

ALGORITHMIC ANALYSIS OF THE MAXIMUM LEVEL LENGTH IN GENERAL-BLOCK TWO-DIMENSIONAL MARKOV PROCESSES

JESUS R. ARTALEJO AND SRINIVAS R. CHAKRAVARTHY

Received 28 February 2005; Revised 11 May 2005; Accepted 20 June 2005

Two-dimensional continuous-time Markov chains (CTMCs) are useful tools for studying stochastic models such as queueing, inventory, and production systems. Of particular interest in this paper is the distribution of the maximal level visited in a busy period because this descriptor provides an excellent measure of the system congestion. We present an algorithmic analysis for the computation of its distribution which is valid for Markov chains with general-block structure. For a multiserver batch arrival queue with retrials and negative arrivals, we exploit the underlying internal block structure and present numerical examples that reveal some interesting facts of the system.

Copyright © 2006 J. R. Artalejo and S. R. Chakravarthy. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

It is well-known that two-dimensional Markov processes often provide an adequate initial starting point for describing the behavior of many stochastic models. The detailed analysis of particular structures has been the subject matter of many books and papers [8, 12, 13]. Specifically, these address the importance of Markov chains whose transition probability matrices (discrete case) or infinitesimal generators (continuous case) are of quasi-birth-and-death (QBD), $M/G/1$, or $GI/M/1$ type. For such structures, the matrix-analytic methods provide implementable and numerically stable solutions for the distribution of the system state and other descriptors of the performance quality.

In this paper, we develop an algorithmic analysis of the maximum level visited by a two-dimensional CTMC during a busy period. Unlike the study of the stationary system state, the analysis of extreme values can be done for general-block structures. So, specific well-suited block forms are not assumed here. The analysis of extreme values often provides a performance measure of interest connected with the system congestion. In a queueing model, for example, moderate values of the queue indicate that the system is evolving smoothly, whereas large values indicate the need for one or more actions such as

2 Maximum level length in Markov processes

increasing the service capacity and rescheduling of work, on the part of the system manager. Extreme values of Markov processes can be investigated using different approaches. For example, in [11] the objective is to determine the distribution of the maximum queue length during a busy period, and in [14] an asymptotic approach is proposed. In this paper, we will adopt the former approach.

In addition to the assumption of dealing with a well-posed matrix structure, in the framework of the two-dimensional Markov chains, some typical requirements are the stationary regime and the system homogeneity. In the absence of these assumptions, the analysis becomes extremely intricate or, indeed, impossible. In the sequel, we will show how the maximum level visited in a busy period overcomes these restrictions and can be efficiently computed in a class of general-block CTMC with level-dependent structure operating in both stationary and nonstationary regimes.

On the other hand, in recent years, there has been an increasing interest in the investigation of retrial phenomenon in computer and telecommunication systems. The existing literature includes applications to collision-avoidance star protocols for local area networks (CASLANs) [7], circuit-switched networks with hybrid fiber-coax (HFC) architecture [6], fast reservation protocols for asynchronous transfer mode (ATM) networks [15], and cellular mobile telephone networks (CMTNs) [10, 16]. Thus, the interest of retrial queues in engineering is a motivating factor of this paper that focuses on applications and algorithmic implementation of such queueing models.

As related works, we mention [4] where a study for the maximum level visited is initiated for quasi-birth-and-death processes with applications to call centers. Extensions to other models involving Markovian arrival processes (MAPs) with multiple exponential servers [3] and Markovian arrivals with single-server and phase-type services [2] have recently been studied.

The paper is organized as follows. In Section 2, we describe the class of two-dimensional Markov chains under study. Then, in Section 3, we show how the computation of the maximum level visited by the process reduces to finding some absorption probabilities in an auxiliary CTMC with two absorbing states. To this end, we propose a “forward-elimination, backward-substitution” algorithm. In Section 4, we consider a multiserver retrial queue with batch arrivals which operates under the presence of a second flow of negative batch arrivals. We exploit the special nature of the involved blocks to refine the algorithm for the computation of the maximum number of customers in the retrial queue. Some numerical work for this queueing model is given in Section 5. More specifically, we illustrate the effect of the system parameters on the percentiles and modes of the maximum level visited.

2. The general-block two-dimensional Markov chain

We consider a CTMC $\{(N(t), C(t)); t \geq 0\}$ with state space given by $S = \cup\{(i, j) : i \geq 0, 0 \leq j \leq L_i\}$. In the state (i, j) , the first coordinate i is called the level of the state. For use in the sequel, we will denote by 0^* the state $(0, 0)$. The other vector states are defined as follows:

$$\begin{aligned} \underline{0} &= \{(0, j) : 1 \leq j \leq L_0\}, \\ \underline{i} &= \{(i, j) : 0 \leq j \leq L_i\}, \quad i \geq 1. \end{aligned} \tag{2.1}$$

We assume that the number of states in each level, L_j , is finite.

The infinitesimal generator of the CTMC, $Q = [q_{(i,j)(m,n)}]$ (for later use, we define $q_{(i,j)} = -q_{(i,j)(i,j)}$), is of the form

$$Q = \begin{pmatrix} q_{0^*} & A_{0^*,0} & A_{0^*,1} & A_{0^*,2} & \cdots \\ A_{0,0^*} & A_{0,0} & A_{0,1} & A_{0,2} & \cdots \\ A_{1,0^*} & A_{1,0} & A_{1,1} & A_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (2.2)$$

We further assume that our CTMC is irreducible and regular.

3. Analysis of the maximal level visited in a busy period

We now define the busy period for the class of general-block two-dimensional Markov chains described in Section 2. In this context, the busy period is defined as the duration that starts when the process leaves state 0^* and ends at the first epoch thereafter when the process visits state 0^* again.

