

On the Performance of Bursty and Modulated Sources Subject to Leaky Bucket Rate-Based Access Control Schemes

Khosrow Sohraby, *Senior Member, IEEE*, and Moshe Sidi, *Senior Member, IEEE*

Abstract— In this paper we provide the analysis of a rate-based access control scheme in high speed environments based on a buffered leaky bucket algorithm. The analysis is carried out in discrete time which is representative of an ATM environment. For the cell arrivals to the leaky bucket we consider a general discrete Markovian arrival process which models bursty and modulated sources. The key of our analysis is the introduction of the *deficit function* that allows the reduction of the original problem to a more standard discrete time queueing system with the same arrival process. As an important special case, the detailed analysis of the Binary Markov Source throttled by such rate-based access control schemes is presented. Along with explicit recursions for computation of state probabilities and simple characterization of the asymptotic behavior of the queue build up, some guidelines for the parameter selection of these schemes are provided. Our results indicate that for sources with relatively large active periods, for an acceptable grade-of-service at the input queue, the token generation rate should be chosen to be close to the peak rate of the source, and increasing the bucket size of the leaky bucket does not improve substantially the performance at the input queue.

I. INTRODUCTION

Congestion control is an essential part of any packet switching network architecture in order to guarantee some performance level (packet loss, packet delay, total network throughput) under unpredictable and changing traffic conditions. Traditional data packet networks have employed hop-by-hop and back-pressure mechanisms in order to guarantee that no user will send more packets than the network can absorb. These mechanisms involve a large processing overhead per individual packet and also require large buffers at the input and the intermediate nodes. Furthermore, in a high speed environment, where the end-to-end propagation delays are much larger than the transmission time of a typical packet, the effectiveness of such *closed loop* schemes which rely mainly on some sort of feedback information from the network becomes questionable.

Recently, there has been much emphasis on *open loop* control schemes, where the goal has been shifted to preventing congestion rather than reacting to it. As means of achieving this, “leaky bucket” rate-based congestion con-

trol scheme or variations and/or generalizations of it have been proposed [2], [6], [10], [16]-[19].

Using leaky bucket, the transport network regulates each source which utilizes the network resources so that each source stays on the average within its allowable rate and allocated bandwidth as an attempt to “guarantee” some grade-of-service. Therefore, it is important to understand the limitations of the scheme and its behavior and also the trade-offs involved in selecting its parameters.

In this paper the analysis of a buffered leaky bucket scheme is provided in discrete time which is representative of an ATM (Asynchronous Transfer Mode) environment. The analysis is carried out for a large and rich class of cell arrival processes. In the important special case, where the source is modelled as a Binary Markov Source, detailed results are given. This model is widely used in practice to represent bursty and modulated cell arrivals in an ATM environment, where a source may stay for relatively long durations in active (ON) and inactive (OFF) periods [19]. For this source, we also provide a simple characterization of the asymptotic behavior of the cell build up in the admission and transmission queue. Our results show that the performance is very sensitive to the average ON period even when the average source utilization is very low. This indicates that for an acceptable performance at the input of the network, it is difficult to throttle these sources using leaky bucket rate-based access control.

Most of the analysis of the “leaky bucket” schemes is confined to *continuous* time. Furthermore, it has been assumed that the only source for delaying packets (or “cells” in an ATM environment) at the access of the network, is the absence of tokens, i.e., the transmission time of the packets has not been taken into account (with the exception of [8]). Examples of analysis under this assumption are [16], [17], where a buffered leaky bucket scheme was considered. For a Poisson arrival process to the system, extensive analysis of the queue length, waiting time and inter-departure time distributions were presented for both finite and infinite buffer capacity. In [3] the analysis of [17] was extended to continuous time Markovian Arrival Processes (MAP) [13]. In a more recent work, Heyman [11], presents the analysis of similar models, where the packets may require more than one token to enter the network.

Our model of the system is similar to [1], where a discrete time environment is assumed and the cell transmission time is explicitly modeled. In [1] the cell arrivals in a time slot are characterized by a batch process, where it is assumed

Paper approved by Aurel Lazar, the Editor for Voice/Data Networks of the IEEE Communications Society. Manuscript received: September 18, 1990; revised: August 16, 1991. This paper was presented at the IEEE INFOCOM '91, Bal Harbour, Florida, April 9-11, 1991.

Khosrow Sohraby is with CSTP, University of Missouri-Kansas City, MO 64110. This work was completed while the author was with the IBM T. J. Watson Research Center.

Moshe Sidi is with the Electrical Engineering Department, Technion-Israel Institute of Technology, Haifa, 32000, ISRAEL. This work was done in part while visiting the IBM Research Division, T. J. Watson Research Center, Yorktown Heights, NY 10598.

that the arrivals in successive slots are independent and identically distributed. The solution method in [1] is based on the Matrix Analytic approach [14], which is of numerical nature. However, the special structure of the underlying transition probability matrix of the problem is utilized to achieve substantial savings in the computations.

In this paper the analysis is extended to a general finite state discrete Markovian arrival process. The main contribution of the paper is the introduction of a simple function of the states of the system, referred to as the *deficit function*, and showing that working with the deficit function the problem can be reduced to a standard discrete time queueing system with the same arrival process. The latter problem can be solved by any of the existing methods available in the literature such as Matrix Analytic or transforms. We discuss in detail the solution using the generating function transform method and show that in many cases of interest it is possible to compute the stationary probabilities of interest, once some "boundary" probabilities are obtained. Examples of i.i.d. batch sources and binary Markov sources are worked out in detail.

In the remainder of this section, we outline the organization and some of the contributions of this paper. In Section II the detailed description of the problem and its mathematical formulation are presented. In Section III the definition of the deficit function and its relationship with the original problem is given. In particular, we show that once the stationary distribution of the deficit function is given, the solution of our original system under study can be obtained recursively. In addition, the analysis of the deficit function is discussed in Section III, where it is shown that the solution complexity of this function is essentially the same as that of a single server queue with the same arrival process. In Section IV some examples are given. In particular, we study a Binary Markov Source as an example of a modulated arrival process, where a maximum of one cell may arrive in a slot. We show in detail our recursive approach for finding the stationary probabilities. In this section we also present some closed form expressions for the i.i.d batch arrival processes. In Section V we provide some numerical results for the Binary Markov Source, where the effect of the parameters of the arrival process on the performance is carefully examined. In particular, it is shown that for a fixed source utilization the queue build up is very sensitive to the average active (ON) period. Furthermore, increasing the size of the token pool, may not improve the performance substantially at the input queue unless the token generation rate is increased. For the sources with high average ON period, this rate may have to increase to the peak rate (or close to it) for an acceptable grade-of-service at the input queue. Finally, in Section VI we summarize the paper.

II. THE MODEL AND PRELIMINARIES

We consider a discrete time system, where the time is slotted and the cell arrival process in each slot is governed by a homogeneous finite-state, aperiodic discrete-time Markov chain called the *modulating Markov chain*,

where transitions between states of the chain take place only at the slot boundaries. The probability of a transition from state i to state j is denoted by p_{ij} for $0 \leq i, j \leq N$ (the number of states of the Markov chain is $N + 1$). The transition probability matrix of the chain is denoted by $P = \{p_{ij}\}$. Transitions of the chain are independent from any other event in the system. When the chain is in state l , the source transmits a random number of cells with probability generating function (p.g.f) $B_l(z) = \sum_i b_i^{(l)} z^i$, where $b_i^{(l)}$ denotes the probability of i arrivals in a slot when the Markov chain is in state l . Arriving cells are stored in a buffer of infinite capacity.

To describe the cell transmission process, consider a *token pool* where tokens are generated *periodically*, one token every K slots ($K \geq 1$). The token pool has a finite size $M - 1$ ($M \geq 1$) so that tokens arriving to a full token pool are lost (see Fig. 1). Each token allows for transmission of a single cell in a slot, where following a transmission a token is removed from the token pool. A transmission in a slot takes place only if a token is available in the pool in that slot.

