Operations Management, 1st Edition.

[10] Kelly, F. P. (1979). Reversibility and Stochastic Networks, John Wiley & Sons, New York.

[11] Kendall, W. S. (1998). Perfect simulation for the area-interaction point process. In Probability Towards 2000 (L. Accardi and C. C. Heyde, eds.) 218–234. Springer Lecture Notes in Statistics 128, Springer Verlag, New York.

[12] Li, L. and X. Liling, 2000. Modeling staffing flexibility: case of china. European Journal of Operational Research, 124: 255-266,

[13] Lindley, D. V. (1952). "The theory of queues with a single server", Proc. Comb. Phil. Soc. 48, 277-289.

[14] M. Tabari, Y. Gholipour-Kanani, M. Seifi-Divkolaii and R. T.Moghaddam. Application of the Queuing Theory to Human Resource Management. World Applied Sciences Journal 17 (9): 1211-1218, 2012

[15] McClain, J.O. (1976) Bed planning using queuing theory models of hospital occupancy: a sensitivity analysis. Inquiry, 13, 167-176.

[16] Nosek, Jr., R.A. and Wilson, J.P. (2001) Queuing theory and customer satisfaction: a review of terminology, trends, and applications to pharmacy practice. Hospital Pharmacy, 36, 275-279.

[17] Pollaczek, F. (1934). "Uber das Warteproblem", Math. Zeitschrift, 38, 492-537.

[18] R. Panneerselvann (2002), Operations Research, 1st Edition, Pondichery University, Prentice Hall

[19] William J. Stevenson (2005), Operations Management, 1st Edition.

[20] Worthington, D.J. (1987) Queueing Models for Hospital Waiting Lists. The Journal of the Operation Research Society 38, 413-422.