In queueing theory, the first transition after leaving state 0^* usually corresponds to an arrival of a customer. However, a concrete definition of a busy period depends on the specific stochastic model under study. For example, if customers arrive according to an MAP [5, 9], then the first transition does not necessarily represent an arrival. In fact, stochastic systems modelled with MAP input often need a higher-dimensional representation. In these cases, the definition of busy period must be modified and the analysis in the sequel remains valid with appropriate modifications (see [2, 3]).

Define N_{\max} as the maximum level visited by the Markov chain during a busy period. We first observe that the mass function of N_{\max} can be expressed as

$$P(N_{\max} = k) = \sum_{(m,n) \in S_0} \frac{q_{0^*,(m,n)}}{q_{0^*}} P_{(m,n)}(N_{\max} = k), \quad k \geq 0, \quad (3.1)$$

where $S_0 = \{(m,n) : q_{0^*,(m,n)} > 0\}$, that is, S_0 represents the set of states that the process can visit when a busy period starts.

For initial states (m,n) such that $m > k$, we trivially have $P_{(m,n)}(N_{\max} = k) = 0$. We also note that $P_{(m,n)}(N_{\max} < k)$ corresponds to the probability of the event that starting from the state (m,n) the Markov chain $\{(N(t), C(t)); t \geq 0\}$ will hit state 0^* before hitting level \underline{k} , for $k \geq 1$. Thus, the computation of $P(N_{\max} < k)$, for $k \geq 1$, reduces to finding the probability of absorption to a particular state in a finite-state CTMC with two absorbing states, say, 0^* and k^* , and whose infinitesimal generator is given below:

$$Q(k) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_{0,0^*} & A_{0,0} & A_{0,1} & \cdots & A_{0,k-2} & A_{0,k-1} & A_{0^*}^* \\ A_{1,0^*} & A_{1,0} & A_{1,1} & \cdots & A_{1,k-2} & A_{1,k-1} & A_{1^*}^* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_{k-1,0^*} & A_{k-1,0} & A_{k-1,1} & \cdots & A_{k-1,k-2} & A_{k-1,k-1} & A_{k-1^*}^* \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}, \quad (3.2)$$

4 Maximum level length in Markov processes

where the coefficient matrices appearing in (3.2) are as seen in (2.2), and

$$A_i^* = \sum_{j=k}^{\infty} A_{ij}, \quad 0 \leq i \leq k-1. \quad (3.3)$$

Let $T(k)$ denote the part of the generator $Q(k)$ that corresponds to the levels $\underline{0}, \underline{1}, \dots, k-1$. Since the original CTMC is irreducible, this set of states is transient and consequently $T(k)$ is invertible. Let $\mathbf{T}^0(k)$ denote a column vector obtained by stacking the column vectors $A_{0,0^*}, A_{1,0^*}, \dots$, and $A_{k-1,0^*}$, and let $\mathbf{D}^0(k) = (A_{0^*,0}, A_{0^*,1}, \dots, A_{0^*,k-1})$. Then, it is easy to verify that

$$P(N_{\max} < k) = -\frac{1}{q_{0^*}} \mathbf{D}^0(k) (T(k))^{-1} \mathbf{T}^0(k). \quad (3.4)$$

Of course, once the cumulative mass function $P(N_{\max} < k)$ is computed, we automatically determine the mass function $P(N_{\max} = k)$. We also note that when $L_0 = 0$, the blocks corresponding to level $\underline{0}$ in (3.2) must be deleted.

It should be pointed out that the analysis of a CTMC with a finite number K of levels needs a minor modification. Following the above arguments, we may determine $P(N_{\max} = k)$, for $0 \leq k \leq K-2$, and set $P(N_{\max} = K-1) = 1 - P(N_{\max} \leq K-2)$.

We now define $\mathbf{x}^{(k)} = -(T(k))^{-1} \mathbf{T}^0(k)$, $k \geq 1$, and partition $\mathbf{x}^{(k)} = (\mathbf{x}^{(k)}(0), \dots, \mathbf{x}^{(k)}(k-1))'$. Here and in the sequel the notation “ $'$ ” stands for the transpose of a matrix. The system $T(k)\mathbf{x}^{(k)} = -\mathbf{T}^0(k)$ can be solved using any all-purpose algorithm. For the sake of completeness, we next formulate a “forward-elimination, backward-substitution” algorithm (Algorithm 3.1).

It should be noted that matrices Δ_{jj} , $0 \leq j \leq k-1$, are invertible. The proof relies on an inductive argument that uses the existence of the inverse of $T(k)$, for any $k \geq 1$.

In the case of particular block structures (i.e., QBD-, $M/G/1$ -, or $GI/M/1$ -type generators), the “forward-elimination, backward-substitution” algorithm can be appropriately reformulated. Some simplifications are also obtained by exploiting specific sparse structures of the blocks. Examples can be seen in [2, 3].

4. Application to the $M^{[X]}/M/c$ retrial queue with negative arrivals

We consider a multiserver queueing system where the arrivals occur in batches according to a Poisson process with rate λ . The batch sizes are i.i.d. random variables with probability mass function given by $\{a_k : k \geq 1\}$. Let, for $k \geq 1$, $\lambda_k = \lambda a_k$ and $\lambda_k^* = \sum_{j=k}^{\infty} \lambda_j$. Customers are served by c homogeneous exponential servers each serving at a rate μ . Arriving batch of customers finding any free server will get into service immediately and any customer not getting into service will enter into an orbit of infinite size. Customers in orbit will compete for service by searching for a free server. The search (retrial) times are assumed to be exponentially distributed with parameter θ . In addition, we also assume a second flow of negative arrivals occurring in batches according to a Poisson process with parameter δ . The effect of these arrivals is seen only when all c servers are busy with at least one customer waiting in the orbit. The batch sizes of negative arrivals are i.i.d.

