Journal of Theoretical and Applied Information Technology
31st March 2021. Vol.99. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREEMPTIVE-RESUME PRIORITY QUEUE SYSTEM WITH ERLANG SERVICE DISTRIBUTION

O.R AJEWOLE[1], C.O. MMDUAKOR[2], E.O ADEYEFA[3], J.O. OKORO[4] and *T.O. OGUNLADE[5]

Department of Mathematics, Federal University, Oye-Ekiti, Ekiti State Nigeria[1,2,3]
Physical Science Department, Mathematics Unit, College of Science and Engineering, Landmark University, Omu-Aran, Kwara State, Nigeria (+2348067620663)[4]
Department of Mathematics, Ekiti State University, Ado-Ekiti, Nigeria[5]
Email: oghenekevwe.ajewole@fuoye.edu.ng[1], chika.mmaduakor@fuoye.edu.ng[2], emmanuel.adeyefa@fuoye.edu.ng[3], okoro.joshua@lmu.edu.ng[4] and *topsmatic@gmail.com[5]
*Corresponding author

ABSTRACT

This paper describes a preemptive-resume priority queue system that assumes Poisson arrivals and a single server facility. Various priority queueing models have been proposed in literature to explain the behaviour of different kinds of queues commonly observed at various service facilities. Majority of these models assume exponential service times. However, when the service time of a given facility is processed in more than one stage and service is in a sequential order (an often encountered scenario in practical situation), the need for a service distribution that can represent this becomes necessary. Hence in this study, the service time distribution is assumed to have Erlang service times. It is assumed that there are two classes of priority levels of which one has preemptive-resume priority over the other. The mean value theorem is applied in determining the performance measures of the higher priority queue. The busy period of the higher priority class assuming First Come First Serve principle and its associated moments is derived. We also evaluate the average number of customers in the system for the lower priority level and other performance measures like the mean sojourn time in the system. Subsequently, the impact and significance of preemptive scheduling is investigated with the use of real life data.

Keywords: *Preemptive-Resume Priority Queue, Erlang Service Distribution, Completion Time, Busy Period, Single Server.*

1. INTRODUCTION

Priority queueing systems have received much attention due to its usefulness in modeling practical situations such as mobile telecommunication networks, production systems and manufacturing. The fact that it allows messages of different classes to be given different quality of service has made it an interesting area of research.
In some queueing situations, customers are grouped in priority classes based on some attributes which could include urgency or preferential treatment. This is known as priority service.

Systems where arrival of higher priority customer brings about an interruption in service of lower priority customer are referred to as preemptive-priority queueing systems. In such situations, service of the lower priority customer resumes again after the higher priority customer has been served.

Two categories of preemptive priority queue system exist: preemptive-resume and preemptive non-resume. For the preemptive-resume priority case, after interruptions occur and are cleared, service resumes again and continues from the point of interruption. preemptive non-resume priority is the situation where interrupted service must begin from scratch when it resumes back

The earliest results on priority queueing models is credited to Alan [1], who introduced head-of-the-line priority queue with multiple priority levels and derived the equilibrium expected waiting times. This was done for two types of systems. The system with single channel and the system with multiple channel with both having exponential service times. This study is focused on preemptive priority hence we discuss some results of some earlier authors

Harrison and Lee [2] studied systems of queues with preemptive priority and obtained explicit expressions for some of their statistical properties. The study in [3] is an extension of the

work in [2] which finds convenient computing procedures to the problem of probabilities in equilibrium for the case of two classes. The waiting line process (also known as queue process) for a queue by imbedded Markov chain method and renewal theory was studied in [4]. The rate of arrival of customers is Poisson and the duration of service is arbitrary. The results in [5] was an extension of [4] to the system with interruptions. But in this case, the completion times was substituted for service times for the lower priority class since it accommodates interruptions.

In [6], Fumiaki studied a preemptive priority queue with server vacations. The system has two queues with MAP arrivals with one of them being served with preemptive resume priority over the other. Performance measures which include the distribution of the time of waiting and queue size distributions are represented in matrix exponential forms. The second queue is stable hence the time of waiting and queue size distributions are defined and the mean waiting time and mean queue size are obtained. The work in [7] discussed an M/M/1 priority queue system with two categories of customers:- type 1 and type 2. Type 1 customers are the higher priority customers and type 2 customers are the lower priority customers. Applying the Mean Value Theorem and the PASTA (Poisson Arrival See Time Average) property, results were obtained for the system which follows the preemptive priority rule discipline and the system that follows the priority rule with non-preemptive priority mechanism. For the preemptive-priority rule, the mean sojourn time in the system $E(S_1)$, and average customer size within the system $E(L_1)$, for type 1, are given respectively as

$$E(S_1) = \frac{1/\mu_1}{1-\rho_1} \quad (1)$$

and

$$E(L_1) = \frac{\rho_1}{1-\rho_1} \quad (2)$$

For the type 2 customers, the performance measures obtained, include:

$$E(S_2) = \frac{1/\mu_1}{(1-\rho_1)(1-\rho_1-\rho_2)} \quad (3)$$

and

$$E(L_2) = \frac{\rho_2}{(1-\rho_1)(1-\rho_1-\rho_2)} \quad (4)$$

where $E(S_2)$ denotes the mean sojourn time of type 2 customer in the system and $E(L_2)$ represents the mean number of type 2 customers in the system.

