Stochastic Analysis and Applications

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsaa20

The Busy Period and the Waiting Time Analysis of a MAP/M/c Queue with Finite Retrial Group

J. R. Artalejo [a], S. R. Chakravarthy [b] & M. J. Lopez-Herrero [c]

[a] Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain

[b] Department of Industrial and Manufacturing Engineering, Kettering University, Flint, Michigan, USA

[c] School of Statistics, Complutense University of Madrid, Madrid, Spain

Available online: 07 Mar 2007

To cite this article: J. R. Artalejo, S. R. Chakravarthy & M. J. Lopez-Herrero (2007): The Busy Period and the Waiting Time Analysis of a MAP/M/c Queue with Finite Retrial Group, Stochastic Analysis and Applications, 25:2, 445-469

To link to this article: http://dx.doi.org/10.1080/07362990601139651

Stochastic Analysis and Applications, 25: 445–469, 2007

ISSN 0736-2994 print/1532-9356 online
DOI: 10.1080/07362990601139651

The Busy Period and the Waiting Time Analysis of a MAP/M/c Queue with Finite Retrial Group

J. R. Artalejo
Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain

S. R. Chakravarthy
Department of Industrial and Manufacturing Engineering, Kettering University, Flint, Michigan, USA

M. J. Lopez-Herrero
School of Statistics, Complutense University of Madrid, Madrid, Spain

Abstract: We concentrate on the analysis of the busy period and the waiting time distribution of a multi-server retrial queue in which primary arrivals occur according to a Markovian arrival process (*MAP*). Since the study of a model with an infinite retrial group seems intractable, we deal with a system having a finite buffer for the retrial group. The system is analyzed in steady state by deriving expressions for (a) the Laplace–Stieltjes transforms of the busy period and the waiting time; (b) the probabiliy generating functions for the number of customers served during a busy period and the number of retrials made by a customer; and (c) various moments of quantites of interest. Some illustrative numerical examples are discussed.

Keywords: Busy period; Markovian arrival process; Multi-server queue; Retrials; Waiting time.

Received July 24, 2006; Accepted September 20, 2006

J. R. Artalejo and M. J. Lopez-Herrero thank the support received from the research project MTM2005-01248.

Address correspondence to J. R. Artalejo, Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid 28040, Spain; E-mail: jesus_artalejo@mat.ucm.es

AMS Subject Classification: 60K25; 90B22.

1. INTRODUCTION

Many queueing situations have the feature that customers who find all servers busy upon arrival leave the service area immediately and repeat their requests for service after a random time. Between requests customers are assumed to be in a retrial orbit. Retrial queueing systems are presented as alternatives to clasical waiting lines and loss systems in the monograph by Falin and Templeton [1]. The practical applications of retrial models are wide including telephone systems, call centers, local area networks, communication protocols and queues arising in daily life where retrials occur due to blocking or impatience. In fact, the retrial nature of many queueing systems has been noted since the beginning of queueing theory, and its importance was expressed by Kosten [2] who writes that "*any theoretical result does not take into consideration this repetition effect should be considered suspect*" (p. 33).

The interest and the study of retrial queues is thus justified, however, we have to note that the majority of retrial systems, even the Markovian models, are not analytically tractable without imposing additional restrictions. The reason lies in the spatial heterogeneity of the underlying process obtained as a result of superimposed flow of repeated attempts coming from the orbit with the stream of primary arrivals. Moreover, most retrial queues operate in random order which makes the waiting time analysis difficult.

In the last few decades there has been an increasing interest in the application of matrix-analytic methods to a variety of retrial systems. In particular, many authors have combined matrix-analytic techniques and approximating methods (i.e., truncation and generalized truncation) to investigate the queue length characteristics of multi-server retrial queues. Among others we mention the work by Neuts and Rao [3], Choi and Chang [4], Diamond and Alfa [5], Breuer et al. [6], and Chakravarthy et al. [7]. For a detailed bibliography see Gomez-Corral [8].

In contrast, the study of the busy period and the waiting time in multi-server retrial queues is relatively new. Recently, a number of papers study these descriptors for the $M/M/c$ retrial queue (see Artalejo et al. [9], Artalejo and Gomez-Corral [10], Artalejo et al. [11], and Artalejo and Lopez-Herrero [12]). To the best of our knowledge, there is no literature available on the study of the busy period and the waiting time analysis for multi-server retrial queues with MAP arrivals. The main objective of this paper is to study these descriptors in the context of $MAP/M/c$ retrial queue. We remark again that the existing literature emphasizes the necessity of approximating the performance characteristics of multi-server retrial queues. Thus, in this article, we assume the most natural and traditional approach of restricting the orbit capacity to be of finite size, say K.

The busy period is probably the most important first-passage descriptor of any queueing model. It gives an interesting measure from the service provider's point of view. Performance characteristics of the busy period and the number of customers served are employed to define the cost structure of control queueing problems. Our analysis of the busy period includes the following contributions:

(a) We obtain the Laplace–Stieltjes transforms and the generating functions governing the length of the busy period and the number of customers served, respectively. The numerical inversion of the underlying density and the computation of the probability mass function are carried out for some selected scenarios.

(b) We develop recursive equations for the computation of any arbitrary moment of the busy period and the number of customers served.

On the other hand, the waiting time is the most significant queueing descriptor from the customer's point of view. We will deal with a random order policy among the customers in orbit. Our contributions include the following:

(c) We obtain the Laplace–Stieltjes transform of the waiting time of a tagged customer and perform the subsequent numerical inversion. We also investigate the discrete counterpart of the waiting time consisting in the number of retrials made by the marked customer before getting a free server.

(d) The development of recursive schemes for the computation of arbitrary moments of both the waiting time and the number of retrials made by a customer.

As related work, we mention Artalejo and Chakravarthy [13] who investigated the computation of the maximum number of customers in orbit in the $MAP/M/c$ with an infinite retrial group. The computation of the stationary distribution of the system state for the $MAP/M/c$ retrial queue can be obtained as a particular case of the model considered by Chakravarthy et al. [7], where primary arrivals follow a Markovian arrival process and the servers with a certain probability search for customers.

The rest of the paper is organized as follows. In Section 2, we describe the mathematical model and give a brief presentation of the stationary distribution of the system state. In Section 3, we develop the computational analysis for the length of a busy period. The number of customers served is investigated in Section 4. In Sections 5 and 6, we investigate respectively the waiting time and the number of retrials made by a customer before getting service. A brief review of the Markovian arrival process is given in the appendix. We present some illustrative numerical results.

For use in sequel, let $\mathbf{e}(r)$, $\mathbf{e}_j(r)$, and I_r denote, respectively, the (column) vector of dimension r consisting of 1's, column vector of dimension r with 1 in the jth 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, $\mathbf{e}$ will denote a column vector of 1's of appropriate dimension. The notation "$\prime$" appearing in a matrix will stand for the matrix transpose. 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 $a_{ij}B$. For more details on Kronecker products, we refer the reader to Marcus and Minc [14].

2. DESCRIPTION OF THE MATHEMATICAL MODEL AND STATIONARY DISTRIBUTION

We deal with a multi-server model with c identical servers. The primary customers arrive according to a Markovian arrival process (MAP) with representation (D_0, D_1) of order m. The service times are exponentially distributed with rate μ. Any arriving customer finding all servers busy enters an orbit of capacity K from where the retrial customers compete for service. The interretrial times of each customer in orbit are assumed to be exponentially distributed with rate θ. The Markovian arrival process, the service times, and the retrial times are assumed to be mutually independent.

Let $N(t)$, $C(t)$, and $M(t)$ denote, respectively, the number of customers in the retrial orbit, the number of busy servers, and the phase of the arrival process at time t. The process $\{(N(t), C(t), M(t)); t \geq 0\}$ is a continuous-time Markov chain with state space given by

$$S = \{(i, j, k); 0 \leq i \leq K, 0 \leq j \leq c, 1 \leq k \leq m\}.$$

We partition S as follows

$$\underline{\mathbf{0}}^* = \{(0, 0, k); 1 \leq k \leq m\},$$

$$\underline{\mathbf{0}} = \{(0, j, k); 1 \leq j \leq c, 1 \leq k \leq m\},$$

$$\underline{\mathbf{i}} = \{(i, j, k); 0 \leq j \leq c, 1 \leq k \leq m\}, \quad 1 \leq i \leq K.$$

Then, the infinitesimal generator of the Markov chain $\{(N(t), C(t), M(t)); t \geq 0\}$ has the form

$$Q^{(K)} = \begin{pmatrix} D_0 & \mathbf{e}_1'(c) \otimes D_1 & & & & & \\ \mu \mathbf{e}_1(c) \otimes I & A_{10} & A_{00} & & & & \\ & A_{21} & A_{11} & A_0 & & & \\ & \ddots & & \ddots & & \ddots & \\ & & & & A_{2,K-1} & A_{1,K-1} & A_0 \\ & & & & & A_{2K} & A_{1K}^* \end{pmatrix}, \quad (1)$$

where the coefficient matrices appearing in (1) are given by

$$A_{10} = \begin{pmatrix} D_0 - \mu I & D_1 & & & \\ 2\mu I & D_0 - 2\mu I & D_1 & & \\ & \ddots & \ddots & \ddots & \\ & & & c\mu I & D_0 - c\mu I \end{pmatrix},$$