Step 1 (forward elimination):

$$M_{ij}(0) := A_{ij}, i, j = 0, \dots, k-1;$$

$$\mathbf{f}_i^{(k)}(0) := -A_{i,0^*}, i = 0, \dots, k-1;$$

$$\Delta_{0j} := M_{0j}(0); j = 0, \dots, k-1;$$

$$\mathbf{h}_0^{(k)} := \mathbf{f}_0^{(k)}(0);$$

$$\mathbf{y}_0^{(k)} := \Delta_{00}^{-1} \mathbf{f}_0^{(k)}(0);$$

for $m = 1, \dots, k-1$ compute

$$M_{ij}(m) := M_{ij}(m-1), i, j = 0, \dots, m-1;$$

$$M_{ij}(m) := 0, i = m, \dots, k-1, j = 0, \dots, m-1;$$

$$M_{ij}(m) := M_{ij}(m-1) - M_{i,m-1}(m-1) \Delta_{m-1,m-1}^{-1} M_{m-1,j}(m-1),$$

$$i, j = m, \dots, k-1;$$

$$\mathbf{f}_i^{(k)}(m) := \mathbf{f}_i^{(k)}(m-1), i = 0, \dots, m-1;$$

$$\mathbf{f}_i^{(k)}(m) := \mathbf{f}_i^{(k)}(m-1) - M_{i,m-1}(m-1) \Delta_{m-1,m-1}^{-1} \mathbf{f}_{m-1}^{(k)}(m-1), i = m, \dots, k-1;$$

$$\Delta_{mj} := M_{mj}(m), j = 0, \dots, k-1;$$

$$\mathbf{h}_m^{(k)} := \mathbf{f}_m^{(k)}(m);$$

$$\mathbf{y}_m^{(k)} := \Delta_{mm}^{-1} (\mathbf{f}_m^{(k)}(m-1) - M_{m,m-1}(m-1) \mathbf{y}_{m-1}^{(k)}).$$

Step 2 (backward substitution):

$$\mathbf{x}^{(k)}(k-1) := \mathbf{y}_{k-1}^{(k)};$$

for $j = k-2, \dots, 0$

$$\mathbf{x}^{(k)}(j) := \mathbf{y}_j^{(k)} - \Delta_{jj}^{-1} \Delta_{j,j+1} \mathbf{x}^{(k)}(j+1) - \dots - \Delta_{jj}^{-1} \Delta_{j,k-1} \mathbf{x}^{(k)}(k-1).$$

Algorithm 3.1

random variables with probability mass function given by $\{d_k : k \geq 1\}$. Let, for $k \geq 1$, $\delta_k = \delta d_k$ and $\delta_k^* = \sum_{j=k}^{\infty} \delta_j$. When a negative arrival occurs with a batch of size k at a time when the orbit size is i with all c servers busy, then $r = \min\{k, i\}$ customers are removed from the orbit. These r customers are considered lost. Putting $a_k = 1$, the model reduces to the retrial queue with negative arrivals investigated in [1].

The infinitesimal generator of the $M^{[X]}/M/c$ retrial queue with negative arrivals has a general-block structure with the following blocks:

$$\begin{aligned} q_{0^*} &= \lambda, & A_{0^*,0} &= (\lambda_1, \dots, \lambda_c), & A_{0^*,i} &= \lambda_{c+i} \mathbf{e}'_{c+1}, & i &\geq 1, \\ & & A_{0,0^*} &= \mu \mathbf{e}_1, & A_{i,0^*} &= 0, & i &\geq 1. \end{aligned} \quad (4.1)$$

6 Maximum level length in Markov processes

(\mathbf{e}_j denotes the column vector of appropriate dimension with 1 in the j th position and 0 elsewhere.)

$$A_{0,0} = \begin{pmatrix} -(\lambda + \mu) & \lambda_1 & \lambda_2 & \cdots & \lambda_{c-2} & \lambda_{c-1} \\ 2\mu & -(\lambda + 2\mu) & \lambda_1 & \cdots & \lambda_{c-3} & \lambda_{c-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda + (c-1)\mu) & \lambda_1 \\ 0 & 0 & 0 & \cdots & c\mu & -(\lambda + c\mu) \end{pmatrix}. \quad (4.2)$$

$A_{0,j}$, $j \geq 1$, is a rectangular matrix of dimension $c \times (c+1)$ whose elements are all zero except the last column entries given by $\{\lambda_{c+j-1}, \lambda_{c+j-2}, \dots, \lambda_j\}$.

$A_{i,0}$, $i \geq 2$, is a rectangular matrix of dimension $(c+1) \times c$ whose elements are all zero except the $(c+1, c)$ th entry which is given by δ_i^* ,

$$A_{i,i} = \begin{pmatrix} -(\lambda + i\theta) & \lambda_1 & \cdots & \lambda_{c-1} & \lambda_c \\ \mu & -(\lambda + \mu + i\theta) & \cdots & \lambda_{c-2} & \lambda_{c-1} \\ 0 & 2\mu & \cdots & \lambda_{c-3} & \lambda_{c-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -(\lambda + (c-1)\mu + i\theta) & \lambda_1 \\ 0 & 0 & \cdots & c\mu & -(\lambda + c\mu + \delta) \end{pmatrix}, \quad i \geq 1. \quad (4.3)$$

$A_{i,j}$, $i \geq 3$, $1 \leq j \leq i-2$, is a square matrix of dimension $c+1$ whose elements are all zero except the $(c+1, c+1)$ th entry which is given by δ_{i-j} .

$A_{i,j}$, $i \geq 1$, $j \geq i+1$, is a square matrix of dimension $c+1$ whose elements are all zero except the last column which has entries given by $\{\lambda_{c+j-i}, \lambda_{c+j-i-1}, \dots, \lambda_{j-i}\}$.

