Traffic Shaping of a Tagged Stream in an ATM Network: Approximate End-to-End Analysis *

Randall Landry

Integrated Systems Laboratory
Texas Instruments Inc.
Dallas, TX 75265

Ioannis Stavrakakis

Dept. of Electrical and Computer Engineering
Northeastern University
Boston, MA 02115

Abstract

Traffic streams—originally shaped to comply with desired traffic descriptors—may suffer substantial distortion within an ATM network due to statistical multiplexing. As a consequence, traffic re-shaping within the network may be necessary. In this work the modification of a tagged traffic stream due to a series of multiplexing operations is studied. The standard FCFS service policy, as well as a peak-rate enforcing multiplexing discipline, are considered. Tagged cell delay, jitter and inter-departure processes are considered for the characterization of the tagged traffic stream.

1 Introduction

The Asynchronous Transfer Mode (ATM) is presently considered to be the most promising switching and multiplexing scheme for B-ISDN, as manifested by the large number of CCITT recommendations related to ATM [1]. Although high network efficiency (utilization) can be achieved through the statistical multiplexing of bursty sources, ATM cannot guarantee certain levels of Quality of Service (QoS) since the latter depends on service demands of the other users sharing the same network resources. Since at a given time instant the cumulative peak rate of the supported ATM users may exceed the network capacity, serious congestion may occur, resulting in QoS degradation.

Source traffic shaping and/or regulation has been proposed as a means of reducing traffic burstiness, thereby controlling network congestion. However, the identification of effective descriptors for the regulated source traffic is still a largely unsolved problem. Such descriptors should be simple enough to allow for a quick computation of the projected network performance, yet detailed enough to yield a sufficiently accurate result. Examples of traffic descriptors are peak and average rate and various measures of burstiness of the regulated traffic [2].

Source traffic descriptors can provide important information on the projected traffic load and are necessary for the implementation of any Call Admission Control (CAC). Yet such descriptors may not be accurately describing the tagged traffic stream after one or more multiplexing stages. It is well understood that statistical multiplexing may induce substantial distortion in an ATM cell stream [3]. Thus, characterization of a tagged traffic stream within the network seems to be necessary in order to accurately evaluate potential congestion, as well as determine the QoS delivered to the tagged stream through a per-session end-to-end study [4],[5]. In addition, a corrective action (that is, traffic shaping) may be necessary within the network for the purpose of reducing distortion, which is a potential cause of network congestion.

In this work, the distortion of a well regulated source traffic stream due to a series of multiplexing operations is studied under the standard FCFS, as well as a "corrective" peak-rate enforcing, multiplexing discipline. The tagged traffic stream is characterized at the output of each of a series of nodes in terms of cell delay, delay jitter and inter-departure processes. Past work regarding the characterization of a tagged traffic stream due to a single FCFS multiplexing operation may be found in [6], [7]. In [6] a method is developed for calculating the cell inter-departure time distribution of a general renewal stream under a FCFS policy. Statistics for the inter-departure process of an initially periodic cell stream in a FCFS multiplexer are also computed in [7] by assuming a first-order Markovian

*Research supported by the Advanced Research Project Agency under Grant F49620-93-1-0564 monitored by the Air Force Office of Scientific Research (AFOSR).

structure for the cell delay process.

Finally, the node-by-node approximate description of the transformation of a tagged stream under the considered multiplexing disciplines is used for an approximate study of the end-to-end QoS delivered to the tagged traffic stream. Past work on approximate end-to-end network performance evaluation based on nodal decomposition and FCFS multiplexing may be found in [8],[9],[10].

In the next section the fundamental block of this work is considered, namely a queueing system under a service policy which occasionally departs from the FCFS policy to enforce a peak output rate (minimum cell inter-departure time) for the tagged traffic stream. When the peak-rate enforcer is disabled (by properly setting a key parameter), the proposed policy, which was originally introduced in [11], becomes the standard FCFS policy considered in [6],[7]. In section 4, the approximate end-to-end study is described along with a discussion of important assumptions. Numerical results are presented in section 4 before concluding in the final section.

2 Analysis of a Single Multiplexing Stage

2.1 The Peak Output-Rate Enforcing (PORE) Service Discipline

The PORE service discipline is associated with a tagged traffic stream (Class-1) and treats the rest of the traffic in an indistinguishable manner (Class-2 or background traffic). As it becomes clear below, the PORE service strategy for Class-1 guarantees that cells belonging to this class are transmitted at a minimum spacing of X_{min} slots. The server operates according to the FCFS discipline except for those times when the FCFS server would place consecutive Class-1 cells on the output link at a distance of less than X_{min}. In this paper it is assumed that under the FCFS policy, Class-1 cells are served before the Class-2 cells which arrive over the same slot.

