Neural, Parallel, and Scientific Computations 28 (2020) 1-18

ANALYSIS OF SINGLE SERVER QUEUEING SYSTEM WITH TWO TYPE OF INTERRUPTIONS AND ITS APPLICATION IN WIRELESS NETWORKS

VARGHESE JACOB AND MITHA P M

Department of Mathematics, Government College, Kottayam-686 013, India.
E-mail: *drvarghesejacob@gmail.com*
Department of Mathematics, C.M.S. College, Kottayam-686001, India.
E-mail: mitha_manu@yahoo.com

ABSTRACT. We presents a continuous-time single server queueing system where customers arrive according to a Poisson process and service time follows an exponential distribution. Also two types of interruptions occur at the service facility, namely customer induced interruption (self-interruption) and server induced interruption and are according to Poisson processes and interruption durations follow exponential distribution. The self-interrupted customer will enter into a buffer of finite capacity. Any self-interrupted customer, finding the buffer full, is considered as a lost customer. Those self-interrupted customer who completes their interruption will be moved to another buffer of the same capacity. The self-interrupted customer who completes their interruption are given non-preemptive priority over new customers. Also, server induced interruption occurs only when the system is in busy state and customer in service will be preempted and is placed at the head of the queue or buffer. The customer whose service is interrupted immediately gets into service on completion of the server interruption. We also gave an inter-disciplinary relevance of the model to wireless networks. We analysed this queueing system using the matrix geometric method and obtain the steady-state behaviour of this system. Several important performance measures along with some numerical examples have been discussed. Finally, under a given cost structure, we give an expected profit function to determine the optimal capacity of the space to accommodate the interrupted customers through an illustrative example.

AMS (MOS) Subject Classification. 68K25.

Key Words and Phrases: Customer induced interruption; server interruption, Matrix geometric methods; Quasi-birth-and-death process, steady-state analysis.

Research of Varghese Jacob was supported by the University Grants Commission, Government of India under Grant MRP(S)-0291/12-13/KLMG003/UGC-SWRO dated 23-09-13.
Received March 1, 2020 1061-5369 $15.00
www.dynamicpublishers.org; www.dynamicpublishers.com;
https://doi.org/10.46719/npsc20202811.

1. Introduction

In the classical queueing systems, servers are always available to serve customers. But, there are situations where the servers may unavailable for a random period of time due to many reasons such as server interruptions, server vacations, removal of servers due to catastrophic or negative arrival of events, getting preempted due to the arrival of priority customers etc. A review of recent work on different type of server interruption, readers are referred to Krishnamoorthy *et al.* [9].

In many daily life situation where the customers are often leave the service area in the middle of a service due to not having enough information for completing a service. However, these customers are request their service after some random period of time. These type of interruption is known as ***customer induced interruption***. A more commonly occurring example is the following :- In a doctors clinic, while patient is being examined, the physician may find that one or more tests needed for prescription of medicine. Hence he/she is asked to undergo these and return to the clinic. Such patients can be regarded as interruption induced by the customer. Salient features of this type of interruption as opposed to server interruptions are that the system (a) can have more interrupted customers than the number of servers in the system, and (b) can offer services to other customers while a/some customers is/are undergoing customer induced interruptions.

As far our knowledge goes, the first work dealing queues with customer induced interruptions is [4] reported at 8^{th} International Workshop on retrial queues in 2010 and the paper of Jacob *et al.* [6]. Subsequently, Krishnamoorthy and Jacob [7] extended the work to a multi-server $M/M/c$ model. In the Ph. D thesis, Jacob [8] gave a detailed study of customer induced interruption in queueing system by assuming the arrival process from Poisson to Markovian arrival process (MAP) and service process from exponential distribution to Phase-type distribution. Jacob and Krishnamoorthy [10] discussed $M/PH/1$ queueing system with customer induced interruption in the retrial set up. Recentlty, Punalal and Babu [11] studied a retrial queueing model with self-generation of priorities and customer induced interruption.

The purpose of this work is to introduce both customer induced and server induced interruptions in a queueing systems.

1.1. Inter-disciplinary relevance.

Stochastic modeling especially queueing models have prominent roll in many field such as communication networks, air traffic controls, health care systems. In telecommunication systems this is not really different, packets (customer) that are transmitted on a shared link. In wireless network data transmission is carried out over radio frequency links. But the radio links affect each other through cross-interference, so during these data transmission the packets may get corrupted due to variety of reasons (customer induced interruption) and also base station may undergo some system failure (server induced interruption). Consequently, the performance of such networks may be severely affected due to invoking transmission control protocol (TCP) congestion algorithms whenever a packet is lost and TCP guarantees delivery of data. Darabkh *et al.* [5] discussed that, wireless networks (both LANs and cellular networks) while data transmissions is carried out, the signals or packets (customer) will be corrupted due to lower signal-to-noise ratio. The model proposed by them describes, how received packets are corrected based on Fano decoding mechanism and how the retransmission of corrupted packets is performed. Using the present queueing model we can analyze these type of problems in wireless networks.

The rest of the paper is organized as follows. In Section 2 the model under study is described. Section 3 provides the steady-state analysis of the model, including expected waiting time of a customer in the queue. A few key performance measures and an optimization problem are given in Section 4. Some illustrative examples are discussed in section 5.

2. Model description

We consider a single server queueing system with two types of service interruptions. Customers arrive to the server facility, according to a Poisson process with rate λ. The service time follows an exponential distribution with parameter μ. An arriving customer finding free server enters to service immediately; otherwise the customer is placed into the primary queue of infinite capacity (PQ) and it will be picked up for service according the order of their arrival. Apart from arrival and service processes, two types of interruptions occur at the service facility, ***customer induced interruption*** (CII) and ***server induced interruption*** (SII) occur according to Poisson processes with rates θ_1 and θ_2 respectively. When CII occurs, while his/her service is going on, the customers currently in service is forced to leave the service facility. In this situation, the freed server when CII occur, is ready serve the other customers, if available. The interrupted customer enters into a buffer (referred to as BIP) of finite capacity, K, should there be a space available. Otherwise, such a customer is lost for ever. The interrupted customers spends a random period of time that is independent of other customers and the time is assumed to follow an exponential distribution with parameter η_1. In this paper we assume that no more than one CII is allowed for a customer while in service. That is, an interrupted customer who gets into service again will leave the system with no further interruption. All interrupted customers, upon completing their interruptions enter into a finite buffer (referred to as BIC) whose size is K. Customers in BIC are given non-preemptive priority over new customers but are served in the order in which they enter into this buffer. Thus, a free server will offer service to those customers waiting in BIC before serving new customers by maintaining the order of their arrival. Because of this restriction, the total number of customers in BIC and BIP will never exceed K. When SII occurs, while the system is in a busy state, to the service resulting in the preemption of the customer in service. SII are assumed to occur only when a service is in progress and not when the server is idle. The SII completion time follows an exponential distribution with parameter η_2. We consider only the case in which, when a service is interrupted, no further interruption befall on that until the present interruption is cleared. The preempted customer are placed at the head of the PQ/BIC. The interrupted customers immediately gets into service on completion of the SII.

