International Journal of Research in Advent Technology, Vol.6, No.6, June 2018
E-ISSN: 2321-9637
Available online at www.ijrat.org

Comprehending Single Server Queueing System with Interruption, Resumption and Repeat

Dashrath[1], Dr. Arun Kumar Singh[2]
[1]*Research Scholor, Shri Venkateshwara University , UP*
dyadav551@gmail.com,
[2]*arunsinghgalaxy@gmail.com*

Abstract-In the fast growing field of communication networks like Internet, Queueing models with interruption have an important role to play. Service interruption models studied in the literature include different types of service unavailability that may be due to server taking vacations, server breakdowns, customer induced interruptions, arrival of a priority customer etc. In this paper we consider a queueing model in three aspects: first, two random variables compete on the onset of interruption to decide whether to repeat or resume the interrupted service; second, no repair time is assumed if the decision is to repeat the service after the interruption and finally we assume reneging of customers during a service interruption.

Index Terms -Queueing, Service Interruption, Repetition of service, Markov process

1. INTRODUCTION

Queueing theory deals with one of the most unpleasant experiences of life, waiting. Queueing is quite common in many fields, for example, in telephone exchange, in a supermarket, at a petrol station, at computer systems, etc[1].
The basic features that characterises a queueing system are the following[2]:

a) **Arrival Pattern:** This describes the manner in which the units arrive and join the system. The customers may arrive in single or in batches. Time interval between any two consecutive arrivals is called the inter-arrival time. The arrival pattern is usually represented by the probability distribution of the inter-arrival time.
On arrival, if a customer sees a long queue, he may decide not to join the queue and may leave the station. This customer behaviour is called balking. Some customers join the queue, wait for a while but loosing their patience, may leave the system without waiting further for service. This situation is referred to as reneging.
If there is more than one queue, customers have a tendency to switch from one queue to another. This is called jockeying.

b) **Service Pattern:** This indicates the manner in which the service is rendered. Like the arrivals, the service also is provided in single or in batches. The probability distribution of the service time describes the service pattern.

c) **Queue Discipline:** Queue Discipline tells us the rule by which the customers are taken for service. Some of the commonly used disciplines include first in first out (FIFO), last in first out (LIFO), service in random order (SIRO) and server sharing. In some systems customers may be given priorities so that the service is rendered in the order of their priorities.

d) **Number of service channels:** This refers to the number of servers providing service to the customers in the system.

e) **Capacity of the system:** The capacity of the system is the maximum number of customers it can accommodate. It may be finite or infinite.
A queuing system is often analysed by modelling it as a Markov chain. Some basic concepts employed in this direction are given briefly in the following sections.
While modelling systems with service interruptions, repetition of service on completion of an interruption can be one of the following:

i) Repeat identical: In this case, the service on completion of interruption has the same distribution as the one offered prior to the onset of interruption.
ii) Repeat same: This is much more complex than the first case repeated service has to follow the same pattern as that prior to the interruption. This means one has to keep all relevant information about the earlier service.

In the model described here as well as in the other chapters we consider identical repetition of interrupted service.
One real life situation where the model in this chapter is appropriate is the following. Consider

a person downloading a software from some site. The downloading may be interrupted for some reason; may be the server site becoming jammed by too many users or it may be some ISP problems or virus attack. Now at this point, the downloading is disrupted and the patience of the person who is browsing starts to decay and he/she may decide to repeat the whole downloading process; or it may happen that before his/her patience reaches the threshold, the download may resume from where it stopped. This is a usual phenomenon for sites which have more visitors than its capacity.

2. LITERATURE REVIEW

Queue with service interruption is an M/M/1 queueing model with exponentially distributed service interruption durations. The queueing model analysed by Wang, Liu and Li [3] with disaster and unreliable server can be considered as models with service interruption. Vacation to server can also be considered as a particular type of service interruption. Li and Tian [4] introduces a vacation model where the vacation can be interrupted by assuming that the server can come back to the normal working level, without completing the vacation period. Boxma, Mandjes and Kella [5] study a single server vacation model where the length of vacation depends on the length of the previous active period.
Krishnamoorthy, Pramod and Deepak [6] considers a queueing model with service interruption and repair where the decision on whether to repeat or resume the interrupted service is taken after completion of interruption according to the realization of a phase type distributed random variable.

