Baltzer Journals October 1, 1996

Some Algorithms for Discrete Time Queues with Finite Capacity

VINOD SHARMA[1,*] AND NANDYALA D. GANGADHAR[2]

[1]INRS-Telecommunications, University du Quebec, Quebec H3E 1H6, Canada.
[2]Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012, India.
E-mail: vinod@inrs-telecom.uquebec.ca, ndg@ee.iisc.ernet.in

We consider a discrete time queue with finite capacity and i.i.d. and Markov modulated arrivals. Efficient algorithms are developed to calculate the moments and the distributions of the first time to overflow and the regeneration length. Results are extended to the multiserver queue. Some illustrative numerical examples are provided.

Keywords: Finite capacity queue, discrete time queue, algorithms, overflow time, regeneration cycle length, probability of overflow, Markov modulated arrivals.

1 Introduction

Although queues have traditionally been studied in continuous time (Cohen [9], Asmussen [1], Prabhu [15]), due to the arrival of the digital technology and in particular of the ATM based networks, discrete time queues have become increasingly important (see Bruneel [4], Bruneel and Kim [5] for recent surveys and Chu and Konheim [7], Kobayashi and Konheim [12], and COST report [16] for various applications).

In the ATM networks, not only a switch (node) is driven by a clock, but the messages are also split into equal sized "cells" which are transmitted on the output lines in a synchronised way. Also, the current switch architectures favour output queueing (COST report [16]). This makes each output link of a switch a discrete time queue that we study in this paper. Further, an ATM network is designed to support packetised voice and video which can tolerate very little delay and delay jitter. There may be real time traffic also which requires small delays in the network. Thus in ATM networks, to allow sharing of the channels for efficient utilisation and small delays by various users, small buffers at different switches

*On leave from the Department of Electrical Engineering, Indian Institute of Science, Bangalore 560 012, INDIA.

are being advocated. This makes the approximation by an infinite capacity queue highly erroneous. Thus, one needs to incorporate the finite buffer constraint in the queueing model itself. But then, in contrast to the infinite capacity queue, closed form expressions for various quantities of interest (even for the i.i.d. traffic model) are not available. In this paper we develop algorithms for a discrete time finite buffer queue.

We consider a discrete queue with a finite buffer size and with one or more servers. The service times are of one slot length. The arrival process could be i.i.d. or modulated by a finite state Markov chain. For a buffer length of size M, we obtain algorithms for calculating the mean and the higher moments of the busy period $\tau(M)$, moments of the first time to overflow $T_i^{(M)}$, and $p_i^{(M)}$, the probability of overflow in a busy period, where i is the initial number in the system. For sometime it is known that the asymptotics of these quantities, as $M \to \infty$, are related (Sharma and Gangadhar [18]). However, a surprising result of this work is that even the above algorithms which compute these quantities exactly for finite M are related and have similar structure. An advantage of our algorithms is that if we know these quantities for buffer lengths $0, \ldots, M$, then calculating for $M+1$ does not require much effort. This is particularly useful in the design stage because then one can find the minimum buffer length required to meet certain performance requirements. This can also be useful if one is interested in observing the behaviour of these parameters as M changes. In another study, we will obtain asymptotic behaviour of $\mathbb{E}\left[(T_0^{(M)})^\alpha\right]$, $\mathbb{E}\left[(\tau(M))^\alpha\right]$, and $p_0^{(M)}$ as M tends to infinity.

We elaborate on the theoretical and practical relevance of the results of this paper. The busy periods $\tau(M)$ (rather, the regeneration lengths) are of fundamental importance in queueing theory and have been studied in all works on queueing. Its mean $\mathbb{E}\left[\tau(M)\right]$ is required in the important formulae of calculating the stationary distributions (see, *e.g.*,Asmussen [1, p.126]), while the second moment is required in the asymptotic variance ([1, p.137]). Futhermore, moments of regeneration lengths are involved in obtaining various rates of convergence and bounds on the α–mixing coefficients (Sharma [17]). We also obtain moments of $\tau_i(M)$, the first time the queue becomes empty when the initial queue length is i. In addition to the moments of $\tau(M)$, these provide the practically useful quantity – the amount of time it takes for the effect of congestion (if i is large) to disappear. Its moments are also useful in obtaining various rates of convergence in the limit theorems. The next quantity studied, $T_0^{(M)}$, provides the first time an overflow occurs if the queue starts empty (or, with an initial queue length i; then we study $T_i^{(M)}$, as explained in the next section). For many practical systems, $T_i^{(M)}$, $i \geq 0$, is actually a more useful quantity than the stationary probability of loss. In fact, for voice traffic, $T_i^{(M)}$, along with the fact that the losses occur in clusters provides better information (than the stationary probability of loss) to judge the quality of voice received from the queue (see Cidon *et.al.* [8] and, in particular, Sharma and Gangadhar [18] for much more structure of the loss process). Finally we will study $p_0^{(M)}$, the probability that the queue length exceeds M in a regeneration cycle, which is relevant in obtaining the tail probabilities of the stationary distribution and also the asymptotics of $T_0^{(M)}$ (Sharma and Gangadhar [18]).

Now we describe the related literature. Probability of overflow is usually studied under stationarity (see *e.g.,* Takine *et.al.* [19], Cidon *et.al.* [8], COST report [16], for discrete queues and Bisdikian *et.al.* [3], Tijms [20], and Kofmann *et.al.* [13] for continuous time queues). Algorithms for the busy period and the first passage times of the queue length for a finite buffer discrete time queue with geometrically distributed interarrival and service times are given in Chaudhary and Zhao [6]. Their assumptions, methods (Laplace transform techniques), and the algorithms are entirely different from ours. The quantity $p_0^{(M)}$ considered above is much studied in continuous time because of its relevance to $T_0^{(M)}$ and the tails of the stationary probabilities. For examples and other references see Glasserman and Kou [11], Asmussen and Perry [2], and Sharma and Gangadhar [18]. But we are not aware of any algorithms to calculate $p_i^{(M)}$. Also, a multiserver finite buffer queue which has some essential differences from single server queue, does not appear to have been studied till now.

The rest of the paper is organised as follows. In Section 2 we consider the i.i.d. case. The algorithms are first developed for the single server queue and then extended to the multiserver queue. In Section 3 we develop the algorithms for Markov modulated traffic.

2 Queues with I.I.D. Traffic

We consider a single server queue in discrete time with a finite buffer size M. The time axis is divided into slots of equal length (taken as one unit) with slot k being the time segment $[k, k+1)$. Let X_k packets arrive in slot k which are assumed to arrive just after time k. These packets are eligible for transmission from time $k+1$ onward. The packets wait in the buffer before transmission. Any packet being transmitted in a slot is stored in a separate memory. The number of packets waiting for transmission at time k is denoted by W_k. Then if $(W_k - 1)^+ + X_k$ exceeds M, the excess packets are lost. Thus $\{W_k\}$ evolves as

$$W_{k+1} = \min\{(W_k - 1)^+ + X_k, M\}. \tag{1}$$

In this section, we assume that $\{X_k\}$ are i.i.d. Next section will generalise the results to the Markov modulated case. Also, towards the end of this section we will generalise our results to the case when the number of servers is S, $S \geq 1$.

If $\mathbb{P}(X_1 = 0) > 0$ and $\mathbb{P}(X_1 \leq 1) < 1$, then $\{W_k\}$ is a finite state, irreducible, aperiodic Markov chain. If $M = \infty$ and $\mathbb{E}[X_1] < 1$, then the chain is ergodic and closed form expressions for the stationary mean and higher moments of W_k and busy period under appropriate additional conditions are available (see, for instance, Bruneel and Kim [5] and Gangadhar [10]). For $M < \infty$, these expressions can provide upper bounds but closed form expressions are not available. Then one can solve

$$\pi \mathbf{P} = \pi \quad \text{and} \quad \sum_{\mathrm{i}=0}^{\mathrm{M}} \pi(\mathrm{i}) = 1,$$

where $\mathbf{P}$ is the transition probability matrix of $\{W_k\}$. From this one can obtain the stationary moments of W_k and waiting time. Now let $W_0 = 0$ and let τ be the first time after 0 when W_k again becomes zero. Observing that whenever $W_k = 0$, there is no transmission in the kth slot, one can show that

$$\mathbb{E}[\tau] = 1/\pi(0). \tag{2}$$

Similar expressions for $\mathbb{E}[\tau^\alpha]$, $\alpha > 1$ a positive integer, do not seem to be available. Also, for the i.i.d. case, τ can be considered as the regeneration length for $\{W_k\}$. The moments of regeneration lengths are of considerable interest. For the multiserver and Markov modulated generalisations, (2) does not provide the mean regeneration length. We obtain algorithms to compute the moments of the regeneration length for all these cases. It is also of interest to know when $W_k = 0$ for the first time after $k = 0$, if $W_0 = i$, $i \geq 0$. We will call it the first busy period, and denote it by $\tau_i(M)$.

