A Queueing Study of Peak-Rate Enforcement for Jitter Reduction in ATM Networks *

Randall Landry† Ioannis Stavrakakis‡
Department of Electrical Engineering and Computer Science
University of Vermont
Burlington, VT 05405

Abstract

In this paper, a dual-class queueing system is formulated and applied to study the effects of cell spacing within the network premises as a means of controlling network-induced delay jitter. The adopted service policy regulates the traffic class of interest by enforcing a predetermined peak output rate. Probability distributions for delay and jitter of the regulated traffic class are derived by analyzing the system in discrete time. While maintaining a high degree of simplicity, the proposed policy is shown to provide improved jitter performance for the regulated traffic class over the First-In First-Out (FIFO) service policy.

1 Introduction

The success of future broadband packet-switched networks, based on technologies such as the Asynchronous Transfer Mode (ATM), will be heavily dependent upon their ability to provide quality of service (QOS) guarantees to a wide range of traffic classes. A substantial portion of this traffic is expected to be generated by real-time sources with diversified service requirements, such as those in multimedia-based applications. Unlike in traditional data-based communications, the QOS associated with real-time applications cannot be adequately described in terms of network induced mean cell delay and losses, due to the significant impact on performance of cell loss correlations and delay variability. As a result, second-order statistics for cell losses, tail distributions for cell delays (which can be related to the probability that some predetermined delay deadline is violated) and the probability distribution of delay jitter are considered to be important QOS descriptors for real time applications [1].

The computation of performance measures involving distributions and second order statistics of individual traffic classes is a challenging problem, even in the simple case of a FIFO service policy. Some progress in this direction is reported in [2], where a method is developed for calculating the cell inter-departure distribution of a periodic traffic stream from a FIFO multiplexer. In [3], the periodic traffic assumption is relaxed and the analytical approach is extended to consider the case of a general renewal stream under the same FIFO service policy. Statistics for the inter-departure process of an initially periodic cell stream in a FIFO multiplexer are also computed in [4], where peak rate enforcement is discussed as a means of reducing jitter. However, the impact of such regulation on the delay and jitter of a cell stream has been studied mainly through simulations.

In this paper, a queueing model is developed which can be used to study the effects of regulating the cell stream of a tagged connection within a network switching node. This is accomplished by introducing a service discipline at the queue which occasionally departs from the FIFO policy by enforcing a peak output rate (minimum cell inter-departure time) for one of the traffic classes (Class-1). Note that the formulated queueing system can be applied to model devices such as the Spacer-Controller considered in [5]. The proposed service policy, which is presented in section 2, maintains a high degree of simplicity while at the same time providing rate-based flow control for the tagged connection at the output of the queue. In practice, such a scheme would enable the traffic class of interest to conform to some agreed-upon peak rate at any point within the network. Consequently, the proposed Peak Output-Rate-Enforcement (PORE) policy provides improved jitter performance for Class-1 traffic compared to that of the simple FIFO discipline.

The queueing analysis presented in section 3 provides for the computation of probability distributions for cell delay and jitter of the tagged traffic, which is modeled here as a periodic stream. It will be shown that when the peak-rate enforcer is disabled (by properly setting a key parameter), the PORE policy becomes the standard FIFO policy considered in [2] and [3]. Thus, this paper also provides a numerical approach for obtaining exact results for cell delay and jitter distributions in the M+D/D/1 queueing system considered in [4], which may be more flexible than the analytical approach followed in [2] and [3]. Finally, numerical results are presented and discussed in section 4.

*Research supported in part by the Advanced Research Project Agency under Grant F49620-93-1-0564.

