Neural, Parallel, and Scientific Computations 20 (2012) 153-172

ANALYSIS OF CUSTOMER INDUCED INTERRUPTION IN A MULTI SERVER SYSTEM

A. KRISHNAMOORTHY AND VARGHESE JACOB

Department of Mathematics, Cochin University of Science and Technology,
Cochin-22, India
achyuthacusat@gmail.com
Department of Mathematics, Government College,
Kottayam-686 013, India,
v.varghesejacob@gmail.com

ABSTRACT. We study a multi-server queueing system with customer induced interruption of service to which customers arrive according to Poisson process and service time follows an exponential distribution. The customer induced interruption occurs according to Poisson process and the interruption duration follows an exponential distribution. The self-interrupted customer will enter into a buffer of finite capacity. Any self-interrupted customer, finding the buffer full, is considered lost for ever. Those self-interrupted customers who complete their interruption will be placed into another buffer of same size. The self-interrupted customers waiting for service are given non-preemptive priority over new customers. We investigate the behavior of this queueing system. Several performance measure are evaluated. Numerical illustrations of the system behavior are also provided. Optimization problem to maximize the revenue with respect to number of servers to be employed and optimal buffer size for the self-interrupted customers are discussed through two illustrative examples.

AMS (MOS) Subject Classification. 68K25

1. INTRODUCTION

Service interruptions is common phenomena in queueing system. For example, in production and manufacturing set up the machine (offering services to jobs) can fail in the middle of a service due to wear and tear. For details on queues with server interruptions (other than priority queues) we refer the reader to Krishnamoorthy et al [1]. Almost all work on interruptions deal with server induced ones.

As far as our knowledge goes, the first paper dealing with customer induced interruption is [2] reported at the 8^{th} International Workshop on Retrial Queues in 2010. The present paper deals with interruption that is induced by the customers while in service. A customer who is currently in service can be self-interrupted. The motivation for such interruptions arise from situations seen in practice. A more commonly occurring example is the following: while a patient is being examined, the

Received February 27, 2012

physician may find that one or more tests are needed for prescription of medicine. Hence he/she is asked to undergo these and return to the clinic. Such patients can be regarded as self-interrupted customers. Another example come is : while trying to send a package in post office, the customer may require a service such as confirmation of the delivery of the package which calls for the customer to fill in certain forms (if it is not done earlier). This is to be done by moving away from the server to a place that has all relevant forms. Yet another example is : while in service the customer is interrupted through his/her cell phone and needs immediate attention. Other examples can be found in Online services. 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 interruptions.

The queueing model considered in the present paper is an extension of the work [**?**] to multi-server systems. Miaomiao Yu et al [6] considered an $M/E_k/1$ system wherein a customer, on completion of first phase of service, is required to undergo service elsewhere with a specified probability, before proceeding to second phase of service. With complement probability he proceeds to second phase of service. However, he may have to wait until service of all those customers ahead of him whose service got interrupted after first phase, but have completed interruption during the present customer's service (in the first phase in case service gets interrupted in that phase; else until he/she completes the whole service), are completed. They also assumed that there is no interruption beyond the first phase. Our model differs from the above in that self interruption occurs with in the phase whereas the interruption discussed in Miaomiao Yu et al [6] occur immediately after service in first phase. However, for queueing systems with more than one server the model described has not been studied so far.

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 a few key performance measures and optimization problem. Some illustrative examples are discussed in section 4.

2. MODEL DESCRIPTION

We consider an infinite capacity multi server queuing model to which customers arrive according to Poisson process with rate λ (Figure1). The service facility consists of c servers. All c servers are assumed to be homogeneous and that the service times are exponentially distributed with parameter μ. An arriving customer, finding a free server, enters into service immediately; otherwise the customer is placed into the buffer of infinite capacity and it will be picked up for service according to the order of their arrival. We consider customer induced interruption while his/her service is

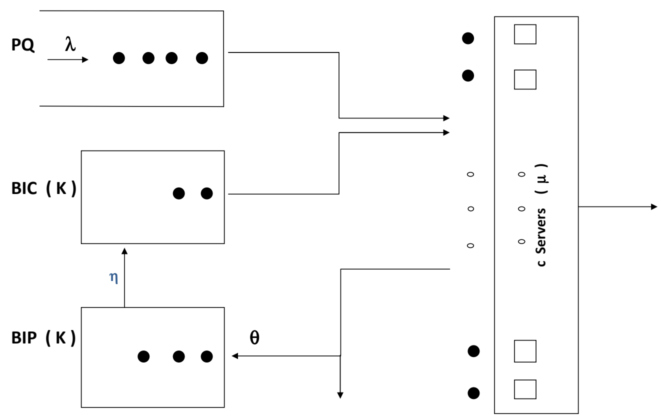

Figure 1: Customer Induced Interruption in a multi-server queueing system

going on. The interruption occurs according to a Poisson process of rate θ. When an interruption occurs, the customer currently in service will be forced to leave the service facility. The freed server is ready to offer services to other customers. 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. An interrupted customer spends a random period of time for completion of interruption, independent of other customers. The duration of an interruption follows an exponential distribution with parameter η. In this paper we assume that no more than one interruption 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 who are 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 services to those customers waiting in BIC before serving new customers by maintaining the order of their arrival. Because of this restriction coupled, with the fact that at most one interruption is allowed for a customer, the total number of customers in BIC and BIP will never exceed the size of BIP and hence we assume the buffer sizes to be the same.

In the sequel we use the following notations.

- $N(t)$= Number of customers in the primary queue at time t
- $N_1(t)$= Number of busy servers at time t

- $N_2(t)$= Number of servers busy with primary customers at time t
- $N_3(t)$ = Number of customers in BIC at time t
- $N_4(t)$ = Number of customers in BIP at time t
- $\boldsymbol{a}_i = (1, 2, \ldots, i);\ 1 \leq i \leq K$
- $\mathbf{e}$ denotes column vector of 1's with appropriate dimension,
- $\mathbf{e}_j(r)$ denote column vector of dimension r with 1 in the j^{th} position and 0 elsewhere
- I_r denote identity matrix of dimension r.
- $\Delta(\boldsymbol{a}_i)$ is a diagonal matrix whose diagonal entries are the components of the vector $\boldsymbol{a}_i$.
- $L = (K+1)(K+2)/2$.
- $LC = L * (c+1)$.