In Subsection 2.1, we develop the algorithms for the moments and distributions of $\tau_i(M)$ while those of $T_i^{(M)}$ and $p_i^{(M)}$ are considered in Subsection 2.2. In Subsection 2.3, we derive an alternate set of algorithms. In Subsection 2.4, algorithms for a multiserver queue are developed.

2.1 Regeneration Lengths

Let $\tau_i(M)$ denote the first time, after zero, $W_k = 0$, if $W_0 = i \leq M$ (then $\tau_0(M) = \tau(M)$). By conditioning on the number of arrivals in the 0th slot,

$$\begin{aligned}
\mathbb{E}[\tau(M)] &= \mathbb{P}(X_0 = 0) + \sum_{k=1}^{M-1} (1 + \mathbb{E}[\tau_k(M)])\mathbb{P}(X_0 = k) \\
&\qquad + \mathbb{P}(X_0 \geq M)(1 + \mathbb{E}[\tau_M(M)]) \\
&= 1 + \sum_{k=1}^{M-1} \mathbb{P}(X_0 = k)\mathbb{E}[\tau_k(M)] + \mathbb{P}(X_0 \geq M)\mathbb{E}[\tau_M(M)]. \qquad (3)
\end{aligned}$$

Now we show the following result which will be repeatedly used in this paper:

For $1 \leq k \leq M$,

$$\tau_k(M) = \sum_{l=1}^{k} \tau(M - l + 1), \tag{4}$$

where $\tau(n), n = M - k + 1, \ldots, M$ can be taken to be independent. Consider the queue with $W_0 = k$. In the zeroth slot, one of these k packets departs while X_0 new packets arrive. From slot one onward we serve only the packets which arrive from the zeroth slot onward till all such packets are served. Now, for these packets, only $M - (k-1)$ buffers are available (since $k - 1$ buffers are occupied by packets which were in the system at time $k = 0$). Thus the time required to serve all the packets arriving from slot 0 onward is $\tau(M - (k - 1))$. In the slot after that, one of the initial $k-1$ packets is served and from then on only the new packets arriving from that slot onwards will be served. This procedure is repeated till all the initial k packets are also exhausted, providing us (4). Using (4), we can rewrite (3) as a set of recursive equations:

$$\begin{aligned}
\mathbb{E}[\tau(1)] &= 1 / \mathbb{P}(X_0 = 0), \\
\mathbb{E}[\tau(2)] &= \frac{1 + \mathbb{P}(X_0 \geq 2)\mathbb{E}[\tau(1)]}{\mathbb{P}(X_0 = 0)}, \\
\mathbb{E}[\tau(M)] &= \frac{1}{\mathbb{P}(X_0 = 0)} \Bigg[1 + \sum_{k=2}^{M-1} \mathbb{E}[\tau(M - k + 1)]\mathbb{P}(M - 1 \geq X_0 \geq k) \\
&\quad + \mathbb{P}(X_0 \geq M) \sum_{l=2}^{M} \mathbb{E}[\tau(M - l + 1)] \Bigg], \quad M > 2. \qquad (5)
\end{aligned}$$

We will obtain similar equations for the higher moments and moment generating function of $\tau(M)$.

Considering (3) we can obtain another algorithm for calculating $\mathbb{E}[\tau_i(M)]$: For a fixed M, in addition to (3), we can also write

$$\begin{aligned}
\mathbb{E}[\tau_1(M)] &= \mathbb{E}[\tau_0(M)] \\
\mathbb{E}[\tau_i(M)] &= 1 + \textstyle\sum_{k=0}^{M-i} \mathbb{P}(X = k)\mathbb{E}[\tau_{i-1+k}(M)] \qquad (6) \\
&\quad + \mathbb{P}(X \geq M - i + 1)\mathbb{E}[\tau_M(M)], \quad M \geq i \geq 2.
\end{aligned}$$

Thus we obtain a system of $M + 1$ equations with equal number of unknowns, $\mathbb{E}[\tau_i(M)]$, $i = 0, 1, \ldots, M$. Defining $a_i = \mathbb{P}(X_0 = i)$, and $\tilde{a}_i = \mathbb{P}(X_0 \geq i)$, this set is of the form $\mathbf{A\Gamma} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1-a_1 & -a_2 & \cdots & -a_{M-2} & -a_{M-1} & -\tilde{a}_M \\ 0 & -a_0 & 1-a_1 & \cdots & -a_{M-3} & -a_{M-2} & -\tilde{a}_{M-1} \\ & \cdots & & & \cdots & & \\ 0 & 0 & 0 & \cdots & -a_0 & 1-a_1 & -\tilde{a}_2 \\ 0 & 0 & 0 & \cdots & 0 & -a_0 & a_0 \end{bmatrix},$$

$\mathbf{\Gamma} = (\mathbb{E}[\tau_0(M)], \ldots, \mathbb{E}[\tau_M(M)])^T$, and $\mathbf{b} = (0, 1, \ldots, 1)^T$. When $a_1 < 1$, the matrix $\mathbf{A}$ is nonsingular and hence we will obtain a unique solution.

Algorithms (5) and (6) could be used at different times. For example, if one only wants $\mathbb{E}[\tau(M)]$ and would like to compare this value for different values of M, (5) should be preferred. On the other hand, if we have a system of buffer length M then (6) may be preferred because $\mathbb{E}[\tau_i(M)]$ are also of practical interest.

Now we obtain algorithms corresponding to (5) and (6) for $\mathbb{E}[(\tau(M))^2]$. The extension to $\mathbb{E}[(\tau(M))^n]$ for n a positive integer will be obvious. To obtain an algorithm corresponding to (5), we again use (4). Then, by independence, we have, for $1 \leq k \leq M$,

$$\begin{aligned} \mathbb{E}[(\tau_k(M))^2] &= \sum_{l=1}^{k} \mathbb{E}[(\tau(M-l+1))^2] \\ &\quad + \sum_{l=1}^{k} \sum_{n=1; n\neq l}^{k} \mathbb{E}[\tau(M-l+1)]\mathbb{E}[\tau(M-n+1)]. \end{aligned} \tag{7}$$

The equation corresponding to (3) now becomes

$$\begin{aligned} \mathbb{E}[(\tau(M))^2] &= 2\mathbb{E}[\tau(M)] - 1 \\ &\quad + \sum_{k=1}^{M-1} \mathbb{P}(X_0 = k)\mathbb{E}[(\tau_k(M))^2] + \tilde{a}_M \mathbb{E}[(\tau_M(M))^2]. \end{aligned} \tag{8}$$

Using (7) in this equation and interchanging the order of the summations, we obtain the following set of recursive equations for computing $\mathbb{E}[(\tau(M))^2]$:

$$\begin{aligned} \mathbb{E}[(\tau(1))^2] &= \frac{1}{\mathbb{P}(X_0 = 0)} [2\mathbb{E}[\tau(1)] - 1], \\ \mathbb{E}[(\tau(2))^2] &= \frac{1}{\mathbb{P}(X_0 = 0)} [2\mathbb{E}[\tau(2)] - 1 + \mathbb{P}(X_0 \geq 2)\mathbb{E}[(\tau(1))^2] \\ &\quad + 2\mathbb{P}(X_0 \geq 2)\mathbb{E}[\tau(1)]\mathbb{E}[\tau(2)]], \end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\left[(\tau(M))^2\right] \\
&= \frac{1}{\mathbb{P}(X_0=0)}\Bigg[2\mathbb{E}[\tau(M)] - 1 + \mathbb{P}(X_0 \geq M)\mathbb{E}\left[(\tau(1))^2\right] \\
&\quad + \sum_{l=2}^{M-1} \mathbb{P}(X_0 \geq l)\mathbb{E}\left[(\tau(M-l+1))^2\right] \\
&\quad + \sum_{k=1}^{M-1}\sum_{l=1}^{k}\sum_{n=1;n\neq l}^{k} \mathbb{P}(X_0 = k)\mathbb{E}[\tau(M-l+1)]\mathbb{E}[\tau(M-n+1)] \\
&\quad + \mathbb{P}(X_0 \geq M)\sum_{l=1}^{M}\sum_{n=1;n\neq l}^{M} \mathbb{E}[\tau(M-l+1)]\mathbb{E}[\tau(M-n+1)]\Bigg]. \qquad (9)
\end{aligned}$$