The study [8] gave results of a priority queueing system that uses the preemptive priority rule for service with retrials and single working vacation. The model is solved by using Matrix geometric technique. The average customer size in the orbit and the probability that the server is either idle or busy are obtained. [9] discussed preemptive-resume priority queue with Markovian arrival process and the concept of retrial queue. The effect of the preemptive rule and the correlation in the inter arrival times was considered. [10] modeled a call-centre with priority service. In this study, the type 1 and type 2 calls had a retrial orbit they can access whenever the system appears busy upon arrival to the system. Whenever the system is free again, type 1 calls compete for service hence it does not follow a FCFS (first come first served) principle but type 2 customers have lower priority because of the preemptive priority rule. The service duration was exponential with different rates. Matrix-analytic method was used in analyzing the system.

In 2014, Richa [11] discussed an M/M/1 two priority class queue with poisson arrivals and exponential service duration. In the study, Runge-Kutta (R-K) method of fourth order was applied to examine the performance measures of the system some of which include, probability of the server's idle state, busy state or broken down state. At time t, the average customer size in the system was also considered. Analysis of an M/M/1 priority queue with preemptive-resume and priority classes was carried out in [12]. The classes have exponentially distributed service duration. A study on the customer behaviour that relies on the parameters of the system and the introduction of priority costs was also carried out.

The study [12] applied the notion of completion times introduced in [5] to describe the service duration of the lower priority customer for a preemptive-resume priority queue system. That is, the sum of the corresponding time of service and interruptions of the higher priority customer. It was denoted by

$$C = S + \sum_{n=0}^{N} S_{b,n} \quad (5)$$

where, N represents the total sum of interruptions during customers service, $S_{b,n}$ denotes the duration of nth interruption (with

$S_{b,0} = 0$) and S, the corresponding service time. For this queue system, the average number of lower priority customers and the mean sojourn time were obtained.

The models discussed, are only suitable for systems which have one phase. When the service time is in more than one stage, the need to use a distribution that describes such situation becomes necessary, hence the study of priority system with Erlang service distribution. The aim of this study therefore is to describe a preemptive priority queue with Erlang distribution service times (an example of phase distribution). The Laplace transform representation for the generating function of duration of busy period for M/E_r/1 (the higher priority queue) including the associated moments will be obtained. This is a new result.

Subsequently, the performance measures for the system size and sojourn time for the lower priority queue in equilibrium will be determined. The results obtained are applied to a real life data.

2. MODEL DESCRIPTION:

We study the preemptive priority queue system with service times having identical exponential phases (also referred to as the Erlang service). It is assumed that two queues arrive the system and their jobs processed by a single channel facility. The first queue represented by the random variable $X_1(t)$, is the M/E_r/1 queueing system that the priority has no effect on since the customers in that category are served on first come first served basis. The arrival rate λ_1 is poisson and the mean time of service is Erlang with mean E(S_1) = $\frac{r}{\mu_1}$.The second moment is given by $Var(S_1)$=$\frac{r}{(\mu_1)^2}$ with $E\left(S_1{}^{(2)}\right) = \frac{r(r+1)}{(\mu_1)^2}$ Customers in the first queue are referred to as the higher priority customers. For the second queue $X_2(t)$, the arrival rate λ_2 is poisson. The service duration is the sum of the corresponding service duration of the customers together with the busy period of interruptions. This is called the completion time and will be shown in the methodology. The corresponding service is Erlang distributed with mean E(S_2) =$\frac{r}{\mu_2}$ and second derivative $E\left(S_2{}^{(2)}\right) = \frac{r(r+1)}{(\mu_2)^2}$ .Whenever an $X_1(t)$ arrives the system during the service of $X_2(t)$, the service of $X_2(t)$ is halted until the service of $X_1(t)$ is completed. When the service of $X_1(t)$ is completed, the service of $X_2(t)$ is continued from where it was left off. This makes $X_2(t)$ a queueing process which accommodates interruption. The customers in the second queue are referred to as lower priority customers.

3. METHODOLOGY

We give the definitions of some useful terminologies

3.1 Busy period of the higher priority M/E_r/1 queue in equilibrium

In this study, the busy period is used to describe the duration of interruption of the higher priority customer. This is needed to investigate the completion time of lower priority queue. The Busy period for an M/G/1 queue as given in [13] is expressed in the form:

$$\tilde{B}(s) = \tilde{X}[S + \lambda_1 - \lambda_1\tilde{B}(s)] \quad (6)$$

with $\tilde{X}(s)$ representing the laplace transform of time of service

The laplace transform for Erlang distribution as used in [7] is:

$$\tilde{X}(s) = \left(\frac{\mu_1}{\mu_1+s}\right)^r \quad (7)$$

Applying (7) to (6), we have

$$\tilde{B}(s) = \left(\frac{\mu_1}{\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s)}\right)^r \quad (8)$$

which is the laplace Steiltjes transform representation of the M/E_r/1 busy period.