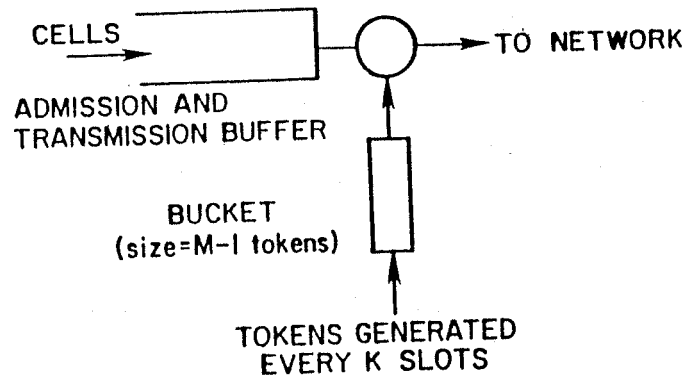


Fig. 1: A Leaky Bucket with Admission and Transmission Buffer

The evolution of the system can be described by a four dimensional Markov chain (Q_n, X_n, T_n, J_n) , where Q_n denotes the the number of cells in the queue at slot n ($Q_n \geq 0$), X_n denotes the number of tokens in the pool at slot n ($0 \leq X_n \leq M - 1$), T_n denotes the "phase" of the periodic token arrival process at slot n ($0 \leq T_n \leq K - 1$), and a token arrives at the token pool in slot n only if $T_n = 0$. Finally, J_n denotes the state of the modulating Markov chain at slot n . Based on the description of the system, the state at slot $n + 1$ is related to the state at slot n according to the following equations:

$$Q_{n+1} = (Q_n - 1 + 1_{X_n=0})^+ + A_{n+1}$$

$$X_{n+1} = \min[(X_n - 1 + 1_{Q_n=0})^+ + 1_{T_n=0}, M - 1] \quad (1)$$

$$T_{n+1} = (T_n + 1) \bmod(K)$$

J_{n+1} = evolves according to P

In the above equation A_{n+1} denotes the number of arrivals in slot $n + 1$, 1_E is the indicator function of the event E and is equal to 1 if E is true, and zero otherwise. The first equation indicates that a cell from the queue is removed in slot $n + 1$, if both the token pool and the queue are not empty. None of the arrivals in slot $n + 1$ are transmitted in the same slot. The second equation shows that following a transmission of a cell in a slot, a token is removed from the pool. When T_n is zero (every K slots), a token is added to the token pool if the pool (of capacity $M - 1$) is not full. The third equation represents the phase of the token arrival process which is periodic, taking values in the range $0, 1, \dots, K - 1$. The final equation represents the evolution of the modulating Markov chain that governs the arrival process.

Of interest is the stationary distribution of the system (1). For future reference, we define the following notations for $i \geq 0$, $0 \leq j \leq M - 1$, $0 \leq k \leq K - 1$, $0 \leq l \leq N$, assuming that T_0 is uniformly distributed on $[0, K - 1]$:

$$p(i, j, k, l) \triangleq \lim_{n \rightarrow \infty} \Pr(Q_n = i, X_n = j, T_n = k, J_n = l) \quad (2)$$

$$\bar{p}(i, j, k) \triangleq [p(i, j, k, 0), p(i, j, k, 1), \dots, p(i, j, k, N)]^T$$

where $[\]^T$ stands for the transpose operation.

One may show that the transition probability matrix of the quadruplet (Q_n, X_n, T_n, J_n) is of the $M/G/1$ type queues studied extensively by Neuts [14]. However, a direct application of Neuts method will result in the determination of $M \times K \times (N + 1)$ boundary probabilities by iterative approach. For the special case of i.i.d batch arrival process, i.e., $N = 0$, the special structure of the transition probability matrix of system (1) has been exploited in [1] and has resulted in a major simplification in Neuts' approach, where most of the computations are carried out recursively in blocks of matrices of size $K \times K$, instead of iterative computations of blocks $MK \times MK$.

Our approach in this paper is not to solve system (1) directly. Instead, we reduce system (1) into a simpler system and solve the latter. The key in this approach is to work with a *simple* function of the queue and the bucket occupancy, referred to as the *deficit function*, which has a simple relationship to the solution of the original system of interest (1). In fact, we will show that by utilizing the deficit function the analysis of system (1) can be reduced to the analysis of a discrete-time queue with arrivals governed by the modulating Markov chain, and with a deterministic service time (equal to one slot), where the server is only available periodically once every K slots. This discrete-time queueing system with a server available in every slot has been considered in [9], where it has been argued that steady-state queue length probabilities can be determined once $(N + 1)$ boundary probabilities are computed. The periodic availability of the server does not increase the complexity of the problem, and one still has to compute $(N + 1)$ boundary probabilities (as in [9]) in order to obtain the

complete queue length distribution.

III. THE DEFICIT FUNCTION AND ITS RELATION TO SYSTEM (1)

In this section we formally define the deficit function and we show how its stationary behavior relates to that of system (1). In particular, we show that once the stationary distribution of the deficit function is given, the special structure of system (1) can be utilized to yield a recursive solution of its stationary behavior.

A. The Deficit Function

The reduction of system (1) to a simple discrete-time queue is obtained via the deficit function defined in slot n by

$$D_n \triangleq Q_n - X_n + (M - 1) \quad (3)$$

Note that the queue size in any slot is always bounded by the value of the deficit function in that slot as long as the bucket size is non-zero ($M \geq 2$).

From (1) and (3) we can show that the evolution of the deficit function is governed by

$$D_{n+1} = (D_n - 1_{T_n=0})^+ + A_{n+1}$$

$$T_{n+1} = (T_n + 1) \bmod(K) \quad (4)$$

J_{n+1} = evolves according to P

Obviously, the analysis of the triplet chain (D_n, T_n, J_n) is much simpler than the quadruplet chain (Q_n, X_n, T_n, J_n) . In fact, (4) has the same structure as a discrete queue with deterministic service time (equal to one slot), where the server is only available periodically once every K slots. The special case of this system in which the server is available in every slot, i.e., $K = 1$, has been considered in [9].

In what follows we first show that a *recursive* solution of system (1) is provided based on the solution of the deficit function (system (4)). Then we discuss the analysis of the deficit function, i.e., the solution of system (4).

B. Recursive Solution of System (1) Based on System (4)

In the following we assume that the stationary probabilities of the deficit function (system (4)) are available. These probabilities are defined for $i \geq 0$, $0 \leq k \leq K - 1$, $0 \leq l \leq N$ as (recall that we assume T_0 is uniformly distributed on $[0, K - 1]$):

$$f_{i,k,l} \triangleq \lim_{n \rightarrow \infty} \Pr(D_n = i, T_n = k, J_n = l) \quad (5)$$

$$\bar{f}_{i,k} \triangleq [f_{i,k,0}, f_{i,k,1}, \dots, f_{i,k,N}]^T$$

We provide a *recursive* procedure to find the stationary distribution of system (1). The important relationship which follows directly from the definition of the deficit function and the law of total probability is

$$\bar{f}_{i,k} = \sum_{j=\max(0, M-1-i)}^{M-1} \bar{p}(i+j-(M-1), j, k) \quad (6)$$

In order to provide the recursions, we need the following notation.

$$b_i \triangleq \begin{bmatrix} p_{00}b_i^{(0)} & p_{10}b_i^{(0)} & \cdots & p_{N0}b_i^{(0)} \\ p_{01}b_i^{(1)} & p_{11}b_i^{(1)} & \cdots & p_{N1}b_i^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{0N}b_i^{(N)} & p_{1N}b_i^{(N)} & \cdots & p_{NN}b_i^{(N)} \end{bmatrix} \quad (7)$$

Recall that $b_i^{(l)}$ denotes the probability of having i arrivals in state l .