Corresponding to (6), we can get the following set of equations to solve for $\mathbb{E}\left[(\tau_i(M))^2\right]$:

$$\begin{aligned}
\mathbb{E}\left[(\tau_1(M))^2\right] &= \mathbb{E}\left[(\tau_0(M))^2\right] \\
&= 2\mathbb{E}[\tau_0(M)] - 1 + \textstyle\sum_{k=1}^{M-1} \mathbb{P}(X=k)\mathbb{E}\left[(\tau_k(M))^2\right] \\
&\qquad + \mathbb{P}(X > M-1)\mathbb{E}\left[(\tau_M(M))^2\right], \qquad (10) \\
\mathbb{E}\left[(\tau_i(M))^2\right] &= 2\mathbb{E}[\tau_i(M)] - 1 + \textstyle\sum_{k=0}^{M-i} \mathbb{P}(X=k)\mathbb{E}\left[(\tau_{i-1+k}(M))^2\right] \\
&\qquad + \mathbb{P}(X > M-i)\mathbb{E}\left[(\tau_M(M))^2\right], \quad M \geq i \geq 2.
\end{aligned}$$

This can be written as $\mathbf{A}\mathbf{\Gamma}_2 = \mathbf{b}_1$, where $\mathbf{\Gamma}_2 = \left(\mathbb{E}\left[(\tau_0(M))^2\right], \ldots, \mathbb{E}\left[(\tau_M(M))^2\right]\right)^T$ and $\mathbf{b}_1 = (0, 2\mathbb{E}[\tau_1(M)] - 1, \ldots, \mathbb{E}[\tau_M(M)] - 1)^T$. To solve this system of equations, we first have to obtain $\mathbb{E}[\tau_i(M)]$ from (6).

Once we have $\mathbb{E}[\tau(M)]$, we can also calculate the stationary probability of loss which equals (see Appendix)

$$1 - \frac{1}{\mathbb{E}[X]}\left[1 - \frac{1}{\mathbb{E}[\tau(M)]}\right].$$

Using Little's law, if $\mathbb{E}[W]$ is available, we can also calculate the mean waiting time of the packets that *enter* the queue.

For obtaining the distributions of $\tau_i(M)$, we proceed as follows: We can write

$$\begin{aligned}
\mathbb{P}\left(\tau_i(M)=1\right) &= \mathbb{P}\left(X_0=0\right) \quad \text{for } i=0 \text{ or } i=1 \\
&= 0 \quad \text{for } i \geq 2,
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left(\tau_i(M)=n\right) &= \textstyle\sum_{j=1}^{M-(i-1)^+} \mathbb{P}\left(X_0=j\right)\mathbb{P}\left(\tau_{(i-1)^++j}(M)=n-1\right) \qquad (11) \\
&\quad + \mathbb{P}\left(X_0 \geq M-(i-1)^+\right)\mathbb{P}\left(\tau_M(M)=n-1\right), \\
&\qquad n \geq 2,\ i \geq 0.
\end{aligned}$$

It is possible to solve this system of equations but unlike the case of $\mathbb{E}\left[(\tau_i(M))^n\right]$, now the number of equations to solve increases with n. A similar algorithm can be provided for $T_i^{(M)}$. Actually for M large, the distribution of $T_i^{(M)}$ can be approximated by an exponential distribution with mean $\mathbb{E}\left[T_i^{(M)}\right]$ which has already been calculated (Sharma and Gangadhar [18]). For $\tau_0(M)$, we can combine these equations with (4) to obtain moment generating functions efficiently as follows:

$$\begin{aligned}
\mathbb{E}\left[z^{\tau_0(1)}\right] &= \sum_{n=0}^{\infty} \mathbb{P}\left(\tau_0(1)=n\right)z^n \\
&= \sum_{n=0}^{\infty} a_0(1-a_0)^{n-1}z^n \\
&= \frac{a_0}{1-a_0}\cdot\frac{1}{1-(1-a_0)z},
\end{aligned}$$

$a_0 \neq 1$. For $M>1$, from (11),

$$\begin{aligned}
&\mathbb{E}\left[z^{\tau_0(M)}\right] \\
&= z\mathbb{P}\left(X_0=0\right) + \sum_{n=2}^{\infty} z^n \sum_{i=1}^{M-1} \mathbb{P}\left(X_0=i\right)\mathbb{P}\left(\tau_i(M)=n-1\right) \\
&\quad + \sum_{n=2}^{\infty} z^n \mathbb{P}\left(X_0 \geq M\right)\mathbb{P}\left(\tau_M(M)=n-1\right) \\
&= z\mathbb{P}\left(X_0=0\right) + z\sum_{i=1}^{M-1} \mathbb{P}\left(X_0=i\right)\sum_{n=2}^{\infty} z^{n-1}\mathbb{P}\left(\tau_i(M)=n-1\right) \\
&\quad + z\mathbb{P}\left(X_0 \geq M\right)\sum_{n=2}^{\infty} z^{n-1}\mathbb{P}\left(\tau_M(M)=n-1\right) \\
&= z\mathbb{P}\left(X_0=0\right) + z\sum_{i=1}^{M-1} \mathbb{P}\left(X_0=i\right)\mathbb{E}\left[z^{\tau_i(M)}\right] + z\mathbb{P}\left(X_0 \geq M\right)\mathbb{E}\left[z^{\tau_M(M)}\right].
\end{aligned}$$

From (4) we get

$$\mathbb{E}\left[z^{\mathcal{T}_i(M)}\right] = \prod_{l=1}^{i} \mathbb{E}\left[z^{\mathcal{T}_0(M-l+1)}\right]$$

and hence we obtain a recursive algorithm for calculating $\mathbb{E}\left[z^{\mathcal{T}_0(M)}\right]$.

2.2 $T_i^{(M)}$ and $p_i^{(M)}$

In this section we derive algorithms to compute the moments of the first time to overflow and the probability of an overflow in a regeneration cycle. Let $T_i^{(M)}$ denote the first time, after zero, $(W_k - 1)^+ + X_k$ exceeds M (hence overflow occurs), when $W_0 = i \leq M$. Let $p_i^{(M)}$ denote the probability that in the first busy period overflow occurs, with $W_0 = i \leq M$. Conditioning on the number of arrivals in the 0th slot, we get

$$\begin{aligned} \mathbb{E}\left[T_0^{(M)}\right] &= 1 + \textstyle\sum_{i=0}^{M} a_i \mathbb{E}\left[T_i^{(M)}\right] = \mathbb{E}\left[T_1^{(M)}\right] \\ \mathbb{E}\left[T_i^{(M)}\right] &= 1 + \textstyle\sum_{j=0}^{M-i+1} a_j \mathbb{E}\left[T_{i-1+j}^{(M)}\right], \ 2 \leq i \leq M. \end{aligned} \tag{12}$$

Equations (12) are in the form $\mathbf{RT} = \mathbf{b}$ where

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -a_0 & 1 - a_1 & -a_2 & \cdots & -a_{M-1} & -a_M \\ 0 & -a_0 & 1 - a_1 & \cdots & -a_{M-2} & -a_{M-1} \\ & \cdots & & \cdots & & \\ 0 & 0 & 0 & \cdots & -a_0 & 1 - a_1 \end{bmatrix},$$

$\mathbf{T} = \left(\mathbb{E}\left[T_0^{(M)}\right], \ldots, \mathbb{E}\left[T_M^{(M)}\right]\right)^T$, and $\mathbf{b} = (0, 1, \ldots, 1)^T$.

For $p_i^{(M)}$, we obtain,

$$\begin{aligned} p_0^{(M)} &= p_1^{(M)} = \tilde{a}_{M+1} + \textstyle\sum_{i=1}^{M} a_i p_i^{(M)} \\ p_i^{(M)} &= \tilde{a}_{M-i+2} + \textstyle\sum_{j=0}^{M-i+1} a_j p_{i-1+j}^{(M)}, \ 2 \leq i \leq M. \end{aligned} \tag{13}$$

These equations can be written as $\mathbf{Q\Pi} = \tilde{a}$, where

$$\mathbf{Q} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 - a_1 & -a_2 & \cdots & -a_{M-1} & -a_M \\ 0 & -a_0 & 1 - a_1 & \cdots & -a_{M-2} & -a_{M-1} \\ & \cdots & & \cdots & & \\ 0 & 0 & 0 & \cdots & -a_0 & 1 - a_1 \end{bmatrix},$$

$\mathbf{\Pi} = \left(p_0^{(M)}, \ldots, p_M^{(M)}\right)^T$, and $\widetilde{a} = (0, \widetilde{a}_{M+1}, \ldots, \widetilde{a}_2)^T$.

We can check that when $a_0 > 0$ and $a_1 < 1$, $\mathbf{R}$ and $\mathbf{Q}$ are nonsingular and hence the above systems of equations have unique solutions.

