Neural, Parallel, and Scientific Computations 24 (2016) 107-120

# $MAP/PH/1$ RETRIAL QUEUE WITH CONSTANT RETRIAL RATE, WORKING VACATIONS AND A FINITE BUFFER FOR ARRIVALS

C. SREENIVASAN AND A. KRISHNAMOORTHY

Department of Mathematics, Government College Chittur
Palakkad - 678104, India
sreenikoonathara@gmail.com
Department of Mathematics, Cochin University of Science and Technology,
Cochin-22, India.
achyuthacusat@gmail.com

**ABSTRACT.** In this paper we study a $MAP/PH/1$ retrial queueing model in which the server is subject to taking vacations and serving at a lower rate during vacation. The service is returned to normal rate whenever the vacation gets over. If an arriving customer finds the server busy he joins a finite buffer. If the buffer is also full he joins a pool of unsatisfied customers called orbit. There from he makes retrial for a place in the server or buffer. Inter-retrial times are exponentially distributed with intensity independent of the number of customers in the orbit. The model is analyzed in steady state using matrix analytic methods. Illustrative numerical examples are presented.

**AMS (MOS) Subject Classification.** 68K25.

## 1. Introduction

In a retrial queueing system, customers entering a busy service system join a group of blocked customers called orbit. From the orbit each unit tries to access a free server independent of the other after a random amount of time. Such systems occur in communication and computer networks. For a nearly exhaustive account of development up to 2008 in this area we refer to surveys by Yang and Templeton [31] and Falin [12], bibliographic works Artalejo [1], [2] and the books Falin and Templeton [13] and Artalejo and Gomez-Corral [3].

In classical retrial queueing systems server idle time is very high. In the modern scenario it is not desirable from the service system's point of view to have a long idle time. To this end Artalejo et al. [4] introduced a new technique called orbital search where the server looks out for potential customers from the orbit after every service completion. Dudin et al. [11] and Krishnamoorthy et al. [15] also consider orbital search with different arrival streams and different service time distributions.

Received March 30, 2016 

Chakravarthy et al. [8] considers the multi server case. Retrial models have tremendous applications in communication systems. An application of retrial queueing model with finite buffer to Internet data traffic has been given in [5].

But even with the search option systems may not be able to utilize the entire server idle time because there may be situations when the orbit is empty. It is from this point of view one explores the possibility of retrial queueing systems with vacations and working vacations. During vacations the idle server may attend some less urgent secondary task.We may also consider the notion of working vacation depending upon the nature of the secondary job attended. In the latter case the server returns to attend the primary job as and when a customer arrives in the system. In this paper we consider a retrial queueing model with working vacations and with a finite buffer for customers (primary as well as orbital). This enhances the server utilization to the extent that the server in this retrial model has no idle time at all.

Queues with vacations have been extensively studied by several authors. Doshi [10] provides an exhaustive survey of such work through 1985. Since then the vacation models have been studied in different contexts. Among these include stochastic decomposition of queue length and that of stationary waiting time and we refer the reader to the recent book by Tian and Zhang [27] for details. Recently vacation models have gained significance in telecommunication networks. However compared to continuous time models discrete time models are more appropriate for modelling computer and telecommunication systems. Servi and Finn [25] introduced a working vacation model with the idea of offering services but at a lower rate whenever the server is on vacation. Their model was generalized to the case of $M/G/1$ in ([14], [30]), and to $GI/M/1$ model in [6]. A survey of working vacation models with emphasis on the use of matrix analytic methods is given in Tian and Li [28] . Working vacation models have a number of applications in practice. Two such examples are given in [28].

Recently, Li and Tian [18] studied an $M/M/1$ queue with working vacations in which vacationing server offers services at a lower rate for the first customer arriving during a vacation. Upon completion of the service at a lower rate the server will (a) continue the current vacation (if not finished) or take another vacation (if the working vacation expired) if there are no customers waiting; or (b) resume at a normal rate (irrespective of whether the vacation expired or not) if there are customers waiting. Resuming services at a normal rate while the vacation is still in progress corresponds to the vacation being interrupted. Sreenivasan et al. [26] analyzed $MAP/PH/1$ queue with working vacations, vacation interruptions and $N$ policy. $M/M/1$ retrial queue with working vacations has been discussed by Van Do [29]. But to the best of our knowledge no attempt has been there so far to analyze a $MAP/PH/1$ retrial queuing model with working vacations and a finite buffer for the primary arrivals and orbital customers.