Let t_k denote the service epoch of the k-th Class-1 cell, defined as the beginning of the slot over which this cell is transmitted. Let $X_k = t_{k+1} - t_k$ denote the inter-departure time of Class-1 cells $k+1$ and k. Similarly, $\tilde{t}_k$ will mark the *potential* service epoch of the k-th Class-1 cell, which is defined to be the beginning of the slot at which all other cells that arrived prior to the k-th Class-1 cell have been served. Note that under the FCFS discipline, $\tilde{t}_k = t_k$, $\forall k \geq 1$.

Under the PORE service discipline, whenever $\tilde{t}_k - t_{k-1} < X_{min}$, the server delays the initiation of service to the k-th Class-1 cell by exactly $H_k = X_{min} - (\tilde{t}_k - t_{k-1})$ slots, which will be referred to as the *holding time* of the k-th cell for the remainder of the paper. During this holding time of length H_k, the server will provide FCFS service to Class-2 cells that may be present in the buffer. When the holding time expires, the k-th Class-1 cell is served; thus, $t_k = \tilde{t}_k + H_k$, thereby ensuring that $X_{k-1} = X_{min}$. Note that the PORE service policy reduces to the FCFS policy when $X_{min} = 1$, in which case $H_k = 0$, $\forall k \geq 1$.

Let the Class-1 cell inter-arrival process be denoted by $\{A_k\}_{k\geq 1}$, where A_k represents the inter-arrival time (in slots) between Class-1 cells k and $k+1$. It is assumed that $A_k \in [A_{min}, \infty) \subseteq \mathbf{Z}^+$, where $\mathbf{Z}^+$ denotes the set of positive integers. Class-2 traffic is considered to be the background traffic whose presence stands to corrupt $\{A_k\}_{k\geq 1}$ through the queueing process.

2.2 Jitter (Cell Delay Variation) and the PORE Policy

A number of definitions for delay jitter have appeared in the literature, including those which compute the probability distribution of cell delays and measure the difference between lower and upper quantiles [7],[12]. In this paper a cell-level definition for delay jitter, like the one considered in [7], will be adopted to describe the amount of distortion introduced into the Class-1 cell stream.

Definition 1 *Let D_k denote the total queueing delay (in slots) of the k-th Class-1 cell, including the transmission slot. The delay jitter of the k-th cell is defined as*

$$J_k = D_{k+1} - D_k \quad \textit{for } k \geq 1. \quad \Box \tag{1}$$

In terms of the quantities defined in the previous subsection, an equivalent definition of delay jitter is given by

$$J_k = X_k - A_k \quad \text{for } k \geq 1. \tag{2}$$

Note that J_k measures the degree to which the inter-arrival time A_k of Class-1 cells k and $k+1$ has been distorted, and it can take both positive and negative values. *Negative* delay jitter applies to those instances when Class-1 cells k and $k+1$ are transmitted at a spacing of less than A_k (cell clustering), while *positive* delay jitter applies to those times when cells k and $k+1$ are transmitted at a spacing of greater than A_k (cell spreading). Note that when the process $\{A_k\}_{k\geq 1}$

has a finite region of support (i.e. $A_k \in [A_{min}, A_{max}]$), the PORE policy completely eliminates a portion of the negative delay jitter by guaranteeing that $J_k \geq X_{min} - A_{max}$.

It is also proven in Theorem 1 of [11] that enforcing a peak output rate for Class-1 does not increase the positive delay jitter of this traffic class. The PORE service policy actually reduces the two types of distortion most commonly induced by multiplexers, namely spreading and clustering.

2.3 Queueing Analysis

In this section, an ATM multiplexer operating under the PORE service policy (Fig. 1) is studied in terms of performance measures associated with the Class-1 cell stream. This system was analyzed in [11] under the assumption that the Class-1 cell stream was periodic. The periodic assumption is relaxed in this paper to allow for a Class-1 inter-arrival process $\{A_k\}_{k\geq 1}$ which is i.i.d. with distribution $f_a(m)$, $A_{min} \leq m \leq A_{max}$. Note that a periodic process with period T can be considered when $A_{min} = A_{max} = T$ and $f_a(T) = 1$.

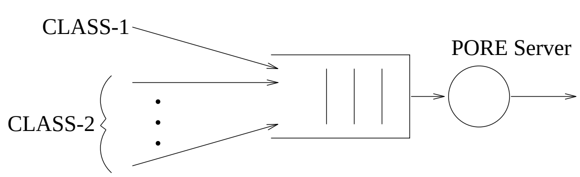

Figure 1: The discrete-time queueing system

Class-2 traffic is modeled in terms of i.i.d. batches of cells of arbitrary distribution. Let the random variable B_j—with probability mass function (PMF) $f_{b_j}(k)$, $k \geq 0$—denote the number of Class-2 cells arriving over an interval of length j (in slots). Clearly, $f_{b_j}(\cdot)$ can be derived as the j-fold convolution of $f_{b_1}(\cdot)$; let N^b be the maximum value of B_1.