Before presenting the recursive solution of system (1), it should be noted that $\bar{p}(*, 0, 1) = 0$, which is an immediate consequence of the evolution equations. By carefully examining the evolution equations in (1), and taking advantage of (6), we get (for $k = 0$ in the left hand side replace $k - 1$ by $K - 1$ in the right hand side):

For $j = M - 1$

$$\bar{p}(0, M - 1, k) = \bar{f}_{0,k} \quad \forall k \quad (8)$$

$$\bar{p}(i, M - 1, k) = b_i \bar{p}(0, M - 1, k - 1) \quad k \neq 1, i \neq 0 \quad (9)$$

$$\begin{aligned} \bar{p}(i, M - 1, 1) &= \sum_{l=0}^i b_l \bar{p}(i + 1 - l, M - 1, 0) \\ &\quad + b_i \bar{p}(0, M - 1, 0) + b_i \bar{p}(0, M - 2, 0) \\ &= \sum_{l=0}^i b_l \bar{p}(i + 1 - l, M - 1, 0) \\ &\quad + b_i \bar{p}(0, M - 1, 0) \\ &\quad + b_i [\bar{f}_{1,0} - \bar{p}(1, M - 1, 0)] \quad i \neq 0 \end{aligned} \quad (10)$$

For $j = M - 2, M - 3, \dots, 1$, we have

$$\bar{p}(0, j, k) = \bar{f}_{M-j-1,k} - \sum_{l=1}^{M-j-1} \bar{p}(M - j - l, M - l, k) \quad \forall k \quad (11)$$

$$\begin{aligned} \bar{p}(i, j, k) &= \sum_{l=0}^i b_l \bar{p}(i + 1 - l, j + 1, k - 1) \\ &\quad + b_i \bar{p}(0, j, k - 1) \quad k \neq 1, i \neq 0 \end{aligned} \quad (12)$$

$$\begin{aligned} \bar{p}(i, j, 1) &= \sum_{l=0}^i b_l \bar{p}(i + 1 - l, j, 0) + b_i \bar{p}(0, j - 1, 0) \\ &= \sum_{l=0}^i b_l \bar{p}(i + 1 - l, j, 0) + b_i [\bar{f}_{M-j,0} \\ &\quad - \sum_{l=1}^{M-j} \bar{p}(M - j - l + 1, M - l, 0)] \\ &\quad j \neq 1, i \neq 0 \end{aligned} \quad (13)$$

The above equations provide a recursive solution of (1) in terms of the deficit function for the stationary distribution for all the states *except* the states for which $(j = 0, k \neq 1)$ and $(j = 1, k = 1)$. The difficulty is that the state probabilities for these two classes of states are related, and state probabilities which are already obtained are not sufficient in finding them. However, we may use the deficit function and the state probabilities which are already determined to find them. Direct consequence of definition (3) gives $\forall i$,

$$\bar{p}(i, 0, k) = \bar{f}_{i+M-1,k} - \sum_{j=1}^{M-1} \bar{p}(i + j, j, k) \quad k \neq 1 \quad (14)$$

$$\bar{p}(i, 1, 1) = \bar{f}_{i+M-2,1} - \sum_{j=2}^{M-1} \bar{p}(i + j - 1, j, 1) \quad (15)$$

which completes the recursive computation of system (1) in terms of the deficit function (system (4)). Note that in general, if the solution of the state probabilities $\bar{p}(i, *, *)$ is desired, the computation of \bar{f} up to $\bar{f}_{i+M-1,*}$ is sufficient.

The evaluation of the moments may be carried out recursively using the probability generating function (p.g.f). Let

$$\bar{Q}_{j,k}(z) \triangleq \sum_i \bar{p}(i, j, k) z^i, \quad 0 \leq j \leq M - 1, \quad 0 \leq k \leq K - 1 \quad (16)$$

Using (9), (10), (12) and finally (13), respectively, we get

$$\bar{Q}_{M-1,k}(z) = B(z) \bar{p}(0, M - 1, k - 1) \quad k \neq 1 \quad (17)$$

$$\begin{aligned} z \bar{Q}_{M-1,1}(z) &= B(z) [\bar{Q}_{M-1,0}(z) + (z - 1) \bar{p}(0, M - 1, 0) \\ &\quad + z \bar{p}(0, M - 2, 0)] \end{aligned} \quad (18)$$

$$\begin{aligned} z \bar{Q}_{j,k}(z) &= B(z) [\bar{Q}_{j+1,k-1}(z) + z \bar{p}(0, j, k - 1) \\ &\quad - \bar{p}(0, j + 1, k - 1)] \\ &\quad 1 \leq j \leq M - 2, \quad k \neq 1 \end{aligned} \quad (19)$$

$$\begin{aligned} z \bar{Q}_{j,1}(z) &= B(z) [\bar{Q}_{j,0}(z) + z \bar{p}(0, j - 1, 0) \\ &\quad - \bar{p}(0, j, 0)] \quad 2 \leq j \leq M - 2, \quad k \neq 1 \end{aligned} \quad (20)$$

where the matrix $B(z)$ is defined by

$$B(z) \triangleq \begin{bmatrix} p_{00}B_0(z) & p_{10}B_0(z) & \cdots & p_{N0}B_0(z) \\ p_{01}B_1(z) & p_{11}B_1(z) & \cdots & p_{N1}B_1(z) \\ \vdots & \vdots & \ddots & \vdots \\ p_{0N}B_N(z) & p_{1N}B_N(z) & \cdots & p_{NN}B_N(z) \end{bmatrix} \quad (21)$$

Recall that $B_i(z) = \sum_i b_i^{(l)} z^i$.

To compute the p.g.f for the set of states for which $(j = 0, k \neq 1)$ and $(j = 1, k = 1)$, we use (14) and (15) directly.

We have

$$\begin{aligned} \bar{Q}_{0,k}(z) = & z^{-(M-1)}\bar{F}_k(z) - \sum_{l=0}^{M-2} [\bar{F}_{l,k}z^{l+1-M} \\ & + \bar{Q}_{l+1,k}(z)z^{-(l+1)} \\ & - \sum_{i=0}^l \bar{p}(i, l+1, k)z^{i-(l+1)}] \end{aligned} \quad (22)$$

and

$$\begin{aligned} \bar{Q}_{1,1}(z) = & z^{-(M-2)}\bar{F}_1(z) - \sum_{l=0}^{M-3} [\bar{F}_{l,1}z^{l+2-M} \\ & + \bar{Q}_{l+2,1}(z)z^{-(l+1)} \\ & - \sum_{i=0}^l \bar{p}(i, l+2, 1)z^{i-(l+1)}] \end{aligned} \quad (23)$$

where an empty sum vanishes, $F_{k,l}(z) \triangleq \sum_{i=0}^{\infty} f_{i,k,l}z^i$ and $\bar{F}_k(z) \triangleq [F_{k,0}(z), F_{k,1}(z), \dots, F_{k,N}(z)]^T$.

Note that the above recursions require the prior computation of $\bar{p}(i, *, *)$, $0 \leq i \leq M-2$ which is readily obtained by the recursions (8)-(13). Once $\bar{Q}_{j,k}(z)$ are obtained, any quantity of interest such as the p.g.f of the queue length distribution, distribution of the bucket fill, etc. can easily be obtained. We omit the details.

C. Discussion on the Analysis of System (4)

It is not difficult to see that the transition probability matrix of the triplet (D_n, T_n, J_n) is still of the $M/G/1$ type [14]. To obtain the stationary probabilities of this system one can use any of the existing methods available in the literature. In what follows we chose to present the solution of system (4) via a generating function approach, as we found it most suitable and useful for the examples that we present in the next section.

From (4), using the standard z transform techniques we obtain:

$$\begin{aligned} [z\mathbf{I} - \mathbf{B}^K(z)]\bar{F}_1(z) &= (z-1)\mathbf{B}(z)\bar{F}_0(0) \\ \bar{F}_k(z) &= \mathbf{B}^{k-1}(z)\bar{F}_1(z) \quad 2 \leq k \leq K-1 \\ \bar{F}_0(z) &= \mathbf{B}^{K-1}(z)\bar{F}_1(z) \end{aligned} \quad (24)$$

Note that $\mathbf{B}(1) = \mathbf{P}^T$, where \mathbf{P} is the transition probability matrix of the underlying modulating Markov chain of the cell arrival process. We have the obvious normalization condition $\bar{F}_1(1) = \frac{1}{K}\Pi^T$, where Π is the stationary distribution of \mathbf{P} satisfying $\Pi = \Pi\mathbf{P}$. The discussion on the determination of the vector $\bar{F}_0(0)$ will follow.

