**Applied Mathematical Sciences, Vol. 7, 2013, no. 52, 2569 - 2589**
**HIKARI Ltd, www.m-hikari.com**

# Preemptive Resume Priority Retrial Queue with Two Classes of MAP Arrivals

**M. Senthil Kumar[1], S. R. Chakravarthy[2] and R. Arumuganathan[3]**

[1,3]Department of Mathematics and Computer Applications
PSG College of Technology, Coimbatore-641004, India

[2] Department of Industrial and Manufacturing Engineering
Kettering University, Flint MI 48504, USA
schakrav@kettering.edu



**Abstract:** Generally in call centers, voice calls (say Type 1 calls) are given higher priority over e-mails (say Type 2 calls). An arriving Type 1 call has a preemptive priority over a Type 2 call in service, if any, and the preempted Type 2 call enters into a retrial buffer (of finite capacity). Any arriving call not able to get into service immediately will enter into the pool of repeated calls provided the buffer is not full; otherwise, the call is considered lost. The calls in the retrial pool are treated alike (like Type 1) and compete for service after a random amount of time, and can preempt a Type 2 call in service. We assume that the two types of calls arrive according to a Markovian arrival process (MAP) and the services are offered with preemptive priority rule. Under the assumption that the service times are exponentially distributed with possibly different rates, we analyze the model using matrix-analytic methods. Illustrative numerical examples to bring out the qualitative aspects of the model under study are presented.

**Keywords:** Markovian Arrival process, retrials, multi-server, preemptive priority, matrix-analytic methods, algorithmic probability, call center

## 1. INTRODUCTION

Call centers have been playing a vital role for many industries and businesses for more than two decades or so now. Traditionally the customers have been contacting the call centers by talking to a customer service representative (CSR) or an agent over the telephone. Now, in addition to contacting over the phone, the customers can contact the

center over the internet either via e-mail or live chat sessions. A traditional center has different components such as an automatic call distributor (ACD), an interactive voice response (IVR) unit, desktop computers and telephones. The ACD is a telephone switch located at conveniently to properly distribute the customer calls. There is only a finite number of trunks connecting the ACD.

As the calls arrive the ACD routes them either to the IVR unit where the customer transactions are handled automatically or to an idle CSR, who provides the necessary service. If no CSR is available, the calls are placed in a queue. The CSR responds to the routed calls either using the telephone and or the computers. For example, if the agent is answering a telephone call, that agent can access the customer information databases through the computer. The heart of a traditional call center is this dynamic routing of a new or pending call by the ACD to the most appropriate and available CSR. Arriving calls are terminated at the ACD switch and are routed to a group of agents (CSRs). In multimedia call center, these calls can be in the form of voice, e-mail, fax or video. Currently, the analytical models applied in practice, are based on some classical queueing models. Here, we study a call center system as *a multi-server (severs are CSRs) retrial queueing model in which two types of calls (or customers)*, say Type 1(high priority) and Type 2(low priority), arrive according to a Markovian arrival process (*MAP*). From henceforth we will interchangeably use customers and calls. The Type-1 customer class consists of voice calls, while Type-2 customer class consists of e-mails. An arriving Type 1 call has a preemptive priority over a Type 2 call in service, and the preempted Type 2 call will enter into a retrial buffer of finite capacity should there be a space; otherwise the call will be lost. Further, all calls entering into the retrial buffer will be treated as Type 1. That is, once a call enters into the retrial buffer it will not be preempted when it gets back into service. In Figure 1 we display a pictorial description of the model.

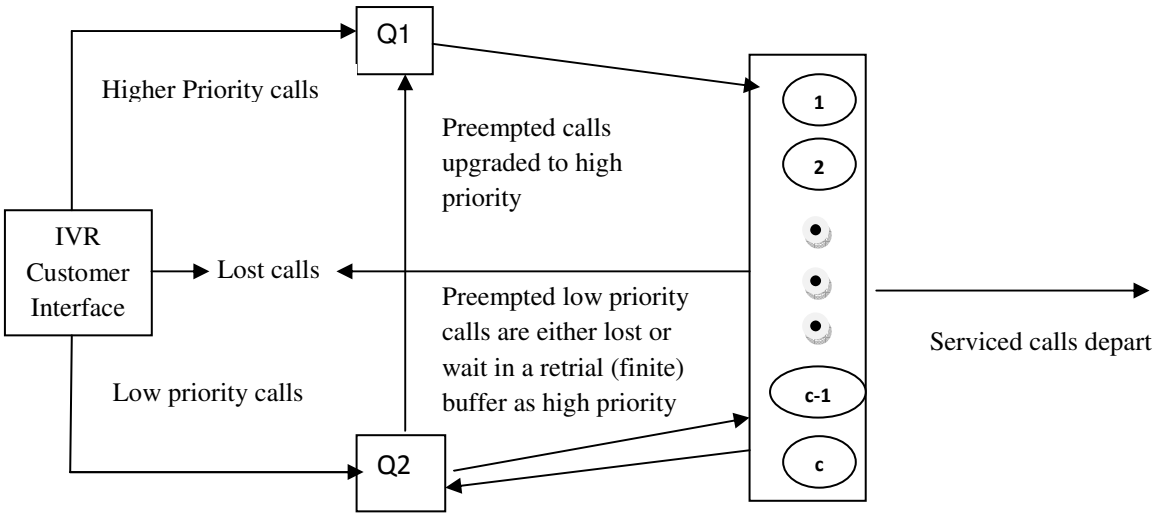



**FIGURE 1: Two class, Preemptive –Resume Priority Queue**

Most of the research on functioning and administration of call centers use queueing theory [22]. Aguir et al. [1] study a call center as a multi-server queueing system where operations such as queueing a client failure, client's impatience, and repeated calls are explicitly modeled. For this queueing system both transient and steady-state analysis are conducted. For the steady-state analysis they used the fluid approximation which facilitates the analysis for the exact mapping of systems of large call centers with heavy traffic [2]. According to Pustova [28] retrial queueing systems are more appropriate and adequate to model call centers.

Retrial queues with two types of customers have been widely used under a variety of scenarios [5, 11-17]. It should be pointed out in Chakravarthy and Dudin [12], Type 1 customers are served in groups of varying sizes (see [9] for the description of the type of group services considered in [12]) and Type 2 customers are served one at a time by one of two servers in the system. Only Type 2 customers enter into a retrial buffer (of infinite size) and the retrial rate is independent of the number in the retrial buffer. Arrivals are modeled using *MAP*.

In the context of two types of customers, Falin [18-20] investigated sufficient conditions for the existence of the stationary distribution of the queue lengths for the $(M_1,M_2)/M/c$ retrial queue. A survey of retrial queues with two types of calls along with some new results is given in Choi and Chang [15]. Choi and Park [13] investigated a retrial queue with two types of calls with no limit on how many such calls can be in the system. Type 2 calls are placed in a retrial buffer should there be no idle server at the time of arrivals. They obtained the joint generating function of queue lengths using supplementary variable method. Kalyanaraman and Srinivasan [21] studied a single server retrial priority queueing system with Type 1 calls, transit Type 2 calls and $K$ recurrent calls. Type 1 calls have a (non-preemptive) priority over the other calls and have their own buffer (of infinite size). An arriving Type 2 call finding the server busy enters into a retrial buffer (of infinites size), and recurrent calls reside in the system and cycle through service and retrial buffer. Assuming general independent services for Type 1 and other non Type 1 calls, the authors derived the joint distribution of the number of calls in the priority queue and in the retrial queue using supplementary variable method.