†email : rlandry@emba.uvm.edu

‡email : stavraka@emba.uvm.edu

2 The PORE Service Policy and Formulated Queueing System

The proposed Peak Output Rate Enforcement (PORE) policy is a service discipline associated with a specific class of interest (Class-1). The rest of the traffic present in the system is treated in an indistinguishable manner under the PORE policy for Class-1 and, thus, is considered to form a single traffic class (Class-2 or background traffic). Note that in this section, no assumptions are made regarding the nature of the Class-1 or Class-2 arrival processes. However, for the queueing analysis in section 3, Class-1 cell arrivals are assumed to be periodic. The presentation of the PORE discipline and the induced delay jitter have been generalized in this section to emphasize the independence between the proposed service policy and the adopted arrival processes.

2.1 The PORE Service Policy

Consider two classes of traffic (Class-1 and Class-2) served by a single server, as shown in Fig. 1. Time is assumed to be slotted. The duration of a slot is equal to the transmission time of a fixed-size cell, and events of interest (cell arrivals and service completions) are assumed to occur at slot boundaries; the buffer capacity is assumed to be infinite.

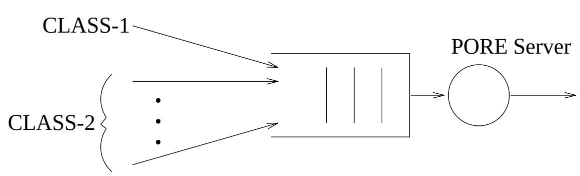

Figure 1: The discrete-time queueing system

Let the Class-1 cell arrival stream be characterized by its peak-rate, defined here as the inverse of the minimum cell inter-arrival time A_{min}, and its mean rate λ. Such a cell arrival stream is said to conform to $(\frac{1}{A_{min}}, \lambda)$. As will become 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}, and therefore, the Class-1 cell inter-departure process from the PORE server conforms to $(\frac{1}{X_{min}}, \lambda)$. The server operates according to the FIFO discipline except for those times when the FIFO 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 FIFO 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. $X_k = t_k - t_{k-1}$ will denote the inter-departure time of Class-1 cells k and $k-1$. Let $\tilde{t}_k$ 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 FIFO discipline, $\tilde{t}_k = t_k$, $\forall k \geq 2$.

The PORE service discipline can be described in terms of the following rules:

(a) 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.

(b) During this holding time of length h_k, the server will provide FIFO service to Class-2 cells that may be present in the buffer. This implies that during the holding interval, up to h_k Class-2 cells can be served before the k^{th} Class-1 cell which arrived at an earlier time.

(c) When the holding time expires, the k^{th} Class-1 cell is served; thus, $t_k = \tilde{t}_k + h_k$, thereby ensuring that $X_k = X_{min}$.

Note that according to rule (b), the server abandons the FIFO discipline whenever a Class-1 cell is held ($h_k \geq 1$), and thus, it provides improved service to Class-2 compared to that under the FIFO policy. The PORE service policy reduces to the FIFO policy when $X_{min} = 1$, in which case $h_k = 0$, $\forall k \geq 2$.

Let the Class-1 cell inter-arrival process be denoted by $\{A_k\}_{k>1}$, where A_k represents the inter-arrival time (in slots) between Class-1 cells k and $k-1$. Note 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 $\{A_k\}_{k>1}$ through the queueing process.

2.2 Discussion of the PORE Policy and the Induced Jitter

In packet-switched networks supporting real-time applications, an important performance measure is *delay jitter*, which describes the degree to which a given traffic stream is distorted in the network. When a call experiences a large amount of jitter, or distortion to its traffic profile, extensive buffering may be required at the receiver in order for the original data stream to be recreated before being played back. Furthermore, an initially well-behaved cell stream can induce queueing problems and QOS degradation for existing connections as it becomes distorted and potentially more bursty within the network. Therefore, controlling jitter achieves both improved QOS and a more uniform distribution of the buffers throughout the network.