A_{00} is a rectangular matrix of dimensions $cm \times (c+1)m$ whose elements are all zero except the $(c, c+1)$th block entry which is given by D_1,

$$A_{1i} = \begin{pmatrix} D_0 - i\theta I & D_1 & & & \\ \mu I & D_0 - (\mu + i\theta)I & D_1 & & \\ & \ddots & \ddots & \ddots & \\ & & (c-1)\mu I & D_0 - ((c-1)\mu + i\theta)I & D_1 \\ & & & c\mu I & D_0 - c\mu I \end{pmatrix},$$

$1 \leq i \leq K - 1,$

A_{1K}^* is defined as A_{1K} except for the $(c+1, c+1)$th block entry that is replaced by $Q^* - c\mu I$,

$$A_{21} = \theta \begin{pmatrix} I & & & \\ & I & & \\ & & \ddots & \\ & & & I \\ & & & 0 \end{pmatrix}, \quad A_{2i} = i\theta \begin{pmatrix} 0 & I & & \\ & & I & \\ & & \ddots & \\ & & & I \\ & & & 0 \end{pmatrix}, \quad 2 \leq i \leq K,$$

and A_0 is a square matrix of dimension $(c+1)m$ whose elements are all zero except for the $(c+1, c+1)$th block entry which is given by D_1.

Let $\mathbf{x}$, partitioned as $\mathbf{x} = (\mathbf{x}^*, \mathbf{x}(0), \ldots, \mathbf{x}(K))$, denote the stationary probability vector of $Q^{(K)}$. That is, $\mathbf{x}$ satisfies

$$\mathbf{x}Q^{(K)} = \mathbf{0}, \quad \mathbf{x}\mathbf{e}((K+1)(c+1)m) = 1.$$

The computation of the stationary distribution is reduced to solving a finite block tridiagonal system. At this point we refer to Chakravarthy et al. [7] from where the details for the computation of the vector $\mathbf{x}$ for our model can be obtained by taking $p = 0$ in their model. In fact, the existence of finite block tridiagonal matrix structures will be a common feature in the sequel, but a comparison among different methods of solution is not our aim in this article. Here, we simply mention that a number of well-known methods can be used such as block forward-elimination-backward substitution, aggregate/disaggregate techniques, block Gaussian–Seidel iteration, etc.

3. THE LENGTH OF THE BUSY PERIOD

The busy period of the $MAP/M/c/K$ retrial queue is the duration commencing when an arriving customer finds the system empty (i.e., the arriving customer sees the state $\underline{\mathbf{0}}^*$) and ends when the system visits state $\underline{\mathbf{0}}^*$ again at a service completion.

First, we introduce some notation:

$T^{(K)}_{(i,j,k)} =$ the first-passage time to the level $\underline{\mathbf{0}}^*$ given that the initial state is (i, j, k),

$$\varphi^{(K)}_{(i,j,k)}(s) = E\Big[\exp\{-sT^{(K)}_{(i,j,k)}\}\Big], \quad \mathrm{Re}(s) \geq 0, \ \ (i, j, k) \in S.$$

The following vectors comprise the above Laplace–Stieltjes transforms partitioned according to the orbit levels:

$$\boldsymbol{\varphi}^{(K)}_0(s) = \big(\varphi^{(K)}_{(0,1,1)}(s), \ldots, \varphi^{(K)}_{(0,c,m)}(s)\big)',$$
$$\boldsymbol{\varphi}^{(K)}_i(s) = \big(\varphi^{(K)}_{(i,0,1)}(s), \ldots, \varphi^{(K)}_{(i,c,m)}(s)\big)', \quad 1 \leq i \leq K,$$
$$\boldsymbol{\varphi}^{(K)}(s) = \big(\boldsymbol{\varphi}^{(K)}_0(s), \ldots, \boldsymbol{\varphi}^{(K)}_K(s)\big)'.$$

Moreover, we have $\boldsymbol{\varphi}^{(K)}_{\underline{\mathbf{0}}^*}(s) = \mathbf{e}(m)$.

Theorem 1. *The Laplace–Stieltjes transforms* $\{\varphi^{(K)}_{(i,j,k)}(s); (i, j, k) \in S\}$ *satisfy the following tridiagonal system*

$$T^{(K)}_L(s)\boldsymbol{\varphi}^{(K)}(s) = \mathbf{f}^{(K)}, \tag{2}$$

where $T^{(K)}_L(s) = \overline{Q}^{(K)} - sI_{cm+K(c+1)m}$ *and* $\mathbf{f}^{(K)} = -(\mu\mathbf{e}_1(c) \otimes \mathbf{e}(m), \mathbf{0})'$. *The matrix* $\overline{Q}^{(K)}$ *is obtained from* $Q^{(K)}$ *by removing the first* (*block*) *column and the first* (*block*) *row.*

Proof. We employ first-step analysis to get

$$\varphi^{(K)}_{(i,j,k)}(s) = \sum_{k'\neq k} \frac{d^0_{kk'}}{\lambda_k + j\mu + i\theta + s}\varphi^{(K)}_{(i,j,k')}(s) + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + j\mu + i\theta + s}\varphi^{(K)}_{(i,j+1,k')}(s)$$

$$+ \frac{j\mu}{\lambda_k + j\mu + i\theta + s}\varphi^{(K)}_{(i,j-1,k)}(s) + \frac{i\theta}{\lambda_k + j\mu + i\theta + s}\varphi^{(K)}_{(i-1,j+1,k)}(s),$$

$$0 \le i \le K,\ 0 \le j \le c-1,\ 1 \le k \le m,\ (i,j) \neq (0,0), \tag{3}$$

$$\varphi^{(K)}_{(i,c,k)}(s) = \sum_{k'\neq k} \frac{d^0_{kk'}}{\lambda_k + c\mu + s}\varphi^{(K)}_{(i,c,k')}(s) + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + c\mu + s}\varphi^{(K)}_{(i+1,c,k')}(s)$$

$$+ \frac{c\mu}{\lambda_k + c\mu + s}\varphi^{(K)}_{(i,c-1,k)}(s), \quad 0 \le i \le K-1,\ 1 \le k \le m, \tag{4}$$

$$\varphi^{(K)}_{(K,c,k)}(s) = \sum_{k'\neq k} \frac{q^*_{kk'}}{q^*_k + c\mu + s}\varphi^{(K)}_{(K,c,k')}(s) + \frac{c\mu}{q^*_k + c\mu + s}\varphi^{(K)}_{(K,c-1,k)}(s), \tag{5}$$

where $q^*_k = \lambda_k(1 - p_{kk}(1))$, for $1 \le k \le m$, and $q^*_{kk'} = \lambda_k(p_{kk'}(0) + p_{kk'}(1))$, for $k \neq k'$.

By expressing Equations (3)–(5) in matrix form we obtain the expression (2). $\square$

Since the busy period starts by visting a state of the sub-level $(\underline{\mathbf{0},\mathbf{1}}) = \{(0,1,k);\ 1 \le k \le m\}$, we next consider the unconditional version with Laplace–Stieljes transform defined as

$$\Psi^{(K)}(s) = \boldsymbol{\pi}(K)\boldsymbol{\varphi}^{(K)}(s), \tag{6}$$

where $\boldsymbol{\pi}(K)$ is the row vector of dimension $cm + K(c+1)m$ given by

$$\boldsymbol{\pi}(K) = \frac{1}{\mathbf{x}^* D_1 \mathbf{e}(m)}(\mathbf{x}^* D_1, \mathbf{0}).$$

Let $f_L(x)$ denote the unconditional density associated with $\Psi^{(K)}(s)$. Its value at point $x = 0$ follows from the Tauberian result: $f_L(0) = \lim_{s\to\infty} s\Psi^{(K)}(s)$. Since

$$\lim_{s\to\infty} s\varphi^{(K)}_{(i,j,k)}(s) = \begin{cases} \mu, & \text{if } (i,j,k) = (0,1,k), \\ 0, & \text{otherwise,} \end{cases}$$

we find that $f_L(0) = \mu$.