Choi et al. [17] investigated the impact of retrials on loss probabilities. They compare the loss probabilities of several channel allocation schemes giving a higher priority to hand-off calls in the cellular mobile wireless network in the context of two types of customers who arrive according to a *MAP*. Wang [29] discussed the $(M_1,M_2)/G_1,G_2/1$ retrial queues with priority subscribers and the server subject to breakdowns and repairs. In all the above models Type 1 calls have a non-preemptive priority over Type 2 calls. But, Artalejo et al. [5] investigated in steady-state a single server retrial queue where customers in the retrial group have preemptive priority over customers waiting in the queue. Recently, Liu and Wu [22] considered *MAP*/*G*/1 *G*-queues with preemptive resume and multiple vacations which gave the importance of preeemptive resume in practical situations.

In [1, 28], call centers are modeled as retrial systems, where the impact of preemptive priority of the customers is not considered. Recently, Pustova [28] studied the effect of retrials in call centers. This study, which takes retrials into account, does not address the effect of the preemption on the low priority customers. Thus, in this paper, we qualitatively study a multi-server retrial queue with two types of customers arriving according to a versatile point process and with Type 1 customers having a preemptive priority over the other type, by looking at the impact of the preemption and the effect of the correlation in the inter-arrival times.

This paper is organized as follows: In Section 2, the mathematical description of the model is presented. The steady state analysis of the model is presented in Section 3 and some selected system performance measures to bring out the qualitative nature of the model under study are given in Section 4. In Section 5 some interesting numerical examples are presented. An optimization problem is discussed in Section 6 and concluding remarks are given in Section 7.

## 2. THE MATHEMATICAL MODEL

We consider a multi-server retrial queueing system in which two types of calls (henceforth, referred to as Type 1 and Type 2 calls) arrive according to a *MAP* with representation ($D_0$, $D_1$, $D_2$) of order $m$. A brief description of *MAP* including the meaning of these parameter matrices are given below. The service facility consists of $c$ identical exponential servers (CSRs). The service times of Type 1 and Type 2 customers are assumed to be exponentially distributed with parameters, respectively, given by $\mu_1$ and $\mu_2$. When all servers are busy with Type 1 customers, an arriving customer (irrespective of the type) enter into the retrial orbit of finite capacity of size, $K$, provided there is a space in the retrial buffer; otherwise the arrival is considered lost. An arriving Type 1 customer finding all servers busy with at least one Type 2 customer in service preempts one of the Type 2 customers and enters into service immediately. The preempted Type 2 customer will join the retrial buffer should there be a space; otherwise this customer is considered lost even though this customer received a partial service. Any arriving Type 2 customer finding all servers busy joins the retrial orbit if there is a space; otherwise this customer is lost. All customers from retrial orbit are treated as Type 1 customers and compete for service at random intervals of time. The retrial times have an exponential distribution with parameter $\theta > 0$. The retrial customers can preempt Type 2 customers and only at the time of retrials.

The *MAP*, a special class of tractable Markov renewal process, is a rich class of point processes that includes many well-known processes such as Poisson, PH-renewal, and Markov – modulated Poisson process. One of the most significant features of the *MAP* is the underlying Markovian structure that fits ideally in the context of matrix-analytic were first introduced and studied by Neuts [26] as versatile point process. As is well known, Poisson processes are the simplest and most tractable ones used extensively in stochastic modeling. The idea of the *MAP* is to significantly generalize the Poisson

processes and still keep tractability for modeling purposes. Furthermore, in many practical applications, notably, in communications engineering, production and manufacturing engineering, the arrivals do not usually form a renewal process. So, *MAP* is a convenient tool to model both renewal and nonrenewal arrivals. While *MAP* is defined for both discrete and continuous times, here we will define only the single arrival case (with two types of customers) and in continuous time.

The *MAP*, a special case of batch Markovian arrival process (*BMAP*), in continuous time is described as follows. Let the underlying Markov chain be irreducible and let $Q^*$ be the generator of this Markov chain. At the end of a sojourn time in state $i$, that is exponentially distributed with parameter $\xi_i$, one of the following two events could occur: with probability $p_{ij}(k), k = 1,2$, the transition corresponds to an arrival of a Type $k$ customer, and the underlying Markov chain is in state $j$ with $1 \leq i, j \leq m$; with probability $p_{ij}(0)$ the transition corresponds to no arrival and the state of the Markov chain is $j$, $j \neq i$. Note that the Markov chain can go from state $i$ to state $i$ only through an arrival. For $0 \leq k \leq 2$, define matrices $D_k = (d_{ij}(k))$ such that $d_{ii}(0) = -\xi_i, 1 \leq i, j \leq m$, $d_{ij}(0) = \xi_i p_{ij}(0)$, $j \neq i, 1 \leq i, j \leq m$, and $d_{ij}(k) = \xi_i p_{ij}(k), 1 \leq i, j \leq m$, $k = 1,2$. Assuming $D_0$ to be a nonsingular matrix guarantees the inter-arrival times will be finite with probability one and hence the arrival process does not terminate. Thus, $D_0$ is a stable matrix. The generator $Q^*$ is given by $Q^* = \sum_k D_k$. Thus, the *MAP* is described by the matrices $\{D_k\}$ with $D_0$ governing the transitions corresponds to *no* arrivals and $D_k$ governing those corresponding to arrivals of type $k$ customers, $1 \leq k \leq 2$. Thus, the representation of this *MAP* is denoted by ($D_0$, $D_1$, $D_2$) of order $m$. Note that here we assume that Type 1 and Type 2 arrivals can be correlated apart from the facts that the inter-arrival times of Type $i$, $i$ = 1, 2, themselves are correlated. However, it is easy to modify this assumption to include the case where these two types of arrivals are not correlated even though within each type the inter-arrival times may be correlated. The details are omitted.

For use in sequel, let $\boldsymbol{e}(r)$, $e_j(r)$ and $I_r$, denote, respectively, the (column) vector of dimension $r$ consisting of 1's, column vector of dimension $r$ with 1 in the $j^{th}$ position and 0 elsewhere, and an identity matrix of dimension $r$. The notation $\otimes$ will stand for the Kronecker product of two matrices. Thus, if $A$ is a matrix of order $m \times$ n and if $B$ is a matrix of order $p \times$ q, then $A \otimes B$ will denote a matrix of order $mp \times nq$ whose $(i, j)^{\text{th}}$ block matrix is given by $a_{ij}B$. The notation "′" stands for the transpose of a vector or a matrix.

Let $\boldsymbol{\eta}$ be the stationary probability vector of the Markov process with generator $Q^*$. That is, $\boldsymbol{\eta}$ is the unique (positive) probability vector satisfying $\boldsymbol{\eta}\, Q^* = 0$ and $\boldsymbol{\eta e} = 1$. Let $\boldsymbol{\delta}$ be the initial probability vector of the underlying Markov chain governing *MAP*; this vector can be chosen in a number of ways, but the most interesting case is the one where we get the stationary version of *MAP* by setting $\boldsymbol{\delta} = \boldsymbol{\eta}$. The constant $\lambda = \boldsymbol{\eta}(D_1 + D_2)\boldsymbol{e}$, referred to as the fundamental rate, gives the expected number of arrivals per unit of time in the stationary version of the *MAP*. The quantity $\lambda_i = \boldsymbol{\eta} D_i \boldsymbol{e}$ gives the arrival rate of $i$-customers, for $i$ =1, 2. Note that $\lambda = \lambda_1 + \lambda_2$. For further details on *MAP* and their

usefulness in stochastic modeling, we refer to [23, 25, 26] and for a review and recent work on *MAP*, we refer to [8, 10-12].