Although a number of definitions have been used to quantify delay jitter in the past, the following definition will be adopted here, as in [4], 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 jitter of the k^{th} cell is defined as*

$$J_k = d_k - d_{k-1} \quad \text{for } k > 1. \quad \square \tag{1}$$

In terms of the quantities defined in the previous subsection, the jitter can also be expressed as

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

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* jitter applies to those instances when Class-1 cells k and $k-1$ are transmitted at a spacing of less than A_k, while *positive* jitter applies to those times when cells k and $k-1$ are transmitted at a spacing of greater than A_k. Note that when the process $\{A_k\}_{k>1}$ 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 jitter by guaranteeing that $J_k \geq X_{min} - A_{max}$. When peak output-rate enforcement is deactivated ($X_{min} = 1$), $J_k \geq 1 - A_{max}$, which implies that Class-1 cells can be placed in consecutive slots on the output link (i.e. peak rate equals 1). Note that for periodic Class-1 traffic, which will be assumed for the queueing analysis, $A_{min} = A_{max} = T$, where T denotes the cell inter-arrival period.

Since providing a minimum spacing of X_{min} eliminates some of the negative jitter, it would be reasonable to investigate whether the PORE policy induces probabilistically larger positive jitter than the FIFO policy. This is addressed in the following theorem.

Theorem 1 *Consider any Class-1 arrival stream which conforms to $(\frac{1}{A_{min}}, \lambda)$. Let J_k^{FIFO} (J_k^{PORE}) denote the jitter of the k^{th} Class-1 cell under the FIFO (PORE) service policy. If $X_{min} < A_{min}$, then for all $k \geq 2$,*

$$Pr\left\{J_k^{PORE} \geq j\right\} \leq Pr\left\{J_k^{FIFO} \geq j\right\} \qquad \forall j \geq 0,$$

with equality when $X_{min} = 1$.

Proof: Recall that if the k^{th} Class-1 cell is held, it is guaranteed to be served at time t_k such that $X_k = X_{min}$. This implies that a held cell will never be placed on the output link at a distance less than X_{min} from the previous Class-1 cell. Since X_{min} is always less than A_{min}, then $X_k < A_k$. From (2), this implies that $J_k = X_k - A_k$ cannot be greater than or equal to zero, or positive jitter is never induced through this output-rate-enforcement action. In fact, the positive jitter is decreased, since while the k^{th} Class-1 cell is being held, the server is attending to Class-2, and consequently, the service time of the $(k+1)^{st}$ cell is moving closer to t_k. $\square$

Theorem 1 establishes the fact that regulating Class-1 by enforcing a peak output rate, does not increase the positive jitter of this traffic class. The PORE service policy actually reduces the two types of distortion most commonly induced by multiplexers, namely the *spreading* (positive jitter) and *clustering* (negative jitter) of cells. Note that a third type of distortion, cell loss, would be present in a finite-capacity queueing system.

One of the prices to be paid for the regulation scheme presented here is, of course, increased cell delays for Class-1 when $X_{min} > 1$. However, this increased delay would be induced under any scheme enforcing an agreed-upon peak rate of $1/X_{min}$ throughout the network. It should also be noted that when $X_{min} > 1$, the service policy is non-work-conserving since in some instances the server will be idle while the system is non-empty. This is the case when a Class-1 cell is in holding and no Class-2 traffic is present in the system. However, the service policy is work-conserving with respect to Class-2, since the server is never idle in the presence of Class-2 traffic.

3 Queueing Analysis

In this section, the PORE service policy is studied, in terms of performance measures associated with both traffic classes. The Class-1 inter-arrival process $\{A_k\}_{k>1}$ is assumed to be periodic with period T. Class-2 traffic is modeled in terms of i.i.d. batches of cells of arbitrary distribution; N^b denotes the maximum batch size for Class-2 arrivals.

Let $\{d_k, h_k\}_{k\geq 1}$ denote the 2-dimensional stochastic process embedded at the sequence of time instants $\{t_k\}_{k\geq 1}$, where 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. It is easy to establish that $\{d_k, h_k\}_{k\geq 1}$ is a Markov chain with state space $\{(i,j) : 0 \leq i < \infty, 0 \leq j \leq X_{min} - 1\}$.