$A_{1,0}$ is a rectangular matrix of dimension $(c+1) \times c$, and $A_{i,i-1}$, $i \geq 2$, is a square matrix of dimension $c+1$ which are given by

$$A_{1,0} = \begin{pmatrix} \theta & & & \\ & \theta & & \\ & & \ddots & \\ & & & \theta \\ & & & & \delta_1^* \end{pmatrix}, \quad A_{i,i-1} = i\theta \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & & \delta_1 \end{pmatrix}, \quad i \geq 2. \quad (4.4)$$

Although the computation of the maximum number of customers in orbit during a busy period can be done for the case of nonstationary regime, this situation is quite unusual because the number of customers in the system grows unbounded. Hence, we will concentrate mainly on the stationary case. The CTMC $\{(N(t), C(t)); t \geq 0\}$ modelling the $M^{[X]}/M/c$ retrial queue with negative arrivals is positive recurrent if and only if

$$\lambda \bar{a} < c\mu + \delta \bar{d}, \quad (4.5)$$

where \bar{a} and \bar{d} denote, respectively, the expected batch size of positive and negative arrivals.

To prove the necessity of (4.5), we partition the state space into diagonal subsets $D(k) = \{(i, j) : i + j = k\}$. By equating the flow rate in and out of the subset $D(k)$ and adding over k , we get

$$\lambda \bar{a} = \sum_{i=0}^{\infty} \sum_{j=0}^c j \mu P_{ij} + \sum_{i=1}^{\infty} P_{ic} \sum_{k=1}^i \delta_k^*, \quad (4.6)$$

where P_{ij} denotes the stationary probability that the system is in state (i, j) .

The first term on the right-hand side of (4.6) is bounded by $c\mu$. On the other hand, by interchanging the order of summation in the second term of the right-hand side and taking into account that $\sum_{i=k}^{\infty} P_{ic} < 1$, we easily conclude that $\lambda \bar{a} < c\mu + \delta \bar{d}$.

The sufficiency can be proved by appealing to the classic Foster criterion for the drifts of the embedded Markov chain $\{Z_n; n \geq 0\}$ at the transition epochs. We may consider the test function $f(i, j) = i + hj$. In the case when $i \geq 0$ and $0 \leq j \leq c - 1$, the mean drifts are given by

$$\gamma(i, j) = \frac{\sum_{r=1}^{c-j} r \lambda_r + \sum_{r=c-j+1}^{\infty} (r - c + j + h(c - j)) \lambda_r - hj\mu + (h - 1)i\theta}{\lambda + j\mu + i\theta} \quad (4.7)$$

and, for $i \geq 1$ and $j = c$,

$$\gamma(i, c) = \frac{\sum_{r=1}^{\infty} r \lambda_r - hc\mu - \sum_{k=1}^{i-1} k \delta_k - i \delta_i^*}{\lambda + c\mu + \delta}. \quad (4.8)$$

For $0 \leq j \leq c - 1$, we have $\gamma(i, j) \rightarrow h - 1$, as $i \rightarrow \infty$, while $\gamma(i, c) \rightarrow (\lambda \bar{a} - hc\mu - \delta \bar{d})(\lambda + c\mu + \delta)^{-1}$.

Let h be any number in the interval $(\max\{(\lambda \bar{a} - \delta \bar{d})/c\mu, 0\}, 1)$. Then, it is clear that we may find an appropriate $\varepsilon > 0$ such that $\gamma(i, j) < -\varepsilon$ (except perhaps for a finite number of states). This proves that condition (4.5) is also sufficient for the positive recurrence.

In what follows, we will describe an efficient algorithm for the computation of the probability $P(N_{\max} < k)$ that is of interest in this paper. Before proceeding with the description of the algorithm, we register a number of observations to ease the understanding of the equations proposed.

(i) To minimize the storage requirements of the number of auxiliary matrices needed, the equations are modified accordingly. For example, where we need matrices of the form $(-A_{ii})^{-1}$, $1 \leq i \leq k - 1$, we simply store $(-A_{k-1, k-1})^{-1}$ and modify the corresponding equations appropriately.

(ii) The equations are displayed in such a way that they are ready for numerical implementation. At this point, we can use a variety of numerical methods such as Gaussian elimination, LU factorization and aggregate/disaggregate method.

Observing that

$$q_{0^*} = \lambda, \quad \mathbf{D}^0(k) = (\lambda_1, \dots, \lambda_c, \lambda_{c+1} \mathbf{e}'_{c+1}, \dots, \lambda_{c+k-1} \mathbf{e}'_{c+1}), \quad \mathbf{T}^0(k) = \mu \mathbf{e}_1, \quad (4.9)$$

formula (3.4) reduces to $P(N_{\max} < k) = \lambda^{-1} \mu \mathbf{D}^0(k) (-T(k))^{-1} \mathbf{e}_1$, $k \geq 1$.

8 Maximum level length in Markov processes

Define

$$\mathbf{b}^{(k)} = \boldsymbol{\eta}_k(-T(k))^{-1}, \quad k \geq 1, \quad (4.10)$$

where

$$\boldsymbol{\eta}_k = \lambda^{-1} \mathbf{D}^0(k) = (a_1, \dots, a_c, a_{c+1} \mathbf{e}'_{c+1}, \dots, a_{c+k-1} \mathbf{e}'_{c+1}). \quad (4.11)$$

Partition $\mathbf{b}^{(k)}$ as

$$\mathbf{b}^{(k)} = (\mathbf{b}^{(k)}(0), \dots, \mathbf{b}^{(k)}(k-1)). \quad (4.12)$$

First note that $P(N_{\max} < k) = \mu b_1^{(k)}(0)$, where $b_1^{(k)}(0)$ is the first component of the vector $\mathbf{b}^{(k)}(0)$. The equation $\mathbf{b}^{(k)} = \boldsymbol{\eta}_k(-T(k))^{-1}$, $k \geq 1$, can be written in terms of matrices of dimension at most equal to $c+1$ that are well suited for numerical implementation. We need to present separately the cases $c=1$ and $c \geq 2$, and the cases $k=1$ and $k \geq 2$.