Let $Q(n;j)$ denote the number of Class-2 cells which arrived before time epoch j and are still in the system at time n, $n \geq j$. The following proposition determines the evolution of $\{Q(n;j)\}_{n\geq j}$ under certain conditions, and will be useful for the study of the queueing system.

Proposition 1 *The process $\{Q(j+m;j)\}_{m=0}^{n}$ evolves as the occupancy process of a Geo/D/1 queue with batch arrivals distributed according to $f_{b_1}(\cdot)$, provided that <u>no</u> Class-1 cells are served over the interval $[j, j+n]$. In this case,*

$$\begin{aligned} P_{k,\ell}^{(m)} &= Pr\{Q(j+m;j) = \ell \mid Q(j;j) = k\} \\ &= \sum_i P_{k,i}^{(m-1)} P_{i,\ell}, \end{aligned} \tag{3}$$

where $P_{k,\ell}^{(m)}$ ($P_{i,\ell}$) is the m-step (1-step) transition probability for the Geo/D/1 *queue; $P_{i,\ell} = f_{b_1}(\ell - i + 1) + f_{b_1}(0)\mathbf{1}_{\{\ell+i=0\}}$.* □

The queueing system will be described in terms of the Markov chain $\{D_k, H_k\}_{k\geq 1}$, embedded at the sequence of time instants $\{t_k\}_{k\geq 1}$; t_k marks the service epoch of the k-th Class-1 cell as defined earlier; D_k and H_k denote the total delay and holding time of the k-th cell, respectively.

The transition probabilities for the Markov chain $\{D_k, H_k\}_{k\geq 1}$ will be derived by first establishing expressions for D_k and H_k given that $(D_{k-1}, H_{k-1}) = (i,j)$, where $i \geq 1$ and $0 \leq j \leq X_{min} - 1$. In order to simplify notation, the quantity $Q(t_{k-1}; t_{k-1} - i + j)$, which appears frequently in the following expressions, will be denoted by Q^*.

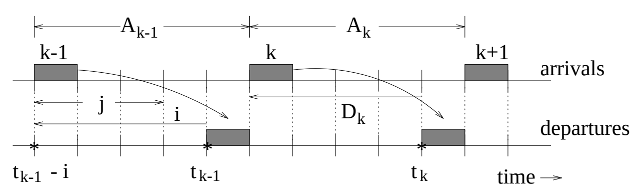

Figure 2: Realization for the derivation of D_k and H_k when $i < A_{max}$.

When $i < A_{max}$, the expressions for D_k and H_k can be derived by referring to Fig. 2, where $i < A_{k-1}$ for the realization shown there. Note that i can also take values greater than or equal to A_{k-1} while still being less than A_{max}, and these cases are also accounted for in the expressions given below:

$$D_k = \begin{cases} 1 \\ \quad \text{if } Q^* = 0 \ ; \ i < A_{k-1} \\ Q^* \\ \quad \text{if } Q^* \geq X_{min} + i - A_{k-1} \ ; \ i < A_{k-1} \\ Q^* + H_k \\ \quad \text{if } Q^* < X_{min} + i - A_{k-1} \ ; \ i < A_{k-1} \\ B_{A_{k-1}-j} + Q^* + 1 + i - A_{k-1} \\ \quad \text{if } Q^* \geq X_{min} - B_{A_{k-1}-j} - 1 \ ; \ i \geq A_{k-1} \\ B_{A_{k-1}-j} + Q^* + 1 + i - A_{k-1} + H_k \\ \quad \text{if } Q^* < X_{min} - B_{A_{k-1}-j} - 1 \ ; \ i \geq A_{k-1} \end{cases} \tag{4}$$

and

$$
H_k = \begin{cases} 0 & \text{if } Q^* = 0 \;;\; i < A_{k-1} \\ 0 & \text{if } Q^* \geq X_{min} + i - A_{k-1} \;;\; i < A_{k-1} \\ X_{min} + i - A_{k-1} - Q^* & \text{if } Q^* < X_{min} + i - A_{k-1} \;;\; i < A_{k-1} \\ 0 & \text{if } Q^* \geq X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \\ X_{min} - B_{A_{k-1}-j} - Q^* - 1 & \text{if } Q^* < X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \end{cases} \tag{5}
$$

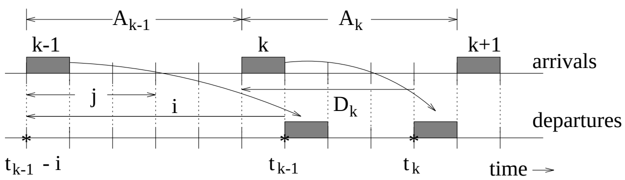

Figure 3: Realization for the derivation of D_k and H_k when $i \geq A_{max}$.