The first moment (expectation) for the queue busy period with Erlang distribution $E(B)$ is obtained by differentiating (8).

$$B^{(k)}(0) = \frac{d^k}{ds^k}B^{(k)}(s)|_{s=0} = (-1)^k E(B^{(k)})$$

k = 1,2,…

where $B^{(k)}(0)$ denotes the kth derivative of $\tilde{B}(s)$

$$\frac{d}{ds}\frac{\tilde{B}(s)}{{\mu_1}^r} = r[1 + \lambda_1 B^{(1)}(s)][\mu_1 + s + \lambda_1 - \lambda_1\tilde{B}(s)]^{-(r+1)}$$

$$= \frac{r[1+\lambda_1 B^{(1)}(s)]}{[\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s)]^{(r+1)}}|_{s=0}$$

$$= \frac{{\mu_1}^r r[1+\lambda_1 E(B)]}{{\mu_1}^{r+1}} \quad (9)$$

$$\mu_1 E(B) = r[1 + \lambda_1 E(B)]$$

$$\mu_1 E(B) - \lambda_1 E(B) = r$$

$$E(B)[\mu_1 - \lambda_1] = r$$

$$E(B) = \frac{r}{\mu_1 - \lambda_1 r} \quad (10)$$

we differentiate (8) twice, substituting s = 0 to obtain $E(B^2)$

$$\frac{d^2}{ds^2}\tilde{B}(s)|_{s=0} = \frac{d^2(\tilde{B}(s))}{ds^2 r{\mu_1}^r} = \frac{d}{ds}\frac{1+\lambda_1 B^{(1)}(s)}{[\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s)]^{(r+1)}}$$

$$\frac{E(B^2)}{r{\mu_1}^r} = \frac{(\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s))^{r+1}(\lambda_1 B^{(2)}(s))+\left(1+\lambda_1 B^{(1)}(s)\right)(r+1)[\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s)]^r\left((1+\lambda_1 B^{(1)}(s)\right)}{(\mu_1+s+\lambda_1-\lambda_1\tilde{B}(s))^{2(r+1)}}|_{s=0} \quad (11)$$

$$\frac{E(B^2)}{r{\mu_1}^r} = \frac{{\mu_1}^{r+1}\lambda_1 E(B^2)+[1+\lambda_1 E(B)]^2(r+1){\mu_1}^r}{{\mu_1}^{2(r+1)}} \quad (12)$$

Where $B^{(2)}(s) = E(B^2)$

Applying (10), we obtain

$$\frac{E(B^2)}{r{\mu_1}^r} = \frac{{\mu_1}^{r+1}\lambda_1 E(B^2)+[\frac{\mu_1}{\mu_1-\lambda_1 r}]^2(r+1){\mu_1}^r}{{\mu_1}^{2(r+1)}}$$

$$E(B^2) = \frac{r\lambda_1 E(B^2)(\mu_1-\lambda_1 r)^2 + r(r+1)\mu_1}{\mu_1(\mu_1-\lambda_1 r)^2}$$

$$E(B^2)\mu_1(\mu_1-\lambda_1 r)^2 - r\lambda_1 E(B^2)(\mu_1-\lambda_1 r)^2 = \mu_1 r(r+1)$$

$$E(B^2)(\mu_1-\lambda_1 r)^2[\mu_1 - r\lambda_1] = \mu_1 r(r+1)$$

$$E(B^2) = \frac{\mu_1 r(r+1)}{(\mu_1-\lambda_1 r)^3}$$

$$E(B^2) = \frac{r(r+1)}{\mu_1^2(1-\rho_1)^3} \quad (13)$$

The second moment (variance) is obtained by using the formular

$$Var(B) = E(B^2) - (E(B))^2 \quad (14)$$

Thus, we have

$$Var[B] = \sigma^2(B) = \frac{r(r+1)}{\mu_1^2(1-\rho_1)^3} - \left[\frac{r}{\mu_1(1-\rho_1)}\right]^2$$

$$= \frac{r}{\mu_1^2(1-\rho_1)^2}\left[\frac{r+1}{(1-\rho_1)} - r\right] \quad (15)$$

The next definition is the notion of completion time for a preemptive-resume priority queue

3.2 Completion time [5]: We define the completion time as the period of service completion for which there are interruptions during service. Thus for the preemptive-resume priority queue, we have

$$C = S_2 + \sum_{i=0}^{N}(B_i) \quad (16)$$

N represents the interruptions that occur while a customer is in service and B_i is busy period for the ith interruption to occur during the service time S_2 (with $B_{(0)} = 0$). The completion time has its laplace Steiltjes transform given below

$$\tilde{C}(s) = E(e^{-sC}) \quad (17)$$

In this study, the completion time is used to describe the lower priority customer's service since its total service duration is the sum of the corresponding service and the interruptions made by the higher priority customer.