The process $\{(N(t), N_1(t), N_2(t), N_3(t), N_4(t)) : t \geq 0\}$ is a continuous-time Markov chain ($CTMC$) whose state space is given by

$$\begin{aligned}\Omega &= \{(0,0,0,0,i_2) : 0 \leq i_2 \leq K\} \\ &\bigcup \{(0,j,m,0,i_2) : 1 \leq j \leq c-1, 0 \leq m \leq j, 0 \leq i_2 \leq K\} \\ &\bigcup \{(n,c,m,i_1,i_2) : n \geq 0, 0 \leq m \leq c, 0 \leq i_1, i_2 \leq K, 0 \leq i_1 + i_2 \leq K\}\end{aligned}$$

A brief description of the above states are given below.

- $(\underline{0,0}) = (0,0,0,0,i_2) : -$ the system has no customers in the primary queue, all servers including primary servers are idle, no customers in the BIC and BIP has i_2 customers.
- $(\underline{j,m}) = (0,j,m,0,i_2) : -$ the system has no customer in the primary queue, there are j servers are busy of which m servers are busy with primary queue customers ($1 \leq j \leq c-1$ and $0 \leq m \leq j$), no customer in BIC and BIP has i_2 customers.
- $(\underline{c,m}) = (n,c,m,i_1,i_2) : -$ there are $n (n \geq 0)$ customers in the primary queue, all c servers are busy of which m servers are busy with primary queue customers, $(0 \leq m \leq c)$, BIC has i_1 customers and BIP has i_2 customers.

Level $l(0,j)$ denotes the union of $(j+1)(K+1)$ states given by

$$l(0,j) = \bigcup_{m=0}^{j} \{(0,j,m,0,i_2) :\ 0 \leq i_2 \leq K\};\ 0 \leq j \leq c-1.$$

Level $l(n,c)$ denotes the union of LC states given by

$$l(n,c) = \bigcup_{m=0}^{c} \{(n,c,m,i_1,i_2) : 0 \leq i_1 + i_2 \leq K, 0 \leq i_1, i_2 \leq K\}\ ;\ n \geq 0.$$

To write down the infinitesimal generator Q, we introduce additionally the following notations:

- $I^* = \begin{bmatrix} \mu & \theta & & & \\ & \mu & \theta & & \\ & & \ddots & \ddots & \\ & & & \mu & \theta \\ & & & & \mu+\theta \end{bmatrix}_{(K+1)\times(K+1)}$
- $\widetilde{I}^* = \begin{bmatrix} I^* \\ O \end{bmatrix}_{L\times(K+1)}$
- $\widetilde{I}^{**} = \begin{bmatrix} I^* & O \\ O & O \end{bmatrix}_{L\times L}$
- $F^* = \eta \begin{bmatrix} \mathbf{0} & 0 \\ \Delta(a_K) & \mathbf{0} \end{bmatrix}_{(K+1)\times(K+1)}$
- $\widehat{F}^* = \begin{bmatrix} F^* & O \end{bmatrix}_{(K+1)\times L}$
- $\widetilde{I}_{K+1} = \begin{bmatrix} I_{K+1} \\ O \end{bmatrix}_{L\times(K+1)}$
- $\widehat{I}_{K+1} = \begin{bmatrix} I_{K+1} & O \end{bmatrix}_{(K+1)\times L}$

For $1 \le p \le K$,

- $F_p = \begin{bmatrix} \mathbf{0} \\ \Delta(a_p) \end{bmatrix}_{(p+1)\times p}$, $J_p = \begin{bmatrix} I_p & \mathbf{0} \end{bmatrix}$, $G_p = \Delta(0 \;\; a_p)$, $G_0 = \; 0$,
- $H_p = \begin{bmatrix} \mu & \theta & & & \\ & \mu & \theta & & \\ & & \ddots & \ddots & \\ & & & \mu & \theta \end{bmatrix}_{p\times(p+1)}$,

If the states in Ω are listed in lexicographical order then the infinitesimal generator of the $CTMC$ governing the system is given by

(2.1)

$$Q = \begin{array}{c} \\ l(0,0) \\ l(0,1) \\ l(0,2) \\ \vdots \\ l(0,c-1) \\ l(0,c) \\ l(1,c) \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccccc} l(0,0) & l(0,1) & l(0,2) & \dots & l(0,c-1) & l(0,c) & l(1,c) & \dots \end{array} \\ \begin{pmatrix} E_0 & C_0 & & & & & & \\ B_1 & E_1 & C_1 & & & & & \\ & B_2 & E_2 & C_2 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & & E_{c-1} & C_{c-1} & & \\ & & & & B_c & A_1 & A_0 & \\ & & & & & A_2 & A_1 & A_0 \\ & & & & & & \ddots & \ddots & \ddots \end{pmatrix} \end{array}$$

where the coefficient matrices appearing in (2.1) are given by

$$E_j = \begin{array}{c} \\ \underline{(j,0)} \\ \underline{(j,1)} \\ \vdots \\ \underline{(j,j-1)} \\ \underline{(j,j)} \end{array} \begin{array}{c} \begin{array}{ccccc} \underline{(j,0)} & \underline{(j,1)} & \dots & \underline{(j,j-1)} & \underline{(j,j)} \end{array} \\ \begin{pmatrix} D_{j,0} & & & & \\ & D_{j,1} & & & \\ & & \ddots & & \\ & & & D_{j,j-1} & \\ & & & & D_{j,j} \end{pmatrix} \end{array}$$

where

$$D_{j,i} = -\Delta\left(\lambda + j\mu + i\theta, \lambda + j\mu + i\theta + \eta, \dots, \lambda + j\mu + i\theta + K\eta\right)$$

$$j = 0, 1, \dots, (c-1), \quad i = 0, \dots, j$$

$$B_j = \begin{array}{c} \\ \underline{(j,0)} \\ \underline{(j,1)} \\ \underline{(j,2)} \\ \vdots \\ \underline{(j,j-1)} \\ \underline{(j,j)} \end{array} \begin{array}{c} \begin{array}{ccccc} \underline{(j-1,0)} & \underline{(j-1,1)} & \dots & \underline{(j-1,j-2)} & \underline{(j-1,j-1)} \end{array} \\ \begin{pmatrix} j\mu I_{K+1} & & & & \\ I^* & (j-1)\mu I_{K+1} & & & \\ & 2I^* & & & \\ & & \ddots & \ddots & \\ & & & (j-1)I^* & \mu I_{K+1} \\ & & & & jI^* \end{pmatrix} \end{array}$$

