Study of the Impact of Temporal / Spatial Correlations on a Tagged Traffic Stream¹

Marco Conti and Enrico Gregori

CNR - Istituto CNUCE

Via Santa Maria 36, 56126 Pisa, Italy

Email: {man@vm. and e.gregori@cnuce.cnr.it} , Fax: +39 50 904052

Ioannis Stavrakakis

Northeastern University

Electrical & Computer Engineering, Boston, MA 02115, USA.

Email: ioannis@cdsp.neu.edu , Fax: (617) 373-8970

Abstract

The problem of evaluating the end-to-end (multi-node) performance of a tagged traffic stream in an ATM environment is known to be difficult and largely open. To simplify the analysis, past studies have assumed that the co-existing (background) traffic is uncorrelated and is diverted after a single multiplexing stage; for this reason, temporal and spatial correlations were assumed to be insignificant and were not investigated. The objective in this work is to consider such correlations and evaluate their impact on the end-to-end performance of a tagged traffic stream. Such correlations can be significant due to temporal correlation in the background traffic or partial commonality in the routing path (background traffic is not necessarily diverted). A binary Queueing Activity Indicator (QAI) is proposed in this work to provide for a simple mechanism to capture these correlations. Results derived for the delay distribution of a tagged traffic stream traversing two consecutive nodes show the substantial impact of spatial/temporal correlations and indicate that these correlations are effectively captured by the proposed QAI. In addition, it is shown that increasing the temporal/spatial correlation results in increasing inaccuracy of the results obtained by ignoring them.

Keywords

Network traffic, jitter, correlations, end-to-end performance.

¹M. Conti's and E. Gregori's work was supported by CNR in the framework of a bilateral project. I. Stavrakakis' research was performed in part while visiting CNR Istituto CNUCE and has been supported in part by the Advanced Research Project Agency (ARPA) under Grant F49620-93-1-0564 monitored by the Air Force Office of Scientific Research (AFOSR); I. Stavrakakis is on leave from the University of Vermont.

1 INTRODUCTION

Asynchronous Transfer Mode (ATM) is considered to be the prevailing switching and multiplexing technique for the implementation of high-speed Broadband - Integrated Services Digital Networks (B-ISDN). Increased network utilization in the presence of diversified applications is expected to be achieved through the adopted statistical multiplexing and the on-demand bandwidth allocation (Onvural, 1993).

Although the maximum network utilization can be achieved by allowing for uncontrolled access to the network of all users, it is clear that in this case congestion will occur and the Quality of Service (QoS) delivered to the applications will suffer significantly. As a result, an optimal compromise between network utilization and delivered QoS is desirable. The Call Admission Control (CAC) and Traffic Regulation (TR) functions are being developed in an effort to achieve this optimal balance (Onvural, 1993).

Unless a very conservative approach is followed – limiting severely the network utilization – statistical multiplexing will surely result in modification of the source traffic profile, which may be severe at periods of network overload. This modification to the source traffic profile (distortion) represents a measure of the reduction of the QoS that the application will experience. When more than one multiplexing stages are along the path from source to destination, the distortion can be significantly increased.

The queueing problems induced by the statistical multiplexing process are well known and studied. Most of the past studies have focused on the determination of the impact of the multiplexing process on a *random* cell (information unit). That is, the underlying assumption has been that all multiplexed applications have the same QoS requirements. A measure of the traffic distortion due to the multiplexing process – the correlation in the *random* cell departure process – has been considered in (Stavrakakis, 1991a), (Stavrakakis, 1991b), (Lau, 1993), where an approximate description of the resulting random cell profile is presented and used for an approximate end-to-end study.

In an ATM environment in which applications may have quite diverse QoS requirements, studies like the above can be of limited usefulness, (Kurose, 1993), (Nagarajan, 1992). It is important that the magnitude of the distortion to a specific traffic profile – in the presence of other multiplexed applications – be determined. Thus, an application needs to be tagged and observed at the output of the multiplexer to evaluate the distortion to its traffic profile.