When $i \geq A_{max}$ (Fig. 3),

$$
D_k = \begin{cases} B_{A_{k-1}-j} + Q^* + 1 + i - A_{k-1} & \text{if } Q^* \geq X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \\ B_{A_{k-1}-j} + Q^* + 1 + i - A_{k-1} + H_k & \text{if } Q^* < X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \end{cases} \tag{6}
$$

and

$$
H_k = \begin{cases} 0 & \text{if } Q^* \geq X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \\ X_{min} - B_{A_{k-1}-j} - Q^* - 1 & \text{if } Q^* < X_{min} - B_{A_{k-1}-j} - 1 \;;\; i \geq A_{k-1} \end{cases} \tag{7}
$$

Let $S^L = \{(i,j)\}$, where $L(A_{max}-1)+1 \leq i \leq (L+1)(A_{max}-1)$ and $0 \leq j \leq X_{min}-1$, represent "level" L of the Markov chain $\{D_k, H_k\}_{k\geq 1}$. Given the relationships in (4)-(7), the transition probabilities for $\{D_k, H_k\}_{k\geq 1}$, which will be denoted by $p(i,j,i',j')$, can be computed in terms of the quantities previously defined. The quantities $p(i,j,i',j')$, for $(i,j) \in S^0$ (transitions from level 0) are given by equations (8) and (9), which are derived from (4) and (5) above, while the transition probabilities for $(i,j) \in S^1$ are given by equations (10) and (11), derived from (6) and (7) above.
For $(i,j) \in S^0$,

$$
\begin{aligned} p(i,j,i',0) \; = \; & \sum_{n=A_{min}}^{A_1} \sum_{m=0}^{N_1} f_a(n) b_{n-j}(m) \\ & P^{(j)}_{0,i'-i+n-m-1} + \sum_{n=A_2}^{A_{max}} \sum_{l=0}^{(N^b-1)j} \sum_{m=0}^{N^b(i-j+1)} f_a(n) P^{(j)}_{0,l} \\ & b_{i-j+1}(m) \left[P^{(n-i-1)}_{m+l,i'} + P^{(n-i-1)}_{m+l,0} \mathbf{1}_{\{i'=1\}} \right], \end{aligned} \tag{8}
$$

where $A_1 = min\,\{i, A_{max}\}$, $A_2 = max\,\{A_{min}, i+1\}$ and $N_1 = min\,\{(n-j)N^b, i'-1-i+n\}$.

$$
\begin{aligned} p(i,j,i',j') \; = \; & \sum_{n=A_{min}}^{A_1} \sum_{m=0}^{N_2} f_a(n) b_{n-j}(m) \\ & P^{(j)}_{0,X_{min}-1-m-j'} + \sum_{n=A_2}^{A_{max}} \sum_{l=0}^{(N^b-1)j} \sum_{m=0}^{N^b(i-j+1)} f_a(n) P^{(j)}_{0,l} \\ & b_{i-j+1}(m) \Big[P^{(n-i-1)}_{m+l,i'-j'} + P^{(n-i-1)}_{m+l,0} \Big] \mathbf{1}_{\{i'=X_{min}-n+i\}}, \end{aligned} \tag{9}
$$

where $N_2 = min\,\{(n-j)N^b, X_{min}-1-j'\}$.
For $(i,j) \in S^1$,

$$
p(i,j,i',0) = \sum_{n=A_{min}}^{A_1} \sum_{m=0}^{N_1} f_a(n) b_{n-j}(m) P^{(j)}_{0,i'-i+n-m-1}, \tag{10}
$$

and

$$
\begin{aligned} p(i,j,i',j') \; = & \sum_{n=A_{min}}^{A_1} \sum_{m=0}^{N_2} f_a(n) b_{n-j}(m) P^{(j)}_{0,X_{min}-1-m-j'} \\ & \mathbf{1}_{\{i'=X_{min}-n+i\}}. \end{aligned} \tag{11}
$$

By lexicographically mapping the quantities $p(i,j,i',j')$ into $p(\ell,\ell')$ through the relations $\ell = (i-1)X_{min}+j$ and $\ell' = (i'-1)X_{min}+j'$, the transition probability matrix $\mathbf{P}$ of the Markov chain $\{D_k, H_k\}_{k\geq 1}$ can be written as a stochastic matrix of the M/G/1 type:

$$
\mathbf{P} = \begin{bmatrix} \mathbf{B_0} & \mathbf{B_1} & \mathbf{B_2} & \mathbf{B_3} & \cdots \\ \mathbf{A_0} & \mathbf{A_1} & \mathbf{A_2} & \mathbf{A_3} & \cdots \\ 0 & \mathbf{A_0} & \mathbf{A_1} & \mathbf{A_2} & \cdots \\ 0 & 0 & \mathbf{A_0} & \mathbf{A_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{12}
$$

where the matrices $\mathbf{B_n}$ and $\mathbf{A_n}$, $n \geq 0$, have dimensions $(A_{max}-1)X_{min} \times (A_{max}-1)X_{min}$. Note that the matrices $\mathbf{B_n}$ consist of the transition probabilities for which $(i,j) \in S^0$ and $(i',j') \in S^n$, while the matrices $\mathbf{A_n}$ consist of the transition probabilities for which $(i,j) \in S^1$ and $(i',j') \in S^n$. The system in (12) can now be solved based on the matrix analytic techniques presented in [13]. The stationary probability vector $\boldsymbol{\pi}$ for $\mathbf{P}$ is mapped back into the space (i,j,i',j') to obtain the stationary probabilities for $\{D_k, H_k\}_{k\geq 1}$, given by $\pi(i,j)$, for $1 \leq i < \infty$, $0 \leq j \leq X_{min}-1$.