$$j = 0, \dots, c-2$$

$$C_j = \begin{array}{c} \\ \underline{(j,0)} \\ \underline{(j,1)} \\ \vdots \\ \underline{(j,j-1)} \\ \underline{(j,j)} \end{array} \begin{array}{c} \begin{array}{cccccc} \underline{(j+1,0)} & \underline{(j+1,1)} & \underline{(j+1,2)} & \dots & \underline{(j+1,j)} & \underline{(j+1,j+1)} \end{array} \\ \begin{pmatrix} F^* & \lambda I_{K+1} & & & & \\ & F^* & \lambda I_{K+1} & & & \\ & & \ddots & \ddots & & \\ & & & & \lambda I_{K+1} & \\ & & & & F^* & \lambda I_{K+1} \end{pmatrix} \end{array}$$

$$j = 0, \dots, c-2$$

$$B_c = \begin{array}{c} \\ \underline{(c,0)} \\ \underline{(c,1)} \\ \underline{(c,2)} \\ \vdots \\ \underline{(c,c-1)} \\ \underline{(c,c)} \end{array} \begin{array}{c} \begin{array}{ccccc} \underline{(c-1,0)} & \underline{(c-1,1)} & \dots & \underline{(c-1,c-2)} & \underline{(c-1,c-1)} \end{array} \\ \begin{pmatrix} c\mu \widetilde{I}_{K+1} & & & & \\ \widetilde{I}^* & (c-1)\mu \widetilde{I}_{K+1} & & & \\ & 2\widetilde{I}^* & & & \\ & & \ddots & \ddots & \\ & & & (c-1)\widetilde{I}^* & \mu \widetilde{I}_{K+1} \\ & & & & c\widetilde{I}^* \end{pmatrix} \end{array}$$

$$C_{c-1} = \begin{array}{c} \\ (\underline{c-1,0}) \\ (\underline{c-1,1}) \\ \vdots \\ (\underline{c-1,c-2}) \\ (\underline{c-1,c-1}) \end{array} \begin{array}{c} \begin{array}{cccccc} (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \dots & (\underline{c,c-1}) & (\underline{c,c}) \end{array} \\ \left(\begin{array}{cccccc} \widehat{F}^{*} & \lambda \widehat{I}_{K+1} & & & & \\ & \widehat{F}^{*} & \lambda \widehat{I}_{K+1} & & & \\ & & \ddots & \ddots & & \\ & & & & \lambda \widehat{I}_{K+1} & \\ & & & & \widehat{F}^{*} & \lambda \widehat{I}_{K+1} \end{array} \right) \end{array}$$

$$A_1 = \begin{array}{c} \\ (\underline{c,0}) \\ (\underline{c,1}) \\ (\underline{c,2}) \\ \vdots \\ (\underline{c,c-1}) \\ (\underline{c,c}) \end{array} \begin{array}{c} \begin{array}{cccccc} (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \dots & (\underline{c,c-1}) & (\underline{c,c}) \end{array} \\ \left(\begin{array}{cccccc} A^{(1)}_{0,0} & & & & & \\ A^{(1)}_{1,0} & A^{(1)}_{1,1} & & & & \\ & A^{(1)}_{2,1} & A^{(1)}_{2,2} & & & \\ & & \ddots & \ddots & & \\ & & & & A^{(1)}_{(c-1),(c-1)} & \\ & & & & A^{(1)}_{c,(c-1)} & A^{(1)}_{c,c} \end{array} \right) \end{array}$$

where

$$A^{(1)}_{i,i} = -\left\{(\lambda + c\mu + i\theta)\ I_L + \eta\ \Delta(G_K \dots G_0)\right\} + \eta \begin{bmatrix} O & \Delta(F_K \dots F_1) \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$+ (c-i)\ \mu \begin{bmatrix} O & \mathbf{0} \\ \Delta(J_K \dots J_1) & \mathbf{0} \end{bmatrix},\ i = 0, \dots, c$$

$$A^{(1)}_{i(i-1)} = i \begin{bmatrix} O & \mathbf{0} \\ \Delta(H_K \dots H_1) & \mathbf{0} \end{bmatrix};\ i = 1, \dots, c$$

$$A_2 = \begin{array}{c} \\ (\underline{c,0}) \\ (\underline{c,1}) \\ (\underline{c,2}) \\ \vdots \\ (\underline{c,c-1}) \\ (\underline{c,c}) \end{array} \begin{array}{c} \begin{array}{cccccc} (\underline{c,0}) & (\underline{c,1}) & (\underline{c,2}) & \dots & (\underline{c,c-1}) & (\underline{c,c}) \end{array} \\ \left(\begin{array}{cccccc} O & A^{(2)}_0 & & & & \\ & \widetilde{I}^{**} & A^{(2)}_1 & & & \\ & & 2\widetilde{I}^{**} & \ddots & & \\ & & & \ddots & & \\ & & & & (c-1)\widetilde{I}^{**} & A^{(2)}_{c-1} \\ & & & & & c\widetilde{I}^{**} \end{array} \right) \end{array}$$

where

$$A^{(2)}_j = (c-j)\ \mu \begin{bmatrix} I_{K+1} & O \\ O & O \end{bmatrix}_{L \times L};\ j = 0, \dots, c-\ 1$$

$$A_0 = \lambda\ I_{LC}$$

3. STEADY-STATE ANALYSIS

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

3.1. **Stability condition.** Let $\boldsymbol{\pi}$ denote the steady-state probability vector of the generator $A_0 + A_1 + A_2$. That is, $\boldsymbol{\pi}(A_0 + A_1 + A_2) = \mathbf{0}$, $\boldsymbol{\pi e} = 1$. The $LIQBD$ description of the model indicates that the queueing system is stable (see Neuts [3]) if and only if

$$\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}. \tag{3.1}$$

The vector, $\boldsymbol{\pi}$, cannot be obtained 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

$$\rho = \frac{\boldsymbol{\pi} A_0 \boldsymbol{e}}{\boldsymbol{\pi} A_2 \boldsymbol{e}}. \tag{3.2}$$

Note that the stability 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 4.