Case 1. $c=1$ and $k=1$.

Here we compute the probability explicitly as

$$P(N_{\max} < 1) = \frac{a_1 \mu}{\lambda + \mu}. \quad (4.13)$$

Case 2. $c=1$ and $k \geq 2$.

In this case, the equations are given by

$$b_1^{(k)}(0) = \frac{1}{\lambda + \mu} \left(a_1 + \theta b_0^{(k)}(1) + \sum_{j=1}^{k-1} b_1^{(k)}(j) \delta_j^* \right), \quad (4.14)$$

and, for $1 \leq i \leq k-2$,

$$\begin{aligned} b_0^{(k)}(i) &= \frac{\mu}{\lambda + i\theta} b_1^{(k)}(i), \\ b_1^{(k)}(i) &= \frac{1}{\lambda + \mu + \delta} \left(a_{i+1} + \sum_{j=0}^{i-1} (b_0^{(k)}(j+1) + b_1^{(k)}(j)) \lambda_{i-j} \right. \\ &\quad \left. + (i+1)\theta b_0^{(k)}(i+1) + \sum_{j=i+1}^{k-1} b_1^{(k)}(j) \delta_{j-i} \right), \\ b_0^{(k)}(k-1) &= \frac{\mu}{\lambda + (k-1)\theta} b_1^{(k)}(k-1), \\ b_1^{(k)}(k-1) &= \frac{1}{\lambda + \mu + \delta} \left(a_k + \sum_{j=0}^{k-2} (b_0^{(k)}(j+1) + b_1^{(k)}(j)) \lambda_{k-1-j} \right). \end{aligned} \quad (4.15)$$

Case 3. $c \geq 2$ and $k=1$.

Here the probability is calculated explicitly as

$$P(N_{\max} < 1) = \mu \mathbf{a}(-A_{0,0})^{-1} \mathbf{e}_1, \quad (4.16)$$

where $\mathbf{a} = (a_1, \dots, a_c)$.

Case 4. $c \geq 1$ and $k \geq 2$.

In this case, the equations are given by

$$\begin{aligned} \mathbf{b}^{(k)}(0) &= \left(a_1 + \theta b_0^{(k)}(1), a_2 + \theta b_1^{(k)}(1), \dots, a_c + \theta b_{c-1}^{(k)}(1) + \sum_{j=1}^{k-1} b_c^{(k)}(j) \delta_j^* \right) (-A_{0,0})^{-1}, \\ \mathbf{b}^{(k)}(i) &= \left(f_0^{(k)}(i), f_1^{(k)}(i), \dots, f_c^{(k)}(i) \right) (-A_{k-1,k-1})^{-1}, \quad 1 \leq i \leq k-2, \end{aligned} \quad (4.17)$$

where

$$\begin{aligned} f_0^{(k)}(i) &= (k-1-i)\theta b_0^{(k)}(i), \\ f_r^{(k)}(i) &= (k-1-i)\theta b_r^{(k)}(i) + (i+1)\theta b_{r-1}^{(k)}(i+1), \quad 1 \leq r \leq c-1, \\ f_c^{(k)}(i) &= a_{i+c} + (i+1)\theta b_{c-1}^{(k)}(i+1) + \sum_{j=i+1}^{k-1} b_c^{(k)}(j) \delta_{j-i} \\ &\quad + \sum_{r=1}^c b_r^{(k)}(0) \lambda_{c+i-r} + \sum_{j=1}^{i-1} \sum_{r=0}^c b_r^{(k)}(j) \lambda_{c+i-j-r}, \quad (4.18) \\ \mathbf{b}^{(k)}(k-1) &= \left(0, \dots, 0, a_{k-1+c} + \sum_{r=1}^c b_r^{(k)}(0) \lambda_{c+k-1-r} \right. \\ &\quad \left. + \sum_{j=1}^{k-2} \sum_{r=0}^c b_r^{(k)}(j) \lambda_{c+k-1-j-r} \right) (-A_{k-1,k-1})^{-1}. \end{aligned}$$

5. Numerical examples

In order to evaluate the performance of the $M^{[X]}/M/c$ retrial queue with negative arrivals, some computational experiments are performed. To this end, we will consider the following three distributions for the batch size.

(a) *Geometric (GEO)*:

$$a_k(d_k) = (1 - \alpha)\alpha^{k-1}, \quad k \geq 1, \alpha \in (0, 1). \quad (5.1)$$

(b) *Deterministic (DET)*:

$$a_k(d_k) = \begin{cases} 1, & \text{if } k = D, \\ 0, & \text{otherwise,} \end{cases} \quad k \geq 1, D \geq 1. \quad (5.2)$$

(c) *Conditional Poisson (POI)*:

$$a_k(d_k) = (1 - e^{-\nu})^{-1} e^{-\nu} \frac{\nu^k}{k!}, \quad k \geq 1, \nu > 0. \quad (5.3)$$

In the following α_1, D_1 , and ν_1 (α_2, D_2 , and ν_2) will indicate parameters associated to the batch of positive (negative) arrivals.

Table 5.1 99th percentile and mode(s) ($c = 5, \lambda = 1.5, \alpha_1 = 0.5, \delta = 1.75, \alpha_2 = 0.3$).