3. THE MATHEMATICAL MODEL

We consider a single server queueing system in which the service time follows PH distribution with representation (α, S) of order m. The service is interrupted at an exponentially distributed duration with parameter θ. At the epoch when an interruption occurs, two renewal processes, namely, resume clock and repeat clock are started, realization times of which follow exponential distribution with parameters γ and δ, respectively. If the realization of the resume clock occurs first, the interrupted service is resumed whereas if the repeat clock realizes first then the interrupted service has to be repeated. The customers arrive to the system according to a Poisson process with rate λ_1 while the service is stopped due to an interruption and with rate λ_0 otherwise. At the stoppage of a service due to interruption the customers, except the one being served, may leave the system without waiting for service. Such reneging of customers is assumed to follow Poisson distribution with rate $k\beta$ when there are k customers waiting for service. Interruptions are assumed to occur only when a service is in progress and not when the server is idle. We consider only the case in which, when a service is interrupted, no further interruption befalls on that until the present interruption is cleared. This situation resembles the type I.

A diagrammatic representation of the model is given in Figure 1.

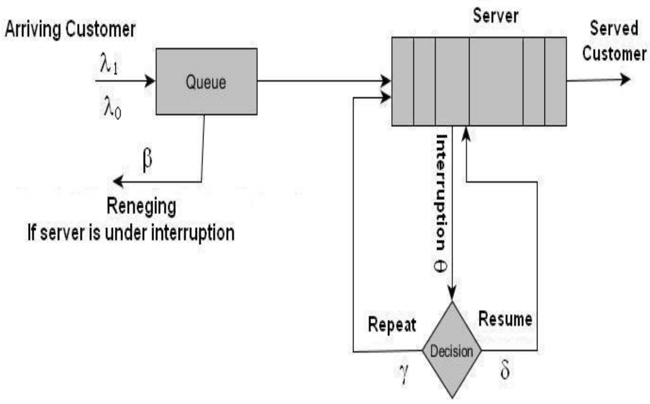


Figure 1: An M/Ph/1 Queue with service interruptions.

Let N (t) be the number of customers in the orbit, S(t) the status of the server which is 0 or 1 according as the server is un- interrupted or interrupted and J (t) the phase of the service pro- cess at time t. Then the above model can be represented by the Markov process $X = \{X(t)/t \geq 0\} = \{(N(t), S(t), J(t)) / t \geq 0\}$. The state space is $\{0\} \cup (\{1, 2, 3, \dots\} \times \{0, 1\} \times \{1, 2, 3, \dots, m\})$.
The one step transitions of the above process from a state are restricted to the states in the same level or to states in one level up or one level down. The level decreases by one when a service completion occurs or reneging occurs while the ongoing service is facing an interruption. The rate at which a service completion occurs at level k, $k \geq 0$ in the phase j is s_j^0 and the probability that of the next service starts in the phase i is α_i. Hence the rate at which transitions happens from $(k, 0, j)$ to $(k -$

1, 0, i) is $\alpha_i\ s_j^0$. The level goes down by one while a reneging occurs consequent to the server being interrupted. Such an event will not alter the phase of the service. For such transitions the rate is $(k - 1)\beta$. These transitions are depicted in Figure 2(a)

The only way the level is increased by one is the arrival of a customer. As this happens at rate λ_1 or λ_0 according as the server is facing an interruption or not and leaves the phase undisturbed, we see that the transition rate from (k, i, j) to (k +1, i, j) is λ_1 while there is an interruption (i = 1) and λ_0 otherwise(i = 0). Figure 2(b) illustrates such transitions.
Now the transitions that will not change the level are those among the phases, given by S, occurrences of interruptions at the rate θ, realizations of resume clock at the rate γ and that of repeat clock at the rate δ as shown in Figure 2(c)