For use in the sequel, let $\boldsymbol{e}(r)$, $\boldsymbol{e}_j(r)$ 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 $r$. When there is no need to emphasize the dimension of these vectors we will suppress the suffix. Thus, $\boldsymbol{e}$ will denote a column vector of 1's of appropriate dimension. 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 $(i,j)^{th}$ block matrix is given by $a_{ij}B$. If A and B are square matrices of order $m$ and $n$ respectively we define the Kronecker sum of A and B as A$\otimes I_n + I_m \otimes$B. In the forthcoming analysis of the model, in the absence of a suffix, the identity matrix which appears as the first factor in a Kronecker product is always of order $n$ and that appears as the second factor is of order $m$. For more details on Kronecker products, we refer the reader to [20] and [21].

In most the earlier works in this topic the input process is assumed to be poisson. But the traffic in modern communication network is highly irregular. Of late to model systems with repeated calls and bursty arrivals $MAP$ is used. The $MAP$ is a tractable class of point processes which is in general non renewal. However by choosing the parameters of the $MAP$ appropriately the underlying arrival process can be made a renewal process. The $MAP$ can represent a variety of processes which includes, as special cases, the Poisson process, the phase- type renewal processes, the Markov Modulated Poisson Process and superpositions of these. A brief discussion of MAP is given below.

A MAP is a Markov process $\{N(t), J(t)\}$ with state space $\{(i,j) : i \geq 0, 1 \leq j \leq m\}$ with infinitesimal generator $Q^*$ having the structure

$$Q^* = \begin{pmatrix} D_0 & D_1 & 0 & 0 & \dots \\ 0 & D_0 & D_1 & 0 & \dots \\ 0 & 0 & D_0 & D_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

Here $D_0$ and $D_1$ are square matrices of order $m$, $D_0$ has negative diagonal elements and non negative off-diagonal elements, $D_1$ has non negative elements and $(D_0 + D_1)\boldsymbol{e}_m = 0$. We define an arrival process associated with this Markov process as follows. An arrival occurs whenever a level state transition occurs into a state in the $D_1$ block, and there is no arrival otherwise. Here $N(t)$ represents the number of arrivals in (0,t], and $J(t)$ the phase of the Markov process at time $t$. Let $\boldsymbol{\delta}$ be the stationary probability vector of the generator $D = D_0 + D_1$. Then the constant $\lambda = \boldsymbol{\delta} D_1 \boldsymbol{e}_m$ is referred to as the fundamental rate gives the expected number of arrivals per unit time in the stationary version of the $MAP$.

Often, in model comparisons, it is convenient to select the time scale of the $MAP$ so that the stationary arrival rate $\lambda$ has a certain value. That is accomplished, in the continuous $MAP$ case, by multiplying the coefficient matrices $D_0$ and $D_1$, by the appropriate common constant. For further details on $MAP$ and their usefulness in stochastic modelling, we refer to [19], [23], [24] and for a review and recent work on $MAP$ we refer the reader to [7] and [9].

This paper is organized as follows. In Section 2 model description is provided.In Section 3 the steady state analysis of the model is presented. In Section 4 we discuss a few illustrative examples.

## 2. Mathematical Model

We consider a single server retrial queueing system in which customers arrive according to a markovian arrival process (MAP) with parameter matrices $D_0$ and $D_1$ of dimension $m$. An arriving primary customer who finds the server free, immediately occupies the server and obtains service. On the other hand if the arriving customer finds the server busy he joins a finite buffer of capacity $L$. If an arrival finds the buffer full he moves to an orbit of infinite size. Each customer in the orbit makes retrial at the rate $\beta$ for a place in the server or buffer. The service times follow phase type distribution with representation $(\boldsymbol{\alpha}, T)$ of order $n$. The server takes vacation when the customer being served depart from the system and no customers are left in the buffer. Duration of vacation is exponentially distributed with parameter $\eta$. During vacation if a customer (primary or orbital) arrives it interrupts the vacation. However the customers who arrive during the vacation is served only at a lower rate compared to the regular service. Precisely the vacation mode service times are also phase type distributed with representation $(\boldsymbol{\alpha}, \theta T)$, with $0 < \theta \leq 1$. Even when the vacation is interrupted by a customer vacation clock continues to tick so that on completion of this service if the vacation clock is not expired, the server continues on vacation in the absence of a customer in the buffer. At the end of each vacation the server takes another vacation if the buffer is empty.

### 2.1. The $QBD$ process.

The model discussed in Section 2 can be studied as a quasi-birth-and-death ($QBD$) process. First, we set up necessary notations. Let $\mu$ denote the service rate and it is easy to verify that $\mu = [\boldsymbol{\alpha}(-T)^{-1}\boldsymbol{e}]^{-1}$. Let $\theta, 0 < \theta \leq 1$, denote the factor by which the normal service rate will be reduced when the server is serving during vacation mode. That is, when the server is serving during the vacation mode, the rate of service is given by $\theta\mu$.