We now turn our attention to the nth moment of $T^{(K)}_{(i,j,k)}$ which is denoted by $m^{(K)}_{(i,j,k)}(n) = E[(T^{(K)}_{(i,j,k)})^n]$, for $n \ge 0$. With the help of Leibnitz's formula for the derivative of a product, we differentiate the expression given in (2) to get

$$T^{(K)}_L(s)\frac{d^n}{ds^n}\boldsymbol{\varphi}^{(K)}(s) - n\frac{d^{n-1}}{ds^{n-1}}\boldsymbol{\varphi}^{(K)}(s) = \mathbf{0}.$$

Let $\mathbf{m}^{(K)}(n)$ denote the vector containing the moments partitioned in accordance with the orbit levels, i.e., we have

$$\mathbf{m}_0^{(K)}(n) = \big(m_{(0,1,1)}^{(K)}(n), \ldots, m_{(0,c,m)}^{(K)}(n)\big)', \quad n \geq 0,$$
$$\mathbf{m}_i^{(K)}(n) = \big(m_{(i,0,1)}^{(K)}(n), \ldots, m_{(i,c,m)}^{(K)}(n)\big)', \quad 1 \leq i \leq K, \ \ n \geq 0,$$
$$\mathbf{m}^{(K)}(n) = \big(\mathbf{m}_0^{(K)}(n), \ldots, \mathbf{m}_K^{(K)}(n)\big)', \quad n \geq 0.$$

Since $\mathbf{m}^{(K)}(n) = (-1)^n \frac{d^n}{ds^n}\boldsymbol{\varphi}^{(K)}(s)|_{s=0}$ and $T_L^{(K)}(0) = \overline{Q}^{(K)}$ we notice that

$$\begin{aligned} \overline{Q}^{(K)}\mathbf{m}^{(K)}(n) &= -n\mathbf{m}^{(K)}(n-1), \quad n \geq 1, \\ \mathbf{m}^{(K)}(0) &= \mathbf{e}(cm + K(c+1)m). \end{aligned} \tag{7}$$

We also notice that the unconditional moments are given by

$$E[L^n] = \boldsymbol{\pi}(K)\mathbf{m}^{(K)}(n), \quad n \geq 0.$$

We next present numerical results involving the first two moments and the numerical inversion of the unconditional transform $\Psi^{(K)}(s)$. In the following examples, we fix $c = 5$ and $\lambda = 1.0$. In addition, we denote the traffic intensity by $\rho = \lambda/c\mu = (5\mu)^{-1}$. We increase successively the truncation level until the first four decimal digits of two successive values of $E[L]$ match. Table 1 summarizes the resulting values of K, for several choices of the traffic intensity, ρ, and the retrial rate, θ. Each block gives the truncation levels corresponding, from left to right, to the arrival processes *ERL*, *EXP*, *HEX*, and *MMPP* (see the appendix).

In Table 2, we consider the three renewal inputs and display the mean, $E[L]$, and the standard deviation, $\sigma(L)$, for different values of ρ and θ. As is to be expected, both the measures are increasing functions of ρ and decreasing functions of θ.

Using Euler and Post-Widder algorithms, we can numerically invert the expression given in (6), and obtain the density function $f_L(x)$. In Figures 1 and 2 we illustrate the influence of ρ and θ for the arrival

Table 1. Truncation levels K associated with $E[L]$

	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$
$\theta = 0.05$	3, 7, 14, 33	11, 25, 59, 75	54, 99, 207, 209
$\theta = 0.5$	1, 4, 9, 18	8, 15, 40, 49	30, 49, 138, 145
$\theta = 1.0$	2, 5, 10, 19	8, 15, 37, 49	31, 50, 121, 130
$\theta = 2.5$	2, 4, 9, 21	9, 15, 34, 46	28, 47, 140, 121
$\theta = 5.0$	1, 5, 9, 19	7, 14, 32, 48	28, 44, 125, 129

Table 2. Moments of the unconditional busy period L

		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$	
		$E[L]$	$\sigma(L)$	$E[L]$	$\sigma(L)$	$E[L]$	$\sigma(L)$
$\theta = 0.05$	*ERL*	2.28930	3.15885	20.84490	40.33858	2334.59355	3817.40231
	EXP	2.94830	6.24091	41.75191	95.75959	3794.35110	6417.54554
	HEX	7.33213	20.21505	155.75531	313.02446	2588.76030	4195.64390
$\theta = 0.5$	*ERL*	2.27990	3.06292	14.02606	20.31492	115.37530	165.90911
	EXP	2.52350	3.15530	12.71971	17.77823	76.19950	113.16275
	HEX	3.50993	4.43020	15.69120	20.54201	57.98240	89.40936
$\theta = 1.0$	*ERL*	2.27950	3.06133	13.76183	19.75304	98.44500	138.96143
	EXP	2.50500	3.09491	12.01540	16.31635	62.66100	92.27115
	HEX	3.37880	4.08798	14.07320	17.60480	49.04640	74.05324
$\theta = 2.5$	*ERL*	2.27930	3.06061	13.62571	19.47443	89.81731	125.35888
	EXP	2.46622	3.07062	11.67232	15.63480	56.24280	79.48115
	HEX	3.31692	3.94726	13.32370	16.27244	44.83870	66.67907
$\theta = 5.0$	*ERL*	2.27924	3.06040	13.58560	19.39336	87.15331	121.17625
	EXP	2.49410	3.06522	11.57512	15.44442	54.35280	76.30562
	HEX	3.30102	3.91341	13.10752	15.88795	43.56590	64.41793

process *MMPP*. First, in Figure 1, we fix $\theta = 1.0$ and display three curves corresponding to $\rho = 0.25, 0.5$, and 0.75. We notice that $f_L(0) = \mu = (5\rho)^{-1}$, in agreement with the Tauberian result. All curves exhibit decreasing shapes with heavier tails for higher values of ρ.

In Figure 2, we plot the density $f_L(x)$ for $\rho = 0.5$ and $\theta = 0.05, 1.0$, and 5.0. We notice that the three densities are graphically indistinguishable in the displayed domain. However, when θ decreases the tail of the distribution becomes heavier. In fact, $E[L] = 124.42$, for the model with $\theta = 0.05$, whereas $E[L] = 9.95$ when $\theta = 5.0$.

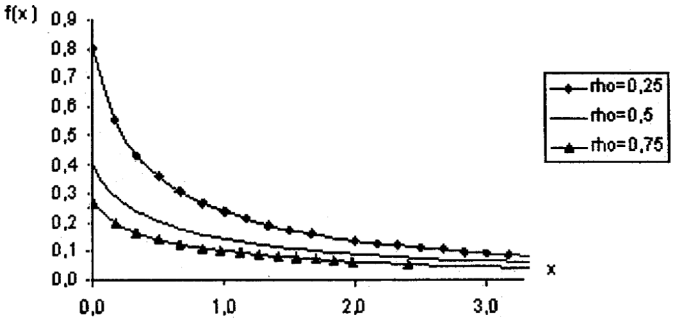

Figure 1. The density $f_L(x)$ versus ρ.

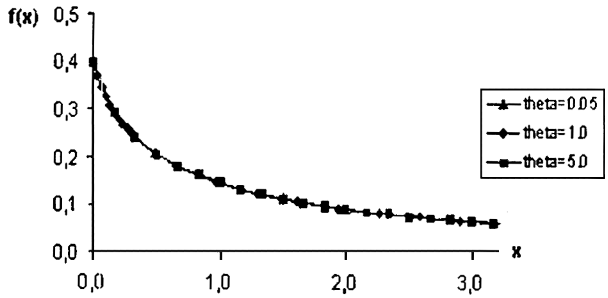

Figure 2. The density $f_L(x)$ versus θ.

4. THE NUMBER OF CUSTOMERS SERVED

In this section we study the number of customers served during a busy period. The study of this descriptor complements the busy period analysis providing a discrete counterpart of the length of the busy period. The methodology is similar to that employed in Section 3, and hence in the sequel we omit the repetitive details.

We next introduce some definitions and notations. We have

$$\begin{aligned}
N_{(i,j,k)}^{(K)} &= \text{the number of customers served during } T_{(i,j,k)}^{(K)},\\
\phi_{(i,j,k)}^{(K)}(z) &= E\big[z^{N_{(i,j,k)}^{(K)}}\big], \quad |z| \le 1, \ \ (i,j,k) \in S,\\
\mathbf{\Phi}_0^{(K)}(z) &= \big(\phi_{(0,1,1)}^{(K)}(z), \ldots, \phi_{(0,c,m)}^{(K)}(z)\big)',\\
\mathbf{\Phi}_i^{(K)}(z) &= \big(\phi_{(i,0,1)}^{(K)}(z), \ldots, \phi_{(i,c,m)}^{(K)}(z)\big)', \quad 1 \le i \le K,\\
\mathbf{\Phi}^{(K)}(z) &= \big(\mathbf{\Phi}_0^{(K)}(z), \ldots, \mathbf{\Phi}_K^{(K)}(z)\big)',\\
\tilde{m}_{(i,j,k)}^{(K)}(n) &= E\big[N_{(i,j,k)}^{(K)} \ldots (N_{(i,j,k)}^{(K)} - n + 1)\big], \ \ n \ge 1, \quad \text{and} \quad \tilde{m}_{(i,j,k)}^{(K)}(0) = 1,\\
\tilde{\mathbf{m}}_0^{(K)}(n) &= \big(\tilde{m}_{(0,1,1)}^{(K)}(n), \ldots, \tilde{m}_{(0,c,m)}^{(K)}(n)\big)', \quad n \ge 0,\\
\tilde{\mathbf{m}}_i^{(K)}(n) &= \big(\tilde{m}_{(i,0,1)}^{(K)}(n), \ldots, \tilde{m}_{(i,c,m)}^{(K)}(n)\big)', \quad 1 \le i \le K, \ \ n \ge 0,\\
\tilde{\mathbf{m}}^{(K)}(n) &= \big(\tilde{\mathbf{m}}_0^{(K)}(n), \ldots, \tilde{\mathbf{m}}_K^{(K)}(n)\big)', \quad n \ge 0.
\end{aligned}$$

Once again we may use the first-step analysis to get the system of equations governing the dynamic of the generating functions $\phi_{(i,j,k)}^{(K)}(z)$. For states of the level $\underline{\mathbf{0}}^*$, we have $\phi_{(0,0,k)}^{(K)}(z) = 1$, for $1 \le k \le m$. The resulting system has the form (3)–(5) for $s = 0$, but the service rates $j\mu$ are replaced by $j\mu z$. The corresponding result is summarized in the following.