## 3. THE STEADY STATE ANALYSIS OF THE MODEL AT AN ARBITRARY EPOCH

Let $N(t)$, $I_1(t)$, $I_2(t)$, and $J(t)$ denote, respectively, the number of customers in the retrial buffer, the number of Type 1 customers in service, number of Type 2 customers in service, and phase of the arrival process at time $t$. Then $\{(N(t), I_1(t), I_2(t), J(t)): t \geq 0\}$ is a continuous time Markov chain whose state space given by

$$\Omega = \{(i, j_1, j_2, l): 0 \leq i \leq K, 0 \leq j_1, j_2 \leq c, 0 \leq j_1 + j_2 \leq c, 1 \leq l \leq m\}.$$

Let the elements of $\Omega$ be ordered lexicographically. Then the infinitesimal generator $Q$ of Markov process $\{N(t), I_1(t), I_2(t), J(t)\}$ is of finite QBD type and is given by

$$Q = \begin{bmatrix} A_0 & B_0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ E_1 & A_1 & B_0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 2E_1 & A_2 & B_0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (K-1)E_1 & A_{K-1} & B_0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & KE_1 & A_K \end{bmatrix},$$

where all the blocks are square matrices of order $(c+2)(c+1)m/2$. The block matrices appearing in $Q$ are given as follows.

$$A_i = \begin{bmatrix} A_{0,0}^{(i)} & A_{0,1}^{(i)} & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_{1,0} & A_{1,1}^{(i)} & A_{1,2}^{(i)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & A_{2,1} & A_{2,2}^{(i)} & A_{2,3}^{(i)} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & A_{c-1,c-2} & A_{c-1,c-1}^{(i)} & A_{c-1,c}^{(i)} \\ 0 & 0 & 0 & 0 & \cdots & 0 & A_{c,c-1} & D_0 - c\mu_1 I \end{bmatrix}, \ 0 \leq i \leq K-1,$$

$$A_K = \begin{bmatrix} A_{0,0}^{(K)} & A_{0,1}^{(K)} & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_{1,0} & A_{1,1}^{(K)} & A_{1,2}^{(K)} & 0 & \cdots & 0 & 0 & 0 \\ 0 & A_{2,1} & A_{2,2}^{(K)} & A_{2,3}^{(K)} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & A_{c-1,c-2} & A_{c-1,c-1}^{(K)} & A_{c-1,c}^{(K)} \\ 0 & 0 & 0 & 0 & \cdots & 0 & A_{c,c-1} & Q^* - c\mu_1 I \end{bmatrix},$$

with

$$A_{j,j}^{(i)} = \begin{bmatrix} D_0 & D_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu_2 I & D_0 - \mu_2 I & D_2 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (c-j-1)\mu_2 I & D_0 - (c-j-1)\mu_2 I & D_2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & (c-j)\mu_2 I & D_0 - (c-j)\mu_2 I + i\theta I \end{bmatrix} - (j\mu_1 + i\theta)I, \quad 0 \le i \le K-1, 0 \le j \le c-1,$$

$$A_{j,j}^{(K)} = \begin{bmatrix} D_0 & D_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu_2 I & D_0 - \mu_2 I & D_2 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & (c-j-1)\mu_2 I & D_0 - (c-j-1)\mu_2 I & D_2 \\ 0 & 0 & 0 & 0 & \cdots & 0 & (c-j)\mu_2 I & Q^* - (c-j)\mu_2 I + K\theta I \end{bmatrix} - (j\mu_1 + K\theta)I, \quad 0 \le j \le c-1,$$

$$A_{j,j-1} = [I_{c-j+1} \quad \mathbf{0}_{c-j+1}] \otimes j\mu_1 I, 1 \le j \le c,$$

$$A_{j,j+1}^{(i)} = \begin{bmatrix} I_{c-j} \otimes D_1 \\ \boldsymbol{e}'_{c-j}(c-j) \otimes i\theta I \end{bmatrix}, 0 \le i \le K, 0 \le j \le c-1,$$

$$B_0 = \begin{bmatrix} \hat{B}_0 & \hat{B}_{0,1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \hat{B}_1 & \hat{B}_{1,2} & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \hat{B}_{c-1} & \hat{B}_{c-1,c} \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & D_1 + D_2 \end{bmatrix},$$

$\hat{B}_j = \begin{bmatrix} 0 \\ \boldsymbol{e}'_{c-j+1}(c-j+1) \otimes D_2 \end{bmatrix}, 0 \le j \le c-1, \quad \hat{B}_{j-1,j} = \begin{bmatrix} 0 \\ \boldsymbol{e}'_{c-j}(c-j) \otimes D_1 \end{bmatrix}, 1 \le j \le c,$

$$E_1 = \begin{bmatrix} 0 & \hat{E}_0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \hat{E}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \hat{E}_{c-1} \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad \hat{E}_j = \begin{bmatrix} I_{c-j} \otimes \theta I \\ 0 \end{bmatrix}, 0 \le j \le c-1.$$

Let $\boldsymbol{x}$ =($\boldsymbol{x}$(0), $\boldsymbol{x}$(1), $\boldsymbol{x}$(2),…. $\boldsymbol{x}$($K$)), denote the steady state probability vector of $Q$. That is, $\boldsymbol{x}$ satisfies $\boldsymbol{x}Q = \mathbf{0}$, $\boldsymbol{x}\boldsymbol{e}$ =1, and is obtained by solving the following set of equations.

$$\begin{gathered} \boldsymbol{x}(0)A_0 + \boldsymbol{x}(1)E_1 = \mathbf{0}, \\ \boldsymbol{x}(i-1)B_0 + \boldsymbol{x}(i)A_1 + (i+1)\boldsymbol{x}(i+1)E_1 = \mathbf{0}, 1 \le i \le K-1, \\ \boldsymbol{x}(K-1)B_1 + \boldsymbol{x}(K)A_K = \mathbf{0}, \end{gathered} \tag{1}$$

subject to the normalizing condition $\sum_{i=0}^{K} \boldsymbol{x}(i)\,\boldsymbol{e} = 1$. The equations in (1) are ideally suited for numerical implementation (after exploiting the special structure of the coefficient matrices with suitably partitioning the vectors $\boldsymbol{x}(i), 0 \leq i \leq K,$ into vectors of smaller dimensions) by any of the well-known methods such as (block) Gauss-Seidel. The details are standard and are omitted.

## 4. SYSTEM PERFORMANCE MEASURES

In this section, we list some key system performance measures useful to bring out the qualitative nature of the model under study.

1. The probability mass function of the number of busy servers with Type 1 customers is
$$P_j^{(1)} = \sum_{i=0}^{K} \sum_{k=0}^{c-j} \boldsymbol{x}_{jk}(i)\,\boldsymbol{e}, \qquad 0 \leq j \leq c.$$
2. The probability mass function of the number of busy servers with Type 2 customers is
$$P_k^{(2)} = \sum_{i=0}^{K} \sum_{j=0}^{c-k} \boldsymbol{x}_{jk}(i)\,\boldsymbol{e}, \qquad 0 \leq k \leq c.$$
3. The probability mass function of number of customers in the retrial orbit is
$$P_i^{(3)} = \sum_{j=0}^{c} \sum_{k=0}^{c-j} \boldsymbol{x}_{jk}(i)\,\boldsymbol{e}, \qquad 0 \leq i \leq K.$$
4. The probability, $P_{block}^{(1)}$, that Type 1 customers are blocked (due to all servers busy with Type 1 customers with at least one space available in the retrial orbit buffer) and the probability, $P_{block}^{(2)}$, that Type 2 customers are blocked (due to all servers busy with either Type 1 or Type 2 customers with at least one space available in the buffer) at an arrival epoch are obtained as
$$P_{block}^{(1)} = \frac{1}{\lambda_1\left[1-P_{loss}^{(1)}\right]} \sum_{i=0}^{K-1} \boldsymbol{x}_{c0}(i) D_1\,\boldsymbol{e},$$
$$P_{block}^{(2)} = \frac{1}{\lambda_2\left[1-P_{loss}^{(2)}\right]} \sum_{i=0}^{K-1} \sum_{j=0}^{c} \boldsymbol{x}_{j,c-j}(i)\, D_2\,\boldsymbol{e}.$$
5. The probability, $P_{loss}^{(i)}, i = 1,2,$ that an arriving Type *i* customer will be lost due to all servers being busy and the retrial orbit size is filled to capacity is obtained as
$$P_{loss}^{(1)} = \frac{1}{\lambda_1} \boldsymbol{x}_{c0}(K) D_1 \boldsymbol{e}, \quad P_{loss}^{(2)} = \frac{1}{\lambda_2} \sum_{j=0}^{c} \boldsymbol{x}_{j,c-j}(K)\, D_2 \boldsymbol{e}.$$
6. The *throughput*, defined as the rate at which the customers depart the system, is given by

$$\gamma = \lambda_1\left[1 - P_{loss}^{(1)}\right] + \lambda_2\left[1 - P_{loss}^{(2)}\right]$$