Denote the average number of arrivals in one slot by ρ which is equal to $\Pi(\mathbf{B}'(1))^T\bar{e}$, where \bar{e} is the unit column matrix defined by $\bar{e} = [1, 1, \dots, 1]^T$ and $\mathbf{B}'(1) = d\mathbf{B}(z)/dz|_{z=1}$. The steady-state condition of system (4) (and hence of system (1)) is $K\rho < 1$, since the average number of arrivals per slot should be less than $1/K$, the average number of tokens per slot.

From (24) we observe that once the $N+1$ components of the vector $\bar{F}_0(0)$ are known, the complete steady-state generating function is available. Then, in principle, the stationary probabilities of system (4) (and hence of system (1), see Section IV.B) can be determined *recursively* as described in [7]. Note that in order to carry out the recursions as in [7] one requires that the matrix $\mathbf{B}(z)$ be invertible at $z=0$. An example how to compute the stationary probabilities when this condition does not hold is provided in Section IV.B.

It is interesting to note that the structure of (24) is very similar to the structure of the corresponding generating function when the server is available in each slot (see [9]). As in [9] the determination of the vector $\bar{F}_0(0)$ for a stable system ($K\rho < 1$) involves the usage of the analytic properties of $\bar{F}_1(z)$ within the unit disk, $|z| \leq 1$. In general, the roots of the determinant $\Delta(z) = \det[z\mathbf{I} - \mathbf{B}^K(z)]$ together with the normalization condition $\bar{F}_1(1) = \frac{1}{K}\Pi^T$, are sufficient to find the $(N+1)$ components of the vector $\bar{F}_0(0)$. The normalization condition yields the simple expected relation

$$\bar{e}^T\bar{F}_0(0) = \frac{1-K\rho}{K} \quad (25)$$

The procedure for computing the vector $\bar{F}_0(0)$ is as follows. One starts with the determination of the roots of $\Delta(z)$ within the unit disk. It can be shown that when $K\rho < 1$ the determinant $\Delta(z)$ has exactly $N+1$ roots $(z_1, z_2, \dots, z_{N+1})$ within the unit disk ($|z_i| \leq 1$), one of them being $z=1$ (without loss of generality we let $z_{N+1} = 1$). For a discussion on the computational experience in determining roots within the unit disk the reader is referred to [4]. Since $\bar{F}_1(z)$ is an analytic function within the unit disk, it follows that for each root z_i we have $\text{Adj}(z_i\mathbf{I} - \mathbf{B}^K(z_i))\mathbf{B}(z_i)\bar{F}_0(0) = 0$ for $i=1, 2, \dots, N$, where $\text{Adj}(z\mathbf{I} - \mathbf{B}^K(z))$ is the adjoint matrix of $z\mathbf{I} - \mathbf{B}^K(z)$. Thus, if all the roots z_i are distinct we have N linear equations for $\bar{F}_0(0)$ that together with relation (25) yield the unknown vector $\bar{F}_0(0)$. If the roots z_i are not distinct, then a standard method is the following. Say z_i is a root of multiplicity m . One obtains m linear equations for $\bar{F}_0(0)$ from $\frac{d^j}{dz^j}[\text{Adj}(z\mathbf{I} - \mathbf{B}^K(z))\mathbf{B}(z)\bar{F}_0(0) = 0]|_{z=z_i} = 0$ for $j=0, 1, \dots, m-1$.

Note that the procedure outlined above for the determination of $\bar{F}_0(0)$ is standard but may be computationally complex. However, as we discuss below and in the next section, in many interesting cases this procedure is simple. For instance, such an interesting special case (discussed also in [9]), where $\bar{F}_0(0)$ may be given *explicitly*, is when the source must transmit at least one cell in all but one state. In this case only one component of $\bar{F}_0(0)$ is non-zero, and we have (without loss of generality, we assume that the zero state is the only state in which the source does not transmit a cell):

$$\bar{F}_0(0) = \frac{1-K\rho}{K}[1, 0, 0, \dots, 0]^T \quad (26)$$

which is an immediate consequence of the system (4), since

the deficit function may not be equal to zero in any other than the zero state.

IV. EXAMPLES

In this section we focus on two special cases as an example of our approach to solve the deficit function recursively and then using the results for a recursive computation of system (1). These cases include independent and identically distributed (i.i.d.) batch arrival processes and Binary Markov Sources.

A. I.I.D. Batch Arrivals

In this section we provide the recursions for computing the stationary probabilities of system (4) when the arrivals are i.i.d. batches, i.e., $N = 0$. This case has been analyzed in [1] using Matrix Analytic approach. In addition, when the bucket size is small ($M = 2$) we provide closed form expressions.

For i.i.d. batches the modulating Markov chain is degenerate and has only one state ($N = 0$). We have from (25) that

$$f_{0,0,0} = \frac{1 - K\rho}{K} \quad (27)$$

Therefore, from (24) we obtain

$$\begin{aligned} [z - B^K(z)]F_1(z) &= (z - 1)B(z)f_{0,0,0} \\ F_k(z) &= B^{k-1}(z)F_1(z) \quad 2 \leq k \leq K - 1 \\ F_0(z) &= B^{K-1}(z)F_1(z) \end{aligned} \quad (28)$$

where $B(z) = \sum_i b_i z^i$. Let $B^k(z) = \sum_i b_{i,k} z^i$ ($1 \leq k \leq K$). Then for any i the quantities $b_{i,k}$ are computed recursively from $b_{i,k} = \sum_{j=0}^i b_{j,k-1} b_{i-j}$ ($k = 1, 2, \dots, K$). Now, as in [7] we equate the coefficients of the same order of z in (28) to obtain $f_{0,1,0} = f_{0,0,0} b_0^{-(K-1)}$ and

$$f_{i,1,0} = (f_{i-1,1,0} + (b_i - b_{i-1})f_{0,0,0} - \sum_{j=0}^{i-1} f_{j,1,0} b_{i-j,K}) / b_0^K \quad (29)$$

In addition,

$$\begin{aligned} f_{i,k,0} &= \sum_{j=0}^i b_{j,k-1} f_{i-j,1,0} \quad 2 \leq k \leq K - 1 \\ f_{i,0,0} &= \sum_{j=0}^i b_{j,K-1} f_{i-j,1,0} \end{aligned} \quad (30)$$

Now we substitute the quantities $f_{i,k,0}$ in (8)-(15) to obtain recursively the stationary probabilities of system (1).

For small values of bucket capacity, it is possible to obtain simple relationships between $\bar{Q}_{j,k}(z)$ and $\bar{F}_k(z)$. To see this, using (17)-(23), or alternatively

$$\bar{F}_k(z) = \sum_{j=0}^{M-1} z^{M-j-1} \bar{Q}_{j,k}(z)$$

(which is a direct consequence of (6)), for $M = 2$ we have ($\bar{Q}_{0,1}(z) = 0$)

$$\begin{aligned} \bar{Q}_{1,1}(z) &= \bar{F}_1(z) \\ \bar{Q}_{1,k}(z) &= B(z)\bar{F}_{k-1}(0) \quad k \neq 1 \\ \bar{Q}_{0,k}(z) &= \frac{\bar{F}_k(z) - B(z)\bar{F}_{k-1}(0)}{z} \quad k \neq 1 \end{aligned} \quad (31)$$

For the special case of i.i.d batch arrivals with p.g.f $B(z)$ we can get the *marginal* densities of the queue size and the bucket size from the above. We have

$$\Pr(X = 0) = \frac{K-1}{K} - \frac{1-K\rho}{K} \frac{1-b_0^{K-1}}{(1-b_0)b_0^{K-1}} \quad (32)$$

$$\Pr(X = 1) = \frac{1}{K} + \frac{1-K\rho}{K} \frac{1-b_0^{K-1}}{(1-b_0)b_0^{K-1}} \quad (33)$$

where b_0 is the probability of no arrivals in a slot. The p.g.f of the queue is given by

$$\begin{aligned} Q(z) &= \left\{ \frac{B(z)[1 - B^{K-1}(z)]}{[1 - B(z)][z - B^K(z)]} \right. \\ &\quad \left. + \frac{z}{z - B^K(z)} + \frac{1 - b_0^{K-1}}{(1 - b_0)b_0^{K-1}} \right\} \\ &\quad \cdot \frac{1 - K\rho}{K} \cdot \frac{z - 1}{z} \cdot B(z) \end{aligned} \quad (34)$$

Note that for a given source utilization, the distribution of the bucket size depends only on b_0 , and not on the other probability components of the batch size distribution. For a general i.i.d source, and a bucket size of $M - 1$, (above result is for $M = 2$), it can be shown that the bucket size distribution depends only on b_0, b_1, \dots, b_{M-2} .