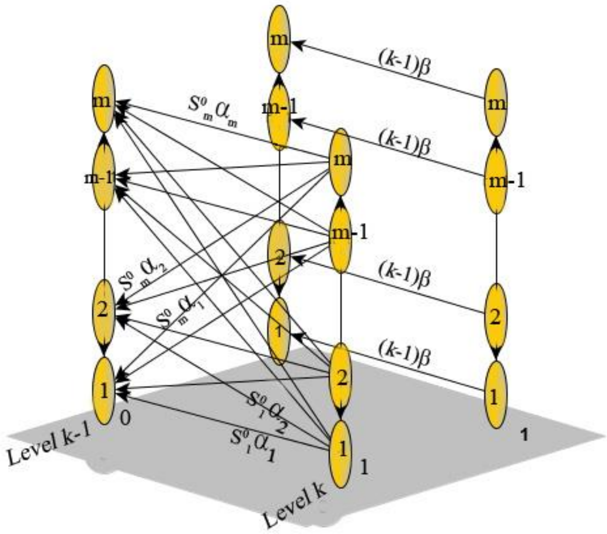
These transitions can be described by the following matrices

$$A_0 = \begin{bmatrix} \lambda_0 I & 0 \\ 0 & \lambda_1 I \end{bmatrix}$$

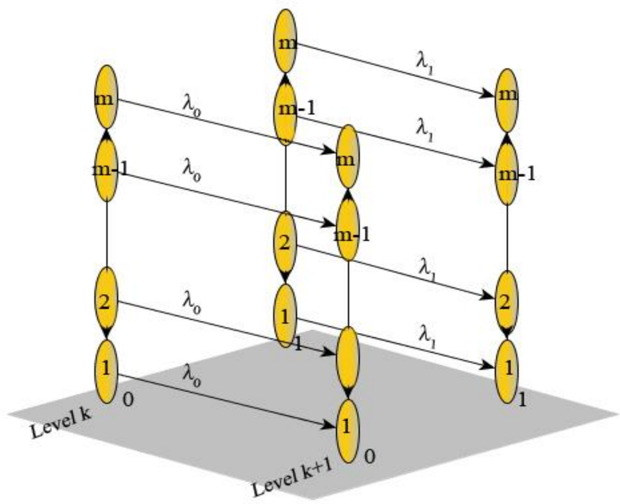
$$A_{1,k} = \begin{bmatrix} S-(\theta+\lambda_0)I & \theta I \\ \gamma I + \delta e\alpha & -(\gamma+\delta+(k-1)\beta+\lambda_1)I \end{bmatrix}$$

and

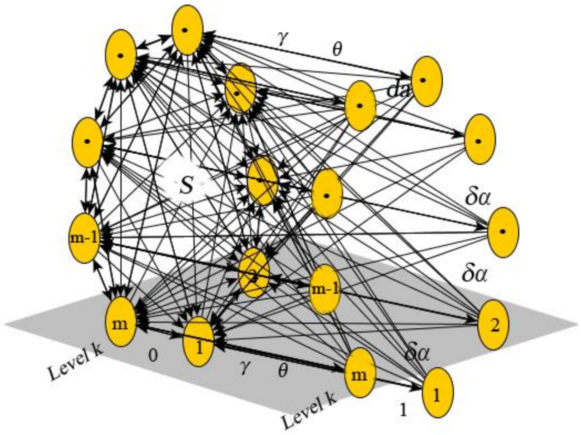
$$A_{2,k+1} = \begin{bmatrix} S^0\alpha & 0 \\ 0 & k\beta I \end{bmatrix}, k = 1, 2, 3, \ldots\ .$$



(a) From k to k-1



(b) From k to k+1



(c) From k to k

Figure 2: State transition diagrams

Hence the infinitesimal generator matrix Q is given by

$$Q = \begin{bmatrix} A_{1,0} & A_{0,0} & & & & \\ A_{2,1} & A_{1,1} & A_0 & & & \\ & A_{2,2} & A_{1,2} & A_0 & & \\ & & \dots & \dots & \dots & \\ & & & \dots & \dots & \dots \end{bmatrix}$$

where

$$A_{1,0} = \begin{bmatrix} -\lambda_0 \end{bmatrix} \quad A_{0,0} = \begin{bmatrix} \lambda_0\alpha & 0\alpha \end{bmatrix} \quad A_{2,1} = \begin{bmatrix} S^0 \\ 0 \end{bmatrix}$$