7. The mean number of customers in the retrial orbit is given by $\mu_{NO} = \sum_{i=0}^{K} i\, P_i^{(3)}$.
8. The rate of preemption of Type 2 customers by (new) Type 1 arrivals, $R_{Preempt}^{(New)}$, and by retrial customers, $R_{Preempt}^{(RT)}$, are given by

$$R_{Preempt}^{(New)} = \frac{1}{\lambda_1}\sum_{i=0}^{K}\sum_{j=0}^{c-1}\sum_{k=1}^{c-j}\boldsymbol{x}_{jk}(i)D_1\boldsymbol{e},$$

$$R_{Preempt}^{(RT)} = \boldsymbol{\theta}\sum_{i=1}^{K} i\sum_{j=0}^{c-1}\sum_{k=1}^{c-j}\boldsymbol{x}_{jk}(i)\boldsymbol{e}.$$

## 5. NUMERICAL RESULTS

In this section, we discuss some interesting numerical examples that qualitatively describe the model under study. For the arrival process, we consider the following five sets of values for $D_0$, $D_1$, and $D_2$. For the arrival process of two types of customers we look at a special class of *MAP* by taking $D_1 = p\ D$, and $D_2 = (1\text{-}p)\ D$, where $0 < p < 1$. This corresponds to the case where arrival of both Type 1 customers and Type 2 customers are correlated among themselves. Thus, for this special *MAP*, we need to specify the matrices $D_0$ and $D$ of order $m$, and the probability $p$. The specific forms of $D_0$ and $D$ are as given below:

1. Erlang (*ERL*): $D_0 = \begin{bmatrix} -4 & 4 \\ 0 & -4 \end{bmatrix},\ D = \begin{bmatrix} 0 & 0 \\ 4 & 0 \end{bmatrix}$.

2. Exponential(*EXP*): $D_0 = -2,\ D = 2$.

3. Hyper-Exponential(*HEX*): $D_0 = \begin{bmatrix} -3.8 & 0 \\ 0 & -0.38 \end{bmatrix},\ D = \begin{bmatrix} 3.420 & 0.380 \\ 0.342 & 0.038 \end{bmatrix}$.

4. *MAP* with Negative Correlation(*MNC*):

$$D_0 = \begin{bmatrix} -2.00442 & 2.00442 & 0 \\ 0 & -2.00442 & 0 \\ 0 & 0 & -451.5 \end{bmatrix},\ D = \begin{bmatrix} 0 & 0 & 0 \\ 0.02004 & 0 & 1.98438 \\ 446.995 & 0 & 4.505 \end{bmatrix}.$$

5. *MAP* with Positive Correlation(*MPC*):

$$D_0 = \begin{bmatrix} -2.00442 & 2.00442 & 0 \\ 0 & -2.00442 & 0 \\ 0 & 0 & -451.5 \end{bmatrix}, D = \begin{bmatrix} 0 & 0 & 0 \\ 1.98438 & 0 & 0.02004 \\ 4.505 & 0 & 446.995 \end{bmatrix}.$$

All these five *MAP* processes are normalized so as to have specific arrival rates $\lambda_1 = 1$ and $\lambda_2 = 1$. Note that the total arrival rate is given by $\lambda = 2$. However, these *MAP*s are qualitatively different in that they have different variance and correlation structure.

Looking only at points at arrivals of customers (irrespective of whether they are Type 1 customers or Type 2 customers), the first three arrival processes correspond to renewal processes and so the correlation is 0. The arrival process labeled *MNC* has correlated arrivals with a correlation value of –0.48891, and the arrivals corresponding to the process labeled *MPC* has a positive correlation value of 0.48891.

***EXAMPLE* 1:** In this example we examine how some system performance measures behave as functions of $c$ and $K$ for the above mentioned five arrival processes. All other parameters are fixed as: $\lambda_1 = 1.0,\ \lambda_2 = 1.0,\ \mu_1 = 1.0, \mu_2 = 1.0,$ and $\theta = 2.$ In Figures 2 through 4, we display the measures: $\mu_{NO}$, $P_{block}^{(k)}$, and $P_{block}^{(k)}$ , for $k$ =1,2. An examination of these figures reveals the following.

- As is to be expected, the measure $\mu_{NO}$ increases as $K$ is increased. Similarly, an increase in $c$ decreases the average number of customers in orbit. This appears to be true for all arrival processes.

- By looking at the three renewal arrivals (namely, *ERL*, *EXP*, and *HEX*) we notice that $\mu_{NO}$ appears to decrease with increasing variability when $c$ = 2 and then for other values of $c$ this measure appears to increase with increasing variability. This phenomenon seems to be the case for all values of $K$.

- Comparing *MNC* and *MPC* arrivals (recall that these have, respectively, negative and positive correlations), there seems to be a cut-off value for $K$, say $K^*$ (which depends on $c$) such that for $K < K^*$, *MNC* has a larger $\mu_{NO}$ compared to *MPC*, and for $K \geq K^*$, *MPC* has larger $\mu_{NO}$. However, for large $c$, the difference in the values for $\mu_{NO}$ for *MPC* and *MNC* arrivals appear to decrease. This value of $K^*$ appears to decrease as $c$ increases.

- It is interesting to see that the values of $\mu_{NO}$ are higher for *HEX* compared to *MPC* for all $c$ considered here, even though *HEX* has a larger variance. This indicates the role of correlation that has been largely ignored when modeling real life applications.