The PMF for cell delay D of Class-1 is easy to calculate in terms of $\pi(i,j)$, and it is given by

$$Pr\left\{D=i\right\} = \sum_{j=0}^{X_{min}-1} \pi(i,j) \quad . \tag{13}$$

By considering the definition in (1), the delay jitter distribution is easily determined to be

$$Pr\left\{J=k\right\} = \sum_{i=1}^{\infty} \sum_{j=0}^{X_{min}-1} \sum_{j'=0}^{X_{min}-1} p(i,j,i+k,j')\pi(i,j) \tag{14}$$

Computing the PMF of the Class-1 cell inter-departure time X is not as straight-forward as computing that of J. From equations (1) and (2), the inter-departure time of cells $k+1$ and k is given as $X_k = D_{k+1} - D_k + A_k$. In order to compute the PMF of X, the system will be described by the Markov chain $\{D_k, H_k, A_k\}_{k\geq 1}$. Let $\tilde{p}(i,j,m,i',j',m')$ denote the transition probabilities and $\tilde{\pi}(i,j,m)$ denote the stationary probabilities for this Markov chain. It is easy to show that the PMF for the cell inter-departure process is given by

$$Pr\left\{X=n\right\} = \sum_{i=1}^{\infty} \sum_{j=0}^{X_{min}-1} \sum_{m=A_{min}}^{A'} \sum_{j'=0}^{X_{min}-1} \sum_{m'=A_{min}}^{A_{max}} \tilde{p}(i,j,m,n+i-m,j',m')\tilde{\pi}(i,j,m), \tag{15}$$

where $A' = min\left\{n+i-1, A_{max}\right\}$.

The quantities $\tilde{\pi}(i,j,m)$ can be computed in the same manner as the quantities $\pi(i,j)$ with the exception that the dimensionality of each level in $\mathbf{P}$ has increased by a factor of A_{max}. Note, however, that upon arrival of the k-th Class-1 cell, D_k is determined. That is, D_k is equal to the queue content at the arrival instant plus the amount of holding time, H_k, to be suffered by the cell. The latter quantity is completely determined by the past (i.e. the service time of the $(k-1)-st$ cell) and neither D_k nor H_k are dependent upon the next Class-1 arrival. Consequently, the random variables D_k and H_k are both independent of the random variable A_k, and the stationary probabilities for the Markov chain $\{D_k, H_k, A_k\}_{k\geq 1}$ can be computed as $\tilde{\pi}(i,j,m) = \pi(i,j) f_a(m)$, where $\pi(i,j)$ and $f_a(m)$ are as given above.

3 Approximate End-to-End Study

The tandem queueing system used for the approximate end-to-end performance evaluation of the tagged session is shown in Fig. 4. The PORE service policy will be adopted at each node so as to allow for the consideration of peak-rate enforcement within the network. The tagged session is assumed to be routed through N nodes and its cell inter-arrival process at node 1 will be periodic, which represents a CBR source at the network edge; cells are assumed to be delivered every T slots. In order to simplify notation, the Class-1 cell inter-arrival (inter-departure) process at node i will be denoted as A^i (X^i). The arrival process at node i ($2 \leq i \leq N$) consists of the tagged (Class-1) inter-departure process from node $i-1$, and some background traffic (Class-2) delivering batches of cells distributed according to the PMF $f_{b_1}(\cdot)$.

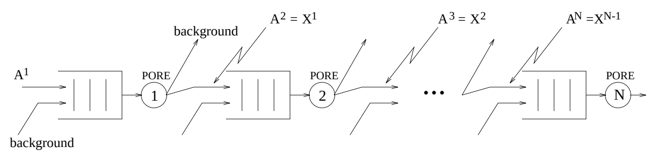

Figure 4: Tandem queueing network for end-to-end performance evaluation.