Let $S^L = \{(i,j)\}$, where $L(T-1)+1 \leq i \leq (L+1)(T-1)$ and $0 \leq j \leq X_{min} - 1$, represent "level" L of the Markov chain $\{d_k, h_k\}_{k\geq 1}$. Conceptually, if the queueing system is at level L, then there are exactly L Class-1 cells in the queue, other than the Class-1 cell which is about to be served. Therefore, level '0' represents the only set of states from which the queue can potentially become empty at the beginning of the time slot over which the next Class-1 cell arrives.

Let the probability that the Markov chain $\{d_k, h_k\}_{k\geq 1}$ moves from state (i,j) at time t_k to state (i',j') at time t_{k+1} be denoted by $p(i,j,i',j')$. By 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:

$$\mathbf{P} = \begin{bmatrix} \mathbf{B_0} & \mathbf{B_1} & \mathbf{B_2} & \mathbf{B_3} & \cdots \\ \mathbf{A_0} & \mathbf{A_1} & \mathbf{A_2} & \mathbf{A_3} & \cdots \\ 0 & \mathbf{A_0} & \mathbf{A_1} & \mathbf{A_2} & \cdots \\ 0 & 0 & \mathbf{A_0} & \mathbf{A_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{3}$$

where the matrices $\mathbf{B_n}$ and $\mathbf{A_n}$, $n \geq 0$, have dimensions $(T-1)X_{min} \times (T-1)X_{min}$. 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 expressions for $p(i,j,i',j')$, where $(i,j) \in S^0 \bigcup S^1$, require lengthy derivations and are not included here.

The system in (3) can now be solved based on the matrix analytic techniques presented in [6]. 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 $1 \leq i < \infty$, $0 \leq j \leq X_{min}-1$.

Because the system state at time t_k is defined as (d_k, h_k), the PMF for cell delays of Class-1 is simply

$$Pr\{d=i\} = \sum_{j=0}^{X_{min}-1} \pi(i,j) \quad . \tag{4}$$

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

$$Pr\{J=k\} = \sum_{i=1}^{\infty} \sum_{j=0}^{X_{min}-1} \sum_{j'=0}^{X_{min}-1} p(i,j,i+k,j')\pi(i,j). \tag{5}$$

4 Numerical Results

In this section, numerical results are presented for the queueing system in Fig. 1 operating under the PORE service policy. Results are provided in terms of the performance measures derived in the previous section, and the relationship between cell spacing and jitter reduction is examined. For all results, Class-2 batch arrivals are assumed to be binomially distributed with rate λ^b and $N^b = 15$, which represents the superposition of 15 identical Bernoulli streams, each with intensity λ^b/N^b.

Fig. 2 depicts scenarios in which the Class-1 arrival process has a period $T=5$, and the Class-2 loads are 0.4 and 0.75, for total offered loads of 0.6 and 0.95, respectively. When the total load is only 0.6, approximately 90% of the Class-1 traffic retains its periodic structure ($J=0$), even under the FIFO service discipline ($X_{min}=1$). In fact, for this load there is virtually no difference between the FIFO policy and the PORE policy when $X_{min}=4$. This result supports the assertion that cell spacing in ATM networks is unnecessary when the network resources are significantly under-utilized. However, when the total load is 0.95, only a small fraction of the Class-1 traffic remains periodic, and a significant amount of both clustering (negative jitter) and spreading (positive jitter) are present in the cell stream. By enforcing a minimum inter-departure time of $X_{min}=4$, all negative jitter less than -1 is eliminated, and the Class-1 output process conforms to $(\frac{1}{X_{min}} = 0.25, \lambda = 0.2)$. Note that complete recovery of the periodicity ($X_{min}=T$) is not possible, since such an attempt would render the multiplexer unstable. For a load of 0.95 and $X_{min}=4$, over