The next theorem shows a relation between the laplace Steiltjes transform of the service time S_2 and the busy period $B(s)$. It was employed in [12] without proof. Here the proof is given

Theorem (1) [12]: The Laplace-Steiljes transform of the completion time is given as

$$\tilde{C}(s) = \tilde{S}_2\{s + \lambda_1 - \lambda_1\tilde{B}(s)\} \quad (18)$$

where $\tilde{S}_2$ denotes Laplace transform of service time

$$E(C) = E(S_2)\{1 + \lambda_1 E(B)\} \quad (19)$$

$$E(C^2) = E(S_2^{(2)})\{1 + \lambda_1 E(B)\}^2 + \lambda_1 E(S_2)E(B^2) \quad (20)$$

$E(C)$ denotes the expectation of the completion duration and $E(C^2)$ the second derivative of the completion time

Proof: From (11), we know that

$$C = S_2 + \sum_{i=o}^{N}(B)_i$$

is the completion time

The laplace Steiltjes transform of the completion time C is also given as

$$\tilde{C}(s) = E[e^{-sC}]$$

By conditioning on S_2 and taking the expectation,

$$\tilde{C}(s) = E(e^{-sC}) = \int_{t=0}^{\infty} E(e^{-sC}\,|\,S_2 = t)\, f_S(t)dt \quad (21)$$

is obtained

Conditioning on N,

$$E(e^{-sC} \mid S_2 = t) = \sum_{n=0}^{\infty} E(e^{-sC}|S_2 = t, N = n)\mathrm{P}(\mathrm{N} = \mathrm{n}|S_2 = \mathrm{t})$$

Where N follows the poisson distribution with

$$P(N = n) = \frac{e^{-\lambda_1 t}\lambda_1 t^n}{n!}$$

Thus,

$$E(e^{-sC} \mid S_2 = t) = \sum_{n=0}^{\infty} E\left(e^{-s(S_2+B_1+B_2+\cdots B_n)}\middle|S_2 = t, N = n\right).\frac{(\lambda_1 t)^{\mathrm{n}}}{\mathrm{n!}}\mathrm{e}^{-\lambda_1 \mathrm{t}} \quad (22)$$

$$= \sum_{n=0}^{\infty} E(e^{-s(t+B_1+B_2+\cdots D_n)})\frac{(\lambda_1 t)^{\mathrm{n}}}{\mathrm{n!}}\mathrm{e}^{-\lambda_1 \mathrm{t}}$$

$$= \sum_{n=0}^{\infty} E(e^{st}.e^{-sB_1}\ldots.e^{-sB_n})\frac{(\lambda_1 t)^{\mathrm{n}}}{\mathrm{n!}}\mathrm{e}^{-\lambda_1 \mathrm{t}}$$

$$= e^{-(s+\lambda_1-\lambda_1 B(s))t} \quad (23)$$

And hence,

$$\tilde{C}(s) = \int_{t=0}^{\infty} e^{-(s+\lambda_1-\lambda_1 B(s))t} f_S\, dt$$
$$= \tilde{S_2}\big(s + \lambda_1 - \lambda_1 B(s)\big)t \quad (24)$$

To obtain the first moment,

$$E(\tilde{C}) = -\frac{d}{ds}\tilde{C}(s)|_{s=0}$$
$$= \tilde{C}^{(1)}(0) = -E(C)$$
$$= S_2{}^{(1)}(0)(1 - \lambda_1 E(B^{(1)}(0))$$
$$= -E(C) = -E(S_2)(1 + \lambda_1 E(B)$$
$$E(C) = E(S_2)(1 + \lambda_1 E(B) \quad (25)$$

To obtain the second moment of the Completion time, we differentiate $\tilde{C}(s)$ twice,

$$E(C^2) = \tilde{C}^{(2)}(0)$$
$$= \tilde{S_2}^{(2)}(0)[1 - \lambda_1 B^{(1)}(0)]^2 + \tilde{S_2}^{(1)}[-\lambda_1 B^{(2)}(0)]$$
$$E(C^2) = E(S_2{}^{(2)})[1 + \lambda_1 E(B)]^2 + \lambda_1 E(S_2)E(B^2) \quad (26)$$

is obtained

3.3 Traffic Intensity for the higher priorityqueue

The traffic intensity or occupation utilization for the higher priority queue is used to describe the busy period of the server for the higher priority queue. This is written as

$$\rho_1 = \lambda_1 E(S_1) \quad (27)$$

where $E(S_1)$, represents the mean sojourn time of the higher priority customer. For a system in equilibrium, it is required that $\rho_1 < 1$ and thus for the time independent system, $X_1(t)$ is represented by X_1

3.4 The performance measures of the higher priority queue

The results for the performance measures for the higher priority queue are known results. The results may as well be derived using the mean value approach. To the higher priority queue, the lower priority queue does not exist because of the preemptive-resume priority rule. An arriving customer has to wait for the customer in the queue, and for the one in service if the server is busy, since the service follows the first come first serve principle (FCFS). According to the PASTA property, the probability that the arriving customer finds a customer in service, is equal to the fraction of time the customer is busy i.e. ρ_1.