The set of possible integer values which can be taken by the Class-1 inter-arrival processes at nodes 2 through N must be truncated in order to facilitate a numerical solution at each node. The minimum (maximum) Class-1 inter-arrival time at node i will be denoted by A^i_{min} (A^i_{max}). That is, $A^i \in \left[A^i_{min}, A^i_{max}\right]$ for $i = 1, 2, \ldots, N$, where $A^1_{min} = A^1_{max} = T$; the minimum inter-arrival time for $i \geq 2$ is simply $A^i_{min} = X^{i-1}_{min}$. The choice of A^i_{max} for $i \geq 2$ will be made according to the rule

$$A^i_{max} = \max\left\{k : Pr\left\{X^{i-1}=k\right\} \geq \epsilon\right\}, \tag{16}$$

where ϵ can be any positive real number less than 1. The PMF for the process A^i is then given by

$$Pr\left\{A^i = k\right\} = \begin{cases} Pr\left\{X^{i-1}=k\right\} & \text{if } 0 < k < A^i_{max} \\ 1 - \displaystyle\sum_{j=1}^{A^i_{max}-1} Pr\left\{X^{i-1}=j\right\} & \text{if } k = A^i_{max} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

It is clear from (16) and (17) that as ϵ approaches zero, A^i_{max} increases and the truncated Class-1 inter-arrival time PMF at a node approaches the actual Class-1 inter-departure time PMF from the previous node. The cost of improving the accuracy of the adopted Class-1 inter-arrival process by increasing A^i_{max}, is a corresponding increase in numerical complexity. Recall that the dimensionality of the square matrices B_n and A_n, $n \geq 0$, which make up the transition probability matrix $\mathbf{P}$ for a single queue, increases linearly with the maximum Class-1 inter-arrival time (see section 2.3), which in turn affects the tractability of the solution technique.

A number of other assumptions will be made in order to facilitate the proposed approximate end-to-end study. First, it will be assumed that only the tagged traffic stream is forwarded from one multiplexer to the next as shown in Fig. 4. All of the remaining traffic (background) at each multiplexer is routed elsewhere.

It is also assumed that the inter-departure process $\{X_k\}_{k\geq 1}$ from each node is an i.i.d. process. This assumption is necessary for employment of the analysis of section 2.2 at each multiplexing stage. The validity of the i.i.d. assumption, as well as the effect of truncating the inter-arrival time PMF at each node, will be examined through a comparative simulation study.

The analysis presented in the previous section for a single queue will be applied successively to each node along the path in Fig. 4 and the PMF for the inter-departure process at node N will be used to establish the amount of delay jitter induced in the cell stream of the tagged session over all N hops.

4 Numerical Results and Discussion

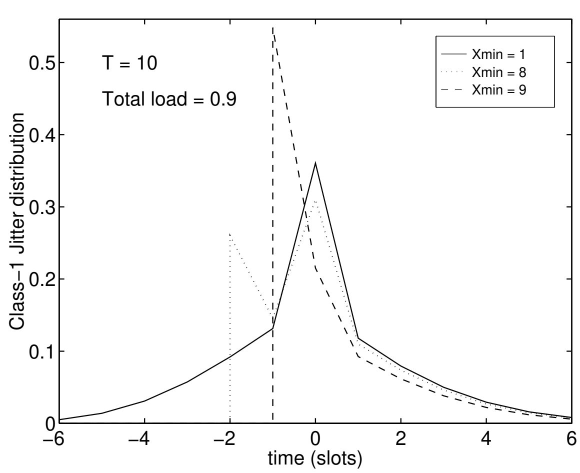

Figure 5: Jitter PMF for $T = 10$ and various values of X_{min} at a single multiplexer.

In this section, some numerical results are presented to characterize the distortion induced in a tagged traffic stream under both FCFS and peak-rate enforcing service strategies. Before considering end-to-end performance, the delay jitter induced at a single ATM multiplexer will be studied.

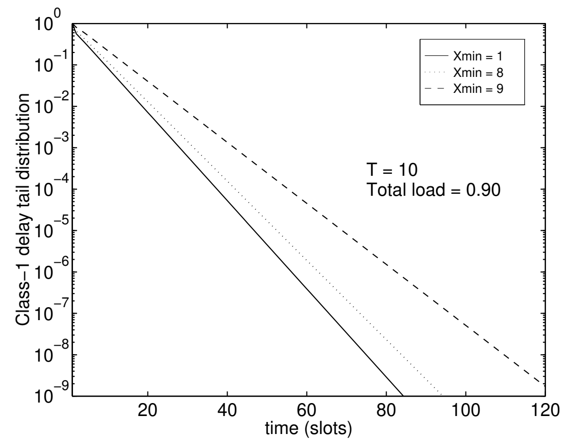

Figure 6: Delay tail distributions for $T = 10$ and various values of X_{min} at a single multiplexer.

In Fig. 5, the delay jitter PMF is considered for different values of X_{min}. The Class-1 inter-arrival process is assumed to be periodic with period $T = 10$; the Class-2 batch arrival process is assumed to be binomially distributed with mean rate $\lambda^b = 0.8$ and maximum batch size $N^b = 15$. The resulting total load is 0.9. When X_{min} is set to $T - 1$ (in this case 9), all negative jitter less than -1 is eliminated and the delay jitter process possesses a very narrow region of support. On the other hand, under a FCFS policy ($X_{min} = 1$) a significant amount of cell clustering is observed; cell spreading is also slightly increased. The cost of providing a less distorted traffic stream at the multiplexer output is illustrated in Fig. 6 in terms of increased cell delays.