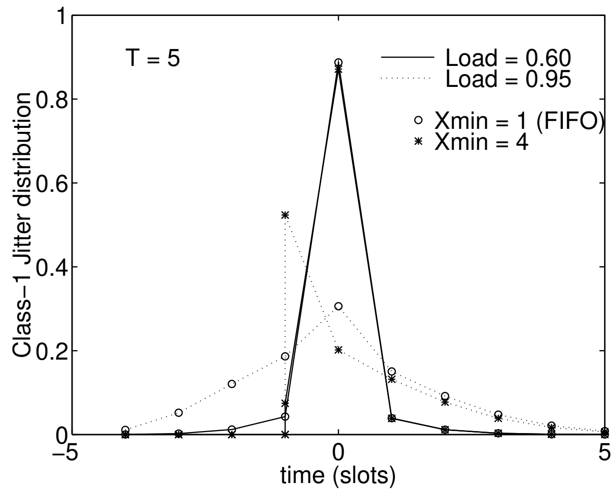

Figure 2: Jitter PMFs with and without PORE for moderate and high loads and $T=5$.

50% of the probability mass of J is located at $J=-1$. While much of this mass may be attributed to the elimination of negative jitter less than -1, a significant portion also comes from the reduction of $Pr\{J \geq 0\}$, in accordance with Theorem 1. Under the PORE policy for $X_{min}=4$, the cell inter-departure process has a much more narrow region of support than that from the FIFO multiplexer, and hence, may represent a more desirable traffic for the network.

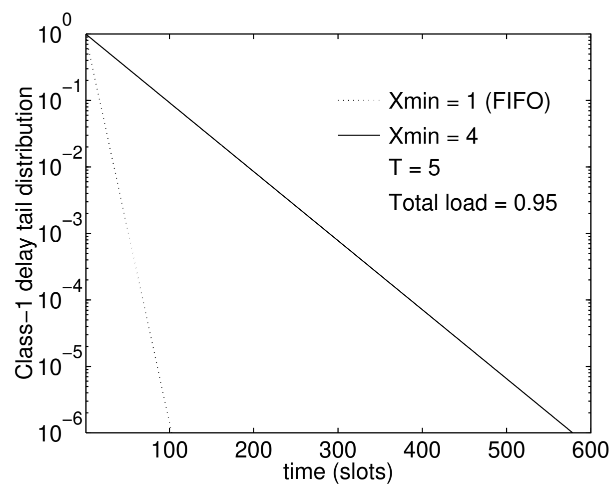

Figure 3: Delay tail distributions for Class-1 under heavy load and $T=5$.

In Fig. 3, the "cost" of improving the traffic profile for Class-1 over the FIFO policy is illustrated in terms of increased cell delays when the total load is 0.95. It is important to note that increased cell delays, quantified through the numerical results of Fig. 3 for the PORE policy, have been shown elsewhere ([5] for example) to be a necessary consequence of improving jitter performance for a given class of traffic.

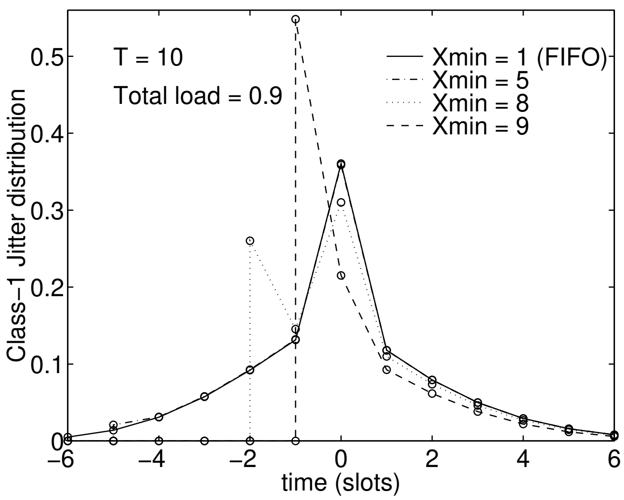

Figure 4: Jitter PMFs for various values of X_{min} and $T = 10$.