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

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

we see that $\boldsymbol{x}$, under the assumption that the stability condition holds, the steady state probability vector is obtained as

$$\boldsymbol{x}(n) = \boldsymbol{x}(0) R^n, \quad n \geq 1, \tag{3.5}$$

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

$$R^2 A_2 + R A_1 + A_0 = \mathbf{O},$$

and the vectors $\boldsymbol{x}^*(0), \boldsymbol{x}^*(1), \ldots, \boldsymbol{x}^*(c-1), \boldsymbol{x}(0)$ are obtained from boundary equations

$$\boldsymbol{x}^*(0) E_0 + \boldsymbol{x}^*(1) B_1 = \mathbf{0},$$

$$\boldsymbol{x}^*(i-1) C_{i-1} + \boldsymbol{x}^*(i) E_i + \boldsymbol{x}^*(i+1) B_{i+1} = \mathbf{0}, \quad 1 \leq i \leq c-2$$

$$\boldsymbol{x}^*(c-2) C_{c-2} + \boldsymbol{x}^*(c-1) E_{c-1} + \boldsymbol{x}(0) B_c = \mathbf{0},$$

$$\boldsymbol{x}^*(c-1) C_{c-1} + \boldsymbol{x}(0)[A_1 + R A_2] = 0,$$

Once R matrix is obtained, from the boundary equation we obtain

$$\boldsymbol{x}(0) = \boldsymbol{x}^*(c-1) R_{c-1}$$

$$\boldsymbol{x}^*(i) = \boldsymbol{x}^*(i-1)R_{i-1}, \quad 1 \leq i \leq (c-1)$$

which gives $\boldsymbol{x}(0) = \boldsymbol{x}^*(0)\prod_{i=0}^{c-1} R_i$ where $R_{c-1} = C_{c-1}\left[-(A_1 + RA_2)\right]^{-1}$ and $R_{i-1} = C_{i-1}\left[-(E_i + R_iB_{i+1})\right]^{-1}$. The component $\boldsymbol{x}^*(0)$ is the steady state distribution of the Markov Chain with generator matrix $[E_0 + R_0B_1]$ subject to the normalizing equation

$$\boldsymbol{x}^*(0)\left(I + \sum_{i=0}^{c-2}\prod_{j=0}^{i} R_j + \prod_{j=0}^{c-1} R_j(I-R)^{-1}\right)\boldsymbol{e} = 1. \tag{3.6}$$

Thus, the vector $\boldsymbol{x}$ can be computed by exploiting the special structure of the coefficient matrices. The details are omitted. One can use logarithmic reduction algorithm for computing R. We will list only the the main steps involved in the logarithmic reduction algorithm for the computation of R. For full details on the logarithmic reduction algorithm we refer the reader to [5].

Logarithmic Reduction Algorithm for R:

Step 0: $H \leftarrow (-A_1)^{-1}A_0$, $L \leftarrow (-A_1)^{-1}A_2$, $G = L$, and $T = H$.

Step 1:

$$\begin{aligned} U &= HL + LH \\ M &= H^2 \\ H &\leftarrow (I-U)^{-1}M \\ M &\leftarrow L^2 \\ L &\leftarrow (I-U)^{-1}M \\ G &\leftarrow G + TL \\ T &\leftarrow TH \end{aligned}$$

Continue Step 1 until $||\boldsymbol{e} - G\boldsymbol{e}||_\infty < \epsilon$.

Step 2: $R = -A_0(A_1 + A_0G)^{-1}$.

3.3. **Stationary waiting time distribution in the queue.** The stationary waiting time distribution for this queueing model, in general, is analytically intractable. However we will obtain the Laplace-Stieltjes transform (LST) of the waiting time of a customer in the queue and derive an expression for its mean. First note that an arriving customer will enter into service immediately with probability $w_0 = \sum_{i=0}^{c-1}\boldsymbol{x}^*(i)\boldsymbol{e}$. With probability $1 - w_0$ the arriving customer has to wait before getting into service. The waiting time may be viewed as the time until absorption in a Markov chain with a highly sparse structure. The state space (that includes the arriving customer in its count) of the Markov chain is given by

$$\tilde{\Omega} = \{*\}\bigcup\{(n, c, j, i_1, i_2) :, 0 \leq j \leq c,\ 0 \leq i_1, i_2 \leq K, 0 \leq i_1 + i_2 \leq K, n \geq 1\}.$$

The state * is obtained by lumping together the states that correspond to at least one of the server being idle. That is, * is obtained by lumping $\{(0, j, m, 0, i_2) :\ 0 \leq$

$j \leq c-1;\ 0 \leq m \leq j;\ 0 \leq i_2 \leq K\}$. Its generator matrix $\tilde{Q}$ is given by

$$\tilde{Q} = \begin{pmatrix} 0 & \boldsymbol{0} & & & \\ \boldsymbol{a} & \tilde{A}_1 & & & \\ & A_2 & \tilde{A}_1 & & \\ & & A_2 & \tilde{A}_1 & \\ & & & \ddots & \ddots \end{pmatrix}, \tag{3.7}$$

where

$$\tilde{A}_1 = A_1 + \lambda I,\ \ \boldsymbol{a} = A_2 \boldsymbol{e}$$

The initial probability vector of $\tilde{Q}$ is denoted by $\boldsymbol{z}$ and in partitioned form is given by

$$\boldsymbol{z} = (w_0, \boldsymbol{x}(0), \boldsymbol{x}(1), \cdots).$$

Define $W(t), t \geq 0$ to be the probability that an arriving customer will enter into service no later than time t. We will now derive the Laplace-Stieltjes transform (LST), $\tilde{w}(s)$, of $W(t)$. This transform is useful in deriving an expression for the mean waiting time. Using the structure of $\tilde{Q}$, it can readily be verified that

Theorem 3.1. *The LST, $\tilde{w}(s)$, of $\widetilde{\mathbf{W}}(t)$ is given by*

$$\tilde{w}(s) = w_0 + \sum_{i=0}^{\infty} \boldsymbol{x}(i)[(sI - \tilde{A}_1)^{-1} A_2]^i (sI - \tilde{A}_1)^{-1} \boldsymbol{a}. \tag{3.8}$$

Corollary 3.2. *The mean waiting time E_W^Q, in the queue of an arriving customer is given by*

(3.9)

$$E_W^Q = [\boldsymbol{x}(0)(I-R)^{-1} - \boldsymbol{x}(0) \sum_{k=0}^{\infty} R^k P^{k+1} + \boldsymbol{x}(0)(I-R)^{-2} \tilde{P}](I - P + \tilde{P})^{-1}(-\tilde{A}_1)^{-1} \boldsymbol{e},$$

where

$$P = (-\tilde{A}_1)^{-1} A_2,\ \ \tilde{P} = \boldsymbol{e}\boldsymbol{p}, \tag{3.10}$$

and $\boldsymbol{p}$ is the invariant probability vector of P. That is,

$$\boldsymbol{p}P = \boldsymbol{p},\ \ \boldsymbol{p}\boldsymbol{e} = 1. \tag{3.11}$$