. In the sequel we use the following notations.

- $N(t)$= Number of customers in the primary queue at time t.
- $S(t) = \begin{cases} 0 & \text{if server is idle at time } t, \\ 1 & \text{if server is busy at time } t, \\ 2 & \text{if server is interrupted at time } t. \end{cases}$

  If $S(t) = 1$ or 2, define
- $S_1(t) = \begin{cases} 1 & \text{if server is busy/interrupted with } PQ \text{ customers,} \\ 2 & \text{if server is busy/interrupted with } BIC \text{ customers,} \end{cases}$
- $N_1(t)$ = Number of customers in BIC at time t.
- $N_2(t)$ = Number of customers in BIP at time t.
- $\mathbf{e}$ denotes column vector of 1's with appropriate dimension.
- $\mathbf{a}'$ denote transpose of the vector $\mathbf{a}$.
- I_r denote identity matrix of dimension r.
- $\mathbf{e}_j(r)$ denote column vector of dimension r with 1 in the j^{th} position and 0 elsewhere.
- $A \otimes B$ denote the Kronecker product of the matrices A and B defined by $A \otimes B = (a_{ij}B)$ where $A = (a_{ij})$.
- $L = (K+1)(K+2)/2$.
- $M = 2 * (K+1)(K+2)$.

. The process $\boldsymbol{X} = \{(N(t), S(t), S_1(t), N_1(t), N_2(t)) : t \geq 0\}$ is a continuous-time Markov chain ($CTMC$).

The state space of the chain is given by

$$\Omega = l^* \cup \left(\bigcup_{n=0}^{\infty} l(n) \right),$$

where level l^* is the union of $K+1$ states

$$l^* = \{ \ (0,0,0,0), (0,0,0,1), \ldots, (0,0,0,K) \},$$

and for $n \geq 0$, level $l(n)$ is the union of M states given by

$$l(n) = \{(n, j_1, j_2, i_1, i_2) : j_1, j_2 = 1, 2; \ i_1 = 0, \ldots, K; \ i_2 = 0, \ldots, K - i_1\}.$$

A brief description of the above states are given below.

- $(0, 0, 0, i_2)$: $-$ the system has no customers in the primary queue (PQ), server is idle, no customers in the BIC and BIP has i_2 customers.
- $(n, 1, 1, i_1, i_2)$: $-$ there are n customers in the primary queue, server is busy with a customer from PQ , BIC has i_1 customers and BIP has i_2 customers.
- $(n, 1, 2, i_1, i_2)$: $-$ there are n customers in the primary queue, server is busy with a customer from BIC , BIC has i_1 customers and BIP has i_2 customers.
- $(n, 2, 1, i_1, i_2)$: $-$ there are n customers in the primary queue, server is interrupted with a customer from PQ , BIC has i_1 customers and BIP has i_2 customers.
- $(n, 2, 2, i_1, i_2)$: $-$ there are n customers in the primary queue, server is interrupted with a customer from BIC , BIC has i_1 customers and BIP has i_2 customers.

Analyzing transitions of the Markov chain $\boldsymbol{X}$ during an interval having an infinitesimal length, we can compute the matrices defining transition rates of this chain.

If we enumerate states in Ω in the lexicographical order, then the infinitesimal generator of the $CTMC$ governing the system is a level independent quasi-birth and death process ($LIQBD$). The infinitesimal generator Q has the following tridiagonal structure.

$$Q = \begin{bmatrix} D_* & E_0 & & & & \\ C_0 & D & E & & & \\ & C & D & E & & \\ & & C & D & E & \\ & & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{2.1}$$

where C, D and E are square matrices of order M; the boundary matrices, D_* is a square matrix of order $K + 1$; E_0 and C_0 are matrices of order $(K + 1) \times M$; and $M \times (K + 1)$ respectively. Now the non-zero entries of these block matrices are described as follows.

- the entries of D_*

  Transitions that the states are unchanged when the system is idle.
  $(0, 0, 0, i_2) \to \ (0, 0, 0, i_2) : \quad -(\lambda + i_2\eta_1) \quad \text{for } i_2 = 0, 1, \ldots, K.$
- The entries of E_0

  The possible transitions are due to arrival of primary customer or CII completion of BIP customers.

  $$(0, 0, 0, i_2) \to \begin{cases} (0, 1, 1, 0, i_2) : & \lambda \quad \text{for } i_2 = 0, 1, \ldots, K. \\ (0, 1, 2, 0, i_2 - 1) : & i_2\eta_1 \quad \text{for } i_2 = 1, 2, \ldots, K. \end{cases}$$

- The entries of C_0

  Transitions due to service completion or CII when PQ is empty.

$$(0,1,1,0,i_2) \rightarrow \begin{cases} (0,0,0,i_2): & \mu \quad \text{for } i_2 = 0,1,\dots,K-1. \\ (0,0,0,i_2+1): & \theta_1 \quad \text{for } i_2 = 0,1,\dots,K-1. \\ (0,0,0,K): & \mu+\theta_1 \quad \text{for } i_2 = K. \end{cases}$$

$$(0,1,2,0,i_2) \rightarrow (0,0,0,i_2): \quad \mu \quad \text{for } i_2 = 0,1,\dots,K.$$

- The entries of A_2

  Transitions due to service completion or CII when server is busy.

$$(n,1,1,0,i_2) \rightarrow \begin{cases} (n-1,1,1,0,i_2): & \mu \\ & \text{for } i_2 = 0,\dots,K-1; n \geq 1. \\ (n-1,1,1,0,i_2+1): & \theta_1 \\ & \text{for } i_2 = 0,\dots,K-1; n \geq 1 \\ (n-1,1,1,0,K): & \mu+\theta_1 \\ & \text{for } i_2 = K;\ n \geq 1 \end{cases}$$

$$(n,1,2,0,i_2) \rightarrow (n-1,1,1,0,i_2): \quad \mu \quad \text{for } i_2 = 0,1,\dots,K; n \geq 1$$

- The entries of A_0

  Transitions due to arrival of primary customers when server is in busy state.

$$(n,j_1,j_2,i_1,i_2) \rightarrow (n+1,j_1,j_2,i_1,i_2): \quad \lambda$$

$$\text{for } j_1,j_2 = 1,2;\ i_1 = 0,1,\dots,K;\ i_2 = 0,1,\dots,K-i_1;\ n \geq 0$$

- The entries of A_1

  Transitions that the levels are unchanged when the system is in busy state.