μ	$\theta = 0.08$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1.0$	$\theta = 10.0$	$\theta = 1000.0$
2.5	(18,0)	(10,0)	(8,0)	(7,0)	(6,0)	(6,0)
1.0	(45,0,28)	(25,0)	(20,0)	(17,0)	(14,0)	(14,0)
0.5	(92,0,62)	(53,0,28)	(43,0,19)	(37,0,14)	(32,0,10)	(31,0,10)
0.25	(232,0,156)	(138,0,82)	(115,0,63)	(102,0,52)	(90,0,41)	(88,0,40)

The first set of numerical results (see Examples 5.1–5.3) shows the influence of the system parameters on the distribution of the maximum orbit length for a model with geometric batch sizes of positive and negative arrivals. More concretely, we will focus on two measures: the 99th percentile and the mode(s) of N_{\max} . In Example 5.4, we analyze the effect of the batch size distributions. Example 5.5 shows how the probability mass function of N_{\max} behaves as a function of the traffic intensity $\rho = \lambda \bar{a}(c\mu + \delta \bar{g})^{-1}$. Finally, in Example 5.6, we consider the $M^{[X]}/M/c$ retrial queue (case $\delta = 0$) and $M/M/c$ retrial queue with negative arrivals (case $a_1 = 1$) which are examples, respectively, of structures of the $M/G/1$ and $GI/M/1$ types.

Example 5.1. The main purpose of this example is to see the influence of μ and θ on the two measures: the 99th percentile and the mode(s) of the random variable N_{\max} . In this example, we fix $c = 5, \lambda = 1.5, \alpha_1 = 0.5, \delta = 1.75, \alpha_2 = 0.3$ and calculate the probability mass function of N_{\max} . μ is varied in such a way that ρ will take the values 0.2, 0.4, 0.6, and 0.8. In Table 5.1 we display the percentile and the mode(s) as ordered set of values. For example, the entry (10,0) will mean that the 99th percentile occurs at $k = 10$ and there is only mode that occurs at 0. The entry (45,0,28) indicates that the 99th percentile occurs at $k = 45$ and the two modes are at $k = 0$ and $k = 28$.

The following observations can be inferred from Table 5.1.

(i) For all values of μ , the 99th percentile appears to decrease as θ increases and seems to approach a constant value. This is to be expected since an increase in θ will result in more success for orbit customers to occupy a free server.

(ii) For all values of θ , the 99th percentile appears to increase as μ decreases. Again, this is as to be expected since a decrease in μ will result in seeing more customers in the orbit.

(iii) In all the cases, there appears to be a mode at 0 and depending on the values of μ and θ there seems to be multiple modes. In the case of multiple modes, it is interesting to see that the second mode appears to be closer to 0 (than to the 99th percentile) when θ is large and seems to be closer to the 99th percentile for small values of θ . The same behavior (with respect to the second mode) is seen as μ is changing. While it is sort of intuitive to see the behavior of the second mode, it is not easy to explain the logic of having a mode at the origin.

Example 5.2. Here we are interested to see the effect of λ and α_1 on the two measures: the 99th percentile and the mode(s) of the random variable N_{\max} . In this example, we fix $c = 5, \mu = 1.0, \theta = 2.0, \delta = 1.75, \alpha_2 = 0.3$ and calculate the probability mass function of N_{\max} . λ is varied in such a way that ρ will take the values 0.2, 0.4, 0.6, and 0.8. Note that as

Table 5.2 99th percentile and mode(s) ($c = 5, \mu = 1.0, \theta = 2.0, \delta = 1.75, \alpha_2 = 0.3$).

α_1	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
0.2	(3,0)	(9,0)	(21,0,7)	(56,0,26)
0.4	(6,0)	(13,0)	(26,0,7)	(67,0,26)
0.6	(10,0)	(19,0)	(36,0)	(85,0,20)
0.8	(25,0)	(38,0)	(63,0)	(131,0)

Table 5.3 99th percentile and mode(s) ($c = 5, \mu = 1.0, \theta = 2.0, \lambda = 3.0, \alpha_1 = 0.5$).

α_2	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
0.2	(16,0,6)	(23,0,9)	(36,0,14)	(73,0,24)
0.4	(16,0,7)	(24,0,10)	(38,0,14)	(76,0,25)
0.6	(18,0,7)	(26,0,11)	(42,0,16)	(82,0,26)
0.8	(21,0,9)	(33,0,13)	(51,0,19)	(99,0,31)

α_1 is varied we need to change λ accordingly so as to get a specific value for ρ . We display the percentile and the mode(s) as ordered sets of values in Table 5.2.

An examination of the above table reveals the following.

(i) For all values of α_1 , as is to be expected, the 99th percentile appears to increase with ρ .

(ii) In the case when α_1 increases, for fixed values of ρ , the 99th percentile appears to increase. The rate of increase appears to be large for large values of ρ . Note that an increase in α_1 results in an increase in the value of the 99th percentile. Hence, it is not surprising to see such a phenomenon.

(iii) There seems to be a mode at 0 in all cases and only in some cases (namely, when ρ is large and α_1 is small) there appears to be a second mode.

Example 5.3. For this example, the varying parameters of interest are δ and α_2 . The other parameters are fixed as follows: $c = 5, \mu = 1.0, \theta = 2.0, \lambda = 3.0, \alpha_1 = 0.5$. δ is varied in such a way that ρ will take the values 0.2, 0.4, 0.6, and 0.8. Again the values of δ are modified according to the values of α_2 . Table 5.3 contains the measures under consideration.

A summary of the observations of Table 5.3 is given below.

(i) Here we see similar behavior (with respect to the 99th percentile) as in Example 5.2. That is, for all values of α_2 , as is to be expected, the 99th percentile appears to increase as ρ increases. As α_2 increases, for fixed values of ρ , the 99th percentile appears to increase. The rate of increase appears to be large for large values of ρ .

(ii) While there is a mode at 0 in all cases (like in Example 5.2), it is very interesting to see that there is also a second mode in all cases. This is different from what we observed in Example 5.2.