NOTE: In the computation of the mean waiting time E_W^Q, we need to evaluate the infinite sum $\sum_{k=0}^{\infty} R^k P^{k+1}$. On noting that P is a stochastic matrix, we get $\boldsymbol{x}(0) \sum_{k=0}^{\infty} R^k P^{k+1} \boldsymbol{e} = 1 - w_0$ and hence in truncating the infinite sum we find N^* such that $|\boldsymbol{x}(0) \sum_{k=0}^{N^*} R^k P^{k+1} \boldsymbol{e} - (1 - w_0)| < \epsilon$, where ϵ is a pre-determined quantity such as 10^{-7}.

3.4. **System performance measures.** In this section we list a number of key system performance measures to bring out the qualitative aspects of the model under study. These are listed below along with their formula for computation. Towards this end, we further partition the vectors $\boldsymbol{x}^*(i)$ and $\boldsymbol{x}(n)$ into smaller vectors as follows:

$$\boldsymbol{x}^*(i) = (\ \boldsymbol{x}^*_{j,0,i_2}(i)\ ),\ i = 0, \ldots, c-1, j = 0, \ldots, i\ ;$$

$$\boldsymbol{x}(n) = (\ \boldsymbol{x}_{c,0}(n), \ldots, \boldsymbol{x}_{c,c}(n)\ )\ ;$$

$$\boldsymbol{x}_{c,m}(n) = (\ \boldsymbol{x}_{c,m,0}(n), \ldots, \boldsymbol{x}_{c,m,K}(n)\ )\ ;$$

$$\boldsymbol{x}(n) = (\ \boldsymbol{x}_{c,m,i_1,i_2}(n)\ )\ ;$$

Note that $\boldsymbol{x}^*(i)$,$\boldsymbol{x}(n)$,$\boldsymbol{x}_{c,m}(n)$ and $\boldsymbol{x}_{c,m,k}(n)$ are respectively of dimension $(i+1)(K+1)$, LC, L and $(K+1)-k$, $n \geq 0,\ m = 0, \ldots c,\ k = 0, \ldots, K$.

- The probability that all servers are idle :
  $P_{idle} = \boldsymbol{x}^*(0)\boldsymbol{e}$.
- The probability that an interrupted customer is lost:
  $P_{loss} = \frac{\theta}{\theta+\mu} \sum\limits_{i=1}^{c-1} \sum\limits_{j=1}^{i} \boldsymbol{x}^*_{j,0,K}(i) + \frac{\theta}{\theta+\mu} \sum\limits_{n=0}^{\infty} \sum\limits_{m=1}^{c} x_{c,m,0,K}(n)$.
- Mean number of idle servers :
  $\mu_{IDS} = \sum\limits_{i=0}^{c-1} (c-i)\ \boldsymbol{x}^*(i)\boldsymbol{e}$.
- Mean number of busy servers :
  $\mu_{BYS} = \sum\limits_{i=1}^{c-1} i\ \boldsymbol{x}^*(i)\boldsymbol{e} + c\ \boldsymbol{x}(0)(I-R)^{-1}\boldsymbol{e}$.
- Mean number of servers busy with primary queue customers :
  $\mu_{SBYP} = \sum\limits_{i=1}^{c-1} \sum\limits_{j=1}^{i} \sum\limits_{i_2=0}^{K} j\ \boldsymbol{x}^*_{j,0,i_2}(i) + \sum\limits_{n=0}^{\infty} \sum\limits_{m=1}^{c} \sum\limits_{i_1=0}^{K} \sum\limits_{i_2=0}^{K-i_1} m\ x_{c,m,i_1,i_2}(n)$.
- Mean number of servers busy with BIC customers :
  $\mu_{SBYI} = \sum\limits_{i=1}^{c-1} \sum\limits_{j=0}^{i-1} \sum\limits_{i_2=0}^{K} (i-j)\ \boldsymbol{x}^*_{j,0,i_2}(i) + \sum\limits_{n=0}^{\infty} \sum\limits_{m=0}^{c-1} \sum\limits_{i_1=0}^{K} \sum\limits_{i_2=0}^{K-i_1} (c-m)\ x_{c,m,i_1,i_2}(n)$.
- Mean number of customers in the primary queue:
  $\mu_{PQ} = \boldsymbol{x}(0)R(I-R)^{-2}\boldsymbol{e}$.
- The mean number of interrupted customers in the BIP buffer:
  $\mu_{BIP} = \sum\limits_{i=0}^{c-1} \sum\limits_{j=0}^{i} \sum\limits_{i_2=0}^{K} i_2\ x^*_{j,0,i_2}(i) + \sum\limits_{n=0}^{\infty} \sum\limits_{m=0}^{c} \sum\limits_{i_2=0}^{K} \sum\limits_{i_1=0}^{K-i_2} i_2\ x_{c,m,i_1,i_2}(n)$.
- The mean number of interrupted customers in the BIC buffer:
  $\mu_{BIC} = \sum\limits_{n=0}^{\infty} \sum\limits_{m=0}^{c} \sum\limits_{i_1=0}^{K} \sum\limits_{i_2=0}^{K-i_1} i_1\ x_{c,m,i_1,i_2}(n)$.
- The mean waiting time in the queue E_W^Q, is as given in (3.9).

3.5. **An Optimization Problem.** In this section we propose an optimization problem and discuss it through illustrative examples (2 and 3 in section 4). To construct an objective function we assume that customer induced interruptions produce revenue to the system in contrast to server induced interruptions. Interrupted customers have to pay more cost than those without interruption. Also idle servers, loss of customers and waiting spaces in primary queue and BIC involve expenditure to the system. Thus we introduce per unit time revenue and cost as follows.