The remainder of this section will deal with the approximate end-to-end performance study. A network of five nodes in tandem will be considered, where the tagged traffic stream originates as a periodic (or CBR) source at node 1 with period $T = 5$. The background traffic at each node is assumed to consist of 10 identical Bernoulli streams each with rate 0.07, resulting in a total offered load of 0.9 at each multiplexer. The Class-1 inter-arrival time PMFs at nodes 2 through 5 will be truncated by setting ϵ equal to 10^{-6} and using the rule presented in the previous section. This choice of ϵ resulted in values of A^i_{max}, $2 \leq i \leq 5$, ranging from 20 to 30.

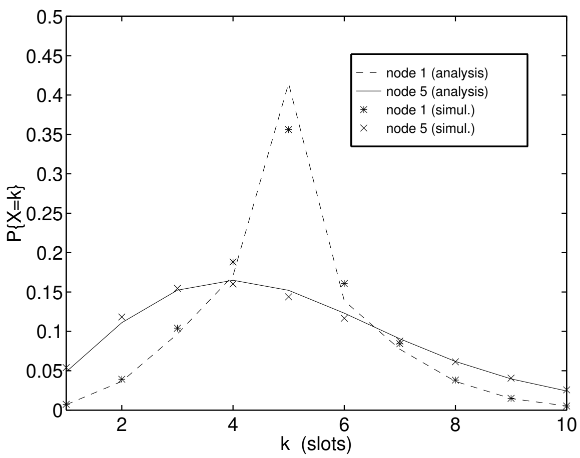

Figure 7: Cell inter-departure time PMFs at nodes 1 and 5 under FCFS policies ($X_{min} = 1$) at each node; $T = 5$.

For the results in Fig. 7, each of the five ATM multiplexers serve arriving cells according to the FCFS policy ($X_{min} = 1$). The PMFs for inter-departure processes at nodes 1 and 5 are computed and compared with simulation results for the identical tandem queueing network. Clearly, as the tagged stream passes

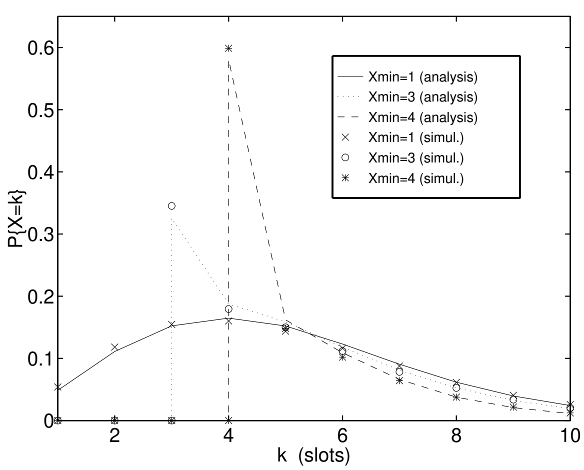

Figure 8: Cell inter-departure time PMFs at node 5 for different X_{min}.

through multiple FCFS multiplexing stages, the inter-departure times become more uniformly distributed. This implies that, in a real-time scenario, significant buffering may be required at the end user in order to achieve an acceptable playback quality. Also, a traffic profile such as that at the output of node 5 will likely induce more severe queueing problems (and possibly QoS degradation) for other users sharing the same network resources.

The validity of an i.i.d. assumption for cell inter-departures (and inter-arrivals) is addressed through the simulation results plotted in Fig. 7 as discrete points. Clearly, under a FCFS policy at each node, this assumption does not significantly affect the accuracy of the end-to-end study, as the numerical results seem to be in good agreement with simulations. Similar agreement is present in Fig. 8, where the traffic has been shaped by the peak-rate enforcing policy at each node.

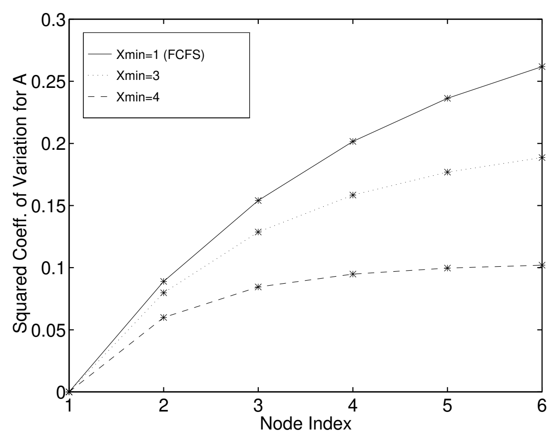

Figure 9: Squared coefficient of variation for the cell inter-arrival process to each node.