We may specialize (34) for a Bernoulli source $B(z) = 1 - p + pz$ and $K = 2$, to obtain *simple* and *closed* form expression for the queue length distribution. For this case, the inversion of (34) yields

$$\begin{aligned} \Pr(Q = 0) &= \frac{(1 - 2p)(2 - 2p + p^2)}{2(1 - p)^2} \\ \Pr(Q = 1) &= \frac{(1 - 2p)(2p - p^2)(1 - p + p^2)}{2(1 - p)^4} \\ \Pr(Q = i) &= \frac{1 - 2p}{2(1 - p)^3} \left(\frac{p}{1 - p} \right)^{2i-1} \quad i \geq 2 \end{aligned} \quad (35)$$

which coincides with the result of [12] using a different approach. The average queue length $E[Q]$ in this case is given by

$$E[Q] = \frac{p(2 - 5p + 4p^2)}{2(1 - p)(1 - 2p)} \quad (36)$$

The above results can be easily extended to other values of M and K , but the expressions become unwieldily very soon.

B. A Binary Markov Source

This is a simple but very important example in the class of Markov sources. It is widely used to model *bursty* and *correlated* sources in which the "OFF" (state 0) and "ON" (state 1) periods can be very large. In this case, $N = 1$, $B_0(z) = 1$, and $B_1(z) = z$. The source utilization is $\rho = p_{01}/(p_{01} + p_{10})$ and we have

$$\begin{aligned} \bar{F}_k(z) &= \left[\sum_i f_{i,k,0} z^i, \sum_i f_{i,k,1} z^i \right]^T \\ B(z) &= \begin{bmatrix} p_{00} & p_{10} \\ zp_{01} & zp_{11} \end{bmatrix} \end{aligned} \quad (37)$$

In what follows, we first determine the stationary probabilities for $k = 1$. Since state zero is the only state that the source does not transmit we have $f_{0,0,0} = (1 - K\rho)/K$ and $f_{0,0,1} = 0$. Therefore,

$$[z\mathbf{I} - \mathbf{B}^K(z)]\bar{F}_1(z) = \frac{1 - K\rho}{K}(z - 1)[p_{00}, zp_{01}]^T \quad (38)$$

Before providing the recursions for $f_{i,1,0}$ and $f_{i,1,1}$, we need to compute the Taylor expansion of the entries of the matrix $\mathbf{B}^k(z)$ recursively. Let

$$\mathbf{B}^k(z) \triangleq \begin{bmatrix} B_{00}^{(k)}(z) & B_{10}^{(k)}(z) \\ zB_{01}^{(k)}(z) & zB_{11}^{(k)}(z) \end{bmatrix} \quad (39)$$

where $B_{mn}^{(k)}(z) \triangleq \sum_{i=0}^k \beta_{i,mn}^{(k)} z^i$. Using the definition of $\mathbf{B}(z)$, a simple recursive expression for computing $\beta_{i,mn}^{(k)}$ for arbitrary k is given by ($\beta_{-1,mn}^{(k)} \triangleq 0$)

$$\begin{aligned} \beta_{0,mn}^{(1)} &= p_{mn} \\ \beta_{i,mn}^{(k)} &= p_{m0} \beta_{i,0n}^{(k-1)} + p_{m1} \beta_{i-1,1n}^{(k-1)} \quad i \geq 1 \end{aligned} \quad (40)$$

Now following [7], it remains to determine the Taylor expansion of (38), and equate the coefficient of z^i on both sides of that equation. We easily get

$$\begin{aligned} f_{0,1,0} &= \frac{1 - K\rho}{K} \cdot \frac{p_{00}}{\beta_{0,00}^{(K)}} = \frac{1 - K\rho}{K p_{00}^{K-1}} \\ f_{0,1,1} &= 0 \end{aligned} \quad (41)$$

and for $i \geq 1$, we get ($1_{i=1} = 1$ if $i = 1$ and zero otherwise):

$$\begin{aligned} \begin{bmatrix} \beta_{0,00}^{(K)} & \beta_{0,10}^{(K)} \\ -\beta_{0,01}^{(K)} & 1 - \beta_{0,11}^{(K)} \end{bmatrix} \cdot \begin{bmatrix} f_{i,1,0} \\ f_{i,1,1} \end{bmatrix} &= \begin{bmatrix} t_{i,0} \\ t_{i,1} \end{bmatrix} \\ &+ \frac{1 - K\rho}{K} \begin{bmatrix} -p_{00} \\ p_{01} \end{bmatrix} 1_{i=1} \end{aligned} \quad (42)$$

where

$$\begin{aligned} t_{i,0} &\triangleq f_{i-1,1,0} - \sum_{l=0}^{i-1} (f_{l,1,0} \beta_{i-l,00}^{(K)} + f_{l,1,1} \beta_{i-l,10}^{(K)}) \\ t_{i,1} &\triangleq \sum_{l=0}^{i-1} (f_{l,1,0} \beta_{i-l,01}^{(K)} + f_{l,1,1} \beta_{i-l,11}^{(K)}) \end{aligned}$$

To find the recursion for $f_{i,k,l}$, $k \neq 1$, we may use (24), and the expansion of $\mathbf{B}^k(z)$. For $l = 0, 1$, and $k \neq 0$, we immediately get

$$f_{i,k,l} = \sum_{m=0}^i (f_{m,1,0} \beta_{i-m,0l}^{(k-1)} + f_{m,1,1} \beta_{i-m,1l}^{(k-1)}) \quad (43)$$

The calculation for $k = 0$ is carried out by replacing k by K in the RHS of above equation.

Equations (41)-(43) provide a convenient recursion for the computation of the stationary distribution of the deficit function, namely $f_{i,k,0}$ and $f_{i,k,1}$. This is an example of

deriving such a recursion when $B(z)$ is not invertible for $z = 0$. To find the stationary distribution of system (1), we utilize (8)-(13), and the fact that $f_{0,k,1} = 0$ for all k to get

$$\begin{aligned} p(0, j, k, 0) &= f_{M-j-1, k, 0} \\ p(1, j, k, 1) &= f_{M-j, k, 1} \quad j \neq 0, \text{ (exclude } (j = 1, k = 1)) \\ p(i, 0, k, l) &= f_{M-1+i, k, l} \quad k \neq 1, l \in \{0, 1\} \\ p(i, 1, 1, l) &= f_{M-2+i, 1, l} \quad l \in \{0, 1\} \end{aligned} \quad (44)$$

The other probabilities are zero. Note that when a maximum of one cell may arrive in a slot, the support of $p(i, j, k, l)$ may be infinite only for $(j = 0, k \neq 1)$ and $(j = 1, k = 1)$.

Asymptotic Behavior

It is widely believed that the tail behavior of most queueing systems with infinite waiting room exhibits a geometric behavior $\Pr(Q > i) \approx \alpha \eta^i$ for large i . This result is rigorously proved for the $GI/PH/c$ queues [15].

From the definition of the deficit function, it is clear that the asymptotic behavior of the deficit function and that of the queue length are identical for a finite size of the token pool $M - 1$. In what follows we provide the asymptotic geometric behavior for the important Binary Markov Source, and we derive a simple approximation which determines the tail behavior.

Examining (24), the asymptotic behavior of the deficit function is obtained by z^* , ($\eta \triangleq 1/z^*$) the smallest positive real root of the determinant $\Delta(z) = |z\mathbf{I} - \mathbf{B}^K(z)|$ outside the unit circle. For small values of K , z^* can be found explicitly. For example for $K = 2$ we can easily show that $z^* = (p_{00}/p_{11})^2$, and for $K = 3$, after some algebra we get

$$z^* = \frac{2p_{00}^3}{p_{01}(p_{01}^2 + 3p_{00}p_{11}) + (p_{01}^2 + p_{00}p_{11})\sqrt{p_{01}^2 + 4p_{00}p_{11}}} \quad (45)$$

In general, the exact evaluation of z^* involves solution of a polynomial of order $K - 1$, which in general is not possible for $K > 4$.