Example 5.4. By fixing $c = 5, \mu = 0.7, \lambda = 1.5, \delta = 5/6, \theta = 2.0$, we vary the batch size distributions for both positive and negative arrivals. Here we consider three cases for each distribution. In the first case we consider geometric batch size, the second one is deterministic, and the third case is conditional Poisson distribution (conditioned on the fact that at least one arrival occurs in a batch). The three cases (for each type of arrivals)

12 Maximum level length in Markov processes

Table 5.4 99th percentile and mode(s) ($c = 5, \mu = 0.7, \theta = 2.0, \lambda = 1.5, \bar{a} = 2.0, \delta = 5/6, \bar{d} = 3.0$).

	Geo(0.5)	Det(2)	Poi(ν_1)
Geo(2/3)	(25,0,6)	(20,0,6)	(23,0,6)
Det(3)	(24,0,6)	(19,0,6)	(23,0,6)
Poi(ν_2)	(25,0,6)	(19,0,6)	(22,0,6)

Table 5.5 $P(N_{\max} = k), k = 0, 1, 5, 10, \text{ and } 20$ ($c = 5, \lambda = 3.0, \alpha_1 = 0.5, \delta = 1.75, \alpha_2 = 0.3, \theta = 2.0$).

ρ	$P(N_{\max} = 0)$	$P(N_{\max} = 1)$	$P(N_{\max} = 5)$	$P(N_{\max} = 10)$	$P(N_{\max} = 20)$
0.2	0.852364	0.048098	0.010300	0.001180	0.000009
0.4	0.633939	0.063892	0.033276	0.011550	0.000592
0.6	0.407895	0.040913	0.033964	0.029674	0.010729
0.8	0.250730	0.016505	0.013068	0.013434	0.015874
0.9	0.197884	0.009513	0.006680	0.005995	0.006150
1.0	0.157656	0.005261	0.003219	0.002411	0.001741
1.1	0.126827	0.002814	0.001500	0.000931	0.000453
1.2	0.102876	0.001460	0.000684	0.000355	0.000118
1.3	0.083949	0.000734	0.000306	0.000134	0.000031
1.4	0.067439	0.000332	0.000123	0.000046	0.000008
1.5	0.056239	0.000165	0.000057	0.000019	0.000002

are such that for positive arrivals the average number of customers per batch is $\bar{a} = 2.0$ and for the negative arrivals the average is set at $\bar{d} = 3.0$. Thus, the parameters for the positive (negative) batch arrivals for these three cases are, respectively, $\alpha_1 = 0.5$ ($\alpha_2 = 2/3$), $D_1 = 2$ ($D_2 = 3$), and $\nu_1 = 1.593624$ ($\nu_2 = 2.821439$). In Table 5.4, the two measures are displayed for various combinations.

The following observations are obtained for the data seen in Table 5.4.

(i) It is interesting to note that the type of distribution considered for the batches of the negative arrivals seems to be some what insensitive to the 99th percentile. However, there seems to be a small dependence when it comes to the distribution of the batch sizes for the regular arrivals.

(ii) With respect to the other measure, there appears to be two modes one at 0 and the other at 6 for all cases.

Example 5.5. The purpose of this example is to see how the probabilities, $P(N_{\max} = k)$, behave as functions of ρ . Here we fix $c = 5, \lambda = 3.0, \alpha_1 = 0.5, \delta = 1.75, \alpha_2 = 0.3$, and $\theta = 2.0$. We vary ρ through changing the values of μ . In Table 5.5, we display some selected probabilities, namely, $P(N_{\max} = k), k = 0, 1, 5, 10, 20$. Note that in order to find the number of customers seen in the retrieval orbit during a busy period, we do not have to assume the system to be stable. However, as is to be expected, when $\rho \geq 1$, the busy period may not terminate (with probability one) and hence the tail of the distributions is heavy.

An examination of the above table indicates the following observations. In the following, let $p_k = P(N_{\max} = k)$, for $k = 0, 1, 5, 10, 20$.

Table 5.6 99th percentile and mode(s) ($c = 5, \mu = 1.0, \alpha_1 = 0.5, \delta = 0.0$).

ρ	$\theta = 0.08$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1.0$	$\theta = 10.0$	$\theta = 1000.0$
0.1	(5,0)	(4,0)	(4,0)	(4,0)	(4,0)	(4,0)
0.2	(9,0)	(7,0)	(6,0)	(6,0)	(6,0)	(6,0)
0.3	(17,0)	(11,0)	(9,0)	(9,0)	(8,0)	(8,0)
0.4	(31,0,18)	(17,0)	(14,0)	(12,0)	(11,0)	(11,0)
0.5	(53,0,36)	(26,0)	(20,0)	(17,0)	(15,0)	(14,0)
0.6	(89,0,67)	(40,0,19)	(29,0)	(24,0)	(20,0)	(20,0)
0.7	(157,0,126)	(66,0,37)	(46,0,14)	(36,0)	(28,0)	(27,0)
0.8	(305,0,258)	(121,0,76)	(79,0,34)	(59,0)	(43,0)	(42,0)

Table 5.7 99th percentile and mode(s) ($c = 5, \mu = 0.5, \alpha_1 = 0.0, \delta = 1.75, \alpha_2 = 0.3$).

ρ	$\theta = 0.08$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 1.0$	$\theta = 10.0$	$\theta = 1000.0$
0.1	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
0.2	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)
0.3	(7,0)	(6,0)	(6,0)	(5,0)	(5,0)	(5,0)
0.4	(13,0)	(11,0)	(10,0)	(9,0)	(8,0)	(8,0)
0.5	(24,0,13)	(18,0)	(16,0)	(14,0)	(13,0)	(12,0)
0.6	(43,0,27)	(30,0,17)	(25,0,13)	(22,0,11)	(19,0,9)	(18,0,8)
0.7	(83,0,60)	(51,0,32)	(41,0,24)	(35,0,19)	(29,0,14)	(28,0,13)
0.8	(184,0,147)	(99,0,68)	(76,0,48)	(63,0,36)	(48,0,24)	(46,0,23)

(i) p_0 decreases as ρ increases. This is as expected since an increase in ρ will cause more and more customers to join the orbit and hence less chance of seeing the orbit to be empty.