In Fig. 8 three networks of servers with X_{min} equal to 1, 3 and 4 are considered. As X_{min} increases, the end-to-end performance of the tagged stream, in terms of the induced distortion, improves dramatically. It is also interesting to note that the approximate analysis agrees very well with simulation results, which validates the i.i.d. assumption for cell inter-departures even when the corrective peak output rate enforcing queueing discipline is adopted.

The squared coefficient of variation for the Class-1 inter-arrival process A^i is plotted in Fig. 9 as a function of the node index i for the three networks of servers considered in Fig. 8. As expected, this quantity increases with the node index for all values of X_{min}. However, depending upon the peak rate being enforced by each of the servers, the rate of increase, as well as the magnitude, of squared coefficient of varia-

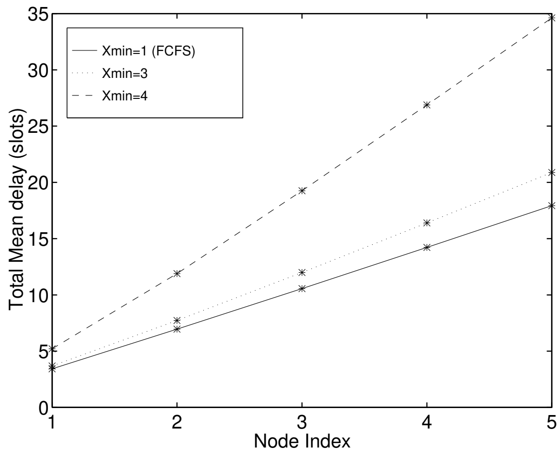

Figure 10: Total mean queueing delay as a function of node index.

tion varies greatly.

The cumulative mean queueing delay for the tagged stream is plotted as a function of the node index in Fig. 10. The total mean delay appears to increase linearly with the node index, and the rate of increase is clearly dependent upon the value of X_{min}. These results indicate that for real-time applications the tradeoff between violating deadline constraints and violating either delay jitter constraints or a declared peak rate, needs to be considered carefully.

5 Conclusions

In this paper, a queueing system was formulated to study the distortion induced in a tagged ATM cell stream; both FCFS and Peak Output Rate Enforcing (PORE) service disciplines were considered. In fact, the FCFS policy was shown to be a special case of the PORE discipline. A discrete-time analysis in the M/G/1 paradigm yielded numerical results for cell delay, delay jitter and inter-departure time probability distributions at an ATM multiplexer served according to the PORE policy.

An approximate end-to-end performance study of the tagged cell stream was also carried out by analyzing a tandem queueing network of PORE servers. Numerical results indicate that a peak rate enforced at the network edge can be severely violated within the network and, thus, traffic re-regulation may be necessary if the peak cell rate traffic descriptor is to be used by the CAC.

References

[1] R. Onvural. *Asynchronous Transfer Mode: Performance Issues.* Artech House, 1993.

[2] ATM Forum. *ATM User-Network Interface Specification.* Version 3.0, 1993.

[3] I. Stavrakakis. Efficient modeling of merging and splitting processes in large networking structures. *IEEE J. Select. Areas Commun.*, 9(8), Oct. 1991.

[4] J. Kurose. Open issues and challenges in providing quality of service guarantees in high speed networks. *Computer Communications Review*, 23(1), Jan. 1993.

[5] D. Towsley. Providing quality of service in packet switched networks. *Performance Evaluation of Computer and Communication Systems*, 1993. Joint Tutorial Papers of Performance'93 and Sigmetrics'93.

[6] W. Matragi, C. Bisdikian, and K. Sohraby. Jitter calculus in atm networks: Part i, single node. In *IEEE INFOCOM'94*, Toronto, Canada, 1994.

[7] J. Roberts and F. Guillemin. Jitter in atm networks and its impact on peak rate enforcement. *Perf. Eval.*, 16(1–3), Nov. 1992.

[8] W. Matragi, K. Sohraby, and C. Bisdikian. Jitter calculus in atm networks: Part ii, multiple nodes. In *IEEE INFOCOM'94*, Toronto, Canada, 1994.

[9] H. Kroner, M. Eberspacher, T. Theimer, P. Khun, and U. Briem. Approximate analysis of end-to-end delay in atm networks. In *IEEE INFOCOM'92*, Florence, Italy, 1992.

[10] J.F. Ren, J. Mark, and J. Wong. End-to-end performance in atm networks. In *ICC'94*, New Orleans, LA, 1994.

[11] R. Landry and I. Stavrakakis. A queueing study of peak-rate enforcement for jitter reduction in atm networks. In *IEEE GLOBECOM,94*, San Fransisco, CA, 1994.

[12] F. Guillemin and W. Monin. Management of cell delay variation in atm networks. In *IEEE GLOBECOM'92*, Orlando, FL, 1992.

[13] M.F. Neuts. *Structured Stochastic Matrices of the M/G/1 Type and their Applications.* Marcel Dekker, 1989.