Hence,

$$E\big(W_{(1,q)}\big) = E\big(L_{(1,q)}\big)E(S_1) + \rho_1 E(S) \quad (28)$$

where $E(S)$denotes the mean service time of the customer in service and $E(S_1)$ the mean service time of the customers in $X_1(t)$. $E(L_{(1,q)})$ and $E(W_{(1,q)})$ represent the mean number of customers and the mean waiting time in queue for $X_1(t)$ respectively. If the server is busy on arrival, then with probability $\frac{1}{r}$ the is busy with the first phase of the service time, also with probability $\frac{1}{r}$ he is busy with the second phase, and so on. Thus,

$$E(S) = \frac{1}{r}.\frac{r}{\mu_1} + \frac{1}{r}.\frac{r-1}{\mu_1} + \cdots + \frac{1}{r}.\frac{1}{\mu_1}$$
$$= \frac{r+1}{2}.\frac{1}{\mu_1} \quad (29)$$

substituting (29) into (28) and replacing $E(S_1)$ with $\frac{r}{\mu_1}$ yields

$$E\big(W_{(1,q)}\big) = E\big(L_{(1,q)}\big)\frac{r}{\mu_1} + \rho_1\frac{r+1}{2}.\frac{1}{\mu_1} \quad (30)$$

And by Little's law

$$E\big(L_{(1,q)}\big) = \lambda_1 E\big(W_{(1,q)}\big) \quad (31)$$

Thus,

$$E\big(W_{(1,q)}\big) = \frac{\rho_1}{(1-\rho_1)}\frac{(r+1)}{2\mu_1} \quad (32)$$

The mean sojourn time of $X_1(t)$ in the system $\big[E\big(W_{(1)}\big)\big]$ is obtained using the relation

$$E\big(W_{(1)}\big) = E\big(W_{(1,q)}\big) + E(S_1)$$
$$E\big(W_{(1)}\big) = \frac{\rho_1}{(1-\rho_1)}\frac{(r+1)}{2\mu_1} + \frac{r}{\mu_1}$$

Using Little's law, the mean number of $X_1(t)$ in the system $\big[E\big(L_{(1)}\big)\big]$is obtained

$$E\big(L_{(1)}\big) = \lambda_1 E\big(W_{(1)}\big) = \frac{\lambda_1\rho_1}{(1-\rho_1)}\frac{(r+1)}{2\mu_1} + \frac{\lambda_1 r}{\mu_1} \quad (34)$$

3.5 Traffic Intensity for the lower priority queue

The traffic intensity ρ_2 for the lower priority class $X_2(t)$is the product of the arrival rate of the lower priority class λ_2 and the completion time since the completion time represents the service time of the lower priority customer. This is denoted by

$$\rho_2 = \lambda_2 E(C) \quad (34)$$

For stability, it is required that $\rho_2 < 1$ and thus for the time independent system, $X_2(t)$ is represented by X_2

Theorem (2) [5]: For a system with interrupted service, the customer size of the lower priority queue is given by the probability generating function

$$Q(z) = \frac{(1-\rho_2)(1-z)\tilde{C}\{\lambda-\lambda z\}}{\tilde{C}\{\lambda-\lambda z\}-z} \quad (35)$$

for a system in steady state.

$\tilde{C}(s)$ is the laplace transform of the Completion time duration and $Q(z)$ is the probability generating function of M/G/1 queue with Completion time C. The mean of (35) represents the queue length for the lower priority queue which is given below

$$E(L_2) = \rho_2 + \frac{\rho_2^2}{2(1-\rho_2)}\frac{E(C^2)}{(E(C))^2} \quad (36)$$

3.6 Performance measures of the lower priority M/E_r/1 queue in equilibrium

The Completion time is the service time of the lower priority queue. For a preemptive-resume priority system, the mean completion time as shown in (19) is given below

$$E(C) = E(S_2)[1 + \lambda_1 E(B)] \quad (37)$$

Hence, to obtain the mean Completion time, we insert the mean service time for the lower priority queue $E(S_2)$, and the mean busy period $E(B)$into(37) and we obtain

$$E(C) = \frac{r}{\mu_2}\left(1 + \frac{r\lambda_1}{\mu_1 - r\lambda_1}\right) = \frac{r}{\mu_2(1-\rho_1)} \quad (38)$$