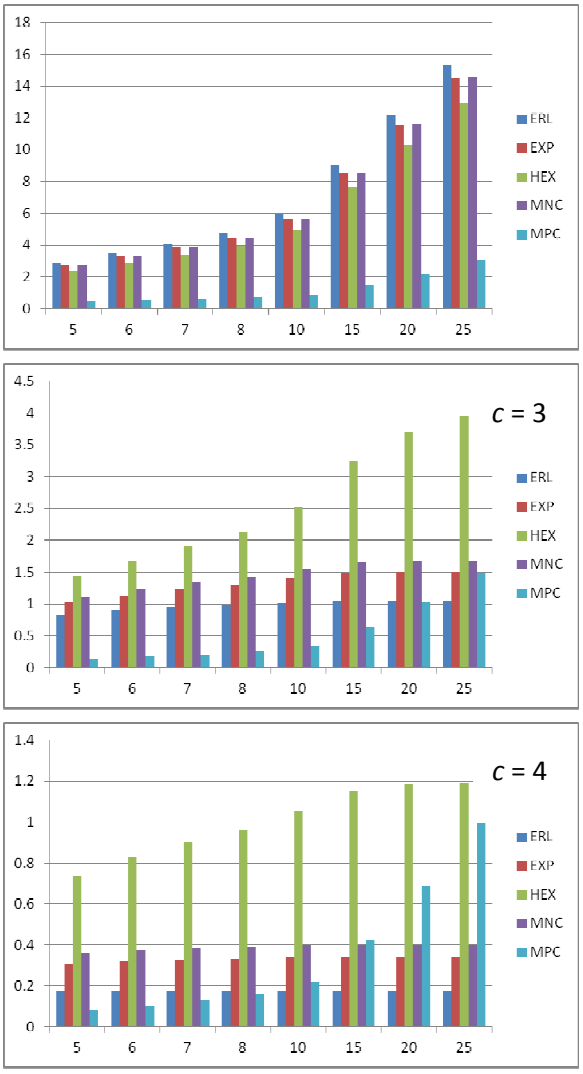
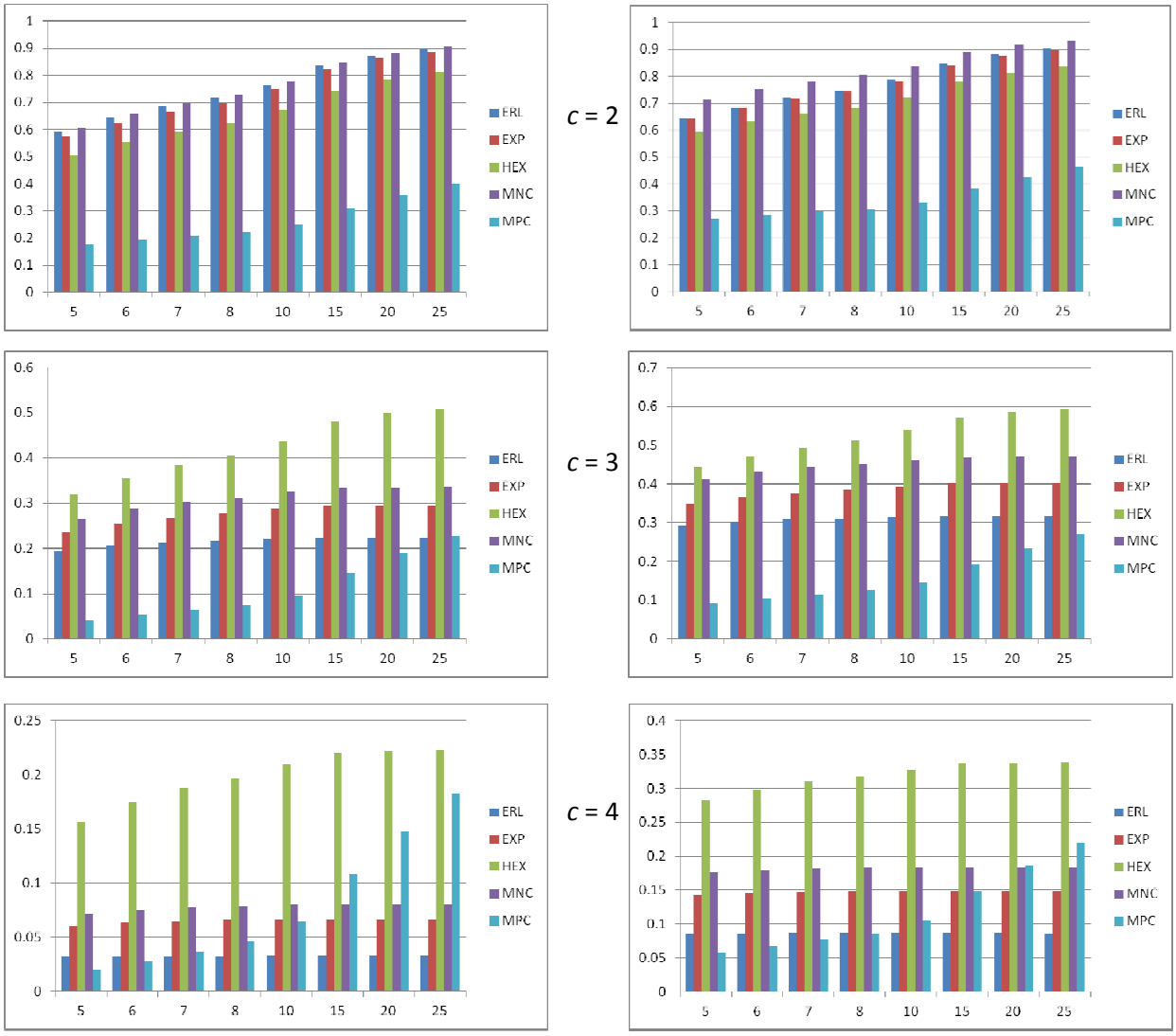



**FIGURE 2: Mean number of customers in orbit**

- As is to be expected, the measures, $P_{block}^{(1)}$ and $P_{block}^{(2)}$, appear to increase as $K$ is increased. This is due to the fact that as $K$ is increased, an arriving customer will more likely to get into the retrial buffer (rather than getting lost) resulting in this phenomenon. Similarly, an increase in $c$ results in a decrease in these two measures. In all cases, we notice that $P_{block}^{(2)} \geq P_{block}^{(1)}$. For some combinations $P_{block}^{(2)}$ exceeds $P_{block}^{(1)}$ by more than 250%.
- With regards to the two measures, $P_{loss}^{(1)}$ and $P_{loss}^{(2)}$, we notice a decreasing trend as $K$ is increased. This is due to the fact that when $K$ is increased, an arriving customer will




**P(Type 1 is blocked)** **P(Type 2 is blocked)**

**FIGURE 3: Blocking probabilities**

more likely to get into the retrial buffer (rather than get lost) resulting in this phenomenon.

- As the number of server increases the loss probabilities of both types of customers are reduced. This is again as expected. In all cases, we notice that $P_{loss}^{(2)} \geq P_{loss}^{(1)}$.