(ii) With respect to $p_k, k = 1, 5, 10,$ and $20,$ we see a pattern in which the probabilities increase initially (as a function of ρ) up to a certain point, say $k^*,$ and then start to decrease. This phenomenon can be explained intuitively as follows. The initial increase is due to the likelihood of seeing a possible mode at k^* when ρ is moderate, and a later decrease indicates that the orbit is likely to build up more customers leading to a longer tail in the distribution. Basically, we see a mode (not counting the possible one at 0) shifting away from the origin as ρ increases.

Example 5.6. In this example, we consider the special cases: $M^{[X]}/M/c$ retrial queue (case $\delta = 0$) and $M/M/c$ retrial queue with negative arrivals (case $a_1 = 0$). For the former, in Table 5.6, we fix $c = 5, \mu = 1.0,$ and $\delta = 0.0.$ We assume geometric batch arrivals with rate $\alpha_1 = 0.5,$ and vary λ to get that $\rho = \lambda \bar{a}(c\mu)^{-1}$ will take values from 0.1 to 0.8. Each entry in the table gives the 99th percentile and the mode(s).

For the model with negative arrivals, in Table 5.7, we display both performance measures. To this end, we fix $c = 5, \mu = 0.5, \delta = 1.75,$ and $\alpha_2 = 0.3.$ Then, we vary λ in such a way that $\rho = \lambda(c\mu + \delta \bar{d})^{-1}$ takes the desired values. Note that both models are normalized to have the global service rate, $c\mu + \delta \bar{d},$ equal to 5. Then, the arrival rate $\lambda \bar{a}$ is fixed as $5\rho.$

Some conclusions obtained from Tables 5.6-5.7 are listed below.

(i) For both queueing models, the 99th percentile is an increasing function of ρ and a decreasing one of θ . While in general the 99th percentile is lesser for the model with negative arrivals, it is interesting to note that the percentile, for fixed values of ρ , varies in a wider domain for the model with batch arrivals.

(ii) The model with negative arrivals appears to have a second mode when ρ is greater than 0.5. In contrast, the model with batch arrivals exhibits a second mode when the system congestion is high (i.e., when ρ is large and θ is small).

(iii) In both models, for all values of ρ , the percentile seems to approach a limit value as far as the retrial rate increases. The convergence is faster in the model with negative arrivals.

Acknowledgments

J. R. Artalejo thanks the support received from the Research Project MTM 2005-01248. The research was conducted while S. R. Chakravarthy was visiting the Complutense University of Madrid, Madrid, Spain, and would like to thank the hospitality of the Department of Statistics and Operations Research.

References

- [1] V. V. Anisimov and J. R. Artalejo, *Analysis of Markov multiserver retrial queues with negative arrivals*, Queueing Systems. Theory and Applications **39** (2001), no. 2-3, 157–182.
- [2] J. R. Artalejo and S. R. Chakravarthy, *Algorithmic analysis of the MAP/PH/1retrial queue*, submitted.
- [3] ———, *Computational analysis of the maximal queue length in the MAP/M/cretrial queue*, submitted.
- [4] J. R. Artalejo, A. Economou, and A. Gomez-Corral, *Applications of maximum queue length to call centers management*, to appear in Computers & Operations Research.
- [5] S. R. Chakravarthy, *The batch Markovian arrival process: A review and future work*, Advances in Probability Theory and Stochastic Processes (A. Krishnamoorthy, N. Raju, and V. Ramaswami, eds.), Notable Publications, New Jersey, 2000, pp. 21–39.
- [6] D. J. Houck and W. S. Lai, *Traffic modeling and analysis of hybrid fiber-coax systems*, Computer Networks and ISDN Systems **30** (1998), no. 8, 821–834.
- [7] G. K. Janssens, *The quasi-random input queueing system with repeated attempts as a model for a collision-avoidance star local area network*, IEEE Transactions on Communications **45** (1997), no. 3, 360–364.
- [8] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Pennsylvania; American Statistical Association, Virginia, 1999.
- [9] D. M. Lucantoni, *New results on the single server queue with a batch Markovian arrival process*, Communications in Statistics. Stochastic Models **7** (1991), no. 1, 1–46.
- [10] M. A. Marsan, G. de Carolis, E. Leonardi, R. Lo Cigno, and M. Meo, *Efficient estimation of call blocking probabilities in cellular mobile telephony networks with customer retrials*, IEEE Journal on Selected Areas in Communications **19** (2001), no. 2, 332–346.
- [11] M. F. Neuts, *The distribution of the maximum length of a Poisson queue during a busy period*, Operations Research **12** (1964), no. 2, 281–285.
- [12] ———, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins Series in the Mathematical Sciences, vol. 2, Johns Hopkins University Press, Maryland, 1981.

- [13] ———, *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, Probability: Pure and Applied, vol. 5, Marcel Dekker, New York, 1989.
- [14] R. F. Serfozo, *Extreme values of birth and death processes and queues*, Stochastic Processes and their Applications **27** (1988), no. 2, 291–306.
- [15] H. Shimonishi, T. Takine, M. Murata, and H. Miyahara, *Performance analysis of fast reservation protocol in ATM networks with arbitrary topologies*, Performance Evaluation **27-28** (1996), 41–69.
- [16] P. Tran-Gia and M. Mandjes, *Modeling of customer retrieval phenomenon in cellular mobile networks*, IEEE Journal on Selected Areas in Communications **15** (1997), no. 8, 1406–1414.

Jesus R. Artalejo: Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid 28040, Spain
E-mail address: jesus_artalejo@mat.ucm.es

Srinivas R. Chakravarthy: Department of Industrial and Manufacturing Engineering and Business, Kettering University, Flint, MI 48439, USA
E-mail address: schakrav@kettering.edu