  (i). When all the states are unchanged.

  ▷ $(n,1,1,i_1,i_2) \rightarrow (n,1,1,i_1,i_2): \quad -(\lambda+\mu+\theta_1+i_2\eta_1+\theta_2)$

  ▷ $(n,1,2,i_1,i_2) \rightarrow (n,1,2,i_1,i_2): \quad -(\lambda+\mu+i_2\eta_1+\theta_2)$

  ▷ $(n,2,1,i_1,i_2) \rightarrow (n,2,1,i_1,i_2): \quad -(\lambda+i_2\eta_1+\eta_2)$

  ▷ $(n,2,2,i_1,i_2) \rightarrow (n,2,2,i_1,i_2): \quad -(\lambda+i_2\eta_1+\eta_2)$

$$\text{for } i_1 = 0,1,\dots,K;\ i_2 = 0,1,\dots,K-i_1;\ n \geq 0.$$

  (ii). Transitions due to CII.

  ▷ $(n,1,1,i_1,i_2) \rightarrow (n,1,2,i_1-1,1,i_2+1): \quad \theta_1$

$$\text{for } i_1 = 1,2,,\dots,K;\ i_2 = 0,1,\dots,K-i_1;\ n \geq 0.$$

  (iii). Transitions due to CII completion.

  ▷ $(n,1,1,i_1,i_2) \rightarrow (n,1,1,i_1+1,i_2-1): \quad i_2\eta_1$

$\triangleright$ $(n,1,2,i_1,i_2) \to \ (n,1,2,i_1+1,i_2-1):\quad i_2\eta_1$

$\triangleright$ $(n,2,1,i_1,i_2) \to \ (n,2,1,i_1+1,i_2-1):\quad i_2\eta_1$

$\triangleright$ $(n,2,2,i_1,i_2) \to \ (n,2,2,i_1+1,i_2-1):\quad i_2\eta_1$

for $i_1 = 0,1,\ldots,K-1;\ i_2 = 1,\ldots,K-i_1;\ n \geq 0.$

(iv). Transitions due to SII.

$\triangleright$ $(n,1,1,i_1,i_2) \to \ (n,2,1,i_1,i_2):\quad \theta_2$

$\triangleright$ $(n,1,2,i_1,i_2) \to \ (n,2,2,i_1,i_2):\quad \theta_2$

for $i_1 = 0,1,\ldots,K;\ i_2 = 0,1,\ldots,K-i_1;\ n \geq 0.$

(v). Transitions due to service completion.

$\triangleright$ $(n,1,1,i_1,i_2) \to \ (n,1,2,i_1-1,i_2):\quad \mu$

$\triangleright$ $(n,1,2,i_1,i_2) \to \ (n,1,2,i_1-1,i_2):\quad \mu$

for $i_1 = 1,2,,\ldots,K;\ i_2 = 0,1,\ldots,K-i_1;\ n \geq 0.$

(vi). Transitions due to SII completion.

$\triangleright$ $(n,2,1,i_1,i_2) \to \ (n,1,1,i_1,i_2):\quad \eta_2$

$\triangleright$ $(n,2,2,i_1,i_2) \to \ (n,1,2,i_1,i_2):\quad \eta_2$

for $i_1 = 0,1,\ldots,K;\ i_2 = 0,1,\ldots,K-i_1;\ n \geq 0.$

3. **Steady-state analysis**

In this section we study the steady-state analysis of the queueing model under study by first establishing the ergodicity condition of the queueing system.

3.1. **Stability condition.** The stability condition can be expressed in terms of the matrices C, D, E as follows.

Let $\boldsymbol{\delta}$ denote the steady-state probability vector of the irreducible generator matrix $A = C + D + E$. That is, $\boldsymbol{\delta} A = \mathbf{0},\ \boldsymbol{\delta}\boldsymbol{e} = 1$. The $LIQBD$ description of the model shows that the queueing system is stable (see Neuts [1]) if and only if

$$\boldsymbol{\delta} E\boldsymbol{e} < \boldsymbol{\delta} C\boldsymbol{e}. \tag{3.1}$$

In other words, the rate of moving down one level in the Markov chain $\boldsymbol{X}$ must exceed the rate of moving up one level.

We cannot obtain the vector, $\boldsymbol{\delta}$, explicitly in terms of the parameters of the model, and hence the stability condition is known only implicitly. For future reference, we define the traffic intensity, ρ as

$$\rho = \frac{\boldsymbol{\delta} E\boldsymbol{e}}{\boldsymbol{\delta} C\boldsymbol{e}}. \tag{3.2}$$

Thus the ergodicity condition in (3.1) is equivalent to $\rho < 1$. We will discuss the impact of the input parameters of the model on the traffic intensity in Section 5.