Following the above procedure, the system of equations for $\mathbb{E}\left[\left(T_i^{(M)}\right)^2\right]$ is :

$$\begin{aligned} \mathbb{E}\left[\left(T_0^{(M)}\right)^2\right] &= 2\mathbb{E}\left[T_0^{(M)}\right] - 1 + \textstyle\sum_{i=0}^{M} a_i \mathbb{E}\left[\left(T_i^{(M)}\right)^2\right] = \mathbb{E}\left[\left(T_1^{(M)}\right)^2\right] \\ \mathbb{E}\left[\left(T_i^{(M)}\right)^2\right] &= 2\mathbb{E}\left[T_i^{(M)}\right] - 1 + \textstyle\sum_{j=0}^{M-i+1} a_j \mathbb{E}\left[\left(T_{i-1+j}^{(M)}\right)^2\right], \; 2 \le i \le M \end{aligned} \tag{14}$$

This set can be written as $\mathbf{R}\mathbf{T}^{(2)} = \mathbf{b}_2$, where $\mathbf{T}^{(2)} = \left(\mathbb{E}\left[\left(T_0^{(M)}\right)^2\right], \ldots, \mathbb{E}\left[\left(T_M^{(M)}\right)^2\right]\right)^T$, $\mathbf{R}$ is as in (12), and $\mathbf{b}_2 = (0, 2\mathbb{E}\left[T_1^{(M)}\right] - 1, \ldots, 2\mathbb{E}\left[T_M^{(M)}\right] - 1)^T$.

The matrices $\mathbf{A}$, $\mathbf{R}$, and $\mathbf{Q}$ have similar structure and these equations can be solved in $\frac{1}{2}(M^2 + 3M)$ multiplications. If we only want $\mathbb{E}\left[\tau_0(M)\right]$ and want to study them as functions of M, (5) provides a more efficient way.

Now we provide efficient algorithms for $\mathbb{E}\left[T_i^{(M)}\right]$ and $\mathbb{E}\left[(T_i^{(M)})^2\right]$. We will require two different algorithms for these two quantities. To calculate $\mathbb{E}\left[(T_i^{(M)})^n\right]$, $n = 1, 2$, once we have $\mathbb{E}\left[(T_i^{(M')})^n\right]$, $M' = 1, \cdots, M-1$, will require $O(M)$ multiplications. First we describe the algorithm for $\mathbb{E}\left[T_i^{(M)}\right]$; the algorithm for computing $p_i^{(M)}$ is obtained by a simple modification of this.

In the following, $\mathbf{A}_M^{(T)}$, and $\mathbf{B}_M$ are $M-1$ dimensional column vectors and $\mathbf{C}_M$ is an $M-1$ dimensional row vector.

By solving the equations for the case $M = 2$, we get

$$\begin{aligned} \mathbb{E}\left[T_0^{(2)}\right] = \mathbb{E}\left[T_1^{(2)}\right] &= \frac{1 + \mathbf{C}_2 \cdot \mathbf{A}_2^{(T)}}{1 - a_0 - a_1 - \mathbf{C}_2 \cdot \mathbf{B}_2}, \\ \mathbb{E}\left[T_2^{(2)}\right] &= \mathbf{A}_2^{(T)} + \mathbf{B}_2 \mathbb{E}\left[T_0^{(2)}\right], \end{aligned}$$

where

$$\mathbf{A}_2^{(T)} = \frac{1}{1 - a_1}; \quad \mathbf{B}_2 = \frac{a_0}{1 - a_1}; \quad \mathbf{C}_2 = a_2.$$

Now we develop equations recursive in M. Consider any $M \ge 3$. Suppose that the following holds for such an M (it does for $M = 3$):

$$\left(\mathbb{E}\left[T_2^{(M-1)}\right], \ldots, \mathbb{E}\left[T_{M-1}^{(M-1)}\right]\right)^T = \mathbf{A}_{M-1}^{(T)} + \mathbf{B}_{M-1} \mathbb{E}\left[T_1^{(M-1)}\right].$$

Then the second equation from the set $\mathbf{RT} = \mathbf{b}$ gives

$$\mathbb{E}\left[T_1^{(M-1)}\right] = \frac{1 + \mathbf{C}_{M-1} \cdot \mathbf{A}_{M-1}^{(T)}}{1 - a_1 - \mathbf{C}_{M-1} \cdot \mathbf{B}_{M-1}} + \frac{a_0}{1 - a_1 - \mathbf{C}_{M-1} \cdot \mathbf{B}_{M-1}} \mathbb{E}\left[T_0^{(M-1)}\right].$$

Hence we obtain

$$\left(\mathbb{E}\left[T_1^{(M-1)}\right], \ldots, \mathbb{E}\left[T_{M-1}^{(M-1)}\right]\right)^T = \mathbf{A}_M^{(T)} + \mathbf{B}_M \mathbb{E}\left[T_0^{(M-1)}\right],$$

where

$$\begin{aligned} \mathbf{A}_M^{(T)} &= \begin{pmatrix} 0 \\ \mathbf{A}_{M-1}^{(T)} \end{pmatrix} + \frac{1 + \mathbf{C}_{M-1} \cdot \mathbf{A}_{M-1}^{(T)}}{1 - a_1 - \mathbf{C}_{M-1} \cdot \mathbf{B}_{M-1}} \begin{pmatrix} 1 \\ \mathbf{B}_{M-1} \end{pmatrix}, \\ \mathbf{B}_M &= \frac{a_0}{1 - a_1 - \mathbf{C}_{M-1} \cdot \mathbf{B}_{M-1}} \begin{pmatrix} 1 \\ \mathbf{B}_{M-1} \end{pmatrix}, \end{aligned} \tag{15}$$

and

$$\mathbf{C}_M = (\mathbf{C}_{M-1},\, a_M), \quad M \geq 3.$$

Now we exploit the fact that the last $M-1$ equations of $\mathbf{RT} = \mathbf{b}$ for the case of $\mathbb{E}\left[T_i^{(M)}\right]$ remain the same as the last $M-1$ equations (but for the new index M) for $\mathbb{E}\left[T_i^{(M-1)}\right]$. Hence,

$$\left(\mathbb{E}\left[T_2^{(M)}\right], \ldots, \mathbb{E}\left[T_M^{(M)}\right]\right)^T = \mathbf{A}_M^{(T)} + \mathbf{B}_M \mathbb{E}\left[T_1^{(M)}\right]. \tag{16}$$

From the second of the equations in the set $\mathbf{RT} = \mathbf{b}$, we obtain

$$\mathbb{E}\left[T_1^{(M)}\right] = \frac{1 + \mathbf{C}_M \cdot \mathbf{A}_M^{(T)} + a_0 \mathbb{E}\left[T_0^{(M)}\right]}{1 - a_1 - \mathbf{C}_M \cdot \mathbf{B}_M}. \tag{17}$$

The first equation then implies

$$\mathbb{E}\left[T_0^{(M)}\right] = \mathbb{E}\left[T_1^{(M)}\right] = \frac{1 + \mathbf{C}_M \cdot \mathbf{A}_M^{(T)}}{1 - a_0 - a_1 - \mathbf{C}_M \cdot \mathbf{B}_M}. \tag{18}$$

Thus we have obtained expressions for $\mathbb{E}\left[T_i^{(M)}\right]$ given the equations for $M-1$.

For $\mathbb{E}\left[(T_i^{(M)})^2\right]$ the above technique does not apply as the right hand sides of these equations change with M. For these, we describe another algorithm. We rewrite $\mathbf{RT}^{(2)} = \mathbf{b}_2$ by shifting the first row to the bottom as

$$\begin{bmatrix} \mathbf{R}_M & \mathbf{r}_{1M} \\ \mathbf{r}_{2M} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{t}_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 \\ 0 \end{bmatrix},$$

where $\mathbf{R}_M$ is an $M \times M$ matrix, $\mathbf{r}_{1M}$ is $(-a_M, -a_{M-1}, \cdots, -a_2, 1-a_1)^T$, and $\mathbf{r}_{2M}$ is $(1, -1, 0, \cdots, 0)$. Similarly, $\mathbf{t}_1$, t_2, $\mathbf{d}_1$ are obviously defined. The last equation of this system of equations simply says that $\mathbb{E}\left[(T_0^{(M)})^2\right] = \mathbb{E}\left[(T_1^{(M)})^2\right]$. From the rest of the equations we get

$$\mathbf{R}_M \mathbf{t}_1 + \mathbf{r}_{1M} t_2 = \mathbf{d}_1$$

and hence

$$\mathbf{t}_1 = \mathbf{R}_M^{-1}(\mathbf{d}_1 - \mathbf{r}_{1M} t_2). \tag{19}$$

Since the first two components of $\mathbf{t}_1$ are the same, we thus get

$$t_2 = \frac{(\mathbf{R}_M^{-1}\mathbf{d}_1)_1 - (\mathbf{R}_M^{-1}\mathbf{d}_1)_2}{(\mathbf{R}_M^{-1}\mathbf{r}_{1M})_1 - (\mathbf{R}_M^{-1}\mathbf{r}_{1M})_2},$$

where $(\mathbf{X})_i$ denotes the ith component of vector $\mathbf{X}$.