4. STEADY STATE ANALYSIS

4.1 Neuts Rao Truncation

Since the model described in the previous section is a level dependent QBD, We use an algorithmic solution based on Neuts-Rao Truncation method (Neuts and Rao [7]) for further analysis. Application of this method modifies the process χ into the process $\overline{\chi}$ with infinitesimal generator

$$\tilde{Q} = \begin{bmatrix} A_{1,0} & A_{0,0} & & & & & & \\ A_{2,1} & A_{1,1} & A_0 & & & & & \\ & A_{2,2} & A_{1,2} & A_0 & & & & \\ & & \dots & \dots & \dots & & & \\ & & & A_{2,N-1} & A_{1,N-1} & A_0 & & \\ & & & & A_2 & A_1 & A_0 & \\ & & & & & A_2 & A_1 & A_0 \\ & & & & & \dots & \dots & \dots \end{bmatrix}$$

where $A_1 = A_{1,N}$ and $A_2 = A_{2,N}$.

4.2 Stability condition for the truncated system

If $\pi = (\pi_0 , \pi_1)$ is the stationary probability vector of the generator matrix $A = A_0 + A_1 + A_2$, then we know that the system is stable if and only if $\pi A_0 e < \pi A_2 e$. For the truncated model described above, we have

$$A = \begin{bmatrix} S - \theta I + S^0\alpha & \theta I \\ \gamma I + \delta e\alpha & -(\gamma+\delta) I \end{bmatrix}$$

A simple arithmetic using the relations

$$\pi_0 S - \theta I + S^0\alpha + \pi_1 (\gamma I + \delta e\alpha) = 0$$
$$\theta\pi_0 - (\gamma + \delta)\pi_1 = 0$$
$$\pi_0 e + \pi_1 e = 1$$

Given us

and

$$E(T) = \xi(-B)^{-1}e = -\left(1+\frac{\theta}{\gamma+\delta}\right)\alpha\left(S+\frac{\theta\delta}{\gamma+\delta}(e\alpha-I)\right)^{-1} e$$

where

5. ANALYSIS OF THE SERVICE PROCESS

5.1 Expected service time

The service process with interruption can be viewed as a Markov process $\Psi = \psi(t) = \{(S(t), J(t)) / t \geq 0\}$ where $S(t)$ is the status of the server which is 0 if the server is uninterrupted and 1 otherwise and $J(t)$ is the phase of the service process at time t. This process has 2m transient states given by $\{0, 1\} \times \{1, 2, 3, ..., m\}$ and one absorbing state Δ. The absorbing state Δ denotes the service completion. Let T be the time until absorption of the process Ψ. The infinitesimal generator $\tilde{Q}$ of this process is given by

$$\tilde{Q} = \begin{bmatrix} B & B^0 \\ 0 & 0 \end{bmatrix}$$

where

$$B = \begin{bmatrix} S - \theta I & \theta I \\ \gamma I + \delta e\alpha & -(\gamma+\delta) I \end{bmatrix} \text{ and } B^0 = \begin{bmatrix} S^0 \\ 0 \end{bmatrix}$$

The $2m \times 2m$ matrix B satisfies $B_{ii} < 0$ for $1 \leq i \leq 2m$ and $B_{ij} \geq 0$ for $i = j$. Also $Be + B^0 = 0$ and the initial probability vector of the process is $(\xi, 0)$, where $\xi = (\alpha, 0\alpha)$.

The probability distribution function $F(.)$ of T

$$F(x) = 1 - \xi.\exp(Bx)\, e,\ x \geq 0$$

Its density function $F^0(x)$ in $(0, \infty)$ is given by

$$F^0(x) = \xi.\exp(Bx)\, B^0.$$

The Laplace-Stieltjes transform $f(s)$ of $F(.)$ is

$$f(s) = \xi(sI - B)^{-1}B^0, \text{ for } Re(s) \geq 0.$$

The non-central moments μ_i' of X are given by

$$\mu_i' = (-1)^i i!(\xi B^{-i} e),\ i = 1, 2, 3, ...$$

In particular we have the expected service time is given by

5.2 Expected number of interruptions during a single service

Expected number of interruptions during a single service $E(i)$ is given by,

$$E(i) = -\frac{\theta}{\gamma+\delta}\alpha(S-\theta I)^{-1}(\gamma I+\delta e\alpha) \times \left[I+\frac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\right]^{-1} e$$

Proof. In any particular service, there are no more interruptions before an interrupted service is repaired. So the number of interruptions faced is independent of the time spend under interruptions. Therefore in evaluating the number of interruptions the information that whether the server is interrupted or not at time t is irrelevant.