3.2. **Steady-state vector.** Let $\boldsymbol{x}$ denote the steady-state probability vector of the generator Q given in (2.1). That is,

$$\boldsymbol{x}\, Q = \mathbf{0},\ \ \boldsymbol{x}\, \boldsymbol{e} = 1. \tag{3.3}$$

Let $\boldsymbol{x}$ be partitioned as

$$\boldsymbol{x} = (\boldsymbol{x}^{*}, \boldsymbol{x}(0), \boldsymbol{x}(1), \ldots) \tag{3.4}$$

we see that $\boldsymbol{x}$, under the assumption that the stability condition holds, there exists a unique steady-state probability vector $\boldsymbol{x}$. Algorithm for computing the stationary probabilities

$$\begin{aligned} \boldsymbol{x}_{j_1,j_2,i_1,i_2}(n) &\\ &= \lim_{t\to\infty} P(N(t) = n, S(t) = j_1, S_1(t) = j_2, N_1(t) = i_1, N_2(t) = i_2) \\ &\qquad (n, j_1, j_2, i_1, i_2) \in \Omega \end{aligned} \tag{3.5}$$

according to the matrix-geometric method (see Neuts [1]).

$$\boldsymbol{x}(n) = \boldsymbol{x}(0)R^{n}, \quad n \geq 1, \tag{3.6}$$

where R is the minimal non-negative solution to the matrix quadratic equation :

$$R^2C + RD + E = O.$$

To start the recursive relation (3.6), the following boundary equations must be solved to obtain the $\boldsymbol{x}^{*}$ and $\boldsymbol{x}(0)$

$$\boldsymbol{x}^{*}D_{*} + \boldsymbol{x}(0)C_0 = \mathbf{0} \tag{3.7}$$

$$\boldsymbol{x}^{*}E_0 + \boldsymbol{x}(0)(D + RC) = \mathbf{0} \tag{3.8}$$

under normalizing condition

$$\boldsymbol{x}^{*}\boldsymbol{e} + \boldsymbol{x}(0)(I - R)^{-1}\boldsymbol{e} = 1. \tag{3.9}$$

We can use logarithmic reduction algorithm [2] for the computation of R matrix in (3.6). Once R matrix and two boundary vectors $\boldsymbol{x}^{*}$ and $\boldsymbol{x}(0)$ are obtained, the vector $\boldsymbol{x}$ can be calculated.

3.3. **Expected waiting time in the primary queue.** First note that an arriving customer will enter into service immediately with probability $p = \boldsymbol{x}^{*}\boldsymbol{e}$. With probability $1-p$, the arriving customer has to wait before getting into service. To compute the expected waiting time in the queue of an arriving (tagged) customer who join as the n^{th} customer in the primary queue until taken for service. Consider the Markov processes $\boldsymbol{X}_1 = \{(\widetilde{N}(t), \widetilde{S}(t), N_1(t), N_2(t)) : t \geq 0\}$ where $\widetilde{N}(t)$ is the rank of a customer in the primary queue, $\widetilde{S}(t)$= 1 or 2 and 3 or 4 according as the server is busy with PQ or BIC customers and server is interrupted with PQ or BIC customers,

$N_1(t)$ = the number of customers in BIC and $N_2(t)$= the number of customers in BIP. Then $\boldsymbol{X}_1$ has the state space

$$\Omega_1 = \{\underline{\mathbf{i}} = (i,j,k,l) : 1 \le i \le n; j = 1,2,3,4; k = 0 \text{ to } K; l = 0 \text{ to } K-k\}$$

The rank $\widetilde{N}(t)$ of the customer is assumed to be n if he joins the n^{th} customer in the primary queue. His rank may decrease to as the customers ahead of him leave the system after completing the service or self interruption. Since the customer who arrives after the tagged customer cannot change his rank. Level changing transitions in $\boldsymbol{X}_1$ can only takes place to one side of the diagonal. We arrange the state space of $\boldsymbol{X}_1$ as $\{\underline{\mathbf{n}}, \underline{\mathbf{n-1}} \ldots \underline{\mathbf{1}}\} \cup \{*\}$ where $\{*\}$ is an absorbing state. On entering the absorbing state, a tagged customer starts receiving service. Thus the infinitesimal generator Q_1 of the process $\boldsymbol{X}_1$ takes the form

$$\tilde{Q}_1 = \begin{bmatrix} \widetilde{T} & \widetilde{T}^0 \\ \mathbf{0} & 0 \end{bmatrix}$$

$$\text{where } \widetilde{T} = \begin{array}{c|cccccc} & \underline{\mathbf{n}} & \underline{\mathbf{n-1}} & \underline{\mathbf{n-2}} & \ldots & \underline{\mathbf{2}} & \underline{\mathbf{1}} \\ \hline \underline{\mathbf{n}} & \tilde{D} & C & & & & \\ \underline{\mathbf{n-1}} & & \tilde{D} & C & & & \\ \underline{\mathbf{n-2}} & & & \tilde{D} & \ddots & & \\ & & & & \ddots & C & \\ \underline{\mathbf{2}} & & & & & \tilde{D} & C \\ \underline{\mathbf{1}} & & & & & & \tilde{D} \end{array} \quad \text{and} \quad \widetilde{T}^0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \boldsymbol{b} \end{pmatrix}$$

C is defined as in (2.1), $\tilde{D} = D + \lambda I$, $\boldsymbol{b} = C\boldsymbol{e} = \begin{bmatrix} (\mu+\theta_1)\boldsymbol{e} & \boldsymbol{0} & \mu\boldsymbol{e} & \boldsymbol{0} \end{bmatrix}'$

In matrix Q_1, the customer arrival rate λ is not required since we are considering $FCFI$ queue discipline and hence, customers that arrive after the tagged customer do not have any impact on the analysis of waiting time. Now the waiting time of a customer until taken for service who joins the primary queue as the n^{th} customer is the time until absorption in the Markov Chain $\boldsymbol{X}_1$. The expected waiting time of a tagged customer in the primary queue is given by the column vector,

$$E_Q^{(n)} = \tilde{I}_{Mn}(-\widetilde{T})^{-1}\ \boldsymbol{e} = (-D)^{-1}\left[\sum_{i=1}^{n}(-1)^{i+1}(CD^{-1})^{i-1}\right]\boldsymbol{e}$$

where $\tilde{I}_{Mn} = \begin{bmatrix} I_M & O \\ O & O \end{bmatrix}_{Mn\times Mn}$