Applying this to the traffic intensity, $\rho_2 = \lambda_2 E(C)$ we have

$$\rho_2 = \frac{\lambda_2 r}{\mu_2(1-\rho_1)} \quad (39)$$

For the computation of the second derivative of the completion time$E(C^2)$, we use (20), which is given by

$$E(C^2) = E\left(E(S_2)^{(2)}\right)[1 + \lambda_1 E(B)]^2 + \lambda_1 E(S_2)E(B^2) \quad (40)$$

Thus plugging the mean and second derivative of the service times and busy period as shown in (10) and (13) into (40), we have

$$E(C^2) = \frac{r(r+1)}{\mu_2^2}\frac{1}{(1-\rho_1)^2} + \frac{r(r+1)}{\mu_2}\frac{\rho_1}{\mu_1(1-\rho_1)^3} \quad (41)$$

For the computation of the average customer number in the lower priority class, we substitute (38), (39) and (41) into (36). Thus, we have

$$E(L_2) = \rho_2 + \frac{r(r+1)\rho_2^2}{2(1-\rho_2)}\left[1 + \frac{\mu_2}{\mu_1}\frac{\rho_1}{(1-\rho_1)}\right] \quad (42)$$

To compute the mean sojourn time in the system for the lower priority class, $E(W_2)$, we use little's law where, $E(W_2) = \frac{E(L_2)}{\lambda_2}$

$$E(W_2) = \frac{r}{\mu_2} + \frac{\lambda_2 r^3(r+1)}{2\mu_2^2(1-\rho_2)}\left[1 + \frac{\mu_2}{\mu_1}\frac{\rho_1}{(1-\rho_1)}\right] \quad (43)$$

With $\rho_1 = \frac{r\lambda_1}{\mu_1}$, and $\rho_2 = \lambda_2 E(C) = \frac{\lambda_2 r}{\mu_2(1-\rho_1)}$ (44)

To obtain $W_{2,q}$the mean waiting in the queue, we have the relation

$$E(W_2) = E(W_{2,q}) + E(C)$$
$$E(W_{2,q}) = E(W_2) - E(C)$$
$$E(W_{2,q}) = \frac{r}{\mu_2}\left(1 + \frac{\rho_2 r^2(r+1)}{2\mu_2(1-\rho_2)}\left[1 + \frac{\mu_2}{\mu_1}\frac{\rho_1}{(1-\rho_1)}\right] - \frac{1}{(1-\rho_1)}\right) \quad (45)$$

Then, the Little's law is used to get the average customer size in the queue

$$E(L_{2,q}) = \lambda_2 E(W_{2,q})$$

$$E(L_{2,q}) = \rho_2\left[1 + \frac{\rho_2 r^2(r+1)}{2\mu_2(1-\rho_2)}\left[1 + \frac{\mu_2}{\mu_1}\frac{\rho_1}{(1-\rho_1)}\right] - \frac{1}{(1-\rho_1)}\right] \quad (46)$$

4. APPLICATION

In this section, the effect of preemptive-resume priority queueing on the queue system comprising of two classes of customers is investigated of which one of the classes has priority over the other. We acquired the data set from the call records of clients who receive frequent calls on Telephone call provider (which we will refer to as the regular calls). A line is referred to as a server to which various telecommunication networks want to hook up with but we consider two of these, the internet call (with little cost) and the regular calls. It was observed that the regular calls have preemptive priority over the internet calls such that if there is an incoming regular call, it interrupts an internet call currently going on at that time. In this scenario, the internet call is suspended and the regular call is received. After the regular call has ended, the internet call is completed provided there is no other incoming regular call.

In this example, the regular calls (X_1) are referred to as the first class customers, while the internet calls $(X_2)$are referred to as second class customers. The five data sets give values for the arrival rate (computed from the arrival times) and service times for each client as extracted from the call records on their lines.

The performance measures which include the mean sojourn time (W_i) and the average customer number (L_i), are given in equations (33),(34),(42) & (43)

Journal of Theoretical and Applied Information Technology
31st March 2021. Vol.99. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For each day, data was collected for over a 15-hour period, from 6am to 9pm. Arrivals before 6am and after 9pm were not considered. From the data

- The performance measures computed include:
  - = average service duration of class i
  - = arrival rate of class i into the system
  - ρ_i= traffic utility of class i
  - $E(L_i)$ = average number of class $i$customers in the system
  - $E(W_i)$= mean waiting time of class $i$customers in the system

where i = 1,2 – class 1 or higher priority customer and class 2 or lower priority customer

4.1 DATA ANALYSIS

Table 4.1: The Arrival rates and Average Service times (in minutes) for 5 data sets

DATA SET	ρ_1	ρ_2	$E(L_1)$	$E(W_1)$	$E(L_2)$	$E(W_2)$
1	0.203	0.102	0.238	0.793	0.229	0.573
2	0.117	0.067	0.227	2.274	0.111	0.553
3	0.176	0.075	0.201	2.012	0.117	0.590
4	0.294	0.096	0.375	1.877	0.324	1.078
5	0.457	0.088	0.713	2.378	0.689	1.722

Table 4.1 shows the arrival rates and average service times for classes 1 and 2 customers. We observe that the mean service times for the class 2, (μ_2) is larger than that of class 1, $(\mu_1)$for the data sets considered. This is because of the interruptions caused by the class 1 customers. The mean service time for the class 2 customers is the sum of the corresponding service time and the duration of interruptions that occurred during the service. The arrival rates for class 1 customers, is also lower than that of the class 2 customers. This is as a result of the small interval between arrivals of class 1 customers since they have a higher priority service over class 2.