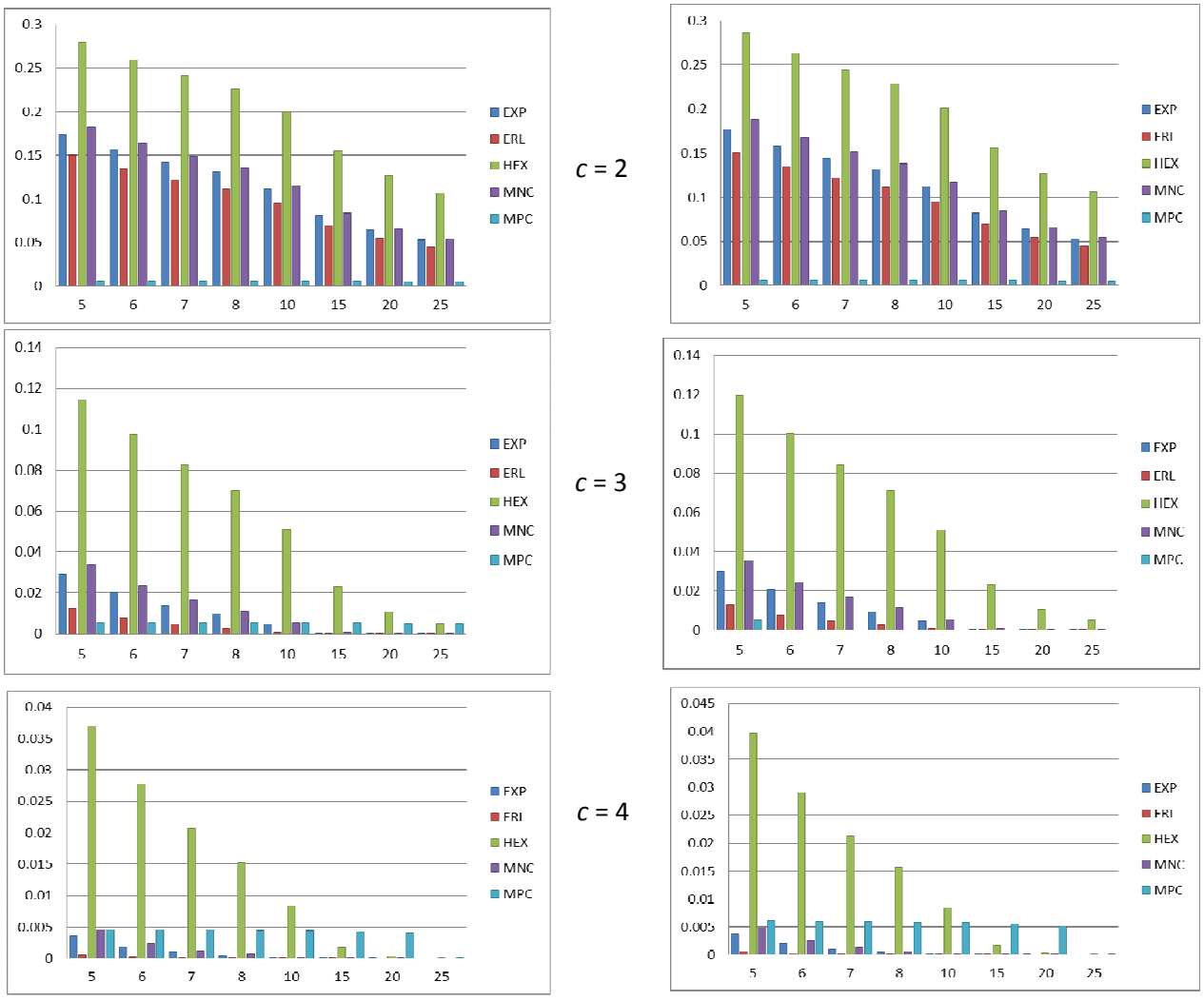



**FIGURE4: Loss probabilities**

Now we will see how correlation (both negative and positive values) plays a role with respect to the three measures: $\mu_{NO}$, $P_{block}^{(k)}$, and $P_{block}^{(k)}, k = 1, 2$.

- Comparing *MNC* and *MPC* processes, there seems to be a cut-off value for *K,* say *K*,* (which depends on *c*) such that for $K < K^*$, *MNC* has larger blocking and loss probabilities for both types of customers as compared to *MPC;* and for $K \geq K^*$, the roles of *MNC* and *MPC* are reversed. Also, *K** appears to decrease as *c* increases.

- Comparing *MNC* and *MPC* processes, there seems to be a cut-off value for *K,* say *K*,* (which depends on *c*) such that for $K < K^*$, *MNC* has a larger $\mu_{NO}$ compared to *MPC* and for $K \geq K^*$, *MPC* has larger $\mu_{NO}$. For large *c*, the difference in $\mu_{NO}$for *MPC* and *MNC* arrivals appear to decrease. Further, *K** appears to decrease as *c* increases.

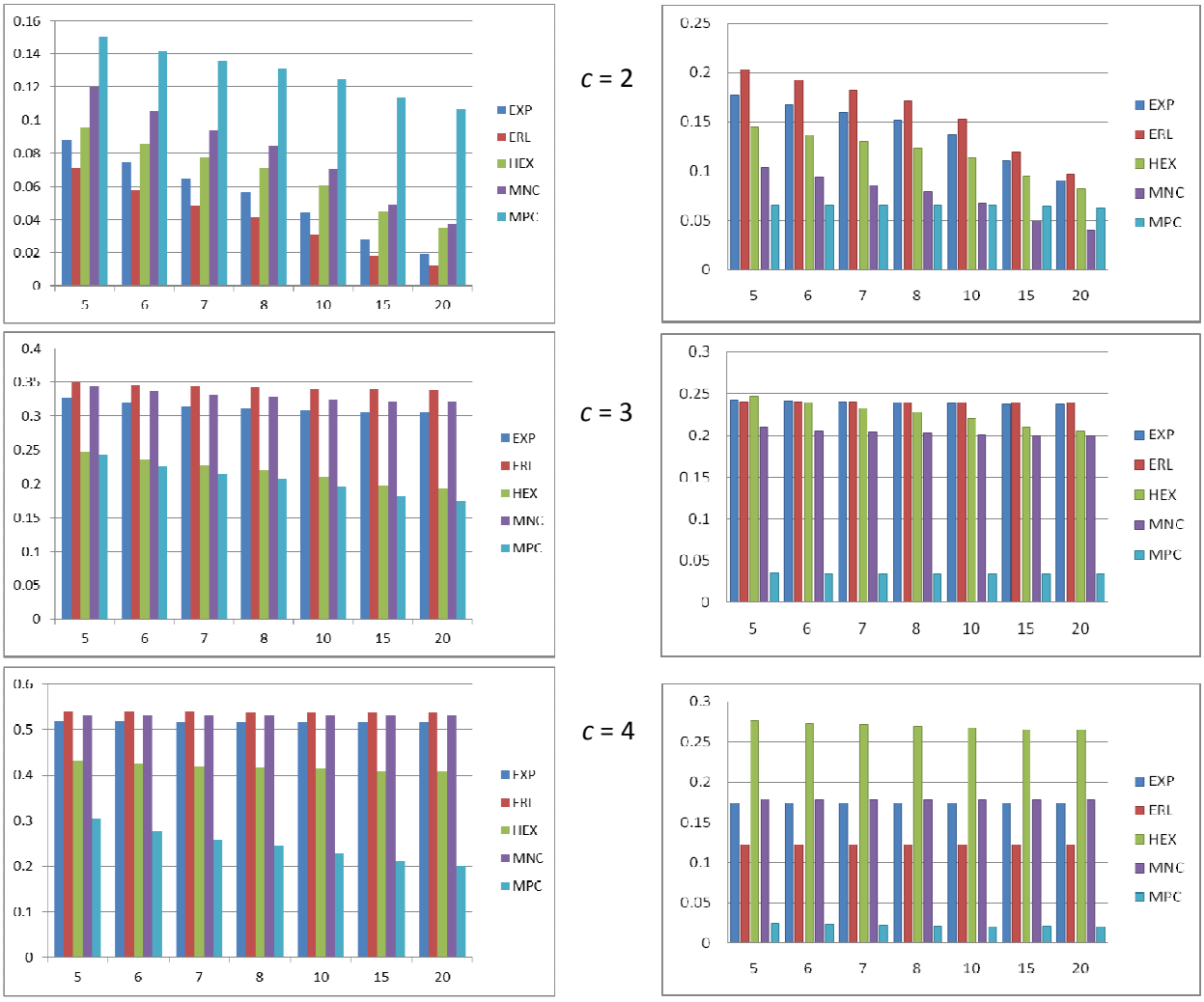
The impact of preemptive priority is now examined. In Figure 5, we display the preemptive probabilities under different scenarios and from this figure we notice the following observations.

- It appears that both $R_{Preempt}^{(New)}$ and $R_{Preempt}^{(RT)}$ are non-increasing functions of *K* when all other parameters are fixed. This is counter-intuitive as one would expect to see more customers to be entering the retrial buffer instead of getting lost from the system. This should result in at least $R_{Preempt}^{(RT)}$to be non-decreasing in *K*.

- As the number of servers *c* increases, $R_{Preempt}^{(New)}$ appears to increase. This is due to the fact that having more servers in the system leads to an increase in accommodating more Type 2 customers, which in turn leads to more preemption.

- Comparing *MNC* and *MPC* processes, there seems to a cut-off value of *K* say, *K** depending on number of servers *c* such that $K < K^*$, *MNC* has larger $R_{Preempt}^{(RT)}$ compared to *MPC* and $K > K^*$, *MPC* has larger $R_{Preempt}^{(RT)}$.

- Comparing *MNC* and *MPC* processes, there seems to a cut-off value of *K* say, *K** depending on number of servers *c* such that $K < K^*$, *MPC* has larger $R_{Preempt}^{(New)}$ compared to *MNC* and $K > K^*$, *MNC* has larger $R_{Preempt}^{(New)}$.