Therefore, the expected waiting time of an arbitrary customer in the primary queue until taken for service is obtained as,

$$E_Q = \sum_{n=1}^{\infty} \boldsymbol{x}(n).E_Q^{(n)} \tag{3.10}$$

4. **Other performance measures**

In this section we provide a few measures of the system performance that are useful for system design. First, we further partition the steady-state probability vector $(\boldsymbol{x}^*, \boldsymbol{x}(0), \boldsymbol{x}(1), \ldots)$ into $\boldsymbol{x}(n) = (\boldsymbol{x}_{1,1}(n), \boldsymbol{x}_{1,2}(n), \boldsymbol{x}_{2,1}(n), \boldsymbol{x}_{2,2}(n))$, $n \geq 0$, where

$$\boldsymbol{x}^* = (x_0^*, x_1^*, \ldots, x_K^*)$$

$$\boldsymbol{x}_{11}(n) = (\boldsymbol{x}_{1,1,i_1,i_2}(n));\ \ \boldsymbol{x}_{12}(n) = (\boldsymbol{x}_{1,2,i_1,i_2}(n))$$

$$\boldsymbol{x}_{21}(n) = (\boldsymbol{x}_{2,1,i_1,i_2}(n));\ \ \boldsymbol{x}_{22}(n) = (\boldsymbol{x}_{2,2,i_1,i_2}(n))$$

$$n \geq 0;\ i_1 = 0 \text{ to } K;\ i_2 = 0 \text{ to } K - i_1;$$

Note that each sub-vectors $\boldsymbol{x}_{lm}(n), l, m = 1, 2$ are of dimension L.

i. Fraction of time the server is idle: $P_{idle} = \boldsymbol{x}^*\boldsymbol{e}$.
ii. Fraction of time the server is busy with a PQ customers :
$P_{BSYP} = \sum_{n=0}^{\infty} \boldsymbol{x}_{11}(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - R)^{-1}(\boldsymbol{e}_1(4) \otimes \boldsymbol{e})$.
iii. Fraction of time the server is busy with BIC customers :
$P_{BSYS} = \sum_{n=0}^{\infty} \boldsymbol{x}_{12}(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - R)^{-1}(\boldsymbol{e}_2(4) \otimes \boldsymbol{e})$.
iv. Fraction of time the server is interrupted with a PQ customer :
$P_{INTP} = \sum_{n=0}^{\infty} \boldsymbol{x}_{21}(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - R)^{-1}(\boldsymbol{e}_3(4) \otimes \boldsymbol{e})$.
v. Fraction of time the server is interrupted with BIC customer :
$P_{INTS} = \sum_{n=0}^{\infty} \boldsymbol{x}_{22}(n)\boldsymbol{e} = \boldsymbol{x}(0)(I - R)^{-1}(\boldsymbol{e}_4(4) \otimes \boldsymbol{e})$.
vi. The probability that an interrupted PQ customer is lost:
$P_{loss} = \dfrac{\theta_1}{\theta_1 + \mu} \sum_{n=0}^{\infty} x_{1,1,0,K}(n)$.
vii. Mean number of customers in the primary queue, PQ :
$\mu_{PQ} = \boldsymbol{x}(0)R(I - R)^{-2}\boldsymbol{e}$.
viii. The mean number of interrupted customers in the BIP buffer:
$\mu_{BIP} = \sum_{i_2=0}^{K} i_2 x_{i_2}^* + \sum_{n=0}^{\infty} \sum_{j_1=1}^{2} \sum_{j_2=1}^{2} \sum_{i_2=0}^{K} \sum_{i_1=0}^{K-i_2} i_2 \boldsymbol{x}_{j_1,j_2,i_1,i_2}(n)$.
ix. The mean number of interrupted customers in the BIC buffer:
$\mu_{BIC} = \sum_{n=0}^{\infty} \sum_{j=1}^{2} \sum_{i_1=0}^{K} \sum_{i_2=0}^{K-i_1} i_1 \boldsymbol{x}_{n,j,i_1,i_2}$.
x. The mean waiting time in the queue, E_Q, is as given in (3.10).

We obtain the following theorem.

Theorem 4.1.

$$P_{BSYP} = \frac{\lambda}{\theta + \mu}.$$

Proof. The steady-state equations given in (3.3) can be written as

$$\boldsymbol{x}^* D_* + \boldsymbol{x}(0)C_0 = \boldsymbol{0}, \tag{4.1}$$

$$\boldsymbol{x}^* E_0 + \boldsymbol{x}(0)D + \boldsymbol{x}(1)C = \mathbf{0}, \tag{4.2}$$

and

$$\boldsymbol{x}(n-1)E + \boldsymbol{x}(n)D + \boldsymbol{x}(n+1)C = \mathbf{0},\ n \geq 1. \tag{4.3}$$

Post-multiplying equations (4.1) through (4.3) by $\boldsymbol{e}$ of appropriate dimensions, we get

$$\lambda \boldsymbol{x}^* \boldsymbol{e} + \eta_1 \sum_{r=0}^{K} r x_r^* = (\mu + \theta_1)\boldsymbol{x}_{1,1,0}(0)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(0)\boldsymbol{e}, \tag{4.4}$$

(4.5)

$$\lambda \boldsymbol{x}^* \boldsymbol{e} + \eta_1 \sum_{r=0}^{K} r x_r^* + (\mu+\theta_1)\boldsymbol{x}_{1,1,0}(1)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(1)\boldsymbol{e} = \lambda \boldsymbol{x}(0)\boldsymbol{e} + (\mu+\theta_1)\boldsymbol{x}_{1,1,0}(0)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(0)\boldsymbol{e},$$

and

$$\begin{aligned} \lambda \boldsymbol{x}(n-1)\boldsymbol{e} + (\mu+\theta_1)\boldsymbol{x}_{1,1,0}(n+1)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(n+1)\boldsymbol{e} \\ = \lambda \boldsymbol{x}(n)\boldsymbol{e} + (\mu+\theta_1)\boldsymbol{x}_{1,1,0}(n)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(n)\boldsymbol{e},\ n \geq 1 \end{aligned} \tag{4.6}$$

Using equation (4.4) in (4.5) and summing (4.6) over $n \geq 1$ and adding,

$$\lambda(1 - \boldsymbol{x}^* \boldsymbol{e}) = (\mu + \theta_1)\sum_{n=1}^{\infty} \boldsymbol{x}_{1,1,0}(n)\boldsymbol{e} + \mu \sum_{n=1}^{\infty} \boldsymbol{x}_{1,2,0}(n)\boldsymbol{e}. \tag{4.7}$$