At time $t$, define

$$N_1(t) = \textit{The number of customers in the orbit,}$$

$$N_2(t) = \textit{The number of customers in the buffer,}$$

$$S_1(t) = \begin{cases} 0, & \textit{if the server is not working,} \\ j, & \textit{if the server is busy in phase } j, \ 1 \le j \le n, \end{cases}$$

If $S_1(t) \neq 0$ then

$$S_2(t) = \begin{cases} 0, & \textit{if the service is in vacation mode,} \\ 1, & \textit{if the service is in normal mode,} \end{cases}$$

and $M(t)$ to be the phase of the arrival process at time $t$. It is easy to verify that $\{(N_1(t), N_2(t), S_1(t), S_2(t), M(t)) : t \ge 0\}$ is a quasi-birth-and-death process ($QBD$) with state space

$$\Omega = \bigcup_{i_1=0}^{\infty} l(i_1)$$

where

$$l(i_1) = \{(i_1, i_2, j_1, j_2, k) : i_1 \ge 0, 0 \le i_2 \le L, 0 \le j_1 \le n, j_2 = 0 \text{ or } 1, 1 \le k \le m\}.$$

Note that when$S_1(t) = 0$, $S_2(t)$ does not play any role and will not be tracked.

The generator $Q$ of the $QBD$ process under consideration is of the form

$$Q = \begin{pmatrix} B_0 & A_0 & & & \\ A_2 & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix},$$

where the (block) matrices appearing in $Q$ are as follows.

$$B_0 = \begin{bmatrix} D_0 & \boldsymbol{\alpha} \otimes D_1 & O & O & O \\ \theta \boldsymbol{T}^0 \otimes I & \theta T \oplus D_0 - \eta I & \eta I & C_1 & O \\ \boldsymbol{T}^0 \otimes I & O & T \oplus D_0 & O & C_1 \\ O & \boldsymbol{e}_L \otimes \theta \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I & O & C_2 & \eta I \\ O & O & \boldsymbol{e}_L \otimes \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I & O & C_3 \end{bmatrix}$$

with $C_1 = \begin{bmatrix} I \otimes D_1 & O \end{bmatrix}$; $C_2$ has the block matrix $\theta T \oplus D_0 - \eta I$ along the diagonal, $I \otimes D_1$ along the superdiagonal and O matrices elsewhere; and the matrix $C_3$ has the block matrix $T \oplus D_0$ along the diagonal, $I \otimes D_1$ along the superdiagonal and O matrices elsewhere.

$$A_0 = \begin{bmatrix} O & O & O & O & O \\ O & O & O & O & O \\ O & O & O & O & O \\ O & O & O & \boldsymbol{e}_L(L)\boldsymbol{e}'_L(L) \otimes I \otimes D_1 & O \\ O & O & O & O & \boldsymbol{e}_L(L)\boldsymbol{e}'_L(L) \otimes I \otimes D_1 \end{bmatrix};$$

$$A_1 = \begin{bmatrix} D_0 - \beta I & \boldsymbol{\alpha} \otimes D_1 & O & O & O \\ \theta \boldsymbol{T}^0 \otimes I & F_1 & \eta I & F_4 & O \\ \boldsymbol{T}^0 \otimes I & O & F_2 & O & F_4 \\ O & F_3 & O & E_1 & \eta I \\ O & O & F_5 & O & E_2 \end{bmatrix};$$

where $F_1 = \theta T \oplus D_0 - \eta I - \beta I$, $F_2 = T \oplus D_0 - \beta I$, $F_3 = \boldsymbol{e}_L \otimes \theta \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I$, $F_4 = \boldsymbol{e}'_L(1) \otimes I \otimes D_1$, $F_5 = \boldsymbol{e}_L \otimes \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I$. The matrix $E_1$ has the block $\theta T \oplus D_0 - \beta I - \eta I$ along the diagonal except for the last block which is $\theta T \oplus D_0 - \eta I$ as retrials do not make any difference in the system state when the buffer is full, $I \otimes D_1$ along the superdiagonal and O matrices elsewhere. $E_2$ has $T \oplus D_0 - \beta I$ along the diagonal except for the last block which is $T \oplus D_0$, $I \otimes D_1$ along the superdiagonal and O blocks elsewhere.

$$A_2 = \begin{bmatrix} O & \beta(\boldsymbol{\alpha} \otimes I) & O & O & O \\ O & O & O & H_1 & O \\ O & O & O & O & H_1 \\ O & O & O & H_2 & O \\ O & O & O & O & H_2 \end{bmatrix}$$

with $H_1 = \begin{bmatrix} \beta I & O \end{bmatrix}$; $H_2$ has the block matrix $\beta I$ along the superdiagonal and O blocks elsewhere.

## 3. Steady state analysis

In this section we will discuss the steady state analysis of the model under study.