- revenue r_1 per customer leaving the system with an uninterrupted service,
- revenue $r_2(> r_1)$ per customer leaving the system on completion of service after an interruption,
- holding cost c_1 per unit time that a customer has to wait in the primary queue,
- holding cost c_2 per unit time that a customer has to wait in the BIC buffer,
- cost c_3 per unit time that each customer lost due to BIP buffer being full at the time an interruption occurs.
- cost c_4 per unit time for each idle servers,

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

$$\boldsymbol{ETP} = r_1\mu_{SBYP} + r_2\mu_{SBYI} - c_1\mu_{PQ} - c_2\mu_{BIC} - c_3(\theta+\mu)P_{loss} - c_4\mu_{IDS}.$$

4. NUMERICAL EXAMPLES

Now we present numerical results for implementing the qualitative nature of the model under study. The correctness and the accuracy of the code are verified by a number of accuracy checks. We consider a few representative examples.

Example 1: The purpose of this example is to see the impact of parameter θ for the case when $c = K = 2, 4, 6, 8$ on some measures. In this example, by fixing $\lambda = 15, \mu = 8$ and $\eta = 2$, we look at the effect of varying θ on some selected measures. These are displayed in Figure 2 and Figure 3. Looking at these figures, we summarize the following observations.

- As θ increases, the traffic intensity ρ, appears to decrease for all values of c and K. The rate of decrease is small for higher values of c and K. ρ is largest for the case when $c = K = 2$. This is as expected since increasing θ will cause an increase in the customers getting lost due to BIP being full for small values of c and K and for higher values of c and K, that is with more servers and more waiting space in BIP, help to clear the customers at a faster rate. When θ is progressively decreased and comes closer and closer to zero,our model converges to the classical queueing problem without interruption. Thus the ratio $\frac{\pi A_0 e}{\pi A_2 e}$ converges to the traffic intensity ρ of the classical situation.
- As is to be expected the measure P_{idle} is a non-decreasing function of θ when all other parameters are fixed.
- From Figure 3a we see that P_{loss} increases with increase in the interruption rate θ and the rate of increase is small for higher values of c and K, of course this is as expected.

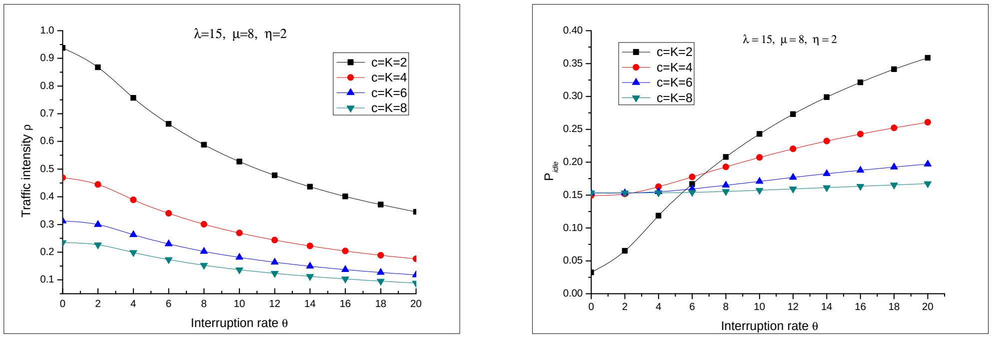

Figure 2: θ versus ρ and θ versus P_{idle}

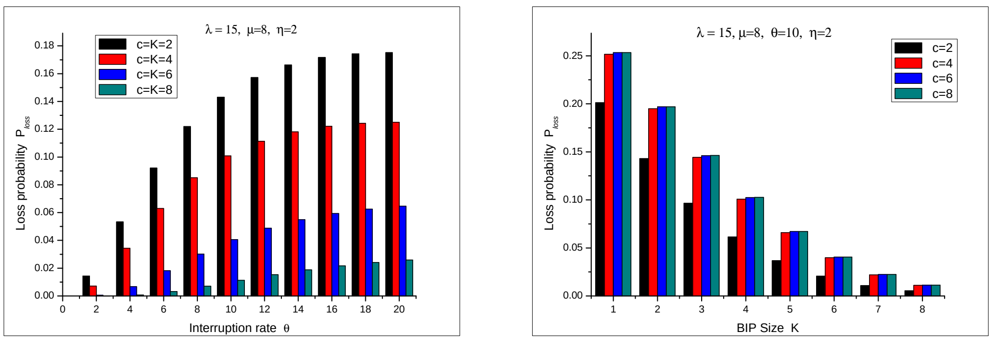

Figure 3: θ versus P_{loss} and K versus P_{loss}

- From Figure 3b it is seen that P_{loss} decrease with increase in the BIP size K for every c fixed, this is as expected. Also for each fixed K, this measure increases as c increases. This is as expected, since for a fixed K, as c increases, more customers may get interrupted from different servers and as a consequence the BIP gets filled (note that $\eta = 2$).
- We notice from Figure 4 that the measure E_W^Q decreases with increase in θ. This measure is largest for the case when $c = K = 2$ and for higher values of c and K, it is quite negligible as to be expected.

Now we discuss two optimization problems associate with Section 3.5.

Example 2: In this example, we fix $K = 5, \lambda = 20, \mu = 11, \eta = 5, r_1 = \$300, r_2 = \$400, c_1 = \$10, c_2 = \$20, c_3 = \$30, c_4 = \$5$. The optimal number of servers, c, that maximizes the expected total profit ETP, for various combinations of θ are displayed in Figure 5a. It is seen from the numerical experiments that ETP increases first and then decreases with increasing θ. The optimum c and the corresponding ETP are given in Table 1.

Example 3: Here we fix $c = 3, \lambda = 15, \mu = 6, \eta = 2, r_1 = \$30, r_2 = \$40, c_1 =$

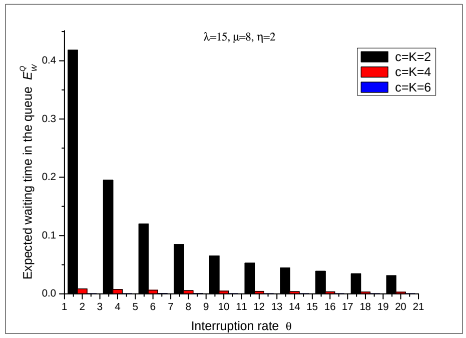

Figure 4: θ versus Expected waiting time in the queue

Table 1: Optimum c and ETP for selected θ

θ	4	8	12	16	20	24
Optimum c	3	4	4	4	4	4
ETP	580.762	599.969	603.297	599.414	592.693	585.104