Hence to get the distribution of the number of interruptions during a single service, we consider the Markov process $\chi = \{(N(t), J(t)) / t \geq 0\}$ where $N(t)$ is the number of interruptions occurred during the service process till time t and $J(t)$ is the phase of the service process at time t. This process has the state space $\{\hat{0}\} \cup \{0, 1\} \times \{1, 2, 3, ..., m\}$ where $\hat{0}$ is the absorbing state denoting the service completion. The infinitesimal generator matrix of this process is given by

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & .... \\ S^0 & S-\theta I & C & 0 & ... \\ S^0 & 0 & S-\theta I & C & ... \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \end{bmatrix}$$

where

$$C = \frac{\theta}{\gamma+\delta}(\gamma I+\delta e\alpha)$$

Let y_k be the probability that the number of interruptions dur- ing a single service is k. Then y_k is the probability that the absorption occurs from the level k for the process χ. Hence y_k are given by

$$y_0 = -\alpha(S - \theta I)^{-1} S^0$$

and for $k = 1, 2, 3, ...,$

$$y_k = -\left(\frac{-\theta}{\gamma+\delta}\right)^k \alpha\left[(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\right]^k (S-\theta I)^{-1} S^0$$

Therefore, the expected number of interruptions during any particular service,

$$E(i) = \sum_{k=0}^{\infty} k y_k$$

$$E(i) = -\frac{\theta}{\gamma+\delta}\alpha(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\left[I+\frac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\right]^{-1} e$$

Since the mean duration of an interruption is $\frac{1}{\lambda+\delta}$, obviously we have expected time spent under interruption during each service is

$$-\frac{\theta}{(\gamma+\delta)^2}\alpha(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\left[I+\frac{\theta}{\gamma+\delta}(S-\theta I)^{-1}(\gamma I+\delta e\alpha)\right]^{-1} e$$

6. PERFORMANCE MEASURES

6.1 Expected waiting time

For computing expected waiting time of a particular customer who joins as the r^{th} customer, $r > 0$, in the queue, we consider the Markov process

$$W(t) = \{(N(t); S(t); J(t))/t \geq 0\}$$

where $N(t)$ is the rank of the customer, $S(t) = 1$ or 0 accord- ing as the service is under interruption or not and $J(t)$ is the phase of the service process at time t. The rank $N(t)$ of the cus- tomer is assumed to be i if he is the i^{th} customer in the queue at time t. His rank may decrease to 1 as the customers ahead of him leave the system either after completing the service or due to reneging. Mean while it may happen that the tagged customer himself may renege from the system. Since the customers who arrive after the tagged customer cannot change his rank, level-changing transitions in W (t) can only take place to one side of the diagonal. We arrange the state space of W (t) as $\{r, r-1, ..., 2, 1\} \times \{0, 1\} \times \{1, 2, ..., m\} \cup \tilde{0}$, where $\tilde{0}$ is the absorbing state denoting that the tagged customer is either selected for service or he leaves the system without waiting for service. Thus the infinitesimal generator W of the process W (t) takes the form

$$W = \begin{bmatrix} \tilde{T} & \tilde{T}^0 \\ 0 & 0 \end{bmatrix}$$

where

$$\tilde{T} = \begin{bmatrix} \tilde{A}_{1,r} & \tilde{A}_{0,r} & & & & & \\ & \tilde{A}_{1,r-1} & \tilde{A}_{0,r-1} & & & & \\ & & \tilde{A}_{1,r-2} & \tilde{A}_{0,r-2} & & & \\ & & & \cdots & \cdots & \cdots & \\ & & & & \tilde{A}_{1,2} & \tilde{A}_{0,2} & \\ & & & & & & \tilde{A}_{1,1} \end{bmatrix} \quad \tilde{T}^0 = \begin{bmatrix} \tilde{B} \\ \tilde{B} \\ \tilde{B} \\ \cdots \\ \tilde{B} \\ \tilde{B}^0 \end{bmatrix}$$