Theorem 2. *The generating functions* $\{\phi^{(K)}_{(i,j,k)}(z); (i,j,k) \in S\}$ *satisfy the following block tridiagonal system*

$$T_N^{(K)}(z)\mathbf{\Phi}^{(K)}(z) = \mathbf{f}^{(K)}(z), \tag{8}$$

where $T_N^{(K)}(z) = \overline{Q}^{(K)} - A^{(K)}(z)$, $\mathbf{f}^{(K)}(z) = z\mathbf{f}^{(K)}$ *and*

$$A^{(K)}(z) = (1-z)\mu \begin{pmatrix} W_0 & 0 \\ 0 & I_K \otimes W \end{pmatrix} \otimes I_m,$$

$$W_0 = \begin{pmatrix} 0 & & & \\ 2 & \ddots & & \\ & \ddots & \ddots & \\ & & c & 0 \end{pmatrix}, W = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & c & 0 \end{pmatrix}.$$

By differentiating the expresion in (8) in Theorem 2, we find that

$$T_N^{(K)}(z)\frac{d^n}{dz^n}\mathbf{\Phi}^{(K)}(z) + nA^{(K)}(0)\frac{d^{n-1}}{dz^{n-1}}\mathbf{\Phi}^{(K)}(z) = \delta_{n1}\mathbf{f}^{(K)},$$

where δ_{n1} denotes Kronecker's function.

Noting that $\tilde{\mathbf{m}}^{(K)}(n) = \frac{d^n}{dz^n}\mathbf{\Phi}^{(K)}(z)|_{z=1}$ and $T_N^{(K)}(1) = \overline{Q}^{(K)}$, we get an appropriate formula for computing any arbitrary vector $\tilde{\mathbf{m}}^{(K)}(n)$ in terms of the vector containing moments of one less order:

$$\begin{aligned} \overline{Q}^{(K)}\tilde{\mathbf{m}}^{(K)}(n) &= \delta_{n1}\mathbf{f}^{(K)} - nA^{(K)}(0)\tilde{\mathbf{m}}^{(K)}(n-1), \quad n \geq 1, \\ \tilde{\mathbf{m}}^{(K)}(0) &= \mathbf{e}(cm + K(c+1)m). \end{aligned} \tag{9}$$

We also observe that the unconditional factorial moment is given by

$$E[N\ldots(N-n+1)] = \boldsymbol{\pi}(K)\tilde{\mathbf{m}}^{(K)}(n), \quad n \geq 0.$$

Our numerical experience indicates the two measures, the mean and standard deviation of the number of customers served during a busy period, behave very similar to the ones reported in Table 2 for L, i.e., both descriptors decrease with increasing retrial rates and increase with increasing values of ρ.

Because the spatial heterogeneity caused by the retrial rates, it seems impossible to deal with the model with infinite orbit capacity (i.e., the case $K = \infty$) and solve the system (8), for the generating functions $\phi^{(K)}_{(i,j,k)}(z)$, or the system (9), for the moments $\tilde{m}^{(K)}_{(i,j,k)}(n)$. The same occurs for the corresponding systems (2) and (7) in Section 3. In contrast, we next show how the probability mass function of the number of customers served, given any initial state (i,j,k), can be recursively computed without truncating the orbit capacity.

Let $N_{(i,j,k)}$ be the number of customers served during a busy period of the $MAP/M/c$ retrial queue, i.e., $N_{(i,j,k)} = N^{(\infty)}_{(i,j,k)}$. For $(i, j) \neq (0, 0)$, we define $x^{(n)}_{(i,j,k)} = P\{N_{(i,j,k)} = n\}$, $n \geq 1$, $i \geq 0$, $0 \leq j \leq c$, $1 \leq k \leq m$. The corresponding partition gives

$$\mathbf{x}^n_0 = \left(x^n_{(0,1,1)}, \ldots, x^n_{(0,c,m)}\right)',$$
$$\mathbf{x}^n_i = \left(x^n_{(i,0,1)}, \ldots, x^n_{(i,c,m)}\right)',$$
$$\mathbf{x}^n = \left(\mathbf{x}^n_0, \mathbf{x}^n_1, \ldots\right)'.$$

We note that the definition can be extended to cover the boundary cases $n = 0$ and $(i, j) = (0, 0)$ as follows

$$x^0_{(i,j,k)} = \begin{cases} 1, & \text{if } (i, j, k) \in \underline{\mathbf{0}}^*, \\ 0, & \text{otherwise,} \end{cases}$$
$$x^n_{(0,0,k)} = 0, \quad n \geq 1.$$

A first-step argument yields

$$\begin{aligned} x^n_{(i,j,k)} = {} & \sum_{k' \neq k} \frac{d^0_{kk'}}{\lambda_k + j\mu + i\theta} x^n_{(i,j,k')} + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + j\mu + i\theta} x^n_{(i,j+1,k')} \\ & + \frac{j\mu}{\lambda_k + j\mu + i\theta} x^{n-1}_{(i,j-1,k)} + \frac{i\theta}{\lambda_k + j\mu + i\theta} x^n_{(i-1,j+1,k)}, \\ & n \geq 1, \ \ i \geq 0, \ \ 0 \leq j \leq c-1, \ \ 1 \leq k \leq m, \ \ (i, j) \neq (0, 0), \end{aligned} \tag{10}$$

$$\begin{aligned} x^n_{(i,c,k)} = {} & \sum_{k' \neq k} \frac{d^0_{kk'}}{\lambda_k + c\mu} x^n_{(i,c,k')} + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + c\mu} x^n_{(i+1,c,k')} \\ & + \frac{c\mu}{\lambda_k + c\mu} x^{n-1}_{(i,c-1,k)}, \quad n \geq 1, \ \ i \geq 0, \ \ 1 \leq k \leq m. \end{aligned} \tag{11}$$

For every fixed $n \geq 1$, we observe that $x^n_{(i,j,k)} = 0$, for $i + j > n$. As a result, the systems given in (10) and (11) involve only a finite number of unknowns corresponding to the orbit levels $i = 0, \ldots, n$. The matrix formulation of Equations (10) and (11), for $0 \leq i \leq n$, leads to the following block tridiagonal system:

$$(\widetilde{Q}^{(n)} - A^{(n)}(0))\mathbf{x}^n(n) = \delta_{n1}\mathbf{f}^{(n)} - A^{(n)}(0)\mathbf{x}^{n-1}(n), \tag{12}$$

where $\mathbf{x}^n(n) = (\mathbf{x}^n_0, \ldots, \mathbf{x}^n_n)'$ and $\mathbf{x}^{n-1}(n) = (\mathbf{x}^{n-1}_0, \ldots, \mathbf{x}^{n-1}_{n-1}, \mathbf{0})'$, and $\widetilde{Q}^{(n)}$ is the square matrix of order $cm + n(c+1)m$ obtained from $\overline{Q}^{(n)}$ by replacing the $(n+1, n+1)$th block entry by A_{1n}.

If we take into account the distribution of the first state visited when the busy period starts, we get the unconditional distribution given by

$$P\{N = n\} = \frac{1}{\mathbf{x}^* D_1 \mathbf{e}(m)} \mathbf{x}^* D_1 \mathbf{x}^n_{(0,1)}, \quad n \geq 1, \tag{13}$$

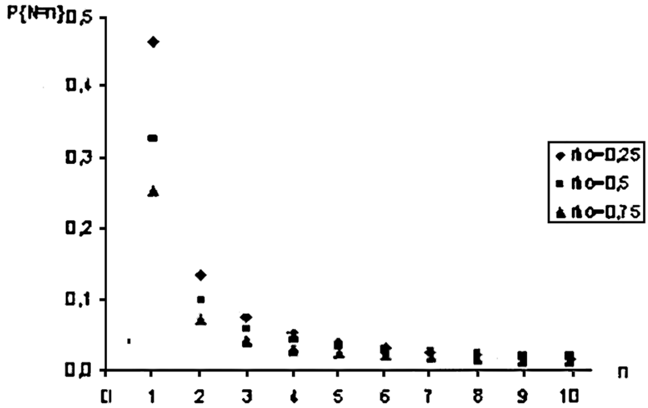

Figure 3. Probability mass function of N versus ρ.

where $\mathbf{x}^n_{(0,1)} = \left(x^n_{(0,1,1)}, \ldots, x^n_{(0,1,m)}\right)'$. Thus, an initial truncation is needed to determine the distribution of the first state visited. Then, the recursive use of formulas (12) and (13) gives the desired unconditional distribution of the number of customers served.