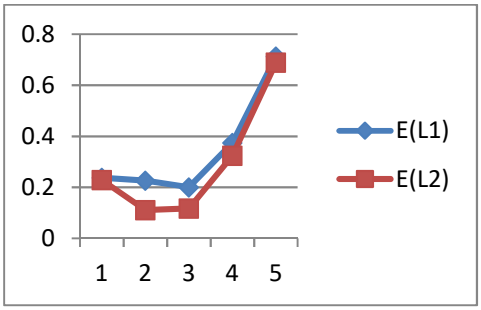
Table 4.2: Traffic intensities and Performance measures for class 1 and class 2

DATA SET	λ_1	λ_2	μ_1	μ_2
1	0.3	0.4	4.42	14.84
2	0.1	0.2	2.56	10.17
3	0.1	0.3	1.70	14.59
4	0.2	0.3	2.04	13.26
5	0.3	0.4	1.97	25.01

Table 4.2 displays the values for the traffic intensities for both classes 1 and 2 customers for the 5 data sets including their mean performance measures. The values for the traffic intensities $\rho_i(i = 1,2)$ are obtained by substituting the values in table (3.1) into (40) and the measures, $E(L_i)$ and $E(W_i)(i = 1,2)$ are obtained, by applying the values in table (3.1) to equations (33), (34), (42) and (43).
Observing the results in table 4.2, we see the effect of priority rule. That is the mean number of class 2 customers in the system, $E(L_2)$ is severely reduced due to the preemptive-resume priority rule for the service when compared with the values obtained for mean number of class 1 customers, $E(L_1)$.

Also, the mean sojourn time of class 1 customers $E(W_i)$ have higher values than that of class 2 customers $E(W_i)$. This is contrary to expectation. The reason may be due to the reneging of the class 2 customers on seeing the length of service time . On that note, they will definitely have higher sojourn time in the system.

4.2 Charts Representing The Performance Measures



Journal of Theoretical and Applied Information Technology
31st March 2021. Vol.99. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4.1 Average Number of Customers $[E(L_i)]$ Plotted against Data Sets 1-5

Figure 4.1 above represents the plot of the mean customer number in the system against the five data sets. It is observed that $E(L_1)$ have greater values than $E(L_2)$. The implication of this is that on the average, the number of customers found on queue 1 is usually more than queue 2 and this is due to the preemptive priority rule. For $E(L_1)$ and $E(L_2)$, the values drop from data set 1 to data set 2 and then to set 3. It increases at data set 4 and then increases again at data set 5.

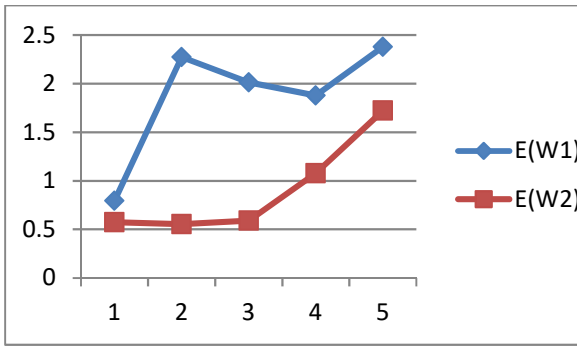


Figure 4.2: Mean Sojourn Times $[E(W_i)]$ Plotted against Data Sets 1-5

Figure 4.2 represents the average waiting period of the two queues being observed. Again, the average waiting period of queue 1, $E(W_1)$ has higher values than that of queue 2. For $E(W_2)$, the line graph for data sets 1 and 2 are on the same level, it increases for data set 3 and then the trend increases from data set 3 to 4 and then to data set 5. For $E(W_1)$, it increases drastically from data set 1 to data set 2, reduces for data sets 3 and 4 and then increases again for data set 5.