Fig. 4 illustrates the convergence of the PMF for jitter under the PORE policy to that under the FIFO policy as $X_{min} \to 1$. Class-1 traffic has $T = 10$ and the total load is 0.9 ($\lambda^b = 0.8$). Again, when X_{min} is set to its maximum value of $T - 1$ (in this case 9), most of the probability mass is located at $J = -1$, and the jitter process possesses a very narrow region of support. When $X_{min} = 8$, the negative jitter is distributed between two possible values (-2 and -1), with $Pr\{J = -2\}$ being the most dominant term. Notice that $Pr\{J = 0\}$ has increased over the case of $X_{min} = 9$ and, in fact, positive jitter is approaching that of the FIFO case. However, the cell inter-departure process created when $X_{min} = 8$ is still far superior to the FIFO inter-departure process in terms of its smoothness, since a significant amount of clustering (for $J < -2$) has been eliminated. It is interesting to note that the jitter PMFs for $X_{min} = 5$ and $X_{min} = 1$ (FIFO) are almost identical, except for values of jitter less than or equal to -5, which have very low probabilities associated with them.

Fig. 5 depicts the delay tail distributions for Class-1 cells when $X_{min} = 1, 5, 8$ and 9. The operating conditions are identical to those of figure 4. As expected, the tail distributions increase with X_{min}.

Note that directions for ongoing research include the generalization of the analytical technique to allow for the relaxation of the periodic traffic assumption for Class-1. The consideration of multiple queueing stages for an end-to-end jitter and delay analysis is also being pursued.

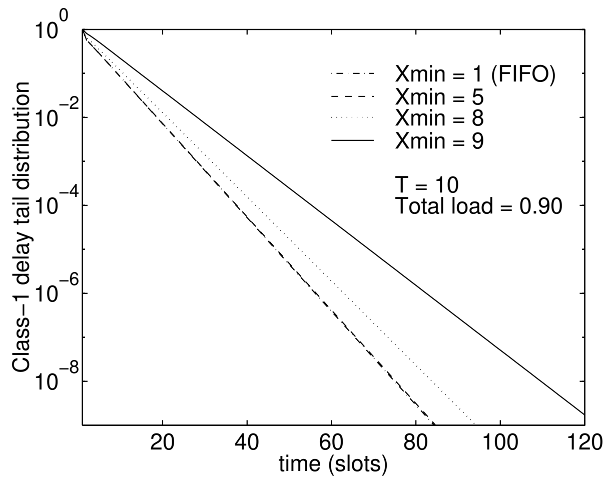

Figure 5: Delay tail distributions for Class-1 for various values of X_{min} and $T = 10$.

References

[1] J. Kurose, "Open Issues and Challenges in Providing Quality of Service Guarantees in High Speed Networks", *Computer Communications Review*, Vol. 23, No. 1, Jan. 1993.

[2] C. Bisdikian, W. Matragi, K. Sohraby, "A Study of the Jitter in ATM Multiplexers", Proceedings of the Fifth International Conference on Data Communication Systems and their Performance", Raleigh, NC, Oct. 25-28, 1993.

[3] W. Matragi, C. Bisdikian, K. Sohraby, "Jitter Calculus in ATM Networks: Part I, Single Node", IEEE Infocom'94, Toronto, Canada (to appear).

[4] J. Roberts, F. Guillemin, "Jitter in ATM Networks and its Impact on Peak Rate Enforcement", *Perf. Eval.*, vol. 16, no. 1-3, pp. 35-48, Nov. 1992.

[5] F. Guillemin, W. Monin, "Management of Cell Delay Variation in ATM Networks", IEEE Globecom'92, Orlando, FL.

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