**3.1. Stability condition.** Define $A = A_0 + A_1 + A_2$. Let $\boldsymbol{\pi} = (\boldsymbol{\pi_0}, \boldsymbol{\pi_1}, \boldsymbol{\pi_2}, \boldsymbol{\pi_3}, \boldsymbol{\pi_4})$ be the steady state probability vector of A, where $\boldsymbol{\pi_0}$ is of dimension $m$, $\boldsymbol{\pi_1}$,$\boldsymbol{\pi_2}$ are of dimension $mn$ and $\boldsymbol{\pi_3}$, $\boldsymbol{\pi_4}$ are of are of dimension $Lmn$. Note that $\boldsymbol{\pi}$ is the unique vector satisfying the condition $\boldsymbol{\pi} A = \boldsymbol{0}$ and $\boldsymbol{\pi e} = 1$. For stability of the queueing model we must have $\boldsymbol{\pi} A_0 \boldsymbol{e} < \boldsymbol{\pi} A_2 \boldsymbol{e}$, (see [22]) which simplifies to $(\boldsymbol{\pi_3} + \boldsymbol{\pi_4})\boldsymbol{e}_L(L) \otimes (\boldsymbol{e}_n \otimes \mathrm{D}_1 \boldsymbol{e}_m) < \beta(\boldsymbol{\pi_0}\boldsymbol{e}_m + (\boldsymbol{\pi_1} + \boldsymbol{\pi_2})\boldsymbol{e}_{mn} + \sum_{j=1}^{Lmn-1}(\boldsymbol{\pi_{3j}} + \boldsymbol{\pi_{4j}}))$. The last inequality suggests that for stability of the queueing system discussed here it is required that the rate of inflow in to the orbit is less than the effective retrial rate.

3.2. **Steady state probability vector.** Let $\boldsymbol{x}$, partitioned as $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots)$, be the steady state probability vector of Q. Note that $\boldsymbol{x}_j$ is of dimension $m + 2mn + 2Lmn$ for $j \geq 0$. The vector $\boldsymbol{x}$ satisfies the condition $\boldsymbol{x}Q = \mathbf{0}$ and $\boldsymbol{x}\boldsymbol{e} = 1$. Apparently when the stability condition is satisfied the sub vectors of $\boldsymbol{x}$ , corresponding to the different level states are given by the equation $\boldsymbol{x}(j) = \boldsymbol{x}_0 R^j$, $j \geq 1$, where $R$ is the minimal non negative solution of the matrix quadratic equation $R^2A_2 + RA_1 + A_0 = 0$, (see [22]). The sub vector $\boldsymbol{x}_0$ is obtained by solving the equations

$$\boldsymbol{x}_0(B_1 + RA_2) = \mathbf{0}$$

subject to the normalizing condition

$$\boldsymbol{x}_0(I - R)^{-1}\boldsymbol{e} = 1.$$

The computation of $R$ matrix can be carried out using a number of well known methods such as logarithmic reduction. We list here only the main steps involved in logarithmic reduction algorithm for computation of $R$. For full details of the logarithmic reduction algorithm we refer the reader to [16].

**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:**

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

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

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

**Remark**: In a model like this it is very significant to know what should be the ideal size of the buffer. Such a discussion invariably depends on how often the buffer overfows and how often the server goes on vacation as the buffer becomes empty. However, it is practically impossible to study these when the orbit level changes. It is in this context we study the following characteristic of the model.

3.3. **Busy server versus empty orbit.** Let $T_B$ be the duration for which the service goes on with no customer in the orbit. It is the duration for which the service goes on once it is started in vacation mode with the arrival of a customer into the empty system. Now $T_B$ can end either by the server taking another vacation as the buffer becomes empty at a departure epoch or as a customer moves to the orbit when the buffer becomes full. Thus $T_B$ can be interpreted as the time until absorption in a finite state continous time Markov chain with two absorbing states. The process starts according to the probability vector $\boldsymbol{\gamma}_M$ given by

$$\boldsymbol{\gamma}_M = c(\boldsymbol{\alpha} \otimes \boldsymbol{x^*}_0 D_1, \boldsymbol{0}),$$

where $\boldsymbol{x^*}_0$ is the row vector formed by the first $m$ components of $\boldsymbol{x}_0$ and the normalising constant $c$ is given by $c = [\boldsymbol{x^*}_0 D_1 \boldsymbol{e}_m]^{-1}$. The matrix $M$ of transient states given by

$$M = \begin{bmatrix} \theta T \oplus D_0 - \eta I & \eta I & C_1 & O \\ O & T \oplus D_0 & O & C_1 \\ \boldsymbol{e}_L \otimes \theta \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I & O & C_2 & \eta I \\ O & \boldsymbol{e}_L \otimes \boldsymbol{T}^0 \boldsymbol{\alpha} \otimes I & O & C_3 \end{bmatrix}$$

is obtained by omitting the first $m$ rows and columns of the block matrix $B_0$ of $Q$. Also let