Now from (19), we can obtain $\mathbb{E}\left[(T_0^{(M)})^2\right] = (\mathbf{t}_1)_1$, once we have $\mathbf{R}_M^{-1}$, in $5M$ multiplications.

We now obtain an efficient algorithm to calculate $\mathbf{R}_M^{-1}$, once $\mathbf{R}_{M-1}^{-1}$ is known. Since

$$\mathbf{R}_M = \begin{bmatrix} -a_0 & 1-a_1 & -a_2 & \cdots & -a_{M-1} \\ 0 & -a_0 & 1-a_1 & \cdots & -a_{M-2} \\ 0 & 0 & -a_0 & \cdots & -a_{M-3} \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & -a_0 \end{bmatrix},$$

it is a triangular Toeplitz matrix. Its inverse is also a triangular Toeplitz matrix (see Yen [21]),

$$\mathbf{R}_M^{-1} \triangleq \begin{bmatrix} d_0 & d_1 & d_2 & \cdots & d_{M-1} \\ 0 & d_0 & d_1 & \cdots & d_{M-2} \\ 0 & 0 & d_0 & \cdots & d_{M-3} \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & d_0 \end{bmatrix}$$

where

$$\begin{aligned} d_0 &= -\frac{1}{a_0}, \\ d_n &= \frac{1}{a_0}\left((1-a_1)d_{n-1} - \sum_{k=0}^{n-2} a_{n-k} d_k\right), \quad 1 \le n \le M-1. \end{aligned} \tag{20}$$

Since $\mathbf{R}_M$ and $\mathbf{R}_{M-1}$ have the same first $M-1$ components, d_n is same for $\mathbf{R}_M^{-1}$ and $\mathbf{R}_{M-1}^{-1}$, for $n = 0, \cdots, M-2$, and thus we only need to calculate d_{M-1}

which requires M multiplications. Thus, our algorithm requires $6M$ multiplications to calculate $\mathbf{R}_M^{-1}$ and $\mathbb{E}\left[(T_0^{(M)})^2\right]$ once we have $\mathbf{R}_{M-1}^{-1}$, $\mathbb{E}\left[(T_0^{(M-1)})^2\right]$, and $\mathbb{E}\left[T_i^{(M)}\right]$. We can, of course, use this algorithm for calculating $\mathbb{E}\left[T_i^{(M)}\right]$ also; but the complexity will be of the same order as for the algorithm given above.

The procedure for computing the distributions, presented in Section 2.1, can be seen to be easily adaptable for finding the distributions of $T_i^{(M)}$.

2.3 Another Set of Algorithms

It is instructive to compare the algorithms presented in the previous subsections with another set of algorithms. Instead of (3) we now write $\mathbb{E}\left[\tau_i(M)\right]$ for $i > 0$ as

$$\mathbb{E}\left[\tau_i(M)\right] = \sum_{n=0}^{\infty} \mathbb{P}\left(\tau_i(M) \geq n\right). \tag{21}$$

Define $\widehat{\mathbf{Q}}$ as a matrix obtained by deleting the zeroth row and column from $\mathbf{P}$. Then we obtain

$$\mathbb{E}\left[\tau_i(M)\right] = \sum_{n=1}^{\infty} \sum_{j=1}^{M} \widehat{\mathbf{Q}}_{ij}^{(n)} = \sum_{j=1}^{M} \mathbf{G}_{ij} \tag{22}$$

where $\mathbf{G} = \left(\mathbf{I} - \widehat{\mathbf{Q}}\right)^{-1}$ exists whenever $\mathbf{P}$ is irreducible. To obtain $\mathbb{E}\left[\tau(M)\right]$ we now use (3). One can similarly show that for $i > 0$,

$$\mathbb{E}\left[(\tau_i(M))^2\right] = 2\sum_{j=1}^{M} \mathbf{G}_{ij}^2 - \sum_{j=1}^{M} \mathbf{G}_{ij} \tag{23}$$

and then obtain $\mathbb{E}\left[(\tau(M))^2\right]$ as above. Similarly one can develop algorithms for $\mathbb{E}\left[(\tau_i(M))^n\right]$. It is easy to show that these algorithms are more expensive than those developed earlier although for the case of $\mathbb{E}\left[T_i^{(M)}\right]$ we can, after some manipulation, obtain $\mathbf{RT} = \mathbf{b}$. It is interesting that we can obtain $p_i^{(M)}$ also from algorithms of the type of (22) using Pitman [14, p.85]. From $\{W_k\}$ define a Markov chain $\{\widetilde{W}_k\}$ as $\widetilde{W}_k = W_k$ if there is no overflow in slot $(k-1)$, otherwise $\widetilde{W}_k = M+1$. From the transition probability matrix of $\{\widetilde{W}_k\}$, obtain $\widetilde{\mathbf{Q}}$ by deleting the last row and column. Then $\mathbb{E}\left[(T_i^{(M)})^n\right]$ is obtained from (22)–(23) if we replace $\widehat{\mathbf{Q}}$ by $\widetilde{\mathbf{Q}}$. Let $\widetilde{\mathbf{G}} = (\mathbf{I} - \widetilde{\mathbf{Q}})^{-1}$ and let $\mathbf{G}_0$ be obtained from $\widetilde{\mathbf{G}}$ by replacing all columns except the zeroth by zeroes. Let N be a r.v. with distribution

$$\mathbb{P}\left(N = n\right) = p_0^{(M)}(1 - p_0^{(M)})^n.$$

Then from Pitman [14],

$$\mathbb{E}\left[N\right] = \frac{1}{p_0^{(M)}} - 1 = \sum_{j=1}^{M} \left(\mathbf{G}_0\right)_{0j}$$

If $\widetilde{\mathbf{G}}$ is known, then this is a less expensive way to calculate $p_0^{(M)}$ than (13).

2.4 Multiserver Queue

Now we extend the algorithms to the multiserver queue with S, $S \geq 1$ servers. The process $\{W_k\}$ evolves as

$$W_{k+1} = \min\{(W_k - S)^+ + X_k,\ M\}.$$

We consider the regeneration epochs as the instants when $W_k \leq S$. An equation corresponding to (4) does not hold for $S > 1$. If we take the regeneration epochs as instants when $W_k < S$, then we would get the inequalities

$$\begin{aligned} \tau_k(M) &= \tau_0(M), \quad \text{for } k \leq S \\ \tau_k(M) &\leq \sum_{j=1}^{k} \tau(\min\{M\,,\ M - k + jS\}). \end{aligned}$$

For S large this bound could be quite weak and also does not provide algorithms for computing $\mathbb{E}[\tau(M)]$. Thus we retain the earlier definition of the regeneration epochs.

The equations (6) get modified to

$$\begin{aligned} \mathbb{E}[\tau_i(M)] &= 1 + \textstyle\sum_{j=1}^{M-1} \mathbb{P}(X = j)\mathbb{E}[\tau_j(M)] \\ &\qquad + \mathbb{P}(X \geq M))\mathbb{E}[\tau_M(M)], \quad 0 \leq i \leq S, \\ &= 1 + \textstyle\sum_{j=0}^{M-i+S-1} \mathbb{P}(X = j)\mathbb{E}[\tau_{i-S+j}(M)] \\ &\qquad + \mathbb{P}(X \geq M - i + S))\mathbb{E}[\tau_M(M)], \quad S < i \leq M \end{aligned} \tag{24}$$

One can similarly develop algorithms corresponding to (10).

Unlike for the single server queue, from $\mathbb{E}[\tau(M)]$ now we can only get bounds on the stationary probability of packet loss (Gangadhar [10]):

$$1 - \frac{S}{\mathbb{E}[X]} \leq \text{Stat. Prob. of Loss } \leq 1 - \frac{S}{\mathbb{E}[X]}\left[1 - \frac{1}{\mathbb{E}[\tau(M)]}\right].$$

and using Little's law, these can provide upper and lower bounds on the mean waiting times of the packets entering the queue.

The equations for $\mathbb{E}\left[T_i^{(M)}\right]$ become

$$\begin{aligned} \mathbb{E}\left[T_i^{(M)}\right] &= 1 + \textstyle\sum_{j=0}^{M} \mathbb{P}(X = j)\mathbb{E}\left[T_j^{(M)}\right], \quad 0 \leq i \leq S \\ &= 1 + \textstyle\sum_{j=0}^{M-(i-S)} \mathbb{P}(X = j)\mathbb{E}\left[T_{i-S+j}^{(M)}\right], \quad S < i \leq M \end{aligned} \tag{25}$$

One can also obtain similar extensions for (14) and (13). One again observes that a triangular Toeplitz matrix can be extracted from the equations for the multiserver queue. We can thus develop efficient algorithms as for the single server case.