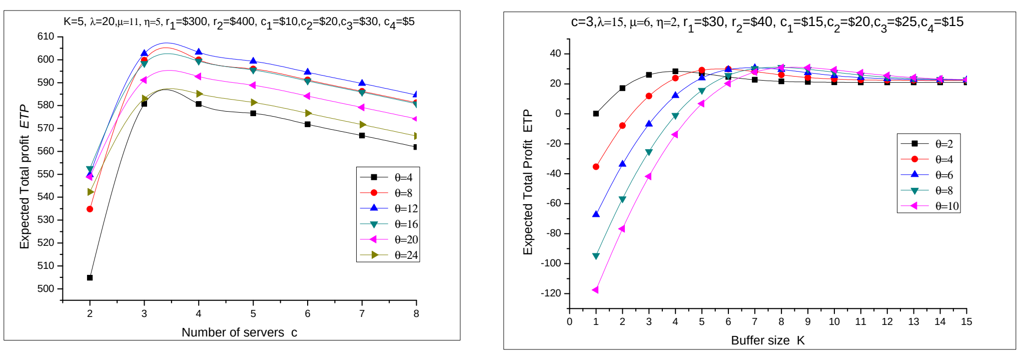

Figure 5: Optimum values of c and K for different θ

\$15, $c_2 = \$20, c_3 = \$25, c_4 = \$15$.The optimum value of K that maximizes the expected total profit ETP, for various combinations of θ are displayed in Figure 5b. The optimum K and the corresponding ETP are given in Table 2.

Table 2: Optimum K and ETP for selected θ

θ	2	4	6	8	10
Optimum K	3	5	6	7	8
ETP	26.90732	29.35727	30.31224	30.78644	31.08506

Example 4: In this example we fix $\lambda = 15, \mu = 7.5, \eta = 2$ and vary the parameters

c, K and θ. In Table 3 and Table 4, we display the measures μ_{IDS}, μ_{BYS}, μ_{SBYP}, μ_{SBYI}, μ_{PQ}, μ_{BIC} and μ_{BIP}. A look at these tables reveal some notable observations.

- For each fixed pair c and K, μ_{IDS} increases and μ_{BYS} decreases as θ increases. This is due to the fact that an increase in θ will cause more customers to be interrupted from different servers leading to an increase in the number of customers leaving the system without getting service.
- For each fixed pair c and K, μ_{SBYI} is a non-decreasing function of θ whereas μ_{SBYP} is a non-increasing function of θ. This is again as expected.
- The measure μ_{PQ} is a non-increasing function of θ for all values of c and K and is largest for the case when $c = 2$ and $K = 5$. This is to be expected since increase in θ results in interrupted customers, for lower values of K, getting lost and for higher values of K they get back to service through BIC buffer.
- Finally, looking at the measures μ_{BIC} and μ_{BIP}, we see some interesting trends. Recall that at any given time the total number of customers in the BIC and BIP buffers cannot exceed K. For all values of c and K, $\mu_{BIC} < \mu_{BIP}$ when θ increases. For higher interruption rate causes more interruption leading to more interrupted customers filling BIP buffer (note that $\eta = 2$) and hence the rate of interrupted customer getting back to service through BIC buffer will be smaller leading to less customers (on the average) in BIC buffer. Also we notice that for each values of c, μ_{BIC} increases initially and then decreases as θ increases further, for higher value K. This is probably due to the fact that as θ reaches a certain value, any further increase in θ will only result in the server being busy with customers in BIC, for higher values of K.

We conclude this section by showing that the mean number of servers busy with primary customers, μ_{SBYP}, is independent of K and c. We are able to prove this only for the case when $c = 1$. Even though the result appear to be true in general, which we verified through numerical computation as we can see in Table 3 and Table 4.

Theorem 4.1. *The server is busy with primary customers is given by*

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

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

$$\boldsymbol{x}^*(0)E_0 + \boldsymbol{x}(0)B_1 = 0, \tag{4.1}$$

$$\boldsymbol{x}^*(0)C_0 + \boldsymbol{x}(0)A_1 + \boldsymbol{x}(1)A_2 = 0, \tag{4.2}$$

and

$$\boldsymbol{x}(i-1)A_0 + \boldsymbol{x}(i)A_1 + \boldsymbol{x}(i+1)A_2 = 0, i \geq 1. \tag{4.3}$$

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

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

and

$$\lambda(\boldsymbol{x}_{1,0}(i)\boldsymbol{e} + \boldsymbol{x}_{1,1}(i)\boldsymbol{e}) = \mu \boldsymbol{x}_{1,0,0}(i+1)\boldsymbol{e} + (\mu+\theta)\boldsymbol{x}_{1,1,0}(i+1)\boldsymbol{e}, \ i \geq 0. \tag{4.5}$$

Now post-multiplying equations (4.2) and (4.3) by $(\boldsymbol{e}_1(2) \otimes \boldsymbol{e})$ and adding over $i = 1$ to ∞, we get

$$\lambda \boldsymbol{x}^*(0)\boldsymbol{e} = (\mu+\theta)\sum_{i=0}^{\infty} \boldsymbol{x}_{1,1}(i)\boldsymbol{e} - \mu \sum_{i=1}^{\infty} \boldsymbol{x}_{1,0,0}(i)\boldsymbol{e} - (\mu+\theta)\sum_{i=1}^{\infty} \boldsymbol{x}_{1,1,0}(i)\boldsymbol{e}. \tag{4.6}$$

The stated result follows by immediately by adding (4.5) over i and (4.6). □

5. CONCLUSION

In this paper we considered a multi-server queueing system with customer induced interruption of service. All underlying distributions are assumed to be exponential that are independent of each other. A finite buffer BIP, of capacity K, for self interrupted customers to wait for completion of interruption and another buffer BIC, of the same capacity, for those who have completed interruptions, are introduced. The combined maximum customers held in BIP and BIC together is K for reasons obvious from the formulation of the model.

The steady state analysis of the model is performed using Matrix Analytic Method. The Lapalace-Stieltjes transform of the waiting time distribution in the primary queue is computed. Several performance measures are derived. Numerical illustration of the system behavior is also performed. Two optimization problems of interest that determine the optimal number of servers to be employed and the optimal capacity of the buffer for interrupted customers, so as to maximize the Expected Total Profit when all other parameters stay put are investigated.