Alternatively, we use spectral decomposition method and take advantage of the closed form expressions for the eigenvalues of the two by two $\mathbf{B}(z)$ matrix in this simple case. We denote the eigenvalues of $\mathbf{B}(z)$ by $\lambda_1(z)$ and $\lambda_2(z)$. We have

$$\begin{aligned} \lambda_{1,2}(z) &= [p_{00} + zp_{11}]/2 \\ &\pm \frac{\sqrt{(p_{00} + zp_{11})^2 - 4z(p_{00} + p_{11} - 1)}}{2} \end{aligned} \quad (46)$$

Then

$$\Delta(z) = (z - \lambda_1^K(z))(z - \lambda_2^K(z)) \quad (47)$$

It can be shown that if the system is stable, i.e., $K\rho < 1$, the root z^* is the smallest positive real root of the equation $z = \lambda_1^K(z)$ outside the unit circle. For numerical solution of this equation, we may use any standard root finding technique. However, a simple approximation is possible. Denote the average ON period of the Binary Markov Source

by $T_{on} \triangleq 1/p_{10}$. We may study the behavior of z^* in terms of the average ON period. We have the following important result:

Theorem: The root z^* and its inverse η allow the following Laurent expansions

$$z^* = 1 + \frac{K(1-K\rho)}{(K-1)(1-\rho)} \frac{1}{T_{on}} + \frac{K(1-K\rho)(-1+2K-K^2\rho)}{2(K-1)^2(1-\rho)^2} \frac{1}{T_{on}^2} + O\left(\frac{1}{T_{on}^3}\right) \quad (48)$$

$$\eta = 1 - \frac{K(1-K\rho)}{(K-1)(1-\rho)} \frac{1}{T_{on}} + \frac{K(1-K\rho)(1-K^2\rho)}{2(K-1)^2(1-\rho)^2} \frac{1}{T_{on}^2} + O\left(\frac{1}{T_{on}^3}\right) \quad (49)$$

Proof: See Appendix.

Although z^* does not provide the complete distribution of the queue length in the admission buffer, it fully characterizes its tail behavior. Based on the above equation, the effect of the ON period on the asymptotic behavior of the queue length distribution is very clear. For a fixed K , $K > 1$ and source utilization ρ , a large T_{on} would result in a z^* close to 1 which in turn would result in a large queue build up in the admission buffer. To see the impact of large ON period on the performance, consider the quantity $1/(1-\eta)$, which may be interpreted as the average queue length assuming that the entire distribution is geometric. Using the above result, we get the following simple expansion of $1/(1-\eta)$

$$\frac{1}{1-\eta} = \frac{1-K^2\rho}{2K(1-K\rho)} + \frac{(K-1)(1-\rho)}{K(1-K\rho)} \frac{1}{T_{on}} + O\left(\frac{1}{T_{on}^2}\right) \quad (50)$$

which indicates that $1/(1-\eta)$ is asymptotically linear in T_{on} .

V. NUMERICAL RESULTS AND CONCLUSIONS

In this section, we provide some numerical results for an important special case where the arrival process is characterized as a Binary Markov Source. The reported results are based on Section IV.B, where a simple recursive procedure for finding the state probabilities was found. In what follows, in general we assume: source utilization $\rho = .02$, bucket size of 19, i.e., $M = 20$, and tokens are assumed to be generated every 10 slots, i.e., $K = 10$, unless otherwise stated.

In Fig. 2, the cumulative probability distribution of the queue length is plotted for different values of average ON period $T_{on} \triangleq 1/p_{10}$. It is evident that the distribution is very sensitive to the average duration of ON period. In fact, for very modest values of T_{on} , the tail of the distribution is very long.

In Fig. 3, the cumulative probability distribution of the queue length is plotted for different values of bucket size $M - 1$. It is evident that the distribution is *not* very sensitive to size of the bucket. In fact for a rather small value of average ON period of 10, even with a bucket size of 24 ($M = 25$), large build up of cells may be anticipated unless a much larger bucket size is used.

In Fig. 4, the cumulative probability distribution of the queue length is plotted for different values of K . (Recall that the tokens are generated every K slots.) It is interesting to note that even for a relatively high value of the token generation rate (ten times the average rate of the source, i.e., $K = 5$), large bucket size of 19 ($M = 20$) and modest average ON period of 15, large build up of cells in the queue may be expected.

In all the above figures, it is clear that the tail behavior of queue distribution becomes geometric rather quickly. We have $\Pr(Q > i) \approx \alpha\eta^i$, for even moderately small values of i . As discussed in Section IV.B, η is the inverse of the smallest positive real root outside the unit circle of the determinant $\Delta(z)$. For the exact calculation of η , we have used bisection method to find this root. Care must be taken for an accurate determination of the root, since in almost all the examples we tried, the value of the determinant is rather small in the range $(1, 1/\eta)$. Extensive numerical experiments indicate that the asymptotic expansion (49) is very accurate even for small values of T_{on} .

Figures 5 and 6 depicts the value of η vs. T_{on} for different values of K and ρ , respectively (the exact result and the asymptotic results are almost very close to each other). It is clear that even for small values of K and ρ , η approaches to 1 rather quickly as T_{on} increases. For example, in Fig. 5, even with a token generation rate of 25 times of the arrival rate (i.e., $K = 2$), the value of η exceeds .95 for the average ON period of 40. Note that although η does not provide the complete distribution of the queue length, it fully characterizes the tail behavior.

As the expansion (50) indicates, $1/(1-\eta)$ is asymptotically linear in T_{on} . This is evident in Figures 7 and 8, where the exact value of $1/(1-\eta)$ is plotted vs. T_{on} for various values of K and ρ . Note that $1/(1-\eta)$ can be interpreted as the average queue length assuming that the entire distribution is geometric with parameter η .

An important question in using the leaky bucket as a rate based access control scheme is the determination of the appropriate bucket size and the rate which tokens have to be generated, in order have acceptable delays at the input queue for a bursty and modulated sources. Based on the results we presented, it is clear that increasing the size of the bucket does little to avoid large queue build ups in the admission and transmission queue (unless the size of the bucket is chosen to be very large, in which case the impact of the leaky bucket on the ON-OFF source is not significant). In fact, increasing the size of the bucket, reduces the amplitude of the cumulative probability distribution of the queue build up, but does not affect its slope, which largely dictates the queue behavior specially for large percentiles. It seems that in order to achieve acceptable delays at the