$$\boldsymbol{M}_1^0 = \begin{pmatrix} \theta(\boldsymbol{T}^0 \otimes \boldsymbol{e}_m) \\ \boldsymbol{T}^0 \otimes \boldsymbol{e}_m \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}, \ \boldsymbol{M}_2^0 = \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{e}_n \otimes D_1 \boldsymbol{e}_m \\ \boldsymbol{0} \\ \boldsymbol{e}_n \otimes D_1 \boldsymbol{e}_m \end{pmatrix},$$

where each $\boldsymbol{0}$ in $\boldsymbol{M}_1^0$ has dimension $Lmn$ and the first and second $\boldsymbol{0}$'s in $\boldsymbol{M}_2^0$ have dimensions $(L+1)mn$ and $(L-1)mn$ respectively. Hence the probability that the server goes on vacation before the buffer overflows is given by $p_v = \boldsymbol{\gamma}_M(-M)^{-1}\boldsymbol{M}_1^0$. Equivalently the probability that the buffer overflows before the server taking a vacation is given by $p_b = \boldsymbol{\gamma}_M(-M)^{-1}\boldsymbol{M}_2^0$. Clearly $p_b = 1 - p_v$.

In the general context, that is if the orbit is not empty, retrials will also come into picture. Natuarally $p_v$ tends to decrease and $p_b$ tends to increase. Hence given the input parameters, the above values of $p_v$ and $p_b$ are respectively the upper and lower bounds for $p_v$ and $p_b$.

3.4. **Key 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. The measures are listed below along with their formulae for computation.

1. Probability that the orbit is empty: $P_{OTY} = \boldsymbol{x}_0 \boldsymbol{e}$.
2. Probability that the buffer is empty: $P_{BUFTY} = \sum_{i_1=0}^{\infty} \boldsymbol{x}_{i_1 0} \boldsymbol{e}_{m+2mn}$.
3. The probability that the server is on vacation: $P_{VACN} = \sum_{i_1=0}^{\infty} \sum_{k=1}^{m} \boldsymbol{x}_{i_1 00.k}$.
4. The probability that the server is busy in vacation mode:
   $P_{BVM} = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{L} \sum_{j_1=1}^{n} \sum_{k=1}^{m} \boldsymbol{x}_{i_1 i_2 j_1 0 k}$.
5. Probability that the server completes a service in vacation mode: $P_{SCSLO} = P(\text{service time in slow mode} < \text{an exponentially distributed random variable with parameter } \eta) = \boldsymbol{\alpha}(\eta I - \theta T)^{-1}\theta \boldsymbol{T}^0$
6. The probability that the server is busy in normal mode:
   $P_{BNM} = \sum_{i_1=0}^{\infty} \sum_{i_2=0}^{L} \sum_{j_1=1}^{n} \sum_{k=1}^{m} \boldsymbol{x}_{i_1 i_2 j_1 1 k}$.

7. The mean number of customers in the orbit:
   $\mu_{MNOBT} = \sum_{i_1=1}^{\infty} i_1 \boldsymbol{x}_{i_1} \boldsymbol{e} = \boldsymbol{x}_0 R(I-R)^{-2} \boldsymbol{e}$
8. The mean number of customers in the buffer: $\mu_{BUF} = \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{L} i_2 \boldsymbol{x}_{i_1 i_2} \boldsymbol{e}_{2mn}$
9. Probability of a successful retrial:
   $P_{SRT} = \beta/(\beta+\lambda) \sum_{i_1=1}^{\infty} \sum_{k=1}^{m} (\sum_{i_2=1}^{L-1} \sum_{j_1=1}^{n} \sum_{j_2=0}^{1} \boldsymbol{x}_{i_1 i_2 j_1 j_2 k} + \boldsymbol{x}_{i_1 00.k})$.

## 4. **Numerical Results**

In order to bring out the qualitative nature of the model under study, we present a few representative examples in this section. For the arrival process we consider the following five sets of matrices for $D_0$ and $D_1$.

1. Erlang (ERA)

$$D_0 = \begin{pmatrix} -5 & 5 & & & \\ & -5 & 5 & & \\ & & -5 & 5 & \\ & & & -5 & 5 \\ & & & & -5 \end{pmatrix} \qquad D_1 = \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ 5 & & & & \end{pmatrix}$$

2. Exponential (EXA)
   $D_0 = (-1), D_1 = (1)$
3. Hyperexponential (HEA)

$$D_0 = \begin{pmatrix} -10 & 0 \\ 0 & -1 \end{pmatrix} \qquad D_1 = \begin{pmatrix} 9 & 1 \\ 0.9 & 0.1 \end{pmatrix}$$

4. MAP with negative correlation ($MNA$)

$$D_0 = \begin{pmatrix} -2 & 2 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -450.5 \end{pmatrix} \qquad D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0.02 & 0 & 1.98 \\ 445.995 & 0 & 4.505 \end{pmatrix}$$