A number of recent works have focused on the evaluation of the distortion of a tagged traffic profile due to the statistical multiplexing process (Guillemin, 1992), (Bisdikian 1993), (Matragi, 1994a), (Matragi, 1994b), (Roberts, 1992), (Landry, 1994). The delay jitter has been adopted as a measure of the distortion and has been determined by developing analytic (Bisdikian 1993), (Matragi, 1994a), (Matragi, 1994b) and numerical (Landry, 1994) approaches. These works evaluate the magnitude of the distortion to a tagged traffic profile and point to the fact that this distortion may be unacceptable for certain applications. In (Landry, 1994) it is attempted to restore in part the original traffic profile by proposing a modification to the FIFO (First-In First-Out) multiplexing discipline and evaluating the reduced distortion to the tagged traffic profile. A typical approach followed for the study of a tagged traffic stream considers all non-tagged traffic to form a single traffic stream called the background traffic.

The above mentioned past work is based on a number of assumptions. First, the back-

ground traffic at a single node is assumed to be a (time) uncorrelated process. The second major assumption is that of the almost complete nodal decomposition in the approximate end-to-end performance study. Although the tagged interarrival process to node $n + 1$ is described in terms of the tagged interdeparture process from node n , consecutive interdepartures are identically distributed and independent. In addition, the background traffic in node $n - 1$ is assumed to follow a different path from that of the tagged traffic and, thus, it is not forwarded to node $n + 1$. As a result, the background traffic in node n is fresh and independent from any other process. In this discussion, $n - 1$, n and $n + 1$ denote three consecutive nodes along the path from the source to the destination of the tagged traffic.

In this work, the major assumptions outlined above will be relaxed to a certain degree. At first, the background traffic will be assumed to be a correlated process, which may be better reflecting a realistic environment. One of the objectives will be to determine the impact of correlations in the background process on the distortion to the tagged traffic stream (temporal correlation). A second departure from the past work is regarding the background traffic of node $n - 1$. By definition, this is the non-tagged traffic which is present at node $n - 1$ and is forwarded to node n together with the tagged traffic. A portion of this traffic will be assumed to be forwarded to node $n + 1$ along with the tagged traffic. This carried-on traffic – which will be allowed to be correlated – will be added to the fresh background traffic at node n . The consideration of the carried-on traffic contributes to a coupling between the queueing processes in nodes $n - 1$ and n (spatial correlation). Again, the existence of carried-on traffic may be better capturing a realistic networking environment.

The objective in this paper is to assess the significance of the temporal/spatial correlation discussed above and evaluate the resulting accuracy in calculating measures of the distortion of the tagged traffic profile. For this reason, results for a system of two consecutive nodes will be derived in the present study. The end-to-end performance when more than two nodes are present can be derived by following the approach proposed in this paper. The present study will help identify the processes which impact significantly on the measures of interest of the distortion of the tagged traffic profile and provide guidance for the larger scale end-to-end study in the presence of arbitrarily many nodes.

In the next section the system to be studied is described; a model is developed and the quantities of interest are defined. An outline of the queueing analysis is presented in section 3, followed by numerical results and discussion in section 4.

2 SYSTEM DESCRIPTION AND MODELING

Consider two consecutive nodes (nodes $n - 1$ and n) along the path from the source to the destination of the tagged traffic. Suppose that: (a) the background traffic is a (time) uncorrelated process and (b) the background traffic at node $n - 1$ is diverted after node n , that is, it is not forwarded to node $n + 1$ along with the tagged traffic.

Let $\{G_k^n\}_k$ denote a process describing some traffic descriptor (such as delay or interdeparture) associated with tagged cell k and node n . Due to the memory present in the queueing process, $\{G_k^n\}_k$ will be a (time) correlated process. Since the queueing activity at node $n - 1$ modulates $\{G_k^{n-1}\}_k$ which shapes the tagged cell arrival process to node n , $\{G_k^{n-1}\}_k$ and $\{G_k^n\}_k$ will be (space) correlated processes. All the past work has ignored such temporal

Figure 1: Input - output quantities for a generic node.

and spatial correlations. Some results suggest that - at least under certain conditions - these correlations may not be significant, (Landry, 1994).