Now post-multiplying equations (4.2) and (4.3) by $(\boldsymbol{e}_1(2)\ \boldsymbol{e}_1(2))' \otimes \boldsymbol{e})$,

$$\lambda \boldsymbol{x}^* \boldsymbol{e} + (\mu + \theta_1)\boldsymbol{x}_{1,1,0}(1)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(1)\boldsymbol{e} = \lambda \boldsymbol{x}_{1,1}(0)\boldsymbol{e} + (\mu + \theta_1)\boldsymbol{x}_{1,1}(0)\boldsymbol{e}, \tag{4.8}$$

and

$$\begin{aligned} \lambda \boldsymbol{x}_{1,1}(n-1)\boldsymbol{e} + (\mu+\theta_1)\boldsymbol{x}_{1,1,0}(n+1)\boldsymbol{e} + \mu \boldsymbol{x}_{1,2,0}(n+1)\boldsymbol{e} \\ = \lambda \boldsymbol{x}_{1,1}(n)\boldsymbol{e} + (\mu+\theta_1)\boldsymbol{x}_{1,1}(n)\boldsymbol{e},\ n \geq 1 \end{aligned} \tag{4.9}$$

Now summing (4.9) over $n = 1$ to ∞ and adding to (4.8), we get

$$\lambda \boldsymbol{x}^* \boldsymbol{e} = (\mu + \theta_1)\sum_{n=0}^{\infty} \boldsymbol{x}_{1,1}(n)\boldsymbol{e} - (\mu + \theta_1)\sum_{n=1}^{\infty} \boldsymbol{x}_{1,1,0}(n)\boldsymbol{e} - \mu \sum_{n=1}^{\infty} \boldsymbol{x}_{1,2,0}(n)\boldsymbol{e}. \tag{4.10}$$

The stated result follows immediately by adding (4.7) and (4.10). □

Note: We observed that the measure P_{BSYP} does not depend on K. This seems to be counter intuitive as one would expect a larger K to increase the server to be busy with serving self-interrupted customers and hence has a bearing on P_{BSYP}. However, it appears that the server's idle probability is reduced when K is increased to off-set the increase in P_{BSYS} and to keep P_{BSYP} the same.

4.1. **Optimisation Analysis.** In this section we study an optimization problem and discuss it through Table 3 in section 5. To construct an objective function, we assume that CII produce revenue to the system in contrast to SII. Interrupted customers have to pay more cost than those without interruption. Also loss of customers, waiting spaces in primary queue and BIC involve expenditure to the system. Moreover, SII and idle server give revenue loss to the system. Thus we introduce per unit time revenue and cost as follows.

The revenue obtained by the system's service is per time unit : the amount of money obtained from the served customers per time unit,

- revenue R_1 per customer leaving the system with an uninterrupted service,
- revenue $R_2(R_2 > R_1)$ per customer leaving the system on completion of service after CII,

The mean costs that occur to system in steady-state per time unit are :

- fixed waiting cost C_1 per unit time that a customer has to wait in the primary queue,
- fixed waiting cost C_2 per unit time that a customer has to wait in the BIC buffer,
- fixed cost C_3 per unit time that each customer lost due to BIP buffer being full at the time an interruption occurs.
- fixed cost C_4 per unit time for revenue loss due to SII.
- cost C_5 per unit time for the revenue loss due to the system is idle.

The problem of interest is to find an optimum value for K (when all other parameters are fixed) that maximizes the expected total profit ETP, as given in the following objective function.

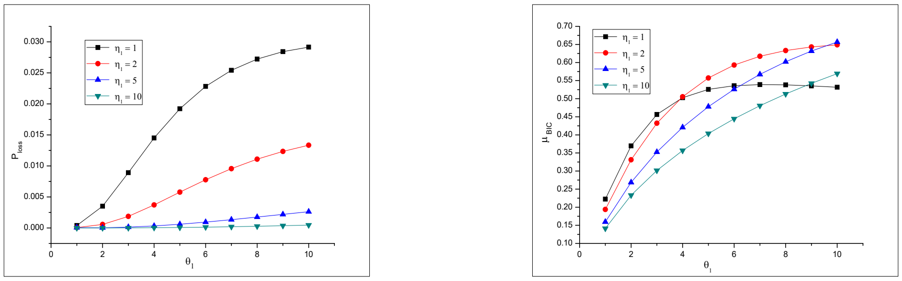
$$
\begin{aligned}
\boldsymbol{ETP} = {} & \mu\left(R_1 P_{BSYP} + R_2 P_{BSYS}\right) - C_1 \mu_{PQ} - C_2 \mu_{BIC} \\
& - C_3(\theta_1 + \mu) P_{loss} + C_4 \theta_2 \left(P_{INTP} + P_{IBIC}\right) + C_5 P_{idle}
\end{aligned}
\tag{4.11}
$$

5. **Numerical Examples**

In this section, we provide some numerical examples that illustrate the qualitative nature of the system under study. The correctness and accuracy of the code are verified by a number of accuracy check.

Example 5.1. This example study the effect of CII rate θ_1 in some performance measures for various η_1. For the fixed parameters $(K, \lambda, \mu, \theta_2, \eta_2) = (4, 3, 5, 0.5, 1)$ we plot the graphs of P_{loss} and μ_{BIC} in figure 1.

From figure 1, we observed the following :-



(A) θ_1 versus P_{loss} (B) θ_1 versus μ_{BIC}

FIGURE 1. Effect of θ_1 on measures P_{loss} and μ_{BIC}

- Both the measures P_{loss} and μ_{BIC} are non-decreasing function of θ_1 for very values of η_1, as expected, .
- From Figure 1(A), we can see that for smaller values of η_1, P_{loss} is high. This is because, for lower values of η_1, the most of interrupted customers remains in BIP. So there is no waiting space in BIP for customers after CII, results in the increase of loss probability. On the other hand, for larger values of η_1 results in a faster clearance of interrupted customers from BIP to BIC and so the loss probability is getting reduced.
- From Figure 1(B), it is interesting to notice that for $\eta_1 = 1$, μ_{BIC} has initially larger values but after some stage of θ_1 it is getting steady and is the smallest when $\theta_1 = 10$. This is due to the fact that, when $\eta_1 = 1$, the loss probability is very high being the BIP full. Recall that at any given time the total number of customers in the BIC and BIP buffers cannot exceed K. So μ_{BIC} becomes the smallest when $\eta_1 = 1$ as CII rate θ_1 increases.

Example 5.2. Through this example, we investigate the influence of SII rate θ_2 in various performance measures by fixing the parameters as $(K, \lambda, \mu, \theta_1, \eta_1) = (4, 1.5, 8, 3, 2)$. A look at the Table 1 shows the following observations.