Table 3. Probability mass function of N, arrivals *EXP* and *MMPP*

		$\theta = 0.05$	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
$n = 1$	*EXP*	0.28571	0.28571	0.28571	0.28571
	MMPP	0.32720	0.32720	0.32720	0.32720
$n = 2$	*EXP*	0.09070	0.09070	0.09070	0.09070
	MMPP	0.09908	0.09908	0.09908	0.09908
$n = 3$	*EXP*	0.05628	0.05628	0.05628	0.05628
	MMPP	0.05958	0.05958	0.05958	0.05958
$n = 4$	*EXP*	0.04260	0.04260	0.04260	0.04260
	MMPP	0.04401	0.04401	0.04401	0.04401
$n = 5$	*EXP*	0.03533	0.03533	0.03533	0.03533
	MMPP	0.03575	0.03575	0.03575	0.03575
$n = 6$	*EXP*	0.03034	0.03055	0.03063	0.03072
	MMPP	0.02874	0.02989	0.03020	0.03046
$n = 7$	*EXP*	0.02642	0.02703	0.02724	0.02745
	MMPP	0.02277	0.02547	0.02615	0.02668
$n = 8$	*EXP*	0.02315	0.02427	0.02463	0.02495
	MMPP	0.01801	0.02212	0.02306	0.02377
$n = 9$	*EXP*	0.02036	0.02204	0.02251	0.02292
	MMPP	0.01432	0.01953	0.02064	0.02143
$n = 10$	*EXP*	0.01797	0.02018	0.02076	0.02123
	MMPP	0.01149	0.01749	0.01867	0.01949

In Figure 3 we fix $\theta = 1.0$ and $c = 5$. Then, we analyze the effect of the traffic intensity in the model with *MMPP* arrivals. We notice that the lowest value at the point $n = 1$ corresponds to the case $\rho = 0.75$ and, consequently, the heaviest tail also is associated with this traffic intensity.

Finally, in Table 3 we compare the probability mass function of a model with renewal input of type *EXP* versus the nonrenewal arrival process described by *MMPP*. To this end, we fix $\rho = 0.5$ and display $P\{N = n\}$, for $1 \leq n \leq 10$. Since the repeated attempts occur only when all servers are busy, the probabilities $P\{N = n\}$, for $1 \leq n \leq 5$, do not depend on the retrial rate. For all choices of θ, we observe that the queue with *MMPP* has a larger mass at the origin, for $1 \leq n \leq 5$, whereas the model with *EXP* arrivals exhibits a heavier tail.

5. THE WAITING TIME

In this section we turn our attention to the waiting time which is defined as the sojourn time of a tagged customer in the retrial orbit. In retrial queues it is typically assumed that customers in the retrial orbit behave independently of each other. It means that the retrial group operates under a random order policy. This assumption makes the analysis difficult because we need to consider not only the system state at the arrival time of the tagged customer, but also the possibility that the customers arriving at later time will compete for free servers.

Define

$W^{(K)}_{(i,j,k)}$ = the residual waiting time of the tagged customer given that the system state is (i, j, k),

$W^{(K)}_{(i,j,k)}(s) = E[\exp\{-sW^{(K)}_{(i,j,k)}\}]$, $\mathrm{Re}(s) \geq 0$, for any $(i, j, k) \in S$ with $i > 0$.

We partition the above Laplace–Stieltjes transforms according to the orbit levels in a similar manner as follows:

$$\mathbf{W}^{(K)}_i(s) = \left(\mathbf{W}^{(K)}_{(i,0,1)}(s), \ldots, W^{(K)}_{(i,c,m)}(s)\right)', \quad 1 \leq i \leq K,$$

$$\mathbf{W}^{(K)}(s) = \left(\mathbf{W}^{(K)}_1(s), \ldots, \mathbf{W}^{(K)}_K(s)\right)'.$$

The following theorem gives a system of linear equations for the Laplace–Stieltjes transforms $W^{(K)}_{(i,j,k)}(s)$.

Theorem 3. *The Laplace–Stieltjes transforms* $\{W^{(K)}_{(i,j,k)}(s); (i, j, k) \in S, i > 0\}$ *satisfy the following block tridiagonal system*

$$T^{(K)}_W(s)\mathbf{W}^{(K)}(s) = \mathbf{g}^{(K)}, \tag{14}$$

where $T_W^{(K)}(s) = \widehat{Q}^{(K)} - sI_{K(c+1)m}$ *and* $\mathbf{g}^{(K)} = -\theta\mathbf{e}(K) \otimes (\mathbf{e}(c+1) - \mathbf{e}_{c+1}(c+1)) \otimes \mathbf{e}(m)$. *The matrix* $\widehat{Q}^{(K)}$ *is of the form*

$$\widehat{Q}^{(K)} = \begin{pmatrix} A_{11} & A_0 & & & & \\ \widehat{A}_{22} & A_{12} & A_0 & & & \\ \ddots & & \ddots & & \ddots & \\ & & & \widehat{A}_{2,K-1} & A_{1,K-1} & A_0 \\ & & & & \widehat{A}_{2K} & A_{1K}^* \end{pmatrix},$$

and $\widehat{A}_{2i} = A_{2i}(i-1)/i$, *for* $2 \le i \le K$.

Proof. Using again the first principles, we find that

$$\begin{aligned} W_{(i,j,k)}^{(K)}(s) = {} & \sum_{k' \neq k} \frac{d_{kk'}^0}{\lambda_k + j\mu + i\theta + s} W_{(i,j,k')}^{(K)}(s) + \sum_{k'} \frac{d_{kk'}^1}{\lambda_k + j\mu + i\theta + s} W_{(i,j+1,k')}^{(K)}(s) \\ & + \frac{j\mu}{\lambda_k + j\mu + i\theta + s} W_{(i,j-1,k)}^{(K)}(s) + \frac{(i-1)\theta}{\lambda_k + j\mu + i\theta + s} W_{(i-1,j+1,k)}^{(K)}(s) \\ & + \frac{\theta}{\lambda_k + j\mu + i\theta + s}, \quad 1 \le i \le K, \ 0 \le j \le c-1, \ 1 \le k \le m, \end{aligned} \tag{15}$$

$$\begin{aligned} W_{(i,c,k)}^{(K)}(s) = {} & \sum_{k' \neq k} \frac{d_{kk'}^0}{\lambda_k + c\mu + s} W_{(i,c,k')}^{(K)}(s) + \sum_{k'} \frac{d_{kk'}^1}{\lambda_k + c\mu + s} W_{(i+1,c,k')}^{(K)}(s) \\ & + \frac{c\mu}{\lambda_k + c\mu + s} W_{(i,c-1,k)}^{(K)}(s), \quad 1 \le i \le K-1, \ 1 \le k \le m, \end{aligned} \tag{16}$$

$$W_{(K,c,k)}^{(K)}(s) = \sum_{k' \neq k} \frac{q_{kk'}^*}{q_k^* + c\mu + s} W_{(K,c,k')}^{(K)}(s) + \frac{c\mu}{q_k^* + c\mu + s} W_{(K,c-1,k)}^{(K)}(s). \tag{17}$$

The contribution due to repeated attempts is explained as follows. The last term on the right-hand side of formula (15) is associated with the case where the first event corresponds to an attempt for service made by the tagged customer. In contrast, if another retrial customer applies for service, we obtain the term $(i-1)\theta(\lambda_k + j\mu + i\theta + s)^{-1} W_{(i-1,j+1,k)}^{(K)}(s)$. Now, after routine (block) identification, we may express the system in (15)–(17) as given in (14). $\square$

The marked customer must wait in orbit if upon arrival he finds the system at any state in the subset $S_W = \{(i, c, k); 0 \le i \le K-1, 1 \le k \le m\}$. Thus, we define the unconditional version of the waiting time, W, as follows

$$\Omega^{(K)}(s) = 1 - \gamma(K)\mathbf{e}(K(c+1)m) + \gamma(K)\mathbf{W}^{(K)}(s), \tag{18}$$

where $\boldsymbol{\gamma}(K)$ is the row vector of dimension $K(c+1)m$ given by

$$\boldsymbol{\gamma}(K) = \eta^{-1}(\mathbf{0}, \mathbf{x}(0, c)D_1, \ldots, \mathbf{0}, \mathbf{x}(K-1, c)D_1), \tag{19}$$

and $\mathbf{x}(i, c) = (x_{(i,c,1)}, \ldots, x_{(i,c,m)})$, for $0 \leq i \leq K-1$, i.e., $\mathbf{x}(i, c)$ is the sub-vector containing the stationary probabilities of the sub-level $(\underline{\mathbf{i}, \mathbf{c}}) = \{(i, c, k);\ 1 \leq k \leq m\}$, and η is the normalizing constant given by

$$\eta = \mathbf{x}^* D_1 \mathbf{e}(m) + \mathbf{x}(0)(\mathbf{e}(c) \otimes D_1 \mathbf{e}(m)) + \sum_{i=1}^{K} \mathbf{x}(i)(\mathbf{e}(c+1) \otimes D_1 \mathbf{e}(m)).$$

We notice that formula (18) includes two contributions: a) $P\{W = 0\} = 1 - \boldsymbol{\gamma}(K)\mathbf{e}(K(c+1)m)$ representing the probability of no-waiting, which occur either when the tagged customer finds a free server or when he sees the sub-level $(\underline{\mathbf{K}, \mathbf{c}})$ and becomes a lost customer, and b) the transform of the continuous contribution with density $f_{W^c}(x)$ on $(0, \infty)$.