Table 3: Some selected measures

K	θ	μ_{IDS}	μ_{BYS}	μ_{SBYP}	μ_{SBYI}	μ_{PQ}	μ_{BIC}	μ_{BIP}
$c = 2$								
1	2	0.26132	1.73868	1.57895	0.15973	5.25534	0.05008	0.59898
	5	0.60368	1.39632	1.20000	0.19632	1.30872	0.03289	0.73620
	20	1.23131	0.76869	0.54545	0.22323	0.14794	0.00682	0.83713
2	2	0.14376	1.85624	1.57895	0.27729	11.10428	0.10525	1.03985
	5	0.43414	1.56586	1.20000	0.36586	2.39902	0.08353	1.37199
	20	1.02358	0.97642	0.54545	0.43096	0.35211	0.02675	1.61612
3	2	0.06821	1.93179	1.57895	0.35284	25.85523	0.15279	1.32316
	5	0.29584	1.70416	1.20000	0.50416	4.27361	0.14768	1.89060
	20	0.83553	1.16447	0.54545	0.61902	0.67847	0.06551	2.32132
4	2	0.02722	1.97278	1.57895	0.39383	69.09705	0.18462	1.47687
	5	0.18978	1.81022	1.20000	0.61022	7.62419	0.21759	2.28834
	20	0.67038	1.32962	0.54545	0.78416	1.15030	0.12495	2.94061
5	2	0.00907	1.99093	1.57895	0.41199	215.16866	0.20109	1.54495
	5	0.11346	1.88654	1.20000	0.68654	14.06905	0.28438	2.57451
	20	0.52979	1.47021	0.54545	0.92476	1.79329	0.20361	3.46783
$c = 4$								
2	2	2.13756	1.86244	1.57895	0.28349	0.11152	0.00737	1.06310
	5	2.42401	1.57599	1.20000	0.37599	0.04801	0.00443	1.40996
	20	3.01456	0.98544	0.54545	0.43998	0.00491	0.00056	1.64993
3	2	2.06197	1.93803	1.57895	0.35909	0.13390	0.01128	1.34657
	5	2.27775	1.72225	1.20000	0.52225	0.07062	0.00877	1.95844
	20	2.81373	1.18627	0.54545	0.64082	0.01154	0.00182	2.40307
4	2	2.02320	1.97680	1.57895	0.39786	0.14831	0.01406	1.49196
	5	2.16591	1.83409	1.20000	0.63409	0.09452	0.01400	2.37785
	20	2.63037	1.36963	0.54545	0.82417	0.02265	0.00442	3.09065
5	2	2.00722	1.99278	1.57895	0.41383	0.15558	0.01556	1.55186
	5	2.08901	1.91099	1.20000	0.71099	0.11628	0.01918	2.66622
	20	2.46836	1.53164	0.54545	0.98619	0.03871	0.00878	3.69821
6	2	2.00190	1.99810	1.57895	0.41915	0.15850	0.01620	1.57183
	5	2.04244	1.95756	1.20000	0.75756	0.13314	0.02346	2.84086
	20	2.33115	1.66885	0.54545	1.12340	0.05917	0.01494	4.21275

Table 4: Continued

K	θ	μ_{IDS}	μ_{BYS}	μ_{SBYP}	μ_{SBYI}	μ_{PQ}	μ_{BIC}	μ_{BIP}
$c = 6$								
2	2	4.13722	1.86278	1.57895	0.28383	0.00501	0.00040	1.06437
	5	4.42355	1.57645	1.20000	0.37645	0.00161	0.00017	1.41169
	20	5.01441	0.98559	0.54545	0.44014	0.00006	0.00001	1.65052
4	2	4.02301	1.97699	1.57895	0.39804	0.00721	0.00089	1.49264
	5	4.16494	1.83506	1.20000	0.63506	0.00384	0.00069	2.38149
	20	4.62924	1.37076	0.54545	0.82530	0.00045	0.00010	3.09488
5	2	4.00715	1.99285	1.57895	0.41391	0.00768	0.00102	1.55215
	5	4.08809	1.91191	1.20000	0.71191	0.00502	0.00105	2.66965
	20	4.46640	1.53360	0.54545	0.98815	0.00092	0.00024	3.70556
6	2	4.00187	1.99813	1.57895	0.41918	0.00788	0.00108	1.57193
	5	4.04176	1.95824	1.20000	0.75824	0.00600	0.00137	2.84339
	20	4.32836	1.67164	0.54545	1.12619	0.00165	0.00050	4.22321
8	2	4.00008	1.99992	1.57895	0.42097	0.00796	0.00111	1.57864
	5	4.00652	1.99348	1.20000	0.79348	0.00708	0.00178	2.97556
	20	4.13472	1.86528	0.54545	1.31982	0.00372	0.00135	4.94933
$c = 8$								
4	2	6.02300	1.97700	1.57895	0.39805	0.00029	0.00004	1.49269
	5	6.16489	1.83511	1.20000	0.63511	0.00012	0.00002	2.38167
	20	6.62922	1.37078	0.54545	0.82533	0.00001	0.00000	3.09497
5	2	6.00714	1.99286	1.57895	0.41391	0.00031	0.00005	1.55217
	5	6.08805	1.91195	1.20000	0.71195	0.00017	0.00004	2.66983
	20	6.46635	1.53365	0.54545	0.98820	0.00002	0.00000	3.70575
7	2	6.00042	1.99958	1.57895	0.42063	0.00033	0.00005	1.57737
	5	6.01749	1.98251	1.20000	0.78251	0.00026	0.00007	2.93440
	20	6.21754	1.78246	0.54545	1.23700	0.00006	0.00002	4.63876
8	2	6.00008	1.99992	1.57895	0.42097	0.00033	0.00005	1.57864
	5	6.00651	1.99349	1.20000	0.79349	0.00028	0.00008	2.97560
	20	6.13460	1.86540	0.54545	1.31994	0.00010	0.00004	4.94979
10	2	6.00000	2.00000	1.57895	0.42105	0.00033	0.00005	1.57894
	5	6.00065	1.99935	1.20000	0.79935	0.00030	0.00009	2.99757
	20	6.04095	1.95905	0.54545	1.41360	0.00018	0.00008	5.30099

REFERENCES

[1] A. Krishnamoorthy, P. K. Pramod, and S. R. Chakravarthy, Queues with interruption: A Survey. (To appear in *TOP-Spanish journal of Statistics & Operations Research*), (2012).

[2] Varghese 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).

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

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

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

[6] Miaomiao Yu, Yinghui Tang, Yonghong Fu and Lemeng Pan, An $M/E_k/1$ queueing system with no damage service interruption. *Mathematical and Computer Modelling*, 54, 1262–1272, (2011).