3 The Markov Modulated Case

In this section we extend the algorithms presented in Section 2 to the queues with Markov modulated arrivals (as described below).

Let $\{Y_k\}$ be a Markov chain with state space $\{0,1,2,\ldots,N-1\}$ and probability transition matrix $\{p_{ij}\}$. We define

$$\mathbb{P}\left(X_n = k; Y_n = j \,|\, X_{n-1},\ldots,X_0, Y_{n-1}=i,\ldots,Y_0\right) = C_k^{ij},$$

for $k \geq 0$. We denote by W_k the workload at time k and the buffer length is M. We consider the single server queue although extension to multiserver queue is possible as in Section 2.

Define the following random variables:

$$\begin{aligned}
\tau_{n,i}^{(M)} &= \inf\{k \geq 0 : W_k = 0 \,|\, W_0 = n, Y_0 = i\},\\
T_{n,i}^{(M)} &= \inf\{k \geq 0 : \text{Queue overflows at time } k \,|\, W_0 = n, Y_0 = i\},\\
p_{n,i}^{(M)} &= \mathbb{P}(\text{overflow occurs at some } k \text{ and } W_j > 0, 0 \leq j \leq k \,|\\
&\qquad\qquad W_0 = n, Y_0 = i).
\end{aligned}$$

The quantities $\mathbb{E}\left[T_{n,i}^{(M)}\right]$ satisfy the following equations (*cf.* (12)):

$$\begin{aligned}
\mathbb{E}\left[T_{0,i}^{(M)}\right] &= 1 + \textstyle\sum_{k=0}^{M}\sum_{j=0}^{N-1} C_k^{ij}\,\mathbb{E}\left[T_{k,j}^{(M)}\right] = \mathbb{E}\left[T_{1,i}^{(M)}\right]\\
\mathbb{E}\left[T_{n,i}^{(M)}\right] &= 1 + \textstyle\sum_{k=0}^{M-n+1}\sum_{j=0}^{N-1} C_k^{ij}\,\mathbb{E}\left[T_{n-1+k,j}^{(M)}\right], \quad 2 \leq n \leq M.
\end{aligned} \tag{26}$$

Similarly, $\{p_{n,i}^{(m)}\}$ satisfies

$$\begin{aligned}
p_{0,i}^{(M)} &= \widetilde{a}_{M+1,i} + \textstyle\sum_{k=1}^{M}\sum_{j=0}^{N-1} C_k^{ij} p_{k,j}^{(M)} = p_{1,i}^{(M)}\\
p_{n,i}^{(M)} &= \widetilde{a}_{M-n+2,i} + \textstyle\sum_{k=0}^{M-n+1}\sum_{j=0}^{N-1} C_k^{ij} p_{n-1+k,j}^{(M)}, \quad 2 \leq n \leq M,
\end{aligned} \tag{27}$$

where $\widetilde{a}_{m,i} = \mathbb{P}\left(X_0 \geq m \,|\, Y_0 = i\right),\ i \geq 0$.

Now we see that the $\{\mathbb{E}\left[T_{n,i}^{(M)}\right]\}$ and $\{p_{n,i}^{(M)}\}$ satisfy $\mathbf{RT} = \mathbf{B}$ and $\mathbf{Q\Pi} = \mathbf{A}$

with

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{b}_0 & \mathbf{I}-\mathbf{b}_1 & -\mathbf{b}_2 & \cdots & -\mathbf{b}_{M-2} & -\mathbf{b}_{M-1} & -\mathbf{b}_M \\ \mathbf{0} & -\mathbf{b}_0 & \mathbf{I}-\mathbf{b}_1 & \cdots & -\mathbf{b}_{M-3} & -\mathbf{b}_{M-2} & -\mathbf{b}_{M-1} \\ & \cdots & & \cdots & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{b}_0 & \mathbf{I}-\mathbf{b}_1 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}-\mathbf{b}_1 & -\mathbf{b}_2 & \cdots & -\mathbf{b}_{M-2} & -\mathbf{b}_{M-1} & -\mathbf{b}_M \\ \mathbf{0} & -\mathbf{b}_0 & \mathbf{I}-\mathbf{b}_1 & \cdots & -\mathbf{b}_{M-3} & -\mathbf{b}_{M-2} & -\mathbf{b}_{M-1} \\ & \cdots & & \cdots & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & -\mathbf{b}_0 & \mathbf{I}-\mathbf{b}_1 \end{bmatrix},$$

$$\mathbf{T} = \left[\left(\mathbb{E}\left[T_{0,1}^{(M)}\right], \cdots, \mathbb{E}\left[T_{0,N}^{(M)}\right]\right), \cdots, \left(\mathbb{E}\left[T_{M,1}^{(1)}\right], \cdots, \mathbb{E}\left[T_{M,N}^{(M)}\right]\right)\right]^T,$$

$$\mathbf{\Pi} = \left[\left(p_{0,1}^{(M)}, \cdots, p_{0,N}^{(M)}\right), \cdots, \left(p_{M,1}^{(M)}, \cdots, p_{M,N}^{(M)}\right)\right]^T,$$

$$\begin{aligned} \mathbf{B} &= \left(\tilde{\mathbf{0}}, \mathbf{1}, \cdots, \mathbf{1}\right)^T, \\ \mathbf{A} &= \left(\tilde{\mathbf{0}}, \tilde{\mathbf{a}}_{m+1}, \tilde{\mathbf{a}}_m, \cdots, \tilde{\mathbf{a}}_2\right)^T, \end{aligned}$$

where $\mathbf{I}$ is an $N \times N$ identity matrix, $\mathbf{0}$ is an $N \times N$ all zeroes matrix, $\tilde{\mathbf{0}} = (0, 0, \cdots, 0)^T$, and $\mathbf{1} = (1, 1, \cdots, 1)^T$ are $N \times 1$ column vectors,

$$\tilde{\mathbf{a}}_m = (\mathbb{P}\,(X_0 \geq m \,|\, Y_0 = i), i \geq 0)^T,$$

and

$$\mathbf{b}_k = \begin{bmatrix} C_k^{0,0} & C_k^{0,2} & \cdots & C_k^{0,N} \\ C_k^{1,1} & C_k^{1,2} & \cdots & C_k^{1,N} \\ \vdots & & \ddots & \vdots \\ C_k^{N-1,1} & C_k^{N-1,2} & \cdots & C_k^{N-1,N-1} \end{bmatrix}.$$

One can similarly write the equations for $\mathbb{E}\left[\tau_{n,i}^{(M)}\right]$ and for the higher moments. An equation corresponding to (4) can also be written and exploited to obtain efficient algorithms.

To compute $\mathbb{E}\left[T_{n,i}^{(M)}\right]$, the algorithm presented in Section 2.2 can be extended to the Markov modulated case as below. This algorithm requires $O(MN^3)$ number of multiplications once the values of $\mathbb{E}\left[T_{n,i}^{(M-1)}\right]$ are known. In the following, we write $\mathbf{T}_n = \left(\mathbb{E}\left[T_{n,0}^{(M)}\right], \cdots, \mathbb{E}\left[T_{n,N-1}^{(M)}\right]\right)^T$.

$\mathbf{M=2}$:

$$\mathbf{D}_2^{(T)} = (\mathbf{I}-\mathbf{b}_1)^{-1}\cdot\mathbf{1}\,;\ \mathbf{E}_2 = (\mathbf{I}-\mathbf{b}_1)^{-1}\cdot\mathbf{b}_0.$$

$$\mathbf{T}_0 = \mathbf{T}_1 = (\mathbf{I}-\mathbf{b}_0-\mathbf{b}_1-\mathbf{C}_2\cdot\mathbf{E}_2)^{-1}\cdot(\mathbf{1}+\mathbf{C}_2\cdot\mathbf{D}_2^{(T)}),$$

$$\mathbf{T}_2 = \mathbf{D}_2^{(T)} + \mathbf{E}_2\cdot\mathbf{T}_1$$

$\mathbf{M\geq 3}$:

$$\begin{aligned}
\mathbf{D}_M^{(T)} &= \begin{pmatrix}\tilde{\mathbf{0}}\\ \mathbf{D}_{M-1}^{(T)}\end{pmatrix}\\
&\quad + \begin{pmatrix}\mathbf{I}\\ \mathbf{E}_{M-1}\end{pmatrix}\cdot(\mathbf{I}-\mathbf{b}_1-\mathbf{C}_{M-1}\cdot\mathbf{E}_{M-1})^{-1}\cdot(\mathbf{1}+\mathbf{C}_{M-1}\cdot\mathbf{D}_{M-1}^{(T)}),\\
\mathbf{E}_M &= \begin{pmatrix}\mathbf{I}\\ \mathbf{E}_{M-1}\end{pmatrix}\cdot(\mathbf{I}-\mathbf{b}_1-\mathbf{C}_{M-1}\cdot\mathbf{E}_{M-1})^{-1}\cdot\mathbf{b}_0,\\
\mathbf{C}_M &= (\mathbf{C}_{M-1}\,,\ \mathbf{b}_M).
\end{aligned}$$

$$\mathbf{T}_0 = \mathbf{T}_1 = (\mathbf{I}-\mathbf{b}_0-\mathbf{b}_1-\mathbf{C}_M\cdot\mathbf{E}_M)^{-1}\cdot(\mathbf{1}+\mathbf{C}_M\cdot\mathbf{D}_M^{(T)}),$$

$$(\mathbf{T}_2,\cdots,\mathbf{T}_M)^T = \mathbf{D}_M^{(T)} + \mathbf{E}_M\cdot\mathbf{T}_1.$$