We notice that

$$\lim_{s \to \infty} s W^{(K)}_{(i,j,k)}(s) = \begin{cases} \theta, & \text{if } 0 \leq j \leq c-1, \\ 0, & \text{if } j = c. \end{cases} \tag{20}$$

Combining (19) and (20), we have

$$f_{W^c}(0) = \lim_{s \to \infty} s \boldsymbol{\gamma}(K) \mathbf{W}^{(K)}(s) = 0.$$

We also define the nth moment of $W^{(K)}_{(i,j,k)}$ which is denoted by $\bar{m}^{(K)}_{(i,j,k)}(n) = E\big[\big(W^{(K)}_{(i,j,k)}\big)^n\big]$, for $n \geq 0$. Now we introduce some notation:

$$\bar{\mathbf{m}}^{(K)}_i(n) = \big(\bar{m}^{(K)}_{(i,0,1)}, \ldots, \bar{m}^{(K)}_{(i,c,m)}(n)\big)', \quad 1 \leq i \leq K, \ \ n \geq 0,$$
$$\bar{\mathbf{m}}^{(K)}(n) = \big(\bar{\mathbf{m}}^{(K)}_{\mathbf{1}}(n), \ldots, \bar{\mathbf{m}}^{(K)}_K(n)\big)', \quad n \geq 0.$$

By differentiating the expression in (14), we get

$$T^{(K)}_W(s) \frac{d^n}{ds^n} \mathbf{W}^{(K)}(s) - n \frac{d^{n-1}}{ds^{n-1}} \mathbf{W}^{(K)}(s) = \mathbf{0}.$$

Since $\bar{\mathbf{m}}^{(K)}(n) = (-1)^n \frac{d^n}{ds^n} \mathbf{W}^{(K)}(s)|_{s=0}$, we obtain

$$\widehat{Q}^{(K)} \bar{\mathbf{m}}^{(K)}(n) = -n \bar{\mathbf{m}}^{(K)}(n-1), \quad n \geq 1,$$
$$\bar{\mathbf{m}}^{(K)}(0) = \mathbf{e}(K(c+1)m).$$

Finally, the moments of the unconditional version satisfy that

$$E[W^n] = \boldsymbol{\gamma}(K) \bar{\mathbf{m}}^{(K)}(n), \quad n \geq 0.$$

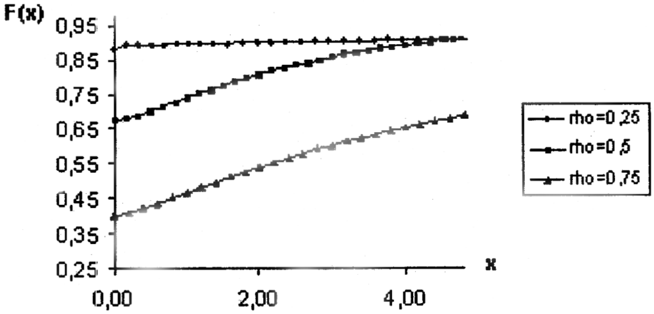

Figure 4. The distribution $F_{W^c}(x)$ versus ρ.

Our numerical experience indicates that $E[W]$ and $\sigma(W)$ increase with increasing values of ρ and, in contrast, they decrease with increasing values of θ.

We next present numerical results regarding the inversion of the distribution function $F_{W^c}(x)$. Once again the displayed curves correspond to those orbit levels that guarantee at least four decimal places of $E[W]$ corresponding to two successive orbit levels are matched.

In Figure 4, we apply the numerical inversion algorithms to get the distribution function for the case of *MMPP* arrivals by fixing $\lambda = 1.0$, $c = 5$, and $\theta = 1$, and varying $\rho = 0.25, 0.5$, and 0.75. The jump at the point $x = 0$ equals the probability $P\{W = 0\}$, and it becomes higher as long as ρ takes smaller values. Obviously, for $\rho = 0.75$ we observe that the distribution function exhibits a heavier tail.

The effect of the retrial rate is illustrated in Figure 5. We keep the same arrival process and $c = 5$. Then, we fix $\rho = 0.5$ and plot the distribution, $F_{W^c}(x)$ for $\theta = 0.05, 1.0$, and 5.0. It should be noticed that when θ decreases the distribution becomes sparser. Moreover, when θ increases, an arriving customer will have more competition (from the retrial customers) to occupy a free server. As a result, $P\{W = 0\}$ decreases when θ increases.

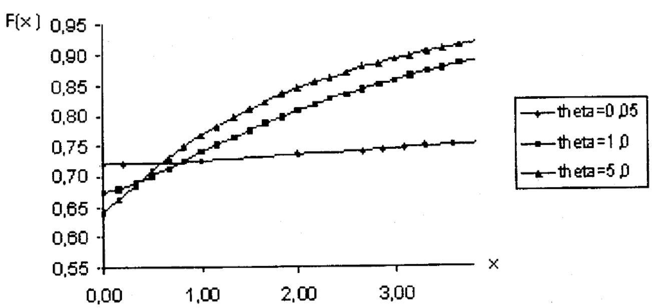

Figure 5. The distribution $F_{W^c}(x)$ versus θ.

6. THE NUMBER OF RETRIALS MADE BY A CUSTOMER

In this section we deal with R, the number of repeated attempts made by a tagged customer until he reaches a free server. This descriptor provides a discrete counterpart of the waiting time W studied in Section 5.

First of all, we define

$R^{(K)}_{(i,j,k)}$ = the number of retrials that a tagged customer will make, given that the system state is (i, j, k),

$R^{(K)}_{(i,j,k)}(z) = E[z^{R^{(K)}_{(i,j,k)}}]$, $|z| \leq 1$, for any $(i, j, k) \in S$ with $i > 0$.

The partition according to the orbit levels gives:

$$\mathbf{R}^{(K)}_i(z) = \left(R^{(K)}_{(i,0,1)}(z), \ldots, R^{(K)}_{(i,c,m)}(z)\right)', \quad 1 \leq i \leq K,$$

$$\mathbf{R}^{(K)}(z) = \left(\mathbf{R}^{(K)}_1(z), \ldots, \mathbf{R}^{(K)}_K(z)\right)'.$$

The following theorem establishes a relationship for the generating functions $R^{(K)}_{(i,j,k)}(z)$.

Theorem 4. *The generating functions* $\{R^{(K)}_{(i,j,k)}(z); (i, j, k) \in S, i > 0\}$ *satisfy the following block tridiagonal system*

$$T^{(K)}_R(z)\mathbf{R}^{(K)}(z) = \mathbf{g}^{(K)}(z) \tag{21}$$

where $T^{(K)}_R(z) = \widehat{Q}^{(K)} + (1 - z)\mathbf{h}^{(K)}$, $\mathbf{h}^{(K)} = -\theta I_K \otimes (\mathbf{e}_{c+1}(c + 1)\mathbf{e}'_{c+1}(c + 1))$ $\otimes I_m$ *and* $\mathbf{g}^{(K)}(z) = z\mathbf{g}^{(K)}$.

Proof. From the first principles we see

$$\begin{aligned} R^{(K)}_{(i,j,k)}(z) = {} & \sum_{k' \neq k} \frac{d^0_{kk'}}{\lambda_k + j\mu + i\theta} R^{(K)}_{(i,j,k')}(z) + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + j\mu + i\theta} R^{(K)}_{(i,j+1,k')}(z) \\ & + \frac{j\mu}{\lambda_k + j\mu + i\theta} R^{(K)}_{(i,j-1,k)}(z) + \frac{(i-1)\theta}{\lambda_k + j\mu + i\theta} R^{(K)}_{(i-1,j+1,k)}(z) \\ & + \frac{z\theta}{\lambda_k + j\mu + i\theta}, \ 1 \leq i \leq K, 0 \leq j \leq c-1, 1 \leq k \leq m, \end{aligned} \tag{22}$$

$$\begin{aligned} R^{(K)}_{(i,c,k)}(z) = {} & \sum_{k' \neq k} \frac{d^0_{kk'}}{\lambda_k + c\mu + \theta} R^{(K)}_{(i,c,k')}(z) + \sum_{k'} \frac{d^1_{kk'}}{\lambda_k + c\mu + \theta} R^{(K)}_{(i+1,c,k')}(z) \\ & + \frac{c\mu}{\lambda_k + c\mu + \theta} R^{(K)}_{(i,c-1,k)}(z) + \frac{z\theta}{\lambda_k + c\mu + \theta} R^{(K)}_{(i,c,k)}(z), \\ & 1 \leq i \leq K-1, \ \ 1 \leq k \leq m, \end{aligned} \tag{23}$$

$$\begin{aligned} R^{(K)}_{(K,c,k)}(z) = {} & \sum_{k' \neq k} \frac{q^*_{kk'}}{q^*_k + c\mu + \theta} R^{(K)}_{(K,c,k')}(z) + \frac{c\mu}{q^*_k + c\mu + \theta} R^{(K)}_{(K,c-1,k)}(z) \\ & + \frac{z\theta}{q^*_k + c\mu + \theta} R^{(K)}_{(K,c,k)}(z). \end{aligned} \tag{24}$$

To derive formulas (22)–(24), we avoid the consideration of vain retrial attempts made by nontagged customers which neither affect the event under study nor modify the current system state. The incidence of a retrial made by the tagged customers depends on the number of free servers. In (22) we have free servers, so the existence of a retrial means that the tagged customer gets service. On the other hand, if $j = c$ and the tagged customer retries, then the system state does not change but we do count that repeated attempt.