5. MAP with positive correlation ($MPA$)

$$D_0 = \begin{pmatrix} -2 & -2 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -450.5 \end{pmatrix} \qquad D_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1.98 & 0 & 0.02 \\ 4.505 & 0 & 445.995 \end{pmatrix}$$

These five $MAP$ processes are qualitatively different in that they have different variance and correlation structure. The first three arrival processes, namely $ERA$, $EXA$, and $HEA$, correspond to renewal processes and so the correlation is 0. The arrival process labelled $MNA$ has correlated arrivals with correlation between two successive inter-arrival times given by $-0.4889$ and the arrival process corresponding to the one labelled $MPA$ has a positive correlation with value 0.4889. The ratio of the standard deviations of the inter-arrival times of these five arrival processes to that of $ERA$ are respectively 1, 2.2361, 5.0194, 3.1518, and 3.1518. Since the mean arrival rate (and hence the mean interarrival time) is taken to be the same for all five $MAP$ processes the coefficients of variation ($CV$) of the interarrival times of the processes

$ERA$, $EXA$, $HEA$, $MNA$, $MPA$ are in the ratio 1, 2.2361, 5.0194, 3.1518, 3.1518 respectively.

For the service time distribution we consider the following three phase type distributions.

1. Erlang (ERS)
   $\boldsymbol{\alpha} = (1, 0)\ T = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}$
2. Exponential (EXS)
   $\boldsymbol{\alpha} = 1.0, T = -1.0$

These two phase type distributions have a service rate of 1. Note that these are qualitatively different in that they have different variances. The ratio of the standard deviation of $EXS$ to that of $ERS$ is 1.4142. Since the mean service time is the same for these two processes the ratios of $CV$ of the service times of $ERS$ and $EXS$ are 1 and 1.4142 respectively.

**ILLUSTRATIVE EXAMPLE 1:** Here we examine the effect of the vacation parameter $\eta$ on $p_v$, the probability that the sever goes on vacation before the buffer overflows, given that the orbit is empty. We fix $\lambda = 0.9$, $\mu = 1$, $\theta = 0.6$ and $L = 3$.

Table 1: The probability $p_v$

| | Erlang services | | | | | Exponential services | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\eta$ | $ERA$ | $EXA$ | $HEA$ | $MNA$ | $MPA$ | $ERA$ | $EXA$ | $HEA$ | $MNA$ | $MPA$ |
| 0.1 | 0.892 | 0.798 | 0.605 | 0.730 | 0.974 | 0.861 | 0.795 | 0.650 | 0.719 | 0.964 |
| 0.2 | 0.925 | 0.829 | 0.637 | 0.765 | 0.977 | 0.89 | 0.820 | 0.673 | 0.749 | 0.969 |
| 0.3 | 0.944 | 0.849 | 0.659 | 0.789 | 0.979 | 0.908 | 0.837 | 0.69 | 0.770 | 0.973 |
| 0.4 | 0.955 | 0.863 | 0.677 | 0.807 | 0.98 | 0.92 | 0.849 | 0.703 | 0.786 | 0.975 |
| 0.5 | 0.962 | 0.873 | 0.690 | 0.820 | 0.981 | 0.928 | 0.858 | 0.714 | 0.798 | 0.976 |

As $\eta$ increases the mean duration of vacation decreases. Hence the server switches from slow service to normal service earlier and hence clears out the customers at a faster rate. Due to this $p_v$ decreases as $\eta$ increases. From the table it is clear that $ERS$ gives the larger values for $p_v$ than $EXS$. Thus as $CV$ of the service time increases the value of $p_v$ decreases. Among the renewal arrival processes $ERA$ has the greatest value and $HEA$ has the least value for $p_v$. This again is the effect of $CV$ of the interarrival times of these processes. Among the correlated arrival processes $MPA$ has the larger value for $p_v$ compared to $MNA$. This shows the difference positive and negative correlation producing on $p_v$.

**ILLUSTRATIVE EXAMPLE 2:** We analyze the effect of change in the buffer size on the measure 'probability of successful retrials $P_{SRT}$', for different arrival and