Now we obtain a result corresponding to the triangular Toeplitz matrix result (20) of Section 2. Let $\widehat{\mathbf{A}}$ be a block triangular Toeplitz matrix of the type obtained from $\mathbf{R}$ in this section. Let $\widehat{\mathbf{D}}$ be its inverse. Let

$$\widehat{\mathbf{A}} = \begin{bmatrix}\mathbf{b}_0 & \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M\\ 0 & \mathbf{b}_0 & \mathbf{b}_1 & \cdots & \mathbf{b}_{M-1}\\ 0 & 0 & \mathbf{b}_0 & \cdots & \mathbf{b}_{M-2}\\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & \mathbf{b}_0\end{bmatrix}$$

Then a matrix

$$\widehat{D} = \begin{bmatrix}\mathbf{d}_0 & \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_M\\ 0 & \mathbf{d}_0 & \mathbf{d}_1 & \cdots & \mathbf{d}_{M-1}\\ 0 & 0 & \mathbf{d}_0 & \cdots & \mathbf{d}_{M-2}\\ & \vdots & & \ddots & \\ 0 & 0 & 0 & \cdots & \mathbf{d}_0\end{bmatrix}$$

is its inverse if it satisfies $\widehat{\mathbf{A}}\widehat{\mathbf{D}} = \mathbf{I}$. Thus

$$\begin{aligned}
\mathbf{d}_0 &= \mathbf{b}_0^{-1}\\
\mathbf{d}_1 &= -\mathbf{b}_0^{-1}\mathbf{b}_1\mathbf{d}_0
\end{aligned}$$

$$\begin{aligned}\mathbf{d}_2 &= -\mathbf{b}_0^{-1}(\mathbf{b}_1\mathbf{d}_1 + \mathbf{b}_2\mathbf{d}_0) \\ &\vdots \\ \mathbf{d}_n &= -\mathbf{b}_0^{-1}(\mathbf{b}_1\mathbf{d}_{n-1} + \cdots + \mathbf{b}_n\mathbf{d}_0)\end{aligned}$$

Therefore, if $\mathbf{b}_0^{-1}$ exists, then $\widehat{\mathbf{D}}$ of given structure is $\widehat{\mathbf{A}}^{-1}$. Then computing the inverse of $\widehat{\mathbf{A}}$ is not expensive once we know the inverse for dimension $M-1$. This can be readily used when applying the algorithms of Section 2 corresponding to $\mathbb{E}\left[(T_{n,i}^{(M)})^2\right]$ to the Markov modulated case.

4 Numerical Examples

In this section we provide some numerical examples. We consider both i.i.d. and Markov modulated cases. First we describe the examples and towards the end of the section we discuss the results and provide some concluding remarks.

4.1 I.I.D. case

a) **An approximate voice source model**

Consider an ATM multiplexer whose outgoing link speed is 1.536Mbits/sec. The incoming traffic is a superposition of, say, N number of voice calls. Each voice source generates ATM cells (53 bytes) in the following manner: The analog speech waveform is sampled at a Nyquist rate of 8 kbits/s. and each sample is encoded into a 8–bit digital signal (thus the traffic generated by an active source is 64 kbits/s). Hence, every 6ms., 48 bytes of information (that can fill an ATM cell) is available from the source. Thus each voice source in its active state generates 166.67 ATM cells/sec. With a buffer to take care of the instantaneous excess (above the server capacity of one cell), arrivals from a maximum of 22 voice calls can be served by the multiplexer and the channel.

By considering the time of transmission of an ATM cell on the outgoing channel (which is 0.276ms.) as a unit of time, the multiplexer and the output channel can be modelled as a discrete time queue. A voice source in active state generates an ATM cell every 22 slots.

In this example, we make the assumption that each of the voice sources can be modelled as a Bernoulli source with parameter p and that the sources are independent. The parameter is fixed by matching the mean number of arrivals from the voice source in a slot (which is, as seen above, 1/22) to the mean of the Bernoulli process, p. Thus $p = 0.045$. With N number of voice sources multiplexed, the load experienced by the multiplexer is Np.

We considered two scenarios: one with $N = 19$, corresponding to a load of

0.855, and another with $N = 14$, which corresponds to a load of 0.63. The corresponding variances are 0.817 and 0.602.

The quantities $\mathbb{E}\left[T_i^{(M)}\right]$, $\mathbb{E}\left[(T_i^{(M)})^2\right]$, $p_i^{(M)}$, and $\mathbb{E}\left[\tau_i(M)\right]$ were computed for $M = 2, \cdots, 80$, $i = 0, \cdots, M$. The results for $N = 14$ and $N = 19$ are plotted in Figures 1 and 2 (The figures include results from next example, for ease of comparison). As $M \to \infty$, $\mathbb{E}\left[(T_i^{(M)})^n\right] \to \infty$, $n = 1, 2$, which gets reflected in the denominator in (18) becoming very close to zero. This will naturally lead to numerical problems for large M.

Figure 3 plots $p_i^{(M)}$ for $M = 2, \cdots, 80$ and $i = 0, 30$, and 60, and Figure 4 plots the computed values of $\mathbb{E}\left[\tau_i(M)\right]$, for the cases $N = 14$ and 19.

We discuss the results towards the end of this section.

b) **Non-Binomial Arrival Process**

In the above example, the overall arrival process to the multiplexer has a Binomial distribution. Next we consider discrete distributions, prob1 and prob9, that are not Binomial, but have (nearly) same means as those in the above example, with 14 and 19 sources, respectively. Table 1 tabulates the two distributions considered.

Figures 1–3 incorporate some of the computational results for this example.

4.2 Markov modulated case

a) **A three state source**

In this example the discrete time queue is fed by a source which has three states and its motion in the state space $\{0, 1, 2\}$ is described by a Markov chain evolving in discrete time with transition probability matrix

$$\begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.46875 & 0.1375 & 0.39375 \\ 0 & 0.7 & 0.3 \end{bmatrix}.$$

State changes of the Markov chain occur at the slot boundaries and when the Markov chain is in state $i, i \in \{0, 1, 2\}$, the source is in the process of generating i packets which it feeds into the queue at the beginning of the next slot. It is easy to check that the stationary mean number of packets per slot is 0.85 and its variance is 0.5775. The expected first time to overflow is $\pi \cdot \mathbf{T}_0$, where π is the stationary probability vector of the Markov chain, and $\mathbf{T}_0$ is defined as in Section 3 ($\pi \cdot \mathbf{T}_0$ is the first time to overflow if the underlying Markov chain starts with stationary distribution and initially the system is empty). Table 2 compares the expected first time to overflow in this example with the i.i.d. (prob9) case.

Figure 5 plots the mean first time to overflow as a function of the buffer size for different initial number, n, in the system.

b) **Superposition of On/Off sources**

This example is related to (infact a refinement of) the approximate voice source model presented in Example 4.1(a). Here, each of the voice sources is modelled as a Markov modulated On/Off source. Hence we consider an arrival process that is generated by a superposition of N On/Off sources each of which is governed by a two state discrete time Markov chain. A source when in On state generates a packet which it feeds to the queue at the beginning of the next slot. When it is in the Off state, it does not generate any packets. The transition probability matrix of each source is

$$\begin{bmatrix} 0.98 & 0.02 \\ 0.42 & 0.58 \end{bmatrix}.$$

The mean interarrival time of packets from each source is 22 slots, the same as in Example 4.1(a).