Rate of preemption by new Type 1 customers)

Rate of preemption by retrial customers)

**FIGURE 5: Preemption rates**

***EXAMPLE 2:*** Suppose that Type 1 (which preempts Type 2) customers arrive not as frequently as Type 2 customers. Fixing $c = 2,\ K = 5, \theta = 2,\ \lambda = 2\ ,\ \ \lambda_1 = \lambda p$ and $\lambda_2 = \lambda(1-p), 0 \leq p \leq 1,$ we study the impact of varying $p$ (from 0.1 to 0.9) on some key measures. In Figure 5 below, we display some key measures as functions of $\lambda_1$ and for the five different MAPs. Before we discuss this figure, it should be pointed out that the three measures: $\mu_{NO}$, $P_{loss}^{(2)}$, and $P_{block}^{(2)}$ remain insensitive to $p$. This is intuitively obvious since the total arrival rate is fixed to be 2, an increase in Type 1 arrivals (which results in a decrease in Type 2 arrivals) will not affect any of these three measures. Now we record the following observations based on Figure 6.

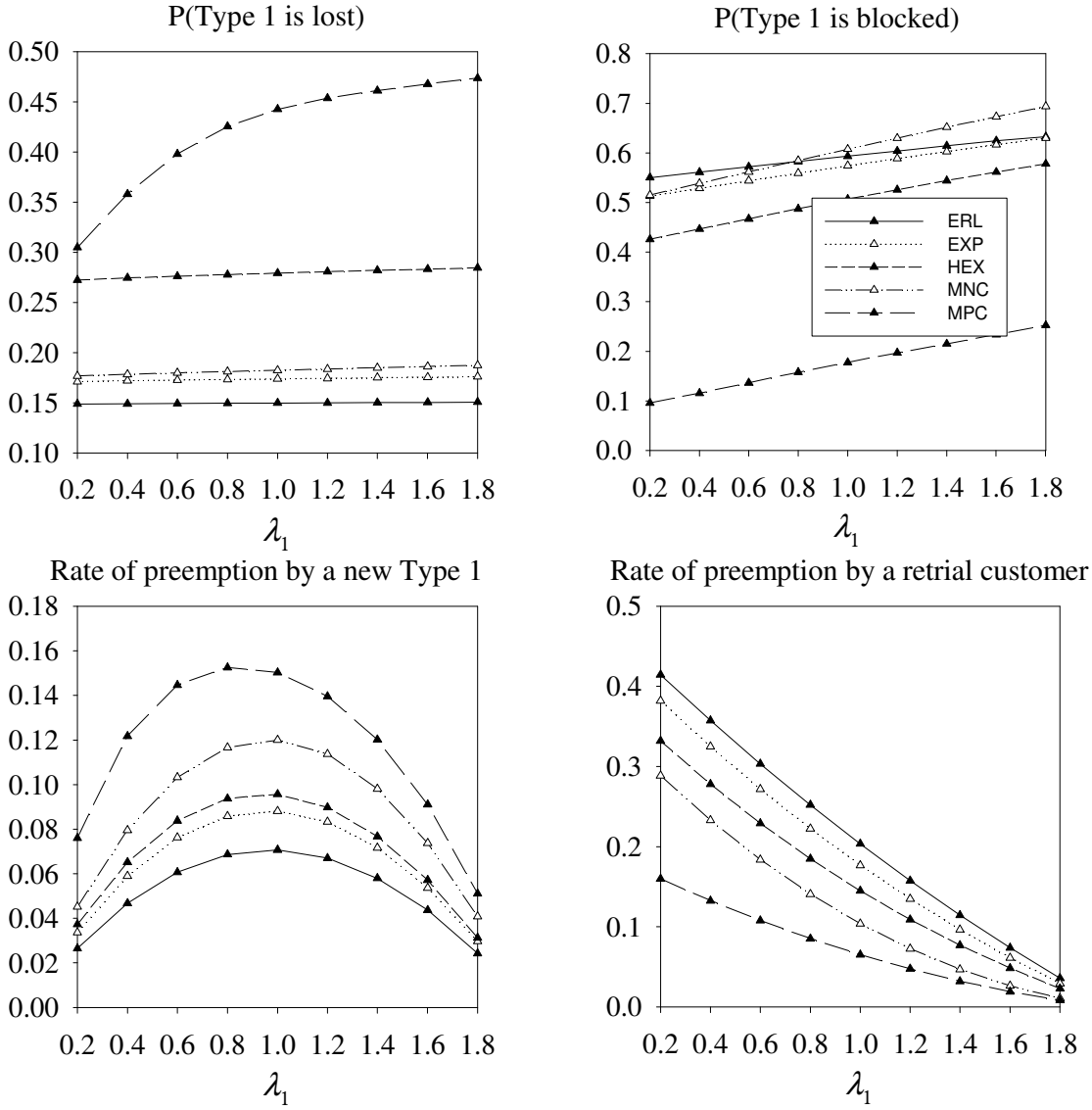



**FIGURE 6: Key measures as functions of $\lambda_1$ and various MAPs**

Only in the case of *MPC* arrivals, we notice a significant change in $P_{loss}^{(1)}$ as $\lambda_1$ is increased.

- As $\lambda_1$ is increased we see an increasing trend in $P_{block}^{(1)}$ for all five *MAPs*. However, this measure is smallest for the *MPC* arrivals for all $\lambda_1$.
- With regards to $R_{Preempt}^{(New)}$, we see an interesting behavior for all five *MAP*s. Initially, this measure appears to increase and then decreases as $\lambda_1$ increases. This is probably due to the fact that when $\lambda_1$ is small the system will have more Type 2 in service resulting in a higher preemption rate; however, for large $\lambda_1$ we will see more Type 1 in service leading to a smaller rate of preemption. We also notice that *MPC* arrivals have the highest value and *ERL* arrivals have the smallest value.
- As $\lambda_1$ is increased, we see a decreasing trend in $R_{Preempt}^{(RT)}$ for all five *MAP*s. Further, *MPC* arrivals have the smallest value while *ERL* arrivals have the largest value.

## 6. AN OPTIMIZATION PROBLEM

In this section we will discuss an optimization problem of interest for the model under study. First we define a number of costs associated with the system. Let $c_w$, $c_o$, $c_{p1}$, and $c_{p2}$ denote, respectively, the holding cost per customer per unit of time, the operating cost of each server per unit of time, the cost per unit of time of preemption by new Type 1 arrivals, and the cost per unit of time of preemption by retrial customers. It is easy to verify that the total expected cost, $TC(\theta)$, per unit of time is calculated as

$$TC(c,\theta) = c_w\mu_{NO} + c_o c + c_{p1}\, R_{Preempt}^{(New)} + c_{p2}\, R_{Preempt}^{(RT)}$$

It is obvious that the study of the total expected cost analytically is difficult and hence one has to resort to its study numerically. Here we report our experimentation to find the local optimal values by considering a small set of decision variables. In Table 1, the total expected cost is played for various scenarios (with bold faced ones indicating the optimum values) involving the type of arrivals and the number of servers. We use the above mentioned five arrival processes and we fix$c_w = 1$, $c_o = 0.4$, $c_{p1} = 2$, and $c_{p2} = 3, \theta = 2$, and $K = 10$.