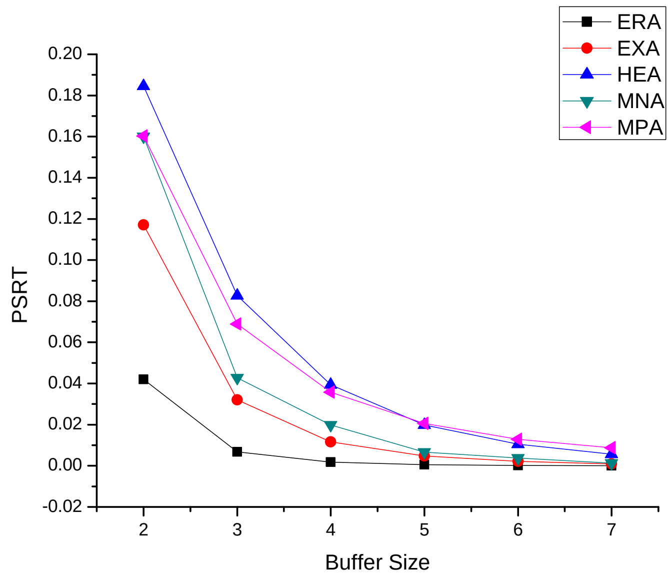


FIGURE 1. Probability of successful retrials - Erlang services

service processes. Figure 1 analyzes the effect of the buffer size with Erlang service and Figure 2 explains its effect with exponential service. We fix $\lambda = 0.9$, $\mu = 1$, $\eta = 0.5$, $\theta = 0.6$ and $\beta = 1$.

- As the buffer size increases more primary arrivals occupy the buffer. This reduces the flow of customers to the orbit and the chance of successful retrial. From the figures it is clear that $P_{SRT}$ increases with $CV$ of the interarrival times of distributions. Note that both $MNA$ and $MPA$ have the same $CV$ for inter-arrival times, but this measure is higher for $MPA$ compared to $MNA$. Observe that $MPA$ has a positive correlation and $MNA$ has a negative correlation. This shows the effect of correlation on this measure.

**ILLUSTRATIVE EXAMPLE 3:** In this example we study the effect of the parameter $\eta$ on the measure probability of a service completion in slow mode ($P_{SCSLO}$). Fix $\lambda = 0.9$, $\mu = 1$, $\beta = 1$ and $\theta = 0.6$ and $L = 3$.

- From the expression for $P_{SCSLO}$ in subsection 3.4, it is clear that this measure is independent of the inter arrival time distributions and that it decreases as $\eta$ increases. So we compare the values for $P_{SCSLO}$ for the two service time distributions. From figure 3, it is clear that $P_{SCSLO}$ increases with the variance of the service time distributions.

**Acknowledgements**

C. Sreenivasan's research is supported by the Kerala State Council for Science, Technology and Environment (KSCSTE), Thiruvananthapuram under the SARD project (Council (P) Order No. 436/2014 KSCSTE, Thiruvananthapuram).

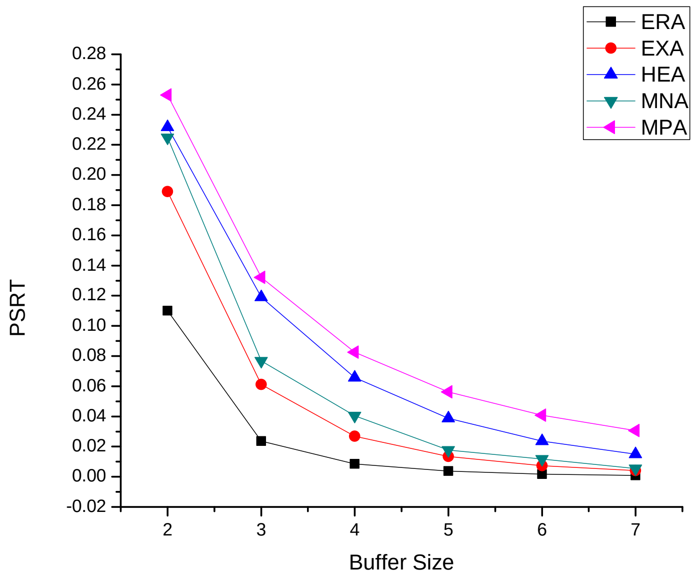


FIGURE 2. Probability of successful retrials - Exponential services

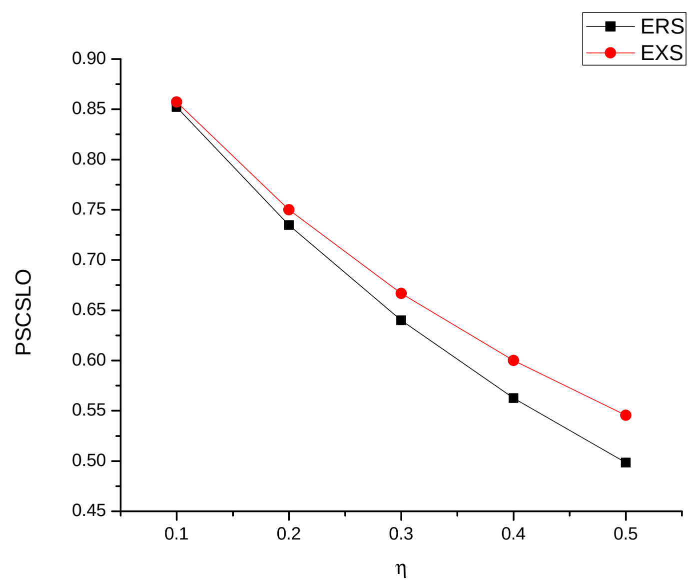


FIGURE 3. Probability of a service completion in slow mode

## REFERENCES

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

[2] J. R. Artalejo, A Classified bibliography of research on retrial queues : Progress in 1990-1999, *TOP*, 7, 187–211, (1999).