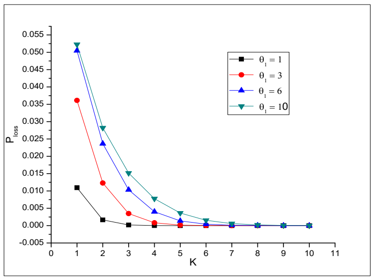
- As θ_2 increases, the measure the traffic intensity ρ, appears to increase for all values of η_2. And the rate of increase is high for smaller values of η_2. This is because for smaller values of η_2, the server is in interrupted state as θ_2 increases. So, on interruption completion epoch, system is in busy mode.
- As one would expect, increasing in θ_2 should result in an increase in the measures μ_{PQ}, μ_{BIC}, P_{INTP} and P_{INTS} and the rate of increase is high for smaller values of η_2. Also, the measure P_{idle} decreases with increase in θ_2 for every η_2. The reason is that, for higher values of η_2, the system can stay in busy state and there by clearing the PQ/BIC customers.
- The measure expected waiting time in queue, E_Q behave similar to ρ as function of θ_2 and η_2.

TABLE 1. Some selected measures when $K = 4, \lambda = 1.5, \mu = 8, \theta_1 = 3, \eta_1 = 2$

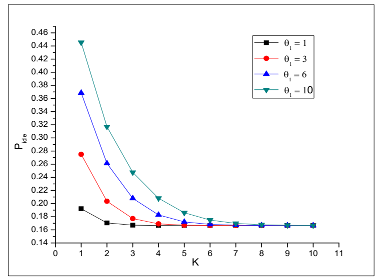
θ_2	ρ	μ_{PQ}	μ_{BIC}	P_{idle}	P_{INTP}	P_{INTS}	E_Q
$\eta_2 = 1$							
0	0.186	0.03639	0.00782	0.8125	0	0	0.02426
0.4	0.2608	0.22791	0.02566	0.73752	0.05455	0.02045	0.15194
1.2	0.4107	0.80723	0.07682	0.5876	0.16364	0.06131	0.53815
2.8	0.7109	4.28067	0.25108	0.288	0.38182	0.14281	2.85378
3.6	0.861	12.3233	0.38073	0.13838	0.49091	0.1834	8.21555
$\eta_2 = 2$							
0	0.186	0.03639	0.00782	0.8125	0	0	0.02426
0.4	0.2235	0.09091	0.01473	0.77501	0.02727	0.01023	0.06061
1.2	0.2985	0.228	0.0312	0.70002	0.08182	0.03068	0.152
2.8	0.4487	0.67177	0.07593	0.55006	0.19091	0.07156	0.44785
3.6	0.5239	1.03504	0.10489	0.47509	0.24545	0.09199	0.69003
$\eta_2 = 5$							
0	0.186	0.03639	0.00782	0.8125	0	0	0.02426
0.4	0.2011	0.049	0.00988	0.7975	0.01091	0.00409	0.03266
1.2	0.2312	0.07727	0.01432	0.76751	0.03273	0.01227	0.05151
2.8	0.2914	0.14797	0.02453	0.70751	0.07636	0.02863	0.09865
3.6	0.3215	0.19174	0.03033	0.67751	0.09818	0.03681	0.12783
$\eta_2 = 6$							
0	0.186	0.03639	0.00782	0.8125	0	0	0.02426
0.4	0.1986	0.04606	0.00945	0.8	0.00909	0.00341	0.03071
1.2	0.2237	0.06743	0.01293	0.775	0.02727	0.01023	0.04496
2.8	0.2739	0.11932	0.02077	0.72501	0.06364	0.02386	0.07955
3.6	0.299	0.15053	0.02515	0.70001	0.08182	0.03068	0.10036
$\eta_2 = 10$							
0	0.186	0.03639	0.00782	0.8125	0	0	0.02426
0.4	0.1935	0.0412	0.00868	0.805	0.00545	0.00205	0.02746
1.2	0.2086	0.05147	0.01047	0.79	0.01636	0.00614	0.03431
2.8	0.2388	0.07487	0.01437	0.76	0.03818	0.01432	0.04991
3.6	0.2539	0.0881	0.01647	0.74501	0.04909	0.01841	0.05874

Example 5.3. In this example, by fixing the parameters $(\lambda, \mu, \eta_1, \theta_2, \eta_2) = (2.5, 6, 2, 1, 1)$, we will discuss the impact of buffer size K in the measures P_{loss}, P_{idle}, μ_{PQ} and μ_{BIC} for various θ_1.

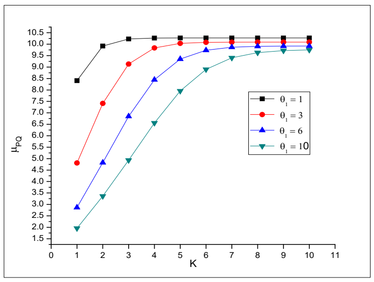
Looking at Figure 2, we summarize the following observations.



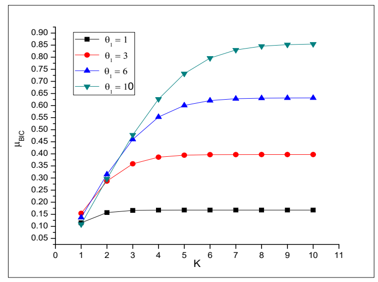
(A) K versus P_{loss}



(B) K versus P_{idle}



(C) K versus μ_{PQ}



(D) K versus μ_{BIC}

FIGURE 2. Effect of buffer size K on various measures

- In Figures 2(A) and 2(B), the measures P_{loss} and P_{idle} are non-increasing function of K for every θ_1. This is because, increasing K leads to an increase in the customers leaving the system with services and hence loss probability and P_{idle} are getting low.
- In Figures 2(C) and 2(D), as one would expect, the μ_{PQ} and μ_{BIC} are non-decreasing functions of K for every values of θ_1. This is due to the fact that, as K increases, the traffic intensity, ρ increases leading to an increase in the number of customers in PQ and BIC. Also, $\mu_{BIC} < \mu_{PQ}$ for every K and θ_1. This is because of priority assumption of BIC customers over BIP customers.

Example 5.4. In this example we look into the optimal value of buffer size K^* that maximize **ETP**, as discussed in subsection 4.1, by varying θ_1 and η_1 when (λ, μ, θ_2, η_2) = (2, 4, 1, 2) and (R_1, R_2, C_1, C_2, C_3, C_4, C_5) = (\$ 45, \$ 60, \$ 5, \$ 7, \$ 15, \$ 20, \$ 7). These are displayed in Tables 2 and 3.