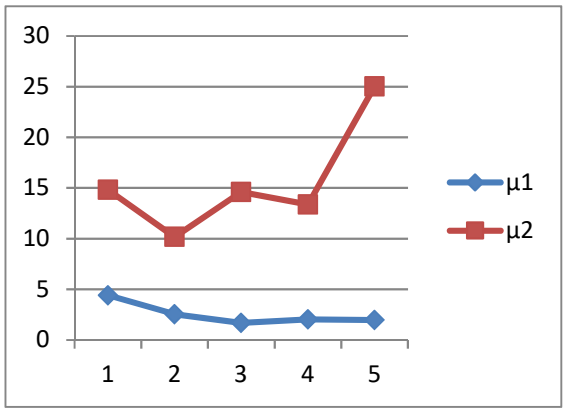


Figure 4.3 Mean Service Times (μi) Plotted against Data Sets 1-5

Figure 4.3 represents the mean service duration for the two classes of customers. The service duration of the lower priority customers μ_2 are higher than that of the higher priority queue μ_1 due to the interruptions that occur during service

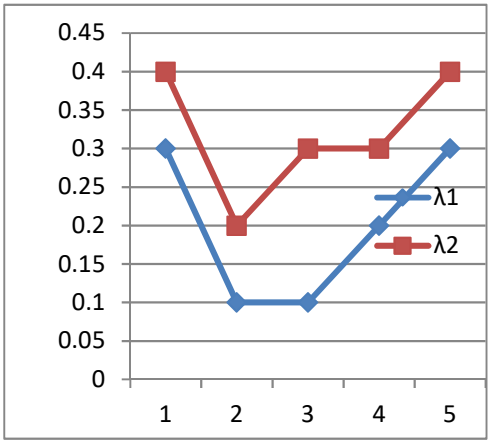


Figure 4.4: Mean Arrival rates (λi) plotted against data sets 1-5

Figure 4.4 represents the values for the arrival rates of class 1 customers λ_1 and arrival rates of class 2 customers λ_2. The arrival rate is higher for the lower priority queue as expected, and lower for the higher priority queue.

5. CONCLUSION

In this study, we discussed in detail a preemptive-resume priority queue with Erlang distribution service times. Two classes of priority customers were assumed. They include, the class 1 (higher priority customers), and the class 2 (lower priority customers). The service is rendered by a single-server facility in stages.

The notion of completion times was used to analyse the service duration of the class 2 customers. The busy period duration for an $M/E_r/1$ queue was derived and its associated moments obtained. This result cannot be found in any literature. The performance measures for the system in steady state for the class 2 customers were obtained. These include, the mean sojourn time and mean number of lower priority queue. The results obtained were further applied to a real life system to illustrate the effect

of the preemptive priority scheduling on the queueing system.

In this study, it was assumed that the system has a single server facility. Thus, the results obtained cannot be applied to a system with multiple servers. A further research can be carried on a situation where there is multiple server facility with the same service time distribution. A system with preemptive non-resume priority queue and Erlang service time can also be considered and the measures derived.

REFERENCES

[1] Alan, C. (1954). Priority Assignment in waiting line problems, Journal of O.R. Soc. Am., (2) 70-76. DOI:10.1287/opre.2.1.70

[2] Harrison, W., and Lee, S. C (1958). Queueing with preemptive priorities or with breakdown, Journal Operations Research Soc. Am. (6) 79-95

[3] Frederick, F. Stephan (1958). Two queues under preemptive priority with poisson arrival and service rates. Journal Operations Research Soc. Am., (6) 399-418. DOI:10.1287/opre.6.3.399

[4] Gaver, D.P. Jr (1959). Imbedded Markov chain analysis of a waiting-line process in continuous time, Annal mathematical Statistics. (31) 86-103

[5] Gaver, D.P. (1962). A Waiting line with interrupted service,Journal of the Royal Statistical Society. Series B (24), No 1 73-90

[6] Fumiaki, M. (1996). A preemptive priority queue as a model with server vacations; Journal of Operations Research Society of Japan. 39(1):118-131. DOI:10.15807/jorsj.39.118

[7] Ivo, A. & Jacques, R. (2001). Queueing theory, Department of Mathematics and Computer Science Eindhoven University of Technology 36, 54-55.

[8] Ayyappan, G., (2010). Retrial queueing system with single working vacation under preemptive priority serice, International journal of computer applications, (2) 28-35. DOI:105120/630-877:

[9] Kumar, M.S. Chakravarthy S. R. and Arumuganathan, R. (2013) Preemptive Resume Priority queue with two classes of MAP arrivals. Applied Mathematical Sciences, 7(52):2569-2589. DOI:10.12988/ams.2013.13231

[10] Senthil, M., Kumar, S. R. Chakravarthy, and Arumuganathan, R. (2013). Preemptive resume priority retrial queue with two classes MAP arrivals, Applied mathematical Science, (7) 2569-2589

[11] Richa S (2014). Mathematical Analysis of Queue with Phase Service: An Overview, Advances in Operations Research. Vol 2014.DOI:101155/2014/240926

[12] Weiger, H. (2014). Preemptive-resume priority queueing; Korteweg-de Vries Institute of Mathematics, Faculty of Sciences, Mathematics and Informatics 8, 20-22

[13] Medhi, J. (2003). Stochastic Models in Queueing theory, Academic press, Amsterdam Boston 277-278