We considered the case when the number of On/Off sources is 19 which corresponds to a load of 0.85 (*cf.* Example 4.1(a)). The state space of the underlying Markov chain has 20 states where state i indicates the number of sources that are in the On state. The array of values $\mathbb{E}\left[T_{n,i}^{(M)}\right]$ for $i = 0, \cdots, 19$, $M = 2, \cdots, 80$, $n = 0, \cdots, M$ were calculated using the matrix operating facilities of MATLAB.

In Figure 5, the quantity $\mathbb{E}\left[T_i^{(M)}\right]$ is plotted as a function of M, for $n = 30$, 60, and 80 (with $i = 0$).

4.3 Discussion

We discuss some features of the queueing phenomenon brought out by the numerical computations done for the examples presented above.

(1) For i and M fixed, as the mean arrival rate increases, the moments of $T_i^{(M)}$ decrease (see Fig. 1 and 2). This simply reflects the fact that the queue builds up more rapidly as the load increases. Also, with the mean fixed, increasing the variance of the arrival process results in the moments of $T_i^{(M)}$ to decrease (Fig. 1 and 2).

(2) For a particular distribution, the moments of $T_i^{(M)}$ are monotonically increasing in M and monotonically decreasing in i. The effect of the initial number in the system, i, on the moments of $T_i^{(M)}$ is very little (The curves for $i = 0, 30$ and 40 merge quickly as M increases; see Figures 1, 2, and 5). The reason for this is that the effect of i vanishes once the queue length decreases to zero before an overflow occurs – this happens with probability $(1 - p_i^{(M)})$, which is sufficiently close to unity for the distributions considered (see Fig. 3).

(3) In contrast to $\mathbb{E}\left[T_i^{(M)}\right]$, even though $\mathbb{E}\left[\tau_i(M)\right]$ has monotonic properties with respect to i and M, the effect of i is quite pronounced. This is to be expected

from Equation (4). Also, for i fixed, $\mathbb{E}[\tau_i(M)]$ reaches an asymptotic value (equal to $\mathbb{E}[\tau_i(\infty)]$, the corresponding value for the infinite buffer case) which depends very much on i (see Fig. 4).

(4) In the Example 4.2(a) of a 3–state Markov modulated source, $\mathbb{E}_\pi\left[T_i^{(M)}\right]$ (π is the stationary distribution of the underlying Markov chain), are considerably higher than the corresponding i.i.d. example (with input distribution prob9) (see Table 2). In the Markov modulated case the variance is lower than that of prob9 and the maximum number of arrivals in the former case is two while as many as ten packets can arrive in the latter. Also, the arrivals from the 3–state source are positively correlated only to a small extent (the covariance between successive arrivals is 0.2325).

(5) Even though the two Markov modulated examples considered have the same mean, the superposition of 19 On/Off sources has a larger variance and a higher peak arrival rate. This resulted in lower values of the mean overflow times (see Fig. 5).

Acknowledgement We are thankful to one of the anonymous reviewers for a close reading of the first version of the paper leading to the elimination of many errors.

References

[1] S. Asmussen. *Applied Probability and Queues.* John Wiley and Sons Ltd., Chichester, 1987.

[2] S. Asmussen and D. Perry. On Cycle Maxima, First Passage Problems and Extreme Value Theory for Queues. *Stoch. Models*, 8:421–458, 1992.

[3] K. Bisidikian, J. S. Lew, and A. N. Tantawi. On the Tail Approximations of the Blocking Probability of Single Server Queues with Finite Capacity. In R. O. Onvural and I. F. Akyildiz, editors, *Queueing Networks with Finite Capacity.* Elsevier Science Publs., 1993.

[4] H. Bruneel. Performance of Discrete Time Queueing Systems. *Comp. and Opns. Res.*, 20(3):303–320, 1993.

[5] H. Bruneel and B. G. Kim. *Discrete–Time Models for Communication Systems Including ATM.* Kluwer Academic Pubs., Dordrecht, The Netherlands, 1993.

[6] M. L. Chaudury and Y. A. Zhao. First Passage Times and Busy Period Distributions of Discrete Time Markovian Queues : Geom(n)/Geom(n)/1/N. *Queueing Systems*, 18:5–26, 1994.

[7] W. W. Chu and A. G. Konheim. On the Analysis and Modeling of a Class of Computer Communication Systems. *IEEE Trans. Comm. COM*, 20(3):645–660, June 1972.

[8] I. Cidon, A. Khamisy, and M. Sidi. On Packet Loss Process in High-speed Networks. In *Proc. INFOCOM 92*, pages 0242–0251. IEEE, 1992.

[9] J. W. Cohen. *The Single Server Queue.* North-Holland, Amsterdam, revised edition, 1982.

[10] N. D. Gangadhar. Analysis of Discrete Time Queues with Applications to ATM based B–ISDNs. M.Sc.(Engg.) Thesis, Department of Electrical Engineering, Indian Institute of Science, Bangalore, India, Apr. 1995.

[11] P. Glasserman and S. G. Kou. Overflow Probabilities in Jackson Networks. In *Proc. IEEE*

Conf. Dec. and Control, pages 3178–3182, 1993.

[12] K. Kobayashi and A. G. Konheim. Queueing Models for Computer Communication System Analysis. *IEEE Trans. Comm. COM*, 25(1):2–29, Jan. 1977.

[13] D. Kofmann and H. Korelioglu. Loss Probabilities and Delay and Jitter Distribution in a Finite Buffer Queue with Heterogenous Batch Markovian Arrival Process. In *Proc. GLOBECOM 93*, pages 830–834. IEEE, 1993.

[14] J. W. Pitman. Occupational Measures For Markov Chains. *Adv. Appl. Prob.*, 9:69–89, 1977.

[15] N. U. Prabhu. *Stochastic Storage Processes : Queues, Insurance Risks and Dams.* Springer-Verlag, New York, 1980.

[16] J. W. Roberts. COST 224 Final Report – Performance Evaluation and Design of Multiserver Networks. Technical report, COST, 1991.

[17] V. Sharma. Reliable Estimation via Simulation. *Queueing Systems*, 19:169–192, 1995.

[18] V. Sharma and N. D. Gangadhar. Asymptotics for Transient and Stationary Probabilities for Finite and Infinite Buffer Discrete Time Queues. Submitted, 1995.

[19] T. Takine, T. Suda, and T. Hasegawa. Cell Loss and Output Process Analysis of a Finite Buffer Discrete-Time ATM Queueing System with Correlated Arrivals. In *Proc. INFOCOM 93*, pages 1259–1269. IEEE, 1993.

[20] H. Tijms. Heuristics for Finite Buffer Queues. *Prob. in Engg. and Info. Sc.*, 6:277–285, 1992.

[21] T. Y. Yen. On The Delay Analysis of a TDMA Channel with Finite Capacity. *IEEE Trans. Comm. COM*, 30:1937–1941, 1982.

Appendix

In this Appendix, we provide a derivation of the expression for stationary probability of packet loss in a finite buffer queue in terms of the mean regeneration cycle length. The latter quantity can be computed using the algorithms presented in this paper.

Let N_L be the number of packets lost in one regeneration cycle (denoted by $\tau(M)$). Then

$$[\tau(M) - 1 + N_L \mid \tau(M) > 1] = \left[\sum_{k=0}^{\tau(M)-1} X_k \mid \tau(M) > 1 \right].$$

Hence

$$\mathbb{E}\left[N_L \mid \tau(M) > 1\right] = 1 - \mathbb{E}\left[\tau(M) \mid \tau(M) > 1\right] + \mathbb{E}\left[\textstyle\sum_{k=0}^{\tau(M)-1} X_k \mid \tau(M) > 1\right].$$

Now

$$\begin{aligned} \mathbb{E}\left[\tau(M) \mid \tau(M) > 1\right] &= \frac{\mathbb{E}\left[\tau(M)\right] - \mathbb{P}\left(\tau(M) = 1\right)}{\mathbb{P}\left(\tau(M) > 1\right)}, \\ \mathbb{P}\left(\tau(M) > 1\right) &= \mathbb{P}\left(X > 0\right), \end{aligned}$$

and

$$\mathbb{E}\left[\sum_{k=0}^{\tau(M)-1} X_k \mid \tau(M) > 1\right] = \mathbb{E}\left[\tau(M)\right] \mathbb{E}\left[X\right] / \mathbb{P}\left(X > 0\right).$$

Therefore, the stationary probability of packet loss

$$\begin{aligned} &= \text{mean fraction of packets lost} \\ &= \mathbb{E}\left[N_L \mid \tau(M) > 1\right] \Big/ \mathbb{E}\left[\sum_{k=0}^{\tau(M)-1} X_k \mid \tau(M) > 1\right] \\ &= 1 - \frac{1}{\mathbb{E}\left[X\right]}\left[1 - \frac{1}{\mathbb{E}\left[\tau(M)\right]}\right]. \end{aligned}$$

∎