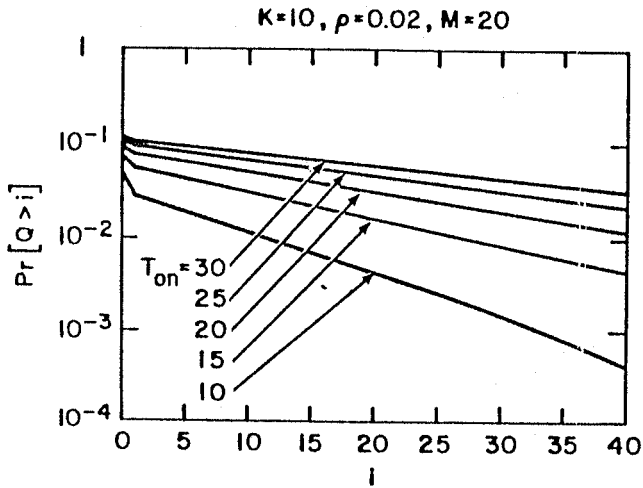


Fig. 2: Sensitivity of Queue Length Distribution to Average ON Period

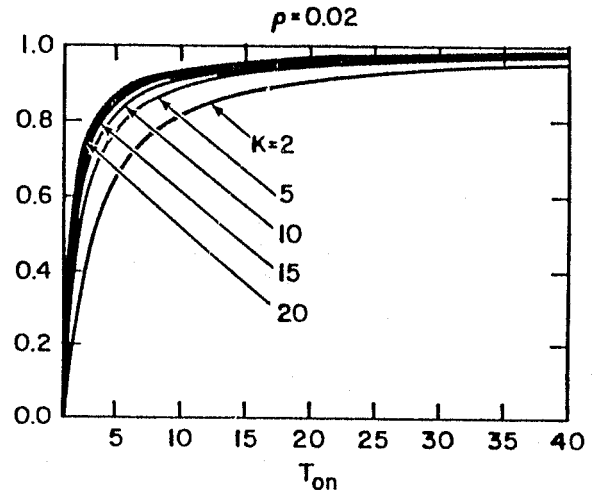


Fig. 5: Sensitivity of η to Average ON Period for Various Token Generation Rate

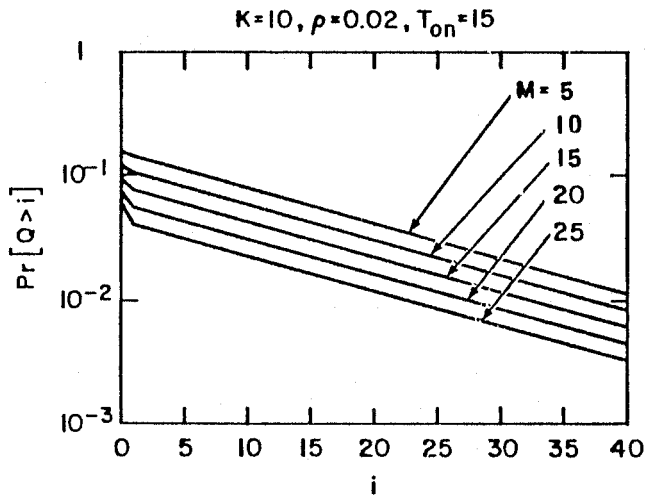


Fig. 3: Sensitivity of Queue Length Distribution to Size of Token Pool

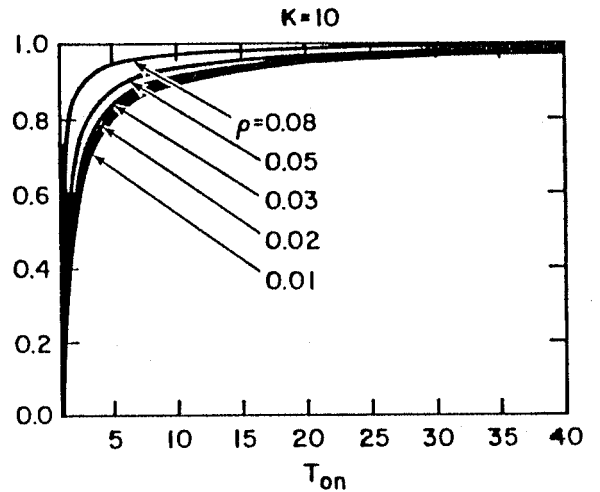


Fig. 6: Sensitivity of η to Average ON Period for Various Source Utilization Values

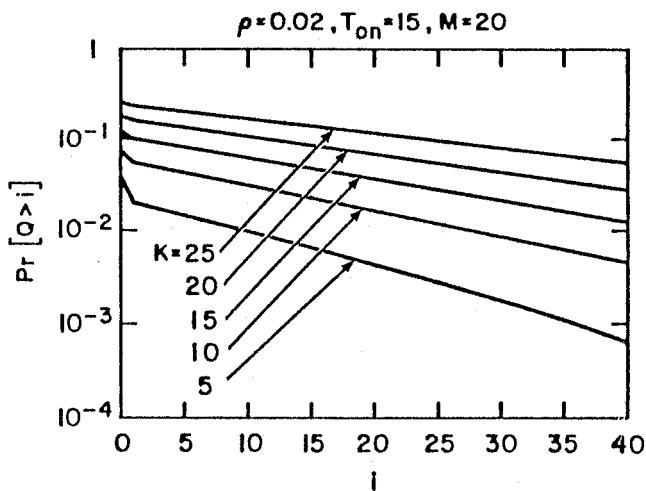


Fig. 4: Sensitivity of Queue Length Distribution to Token Generation Rate

should be generated, should be comparable to the peak rate of the source, i.e., K should be one!

As a guideline in determining the parameters of the leaky bucket, the token generation rate, $1/K$ should be chosen so that the value of η remains acceptably below 1. Once η is determined, the actual distribution of the queue length may be obtained for different values of bucket size, and based on the required delays, an appropriate value of bucket size should be chosen. As stated, the dominant factor is the rate of generation of tokens, and not the size of the bucket.

VI. SUMMARY AND DISCUSSION

In this paper we analyzed the buffered leaky bucket algorithm in discrete time when arrivals are governed by a general finite-state discrete Markov chain. The key of our analysis was the introduction of the deficit function that allows the reduction of the original problem to a more standard discrete time queuing system with the same ar-

input queue for these sources, the rate at which the tokens

REFERENCES

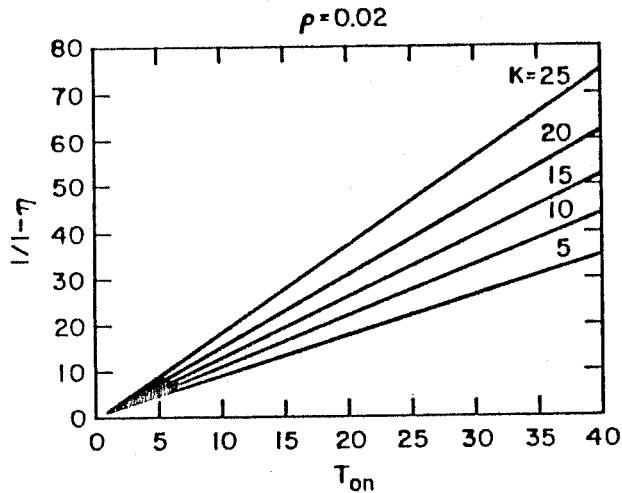


Fig. 7: Sensitivity of $1/(1-\eta)$ to Average ON Period for Various Token Generation Rates

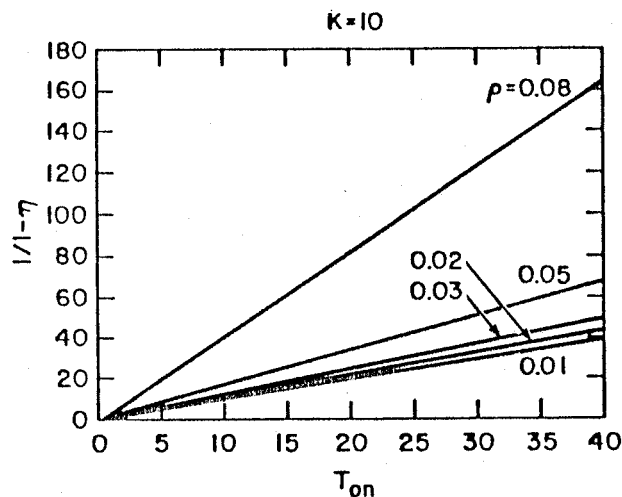


Fig. 8: Sensitivity of $1/(1-\eta)$ to Average ON Period for Various Source Utilization Values

rival process. Our analysis yields a recursive procedure for computing state probabilities. Once these probabilities are known, other quantities of interest such as waiting time distribution and distribution of the time between successive departures can be computed [1].

Note that our analysis is confined to the performance analysis at the access level of the network. Further work should be done to study the effect of the leaky bucket on the network performance. Specifically, the smoothing effect of the leaky bucket [17] on the network performance should be analyzed.

ACKNOWLEDGEMENT

The authors would like to thank H. Ahmadi, R. Guerin, J. Hui and G. Leeman for interesting discussions. They would also like to thank the reviewers for their helpful comments.