Comparing these with the system given in (15)–(17) for the waiting time analysis, we observe some similarities. Putting $s = 0$, formula (15) agrees with (22). Moreover, when $j = c$ we must add the contribution $-\theta(1 - z)$ at the main diagonal of the matrix of coefficients. This yields the matrix form expression (21). □

We also define the unconditional version of the number of retrials made by a customer as

$$\Lambda^{(k)}(z) = 1 - \gamma(K)\mathbf{e}(K(c+1)m) + \gamma(K)\mathbf{R}^{(k)}(z),$$

and the nth factorial moment of $R^{(K)}_{(i,j,k)}$ which is denoted by $\hat{m}^{(K)}_{(i,j,k)}(n) = E[R^{(K)}_{(i,j,k)} \dots (R^{(K)}_{(i,j,k)} - n + 1)]$, $n \geq 1$, and $\hat{m}^{(K)}_{(i,j,k)}(0) = 1$, for $n \geq 0$. We denote

$$\hat{\mathbf{m}}^{(K)}_i(n) = \left(\hat{m}^{(K)}_{(i,0,1)}(n), \dots, \hat{m}^{(K)}_{(i,c,m)}(n)\right)', \quad 1 \leq i \leq K, \;\; n \geq 0,$$
$$\hat{\mathbf{m}}^{(K)}(n) = \left(\hat{\mathbf{m}}^{(K)}_1(n), \dots, \hat{\mathbf{m}}^{(K)}_K(n)\right)', \quad n \geq 0.$$

By differentiating the expression in (21), we get

$$T^{(K)}_R(z)\frac{d^n}{dz^n}\mathbf{R}^{(K)}(z) - n\mathbf{h}^{(K)}(z)\frac{d^{n-1}}{dz^{n-1}}\mathbf{R}^{(K)}(z) = \delta_{n1}\mathbf{g}^{(K)}.$$

Since $\hat{\mathbf{m}}^{(K)}(n) = \frac{d^n}{dz^n}\mathbf{R}^{(K)}(z)|_{z=1}$, we obtain

$$\widehat{Q}^{(K)}\hat{\mathbf{m}}^{(K)}(n) = \delta_{n1}\mathbf{g}^{(K)} + n\mathbf{h}^{(K)}\hat{\mathbf{m}}^{(K)}(n-1), \quad n \geq 1,$$
$$\hat{\mathbf{m}}^{(K)}(0) = \mathbf{e}(K(c+1)m).$$

Then, the moments of the unconditional random variable satisfies

$$E[R \dots (R - n + 1)] = \gamma(K)\hat{\mathbf{m}}^{(K)}(n), \quad n \geq 0.$$

Table 4 gives the values of $E[R]$ and $\sigma(R)$ for the queueing model with $c = 5$ and for the three renewal arrival processes with $\lambda = 1$. Firstly, we determine the values of K for which the first four decimal digits of $E[R]$ agree. Table 4 shows that both the performance measures are increasing

Table 4. Moments of the number retrials R

		$\rho = 0.25$		$\rho = 0.5$		$\rho = 0.75$	
		$E[R]$	$\sigma(R)$	$E[R]$	$\sigma(R)$	$E[R]$	$\sigma(R)$
$\theta = 0.05$	*ERL*	0.00019	0.01417	0.02860	0.18411	0.34060	0.79291
	EXP	0.00776	0.09004	0.10522	0.35730	0.57930	1.03684
	HEX	0.04080	0.20782	0.30191	0.61823	1.06930	1.54204
$\theta = 0.5$	*ERL*	0.00022	0.01720	0.04435	0.29991	0.66210	1.67478
	EXP	0.00942	0.11398	0.17301	0.63143	1.25420	2.58730
	HEX	0.05433	0.28657	0.63720	1.43764	3.56840	6.17578
$\theta = 1.0$	*ERL*	0.00026	0.02041	0.05991	0.41842	0.98970	2.60948
	EXP	0.01130	0.14095	0.24391	0.92948	1.97510	4.29800
	HEX	0.06820	0.37078	0.99150	2.33362	6.28670	11.27336
$\theta = 2.5$	*ERL*	0.00035	0.02957	0.10263	0.75490	1.92423	5.35854
	EXP	0.01651	0.21829	0.44780	1.80771	4.09041	9.38993
	HEX	0.10720	0.61507	2.01590	4.98086	14.33750	26.47920
$\theta = 5.0$	*ERL*	0.00050	0.04425	0.17032	1.30115	3.44221	9.90257
	EXP	0.02490	0.34446	0.77930	3.25581	7.57670	17.84378
	HEX	0.16951	1.01375	3.68880	9.35609	27.67170	51.75798

functions of ρ and θ. It should be noticed that a rapid reattempt for service has a significant chance of being blocked, so $E[R]$ and $\sigma(R)$ increase for increasing values of θ.

The numerical inversion of the expression in (21) can be performed with the help of a Fast Fourier transform algorithm. An alternative approach may be attained introducing the probabilities

$z^r_{l,(i,j,k)}(K) =$ the probability that the tagged customer produces r retrials before entering service, given that he has accumulated l retrials and the current system state is (i, j, k), for $(i, j, k) \in S$, $i > 0$.

We note that the probability mass function of R is given by

$$P\{R = 0\} = 1 - \gamma(K)\mathbf{e}(K(c+1)m), \tag{25}$$

$$P\{R = r\} = \gamma(K)\mathbf{z}^r_0(K), \quad r \geq 1, \tag{26}$$

where $\mathbf{z}^r_l(K)$ is the column vector of dimension $K(c+l)m$ containing the unknowns $z^r_{l,(i,j,k)}(K)$.

A generalization of the arguments given by Artalejo and Lopez–Herrero [12] for the $M/M/c$ retrial queue gives

$$T^{(K)}_R \mathbf{z}^r_{r-1}(K) = \mathbf{g}^{(K)}, \quad r \geq 1, \tag{27}$$

$$T^{(K)}_R \mathbf{z}^r_l(K) = \mathbf{h}^{(K)} \mathbf{z}^r_{l+1}(K), \quad r \geq 1, \ 0 \leq l \leq r-2, \tag{28}$$

where $T^{(K)}_R = \widehat{Q}^{(K)} + \mathbf{h}^{(K)}$.

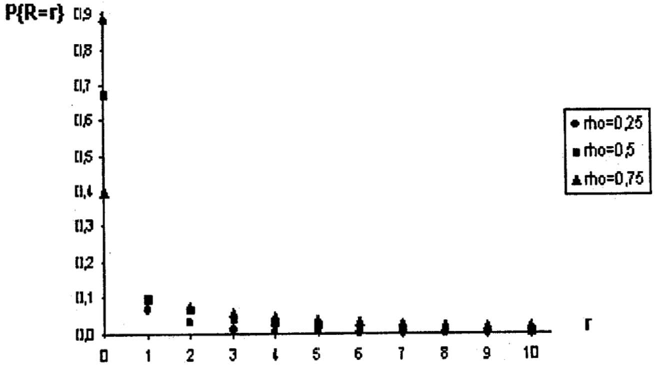

Figure 6. Probability mass function of R versus ρ.

For any fixed $r \geq 1$, Equations (27) and (28) can be recursively solved to get $\mathbf{z}_l^r(K)$, from $l = r - 1$ to $l = 0$.

We next present numerical examples on R. In Figure 6 we consider the $MMPP/M/5$ retrial queue with $\theta = 1.0$ and display $P\{R = r\}$ as a function

Table 5. Probability mass function of R, arrivals *EXP* and *MMPP*

		$\theta = 0.05$	$\theta = 0.5$	$\theta = 1.0$	$\theta = 5.0$
$r = 0$	*EXP*	0.90912	0.89748	0.89124	0.87803
	MMPP	0.72031	0.68845	0.67337	0.64173
$r = 1$	*EXP*	0.07856	0.06247	0.05151	0.02300
	MMPP	0.02240	0.13638	0.09757	0.03547
$r = 2$	*EXP*	0.01056	0.02339	0.02573	0.01786
	MMPP	0.04406	0.07364	0.06574	0.03014
$r = 3$	*EXP*	0.00148	0.00935	0.01349	0.01405
	MMPP	0.00904	0.04041	0.04457	0.02605
$r = 4$	*EXP*	0.00021	0.00395	0.00738	0.01117
	MMPP	0.00196	0.02308	0.03084	0.02279
$r = 5$	*EXP*	0.00003	0.00175	0.00418	0.00897
	MMPP	0.00044	0.01370	0.02181	0.02014
$r = 6$	*EXP*	5.0×10^{-6}	0.00080	0.00244	0.00727
	MMPP	0.00010	0.00840	0.01573	0.01792
$r = 7$	*EXP*	7.7×10^{-7}	0.00038	0.00146	0.00595
	MMPP	0.00002	0.00530	0.01154	0.01604
$r = 8$	*EXP*	1.2×10^{-7}	0.00018	0.00090	0.00490
	MMPP	6.4×10^{-6}	0.00341	0.00860	0.01443
$r = 9$	*EXP*	1.9×10^{-8}	0.00009	0.00056	0.00407
	MMPP	1.6×10^{-6}	0.00224	0.00648	0.01303
$r = 10$	*EXP*	3.1×10^{-9}	0.00004	0.00035	0.00340
	MMPP	4.3×10^{-7}	0.00150	0.00495	0.01180

of ρ. The probability $P\{R = 0\}$ decreases as ρ increases. This behavior is as expected since an increase in ρ causes more congestion and consequently more retrials. Heaviest tail corresponds to the case when $\theta = 0.75$.