With

$$\tilde{A}_{1,i} = \begin{bmatrix} S - \theta I & \theta I \\ \gamma I + \delta e \alpha & -(\gamma + \delta + (i-1)\beta) I \end{bmatrix} i = 1, 2, 3, ..., r.$$

$$\tilde{A}_{0,i} = \begin{bmatrix} S^0 \alpha & 0 \\ 0 & (i-1)\beta I \end{bmatrix} i = 2, 3, ..., r. \quad \tilde{B} = \begin{bmatrix} 0 \\ \beta e \end{bmatrix} \quad \tilde{B}^0 = \begin{bmatrix} S^0 \\ \beta e \end{bmatrix}.$$

Now, the waiting time W of a customer, who joins the queue as the r^{th} customer is the time until absorption of the Markov chain W (t). Thus the expected waiting times of this particular customer according to the phase of the service process at the time of his arrival are given by the column vector,

$$E_W^{(r)} = \left[-A_{1,r}^{-1} \left(I + \sum_{i=1}^{r-1} (-1)^i \prod_{j=1}^{i} A_{0,r+1-j} A_{1,r-j}^{-1} \right) \right] e.$$

The second moments of waiting times of the tagged customer are given by the column vector $E^r_{W^2}$ which is the first block of the matrix $2(-\tilde{T})^{-2}e$.

Hence, the expected waiting time of a general customer in the queue is,

$$W_L = \sum_{r=1}^{\infty} x(r) E_W^{(r)}$$

The second moment of W is

$$W_L^{(2)} = \sum_{r=1}^{\infty} x(r) E^r_{W^2}$$

6.2 Expected waiting time of customer who was served

The expected waiting time of a customer who waited till he gets the service, is obtained as in the previous section. In this case the W can be obtained by replacing $\tilde{B}$ and $\tilde{B}^0$ with zero vector and $\begin{bmatrix} S^0 \\ 0 \end{bmatrix}$ respectively and adjusting the diagonals so as to form a generator. Here again we can calculate the first two moments $E^{(r)}_{w_s}$ and $E^r_{w^2}$ of waiting time as done earlier. Using these the expected waiting time W_L^s and the variance V_L^s are computed as

$$W_L^S = \sum_{r=1}^{\infty} x(r) E_{W_S}^{(r)}$$

$$V_L^S = W_L^{S(2)} - \left(W_L^S \right)^2$$

7. CONCLUSION

In this paper we consider a single server queueing system where the service time distribution is phase type. The service process may face some interruptions during the service. The interruption occurs according to a Poisson process. Interruptions are assumed to occur only when a service is in progress and not when the server is idle. The interrupted service is either resumed or repeated based on which of the two renewal processes, started simultaneously with the interruption, renews first. Customers arrive according to a Poisson process with different means depending on whether the server is interrupted or not. The customers waiting in the queue for service may leave the system without waiting further for service while the server is interrupted. Stability of the above system is analysed and steady state vector is calculated using Neuts-Rao truncation

REFERENCES

[1]. Sundari, S. Maragatha, and Miriam Cathy Joy. "Queueing model of optional type of services with service stoppage and revamp process in web hosting queueing." International Journal of Knowledge Management in Tourism and Hospitality1.2 (2017): 241-262.
[2]. Chen P, Zhu Y and Zhang Y, A Retrial Queue with Modified Vacations and Server Breakdowns, Computer Science and Information Technology (ICCSIT), 3rd Inter- national Conference on, pp. 26-30, DOI: 10.1109/ICC- SIT.2010.5563560, 2010
[3]. J Wang, B Liu and J Li, Transient analysis of an M/G/1 retrial queue subject to disasters and server failures, European Journal of Operational Research, Vol. 189, Issue 3, 1118-1132, 2008.
[4]. J Li, N Tian, The M/M/1 queue with working vacations and vacation interruptions, Journal of Systems Science and Systems Engineering, Vol. 16, No.1, pp. 121-127, 2007.
[5]. Boxma, M Mandjes and O Kella, On a queueing model with service interruptions, Probability in the Engineering and Informational Sciences, Vol. 22, pp 537-555, 2008.
[6]. A Krishnamoorthy and P K Pramod, Queues With Interruption and Repeat/Resumption of Service- A Survey and Further Results, Fourth International Conference on Neural, Parallel & Scientific Computations, August 11-14, Morehouse College, Atlanta, Georgia, USA, 2010.
[7]. M F Neuts and B M Rao, Numerical investigation of a multiserver retrial model. Queueing systems, 7:169-190,1990.