- [1] H. Ahmadi, R. Guerin and K. Sohrawy, "Analysis of a Rate-Based Access Control Mechanism for High-Speed Networks," to be presented at *GLOBECOM'90*, also an IBM Research Report, RC-15831, June 1990.
- [2] K. Bala, I. Cidon and K. Sohrawy, "Congestion Control for High-Speed Packet Switch," *INFOCOM'90*, June 1990.
- [3] A. W. Berger, "Performance Analysis of a Rate Control Throttle where Tokens and Jobs Queue," *INFOCOM'90*, pp. 30-38, June 1990.
- [4] M. L. Chaudhry, C. M. Harris and W. G. Marchal, "Robustness of Rootfinding in Single-Server Queuing Models," *ORSA Journal on Computing*, Vol. 2, pp. 273-286, 1990.
- [5] M. C. Chuah and R. L. Cruz, "Approximate Analysis of Average Performance of (σ, ρ) Regulators," *INFOCOM'90*, pp. 874-880, June 1990.
- [6] I. Cidon and I. S. Gopal, "PARIS: An Approach to Integrated High-Speed Private Networks," *International Journal of Digital & Analog Cabled Systems*, Vol. 1, No. 2, pp. 77-86, April-June 1988.
- [7] I. Cidon and M. Sidi, "Recursive Computation of Steady-State Probabilities in Priority Queuing Systems," *Operations Research Letters*, Vol. 9, No. 4, pp. 249-256, 1990.
- [8] R. L. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation; Part II: Network Analysis," *IEEE Transactions on Information Theory*, pp. 114-141, January 1991.
- [9] J. N. Daigle, Y. Lee and M. N. Magalhaes, "Discrete Time Queues with Phase Dependent Arrivals," *INFOCOM'90*, pp. 728-732, June 1990.
- [10] A. E. Eckberg, D. T. Luan and D. M. Lucantoni, "Meeting the Challenge: Congestion and Flow Control Strategies for Broadband Information Transport," *GLOBECOM'89* pp. 49.3.1-49.3.5, 1989.
- [11] D. P. Heyman, "A Performance Model of the Credit Manager Algorithm," *ORSA/TIMS*, May 1990.
- [12] G. B. Leeman, private communication.
- [13] D. M. Lucantoni, K. S. Meier-Hellstern and M. F. Neuts "A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes," to appear in *Advances in Applied Probability*.
- [14] M. F. Neuts, *Structured Stochastic Matrices of the M/G/1 Type and Their Applications*, New York, Marcel Dekker Inc., 1989.
- [15] M. F. Neuts and Y. Takahashi, "Asymptotic Behavior of the Stationary Distributions in the GI/PH/c Queue with Heterogeneous Servers," *Wahrscheinlichkeitstheorie*, Vol. 57, pp. 441-452, 1981.
- [16] E. P. Rathgeb, "Modeling and Performance Comparison of Policing Mechanisms for ATM Networks," *IEEE Journal on Selected Areas on Communications*, Vol. 9, No. 3, pp. 325-334, April 1991.
- [17] M. Sidi, W. Z. Liu, I. Cidon and I. Gopal, "Congestion Control Through Input Rate Regulation," *GLOBECOM'89*, pp. 1764-1768, November 1989 (see also in the *IEEE Transactions on Communications*, Vol. COM-41, No. 3, pp. 471-477, March 1993).
- [18] J. Turner, "New Directions in communications (or Which Way to the Information Age?)," *IEEE Communications Magazine*, October 1986, Vol. 24, No. 10.
- [19] G. M. Woodruff, R. G. H. Rogers and P. S. Richards, "A Congestion Control Framework for High-Speed Integrated Packetized Transport," *GLOBECOM'88* pp. 203-207, 1988.

APPENDIX

Let $t = 1/T_{on}$. The parameters of the Binary Markov Source are expressed as follows

$$p_{00} = 1 - \frac{\rho t}{1 - \rho}, \quad p_{11} = 1 - t$$

and obviously $p_{01} = 1 - p_{00}$ and $p_{10} = 1 - p_{11}$.

In what follows, we study the root $z^*(t)$ of the equation

$z = \lambda_1^K(z)$ as a function of t . Let

$$z^*(t) = \sum_{n=0}^{\infty} x_n \frac{t^n}{n!}$$

so that we get a system of equations satisfying x_i from

$$x_i = \left. \frac{\partial^i}{\partial t^i} \lambda_1^K \left(\sum_{n=0}^{\infty} x_n \frac{t^n}{n!} \right) \right|_{t=0}$$

It is easy to see that $\lambda(1) = 1$ so that $x_0 = 1$. Using these information and taking derivative w.r.t t , and letting $t = 0$, after some algebra we get the following equation for x_1

$$x_1 = K \left(\frac{x_1}{2} - \frac{1}{2(1-\rho)} + f(x_1) \right)$$

where

$$f(x_1) = \frac{\sqrt{(1-\rho)^2 x_1^2 - 2(1-\rho)(1-2\rho)x_1 + 1}}{2(1-\rho)}$$

which has the non-zero solution of

$$x_1 = \frac{K(1-K\rho)}{(K-1)(1-\rho)}$$

Similarly we can solve for x_2 . The asymptotic expansion of η follows easily from the expansion of $1/z^*$.

KHOSROW SOHRABY (S'79-M'84-SM'89) received B.Eng and M.Eng degrees from the McGill University, Montreal, Canada in 1979 and 1981, respectively, and Ph.D degree in 1985 from the University of Toronto, Toronto, Canada, all in Electrical Engineering.

During 1984-1986 he was a research associate at the L'institute national de la recherche scientifique - INRS Telecommunications, Montreal, Canada. In 1986 he joined AT&T Bell Laboratories, Holmdel, NJ as a Member of Technical Staff in the Teletraffic Theory and System Performance Analysis Department. In 1989 he joined IBM T.J. Watson Research Center, Yorktown Heights, NY as a Research Staff Member in the Communications Systems Department. While at IBM research, he was an adjunct faculty at Polytechnic University and Columbia University in the Department of Electrical Engineering teaching undergraduate and graduate courses in telecommunications. Since 1994 he has been a professor with the CSTP at the University of Missouri-Kansas City.

He is the former associate editor of the *IEEE TRANSACTIONS ON EDUCATION* in the area of Communications and DSP and a technical editor of the *NETWORK MAGAZINE*, member of editorial board of the *JOURNAL OF COMPUTER NETWORKS AND ISDN SYSTEMS*, *INTERNATIONAL JOURNAL OF WIRELESS INFORMATION SYSTEMS*, and *WIRELESS NETWORKS*. He has served as the guest editor of two special issues of the *JOURNAL OF SELECTED AREAS IN COMMUNICATIONS* on "Congestion Control in High Speed Packet Switching Networks" and "Wireless and Mobile High Speed Communication Networks: Architecture, Modelling and Analysis". He was the guest editor of a special issue of the *COMMUNICATIONS MAGAZINE* on "Congestion Control in High Speed Networks". He is an active member of the IEEE Communications Society and has served on the technical program committee of many IEEE conferences and workshops. He is the technical program vice-chair for Infocom '94 and the technical program co-chair for Infocom '95.

His current research interests are in design and analysis of high speed communication networks, networking aspects of wireless communications, performance analysis and probabilistic analysis of algorithms.

MOSHE SIDI (S'77-M'82-SM'87) received the B.Sc., M.Sc. and the D.Sc. degrees from the Technion - Israel Institute of Technology,

Haifa, Israel, in 1975, 1979 and 1982, respectively, all in electrical engineering. In 1982 he joined the faculty of Electrical Engineering Department at the Technion. During the academic year 1983-1984 he was a Post-Doctoral Associate at the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology, Cambridge, MA. During 1986-1987 he was a visiting scientist at IBM, Thomas J. Watson Research Center, Yorktown Heights, NY. He received the New England Academic Award in 1989. He coauthors the book "Multiple Access Protocols: Performance and Analysis," Springer Verlag 1990. He served as the Editor for Communication Networks in the *IEEE TRANSACTIONS ON COMMUNICATIONS*. Currently he serves as an Editor in the *IEEE/ACM TRANSACTIONS ON NETWORKING* and as the Associate Editor for Communication Networks and Computer Networks in the *IEEE TRANSACTIONS ON INFORMATION THEORY*. His current research interests are in queuing systems and in the area of computer communication networks.