Finally, in Table 5 we compare the *EXP* and *MMPP* arrival processes. We consider $c = 5$, $\rho = 0.5$, and display $P\{R = r\}$, for $0 \leq r \leq 10$. The mass functions are decreasing, but the queue with *EXP* arrivals has a larger mass at $r = 0$. However, *MMPP* arrivals have a heavier tails as compared to *EXP* arrivals.

APPENDIX

We next give a brief description of the *MAP* and introduce some notation. The *MAP* is a tractable class of Markov renewal processes that includes many well-known processes such as Poisson, Markov modulated Poisson process, and *PH*-renewal processes. For appropriate particularizations of the *MAP* parameters, the underlying arrival process becomes a renewal process. The idea of the *MAP* is to generalize the Poisson processes and still keep the tractability for stochastic modelling purposes. Since in many practical applications the arrival input do not form a renewal process, the *MAP* is a versatile tool to model both renewal and non-renewal input streams.

In this paper, we need only the *MAP* in continuous time which is described as follows. Let the underlying Markov chain be irreducible and let Q^* be the corresponding infinitesimal generator. At the end of a sojourn time in state i, that is exponentially distributed with parameter λ_i, one of the following two events could occur: i) with probability $p_{ij}(1)$ the transition corresponds to an arrival and the underlying Markov chain is in state j with $1 \leq i, j \leq m$, and ii) 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. Also, we have

$$\sum_{j=1}^{m} p_{ij}(1) + \sum_{j=1, j\neq i}^{m} p_{ij}(0) = 1, \quad 1 \leq i \leq m.$$

Define matrices $D_0 = (d_{ij}^0)$ and $D_1 = (d_{ij}^1)$ such that $d_{ii}^0 = -\lambda_i$, $1 \leq i \leq m$, $d_{ij}^0 = \lambda_i p_{ij}(0)$, for $j \neq i$ and $d_{ij}^1 = \lambda_i p_{ij}(1)$, $1 \leq i, j \leq m$. By assuming D_0 to be a non-singular matrix, the interarrival times will be finite with probability one and the arrival process does not terminate. Hence, we see that D_0 is a stable matrix. The generator Q^* is then given by $Q^* = D_0 + D_1$. Thus, D_0 governs the transitions corresponding to no arrival and D_1 governs those corresponding to an arrival.

Let $\boldsymbol{\vartheta}$ be the stationary probability vector of the Markov process with generator Q^*. That is, $\boldsymbol{\vartheta}$ is the unique positive probability vector satisfying.

$$\boldsymbol{\vartheta} Q^* = \mathbf{0}, \quad \boldsymbol{\vartheta}\mathbf{e}(m) = 1.$$

Let $\boldsymbol{\alpha}$ be the initial probability vector of the underlying Markov chain governing the *MAP*. Then, by choosing $\boldsymbol{\alpha}$ appropriately we can model the time origin to be: a) an arbitrary arrival point, b) the end of an interval during which there are at least k arrivals, and c) the point at which the system is in specific state such as the busy period ends or busy period begins. The most interesting case is the one where we get the stationary version of the *MAP* by $\boldsymbol{\alpha} = \boldsymbol{\vartheta}$. The constant $\lambda = \boldsymbol{\vartheta} D_1 \mathbf{e}(m)$, referred to as the fundamental rate, gives the expected number of arrivals per unit of 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 λ 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 Lucantoni [15] and Neuts [16], and for a review and recent work on *MAP* we refer the reader to Chakravarthy [17].

For the numerical examples along the article, we consider the following set of values for D_0 and D_1.

1. Erlang (*ERL*):

$$D_0 = \begin{pmatrix} -5 & 5 & 0 & 0 & 0 \\ 0 & -5 & 5 & 0 & 0 \\ 0 & 0 & -5 & 5 & 0 \\ 0 & 0 & 0 & -5 & 5 \\ 0 & 0 & 0 & 0 & -5 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2. Exponential (*EXP*):

$$D_0 = (-1), \quad D_1 = (1).$$

3. Hyperexponential (*HEX*):

$$D_0 = \begin{pmatrix} -1.90 & 0 \\ 0 & -0.19 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 1.710 & 0.190 \\ 0.171 & 0.019 \end{pmatrix}.$$

4. Markov Modulated Poisson Process (*MMPP*):

$$D_0 = \frac{55}{86}\begin{pmatrix} -1.3 & 0.5 & 0.3 \\ 1 & -2.5 & 0.5 \\ 2.4 & 0 & -10.4 \end{pmatrix}, \quad D_1 = \frac{55}{86}\begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

The above four *MAP* processes have arrival rate $\lambda = 1$. The first three arrival processes are renewal processes, whereas the *MMPP* has correlated arrivals. On the other hand, the ratio of the standard deviations of these four arrival processes with respect to ERL are 1.0, 2.236067, 5.019353, and 2.181479, respectively.

REFERENCES

1. Falin, G.I., and Templeton, J.G.C. 1997. *Retrial Queues*. Chapman and Hall, London.
2. Kosten, L. 1973. *Stochastic Theory of Service Systems*. International Series of Monographs in Pure and Applied Mathematics. Vol. 103, Pergamon Press, Oxford.
3. Neuts, M.F., and Rao, B.M. 1990. Numerical investigation of a multiserver retrial model. *Queueing Systems* 7:169–190.
4. Choi, B.D., and Chang, Y. 1999. MAP_1, $MAP_2/M/c$ with retrial queue with the retrial group of finite capacity and geometric loss. *Mathematical and Computer Modelling* 30:99–113.
5. Diamond, J.E., and Alfa, A.S. 1999. Matrix-analytic methods for a multi-server retrial queue with buffer. *Top* 7:249–266.
6. Breuer, L., Dudin, A.N., and Klimenok, V.I. 2002. A retrial $BMAP/PH/N$ system. *Queueing Systems* 40:433–457.
7. Chakravarthy, S.R., Krishnamoorthy, A., and Joshua, V.C. 2006. Analysis of a multi-server queue with search of customers from the orbit. *Performance Evaluation* 63:776–798.
8. Gomez-Corral, A. 2006. A bibliographical guide to the analysis of retrial queues through matrix analytic techniques. *Annals of Operations Research* 141:163–191.
9. Artalejo, J.R., Economou, A., and Lopez-Herrero, M.J. 2005. Algorithmic analysis for the number of customers served in a busy period of the $M/M/c$ retrial queue. In Proceedings of the National Conference on Mathematical and Computational Models, December 15–16, Coimbatore, India. Arumugnathan, R., and Nadarajan, R. (Eds.), Allied Publishers, New Delhi, 3–15.
10. Artalejo, J.R., and Gomez-Corral, A. 2005. Waiting time in the $M/M/c$ queue with finite retrial group. *Bulletin of Kerala Mathematics Association* 2:1–17.
11. Artalejo, J.R., Economou, A., and Lopez-Herrero, M.J. 2007. Algorithmic approximations for the busy period of the $M/M/c$ retrial queue. *European Journal of Operational Research* 176:1687–1702.
12. Artalejo, J.R., and Lopez-Herrero, M.J. 2007. On the distribution of the number of retrials. *Applied Mathematical Modelling* 31:478–489.
13. Artalejo, J.R., and Chakravarthy, S.R. 2006. Computational analysis of the maximal queue length in the $MAP/M/c$ retrial queue. *Applied Mathematics and Computation* 183:1399–1409.
14. Marcus, M., and Minc, H. 1964. *A Survey of Matrix Theory and Matrix Inequalities*. Allyn and Bacon, Boston, MA.
15. Lucantoni, D.M. 1991. New results on the single server queue with a batch Markovian arrival process. *Stochastic Models* 7:1–46.

16. Neuts, M.F. 1992. Models based on the Markovian arrival process. *IEICE Transactions on Communications* E75B:1255–1265.
17. Chakravarthy, S.R. 2001. The batch Markovian arrival process: A review and future work. In *Advances in Probability Theory and Stochastic Processes*. Krishnamoorthy, A., et al. (Eds.), Notable Publications Inc., New Jersey, 21–39.