**Table 1: Optimal Total expected cost *TC*($c$, θ) by varying $c$ and *MAP***

| $c$ | *ERL* | *EXP* | *HEX* | *MNC* | *MPC* |
|---|---|---|---|---|---|
| 2 | 7.0352 | 6.9201 | 6.4462 | 6.4325 | 7.7652 |
| 3 | 3.4229 | 3.9287 | 5.2006 | 4.7151 | 4.4669 |
| 4 | **3.1427** | **3.492** | 4.7351 | 4.6044 | 3.6941 |
| 5 | 3.3117 | 3.5369 | **4.5658** | **4.5898** | **3.6284** |
| 6 | 3.6353 | 3.7616 | 4.4799 | 4.5909 | 3.8168 |
| 7 | 4.0172 | 4.093 | 4.5336 | 4.6489 | 4.1361 |
| 8 | 4.414 | 4.4715 | 4.7419 | 4.8108 | 4.5107 |

From the table 1, it can be noted that in the case of renewal arrivals, the larger the variability in the inter-arrival times, the higher the number of servers needed to arrive at an optimum. With regard to correlated arrivals, both positive and negative ones appear to have the same optimum number of servers but with different optimum total expected costs.

For the next example, we once again consider the above mentioned five arrival processes and take $c_w = 1, c_o = 0.4,\ c_{p1} = 2,$ and $c_{p2} = 3, c = 4,$ and $K = 5$, and see the effect of the retrial rate, $\theta$, on the optimum cost. In Table 2 we display the total expected cost for various scenarios (with bold faced ones indicating the optimum values). Unlike what we observed in Table 1, here we notice that the ERL arrivals require the highest retrial rate to arrive at an optimum total expected cost. However, the optimum value is still the smallest among all arrivals.

**Table 2: Optimal Total expected cost *TC*($c$, θ) by varying θ and *MAP***

| $\theta$ | *ERL* | *EXP* | *HEX* | *MNC* | *MPC* |
|---|---|---|---|---|---|
| 2 | 3.1385 | 3.4671 | 4.5023 | **4.4256** | 3.6525 |
| 3 | 3.1103 | **3.4456** | **4.4477** | 4.7248 | **3.6354** |
| 4 | **3.1022** | 3.4547 | 4.4501 | 5.0843 | 3.6588 |
| 5 | 3.1038 | 3.4775 | 4.4814 | 5.4691 | 3.7005 |
| 6 | 3.1109 | 3.5076 | 4.5297 | 5.8661 | 3.7525 |
| 7 | 3.1217 | 3.5422 | 4.589 | 6.2691 | 3.8107 |
| 8 | 3.1349 | 3.5798 | 4.6558 | 6.6748 | 3.8732 |

## 7. CONCLUDING REMARKS

In this paper, we modeled a call center as a multi-server retrial queueing model in which two types of customers arrive according to a Markovian arrival process (*MAP*). The customers who cannot enter into service immediately are allowed to enter into a retrial buffer of finite capacity provided there is enough waiting space. Otherwise they are lost. Arriving Type 1 and all retrial customers can preempt any Type 2 customers in service. The effect of the type of arrivals on the preemption rates is studied. Further an optimization problem is discussed.

## REFERENCES

[1] M.S. Aguir, Z. Karaesman, Aksin and F. Chauvet, The impact of retrials on call center performance, 26 (2004), 353-376.

[2] A.D. Ridley, W. Massey and M. Fu, Fluid approximation of a priority call center with time-varying arrivals, Telecommunications Review (2004) , 69-76.

[3] J.R. Artalejo, Accessible bibliography on retrial queues, Mathematical and Computer Modelling, 30 (1999), 1–6.

[4] J.R. Artalejo, A classical bibliography of research on retrial queues: progress in 1990–1999, TOP, 7 (1999), 187–211.

[5] J.R. Artalejo, A.N. Dudin and V.I. Klimenok, Stationary analysis of a retrial queue with preemptive repeated attempts, Operations Research Letters, 28 (2001), 173-180.

[6] J.R. Artalejo and M. Pozo, Numerical calculation of the stationary distribution of the main multi-server retrial queue, Annals of Operations Research, 116 (2002), 41–56.

[7] J.R. Artalejo, Accessible bibliography on retrial queues: Progress in 2000–2009, Mathematical and Computer Modelling, 51 (2010), 1071-1081.

[8] J.R. Artalejo, A. Gomez-Corral and Q. He, Markovian arrivals in stochastic modelling: a survey and some new results. SORT, 34 (2010), 101-144.

[9] S.R. Chakravarthy, A finite capacity GI/PH/l queue with group services, Naval Research Logistics Quarterly, 39 (1992), 345-35.

[10] S.R. Chakravarthy, The batch Markovian arrival process: A review and future work. Advances in Probability Theory and Stochastic Processes. Eds., A. Krishnamoorthy et al., Notable Publications Inc., NJ, 2001, 21-39.

[11] S.R. Chakravarthy, Markovian Arrival Processes. Wiley Encyclopedia of Operations Research and Management Science. Published Online: 15 JUN 2010.

[12] S.R. Chakravarthy and A.N.Dudin, Analysis of a Retrial Queuing Model with MAP Arrivals and Two Types of Customers, Mathematical and Computer Modelling, 37 (2003), 343-363.

[13] B.D. Choi and K.K. Park, The M/G/1 retrial queue for Bernoulli schedule, Queueing Systems 7 (1990), 219-228.

[14] B.D. Choi and J.W. Kim, Discrete-time $Geo_1$,$Geo_2$/G/l retrial queueing systems with two types of calls. Computers and Mathematics with Applications, 33 (1997), 79-88.

[15] B.D. Choi and Y. Chang, Single server retrial queues with priority calls. Mathematical and Cornputer Modelling, 30 (1999), 7-32.

[16] B.D. Choi, B.D and Y. Chang, $MAP_{1,}$ $MAP_2$/M/ c retrial queue with the retrial group of finite capacity and geometric loss, Mathematical and Computer Modelling, 30 (1999), 99-113.

[17] B.D. Choi, Y. Chang and B. Kim, $MAP_{1,}$ $MAP_2$/M/c retrial queue with guard channels and its applications to cellular networks, Top 7 (1999), 231-248.

[18] G.I. Falin, On sufficient conditions for Ergodicity of multichannel queueing systems with repeated calls, Advances in Applied Probability, 16 (1984), 447–448.

[19] G.I. Falin, Multi-channel queueing systems with repeated calls under high intensity of repetition, Journal of Information Processing and Cybernetics, 1 (1987), 37–47.

[20] G.I. Falin, and J.G.C. Templeton, Retrial Queues, Chapman and Hall, London, 1997.

[21] R. Kalyanaraman, and B.Srinivasan, A Single server retrial queue with Two types of calls and Recurrent repeated calls, International Journal of Information and Management Sciences, 14 (2003), 46-62.

[22] Z. Liu, Z and J. Wu, An MAP/G/1 G-queue with preemptive resume and multiple vacations, Applied Mathematical Modelling, 33 (2009), 1738-1748.

[23] D.M. Lucantoni, New results on the single server queue with a batch Markovian arrival process. Stochastic Models, 7 (1991), 1-46.

[24] A. Mandelbaum, Call centers (centres): Research Bibliography with Abstracts, Version 7 2006. http://iew3.technion.ac.il/serveng/References/US7_CC_avi.pdf.

[25] M.F. Neuts, Matrix-Geometric Solutions in Stochastic Models-An Algorithmic Approach. Dover, (originally published by Johns Hopkins University Press, 1981), (1995).

[26] M.F. Neuts, Structured Stochastic Matrices of M/G/1 type and their Applications. Marcel Dekker, NY, 1989.

[27] M.F. Neuts, Models based on the Markovian arrival process. IEICE Transactions on Communications, E75B (1992), 1255-1265.

[28] S.V. Pustova, Investigation of calls centers as retrial queueing systems, Cybernetics and System Analysis, 46 (2010), 494-499.

[29] J. Wang, On the single server retrial queue with priority subscribers and server breakdowns, J. Systems Science & Complexity, 21 (2009), 304-315.

**Received: February 1, 2013**