It is seen from the numerical experiment that **ETP** increases first and then decreases. We summarize it as follows.

i. From Table 2 we have the optimum value of $K = 7$ and **ETP** $= 96.38369$ when (λ, μ, θ_1, θ_2, η_1, η_2) = (2, 4, 10, 2, 1, 2) and (R_1, R_2, C_1, C_2, C_3, C_4, C_5) = (\$ 45, \$ 60, \$ 5, \$ 7, \$ 15, \$ 20, \$ 7).
ii. The optimum K^* and the corresponding **ETP** are given in Table 3 by varying θ_1 and η_1 when (λ, μ, θ_2, η_2) = (2, 4, 1, 2) and (R_1, R_2, C_1, C_2, C_3, C_4, C_5) =

Table 2. K and $\boldsymbol{ETP}$

K	1	2	3	4	5
$\boldsymbol{ETP}$	55.34053	80.92971	91.28774	94.99600	96.10826
K	6	**7**	8	9	10
$\boldsymbol{ETP}$	96.35804	**96.38369**	96.37222	96.36287	96.35858

(\$45, \$60, \$5, \$7, \$15, \$20, \$7). It is noticed that for a fixed η_1, K^* is a non-decreasing function of θ_1. This is because an increase in CII rate θ_1 results in more customers being interrupted and hence require more waiting space in BIP to avoid the loss of customers.

On other hand, for a fixed θ_1, K^* is a non-increasing function of η_1. A possible explanation is that as η_1 increases, more customers will clear from BIP and so the cost associated with customers waiting in BIP can be reduced.

Conclusion

In this paper, we studied an $M/M/1$ queueing system with two types of service interruptions namely, customer induced interruption (CII) and server induced interruption (SII). We assumed that no more than one CII is allowed for a customer in service and 'non-preemptive priority' for interrupted customers over the primary arrivals. An application of this model in the wireless network are also discussed. The system is exhaustively analyzed. Steady-state analysis of the model is carried out using Matrix-geometric method. Several performance measures including expected waiting time in the queue are derived. The influence of various parameters on the system performance are also investigated. For this model, an optimization problem for the optimal value of buffer size, K is numerically investigated. As extensions of this work, one can consider the same model with infinite number of CII to a customer is permitted to have before service completion. More complex case where a self-interrupted is permitted to have a finite number of interruption during his service.

Acknowlegment

The authors thanks to Professor A. Krishnamoorthy and the editor Professor M. Sambandham for constructive suggestions in the entire work of this paper.

REFERENCES

[1] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*, 2nd ed., Dover Publications Inc., New York, 1994.

[2] G. Latouche and V. A. Ramaswami, A logarithmic reduction algorithm for quasi-birth-and-death processes. *Journal of Applied Probability*, 30 : 650–674, 1993.

Table 3. Optimal K^* and $\boldsymbol{ETP}$ for selected θ_1 and η_1

θ_1	Measure	η_1				
		1	2	3	4	5
1	K^*	7	6	6	5	5
	$\boldsymbol{ETP}$	82.53600	82.64507	82.69092	82.71677	82.73352
2	K^*	7	6	6	6	5
	$\boldsymbol{ETP}$	86.13011	86.36037	86.45718	86.51190	86.54747
3	K^*	7	6	6	6	5
	$\boldsymbol{ETP}$	88.62145	88.97006	89.11680	89.19996	89.25416
4	K^*	7	6	6	6	6
	$\boldsymbol{ETP}$	90.43359	90.89278	91.08624	91.19608	91.26785
5	K^*	8	6	6	6	6
	$\boldsymbol{ETP}$	91.80310	92.36162	92.59810	92.73243	92.82044
6	K^*	8	7	6	6	6
	$\boldsymbol{ETP}$	92.86957	93.51608	93.79200	93.94856	94.05139
7	K^*	8	7	6	6	6
	$\boldsymbol{ETP}$	93.71980	94.44590	94.75651	94.93318	95.04946
8	K^*	8	7	6	6	6
	$\boldsymbol{ETP}$	94.41045	95.20920	95.55043	95.74533	95.87377
9	K^*	8	7	6	6	6
	$\boldsymbol{ETP}$	94.97978	95.84549	96.21418	96.42571	96.56514
10	K^*	9	7	6	6	6
	$\boldsymbol{ETP}$	95.45698	96.38369	96.77645	97.00329	97.15267

[3] G. Latouche and V. Ramaswami, Introduction to Matrix Analytic Methods in Stochastic Modeling. *SIAM.*, Philadelphia, PA, 1999.

[4] V. Jacob, S. R. Chakravarthy and A. Krishnamoorthy, On a Customer induced interruption in a service system. *Abstract of International Workshop on Retrial Queues* (8^{th} WRQ), July 27–29, Beijing, China, 2010.

[5] K. A. Darabkh1, Abu-Jaradeh, and I. F. Jafar, Incorporating automatic repeat request and thresholds with variable complexity decoding algorithms over wireless networks: queuing analysis. *IET Commun.*, 5(10) : 1377-1393, 2011.

[6] V. Jacob, S. R. Chakravarthy and A. Krishnamoorthy, On a Customer Induced Interruption in a service system. *Stochastic Analysis and Applications*, 30 : 949–962, 2012.

[7] V. Jacob and A. Krishnamoorthy, Analysis of Customer induced interruption in a multi server system. *Neural, Parallel, and Scientific Computations*, 20 : 153–172, 2012.

[8] V. Jacob, Queues with Customer Induced Interruption. *Ph. D thesis submitted to Cochin University of Science & Technology*, Cochin-682 022, India, 2012.

[9] A. Krishnamoorthy, P. K Pramod and S. R. Chakravarthy, A survey on queues with interruptions. *TOP-Spanish journal of Statistics & Operations Research*, 22(1) : 290-320, 2014.

[10] V. Jacob, and A. Krishnamoorthy, Analysis of Customer-Induced Interruption and Retrial of Interrupted Customers. *American Journal of Mathematical and Management Sciences*, 34 (4) : 343–366, 2015.

[11] J. Punalal and S. Babu, A single server queueing model with self-generation of priorities, customer induced interruption and retrial of customers. *International Journal of Research in Advent Technology*,7(3 : 1299–1303, 2019.