[3] J. R. Artalejo and A Gomez-Corral, Retrial Queueing Systems : A Computational Approach, *Springer Verlag*, 2008.

[4] J. R. Artalejo, V. C. Joushua, and A. Krishnamoorthy, An $M/G/1$ retrial queue with orbital search by the server, *In Advances in Stochastic Modelling*, eds. J. R. Artalejo and A. Krishnamoorthy, 41–54, Notable Publications, New Jersey, (2002).

[5] K. Avrachenkov and U. Yechiali, Retrial networks with finite buffers and their application to internet data traffic, *Probability in the Engineering and Informational Sciences archive*, 22(4), 519–536, (2008).

[6] Y. Baba, Analysis of a $GI/M/1$ queue with multiple working vacations, *Operations Research Letters*, 33, 201–209, (2005).

[7] S. R. Chakravarthy, The Batch Markovian Arrival Process: A Review and Future Work, *In Advances in Stochastic Modelling, eds. J.R. Artalejo and A. Krishnamoorthy*, 41–54, Notable Publications, New Jersey, (2002).

[8] S. R. Chakravarthy, A. Krishnamoorthy and V. C. Jousua, Analysis of a multi-server retrial queue with search of customers from the orbit, *Performance Evaluation*, 63, 776–798, (2006).

[9] S. R. Chakravarthy, Markovian arrival processes, *Wiley Encyclopedia of Operations Research and Management Science*, 15 JUN (2010).

[10] B. T. Doshi, Queueing systems with vacations- a survey, *Queueing System*, 1, 29–66, (1986).

[11] A. N. Dudin, A. Krishnamoorthy, V. C. Joushua, and G. V. Tsarenkov, Analysis of the $BMAP/G/1$ retrial system with search of customers from the orbit, *European Journal of Operational Research*, 157, 169–179, (2004).

[12] G. I. Falin, A Survey of retrial queues, *Queueing Systems*, 7, 127–168, (1990).

[13] G. I. Falin, and J. G. C. Templeton, Retrial Queues, *Chapman and Hall*, London, (1997)

[14] J. Kim, D. Choi and K. Chae, Analysis of queue-length distribution of the $M/G/1$ queue with working vacations, *International Conference on Statistics and Related Fields*, Hawaii, (2003).

[15] A. Krishnamoorthy, T. G. Deepak and V. C. Joushua, An $M/G/1$ Retrial Queue with Non persistent Customers and Orbital Search, *Stochastic Analysis and Applications*, 23, 975–997, (2005).

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

[17] Y. Levy and U. Yechiali, Utilization of idle time in an $M/G/1$ queueing system, *Management Science*, 22, 202–211, (1975).

[18] J. Li and N. Tian, The $M/M/1$ queue with working vacations and vacation interuptions; *Journal of Systems Science and Systems Engineering* 16(1), 121–127, (2007).

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

[20] A. Graham, Kronecker Products and Matrix calculus with Applications, *Ellis Horwood Ltd.*, Chichester, (1981).

[21] M. Marcus and H. Minc, A Survey of Matrix Theory and Matrix Inequalities *Allyn and Bacon*, Boston, MA, (1964).

[22] M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models : An Algorithmic Approach, *The Johns Hopkins University Press*, Baltimore, MD, (1981) [1994 version is Dover Edition]

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

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

[25] L. Servi and S. Finn, $M/M/1$ queue with working vacations ($M/M/1/WV$). *Performance Evaluation*, 50, 41–52, (2002).

[26] C. Sreenivasan, S. R. Chakravarthy, and A. Krishnamoorthy $MAP/PH/1$ queue with working vacations, vacation interruptions and $N$ policy, *Applied Mathematical Modelling*, 37, 3879–3893, (2013).

[27] N. Tian and Z. G. Zhang,Vacation Queueing Models: Theory and applications, *Springer Publishers*, New York, 2006.

[28] N. Tian, J. Li, Z. G. Zhang, Matrix Analytic Method and Working Vacation Queues - A Survey. *International Journal of Information and Management Sciences*, 20, 603–633, (2009).

[29] T. Van Do, $M/M/1$ retrial queue with working vacations, *Acta Informatica*, 47, 67–75, (2010).

[30] D. Wu and H. Takagi, $M/G/1$ queue with multiple working vacations, *Performance Evaluation*, 63, 654–681, (2006).

[31] T. Yang and J. G. C. Templeton, A Survey on retrial queues, *Queueing Systems*, 2, 201–233, (1987).