In this work, the above assumptions (a) and (b) are being relaxed. In addition to the memory in the queueing process, (temporal) correlation in the background traffic will be present, potentially increasing substantially the temporal correlation in the process $\{G_k^n\}_k$. For instance, if a 2-state background arrival process remains in one of two substantially different (in terms of cell arrivals) states, a bimodal queueing behavior may be induced. Ignoring such correlations may result in very inaccurate performance calculations. By relaxing (b), the spatial correlation will be increased through the contribution of the carried-on traffic which is shaped by the queueing activity in the previous node.

The objective in this work is to study the impact of spatial and temporal correlations (coupling) on the characteristics of a tagged traffic stream. The term *spatial correlation* is adopted in this work to refer to the dependence of the queueing activity at some node n from the queueing activity at node $n - 1$. The term *temporal correlation* refers to the (temporal) correlation in a traffic descriptor associated with two consecutive tagged cells. A simple, binary Queueing Activity Indicator (QAI) is adopted in this work, to provide for a simple mechanism to (approximately) capture these spatial and temporal correlations. The QAI will modulate the carried-on traffic as well as the tagged traffic stream, and provide for a limited coupling of queueing processes associated with consecutive cells and consecutive nodes.

The queueing study of a generic node n will be based on the consideration of the input triplet $\{QAI_k^{n-1}, X_{ok}^{n-1}, C_j^{n-1}\}$ - see Figure 1 - where the involved quantities denote the QAI associated with node $n - 1$ and tagged cell k , the tagged cell interdeparture distribution associated with node $n - 1$ and tagged cell k , and the background traffic at the output of node $n - 1$, respectively.

Process $\{QAI_k^{n-1}\}_k$ will be approximated by a 2-state, first-order Markov process with parameters matched to those of the exact process. The state of this process - which will be part of the state description of node n - will modulate X_{ok}^{n-1} as well as C_j^{n-1} ; that is, the input traffic to node n will be shaped by the queueing activity at node $n - 1$, through the consideration of QAI_k^{n-1} . Furthermore, measures of the queueing behavior at node n - such as the delay D - will be shaped by the QAI associated with node $n - 1$. Given the current state of QAI_k^{n-1} , the conditional probabilities X_{ok}^{n-1} given QAI_k^{n-1} , C_j^{n-1} given QAI_k^{n-1} and D_{ok}^n given QAI_k^{n-1} will be derived. Finally, the queueing analysis of node n will determine the output triplet $\{QAI_k^n, X_{ok}^n, C_j^n\}$ which will be considered in the study of node $n + 1$.

Figure 2: Arrival and departure time axes.

The following definitions will be used in the description and analysis of the system to be studied in this work. Although results for two nodes will be derived in the present work, the general notation n will be used to denote the node under study. A subscript j denotes a quantity associated with time slot j . A subscript ok denotes a quantity associated with the k th tagged cell arriving in the associated node; for simplicity, subscript k will be used instead of ok if the associated quantity is defined only with respect to a tagged cell arrival and not at time slot j . A superscript n denotes a quantity associated with node n . For simplicity, let $LTCA_j^n$ denote the Latest Tagged Cell Arrival to node n which occurred before or at time (slot) j . Similarly, let $FTCA_j^n$ ($STCA_j^n$) denote the First (Second) Tagged Cell Arrival following time (slot) j . Some of the quantities introduced below are depicted in Figure 2.

$A_j^n(\bar{A}_j^n)$: Interarrival time between $LTCA_j^n$ and $FTCA_j^n$ ($FTCA_j^n$ and $STCA_j^n$); $1 \leq A_j^n, \bar{A}_j^n \leq A_{max}^n$; let $f_a^n(k) = Pr\{A_j^n = k/A_j^n > k-1\}$, $1 \leq k \leq A_{max}^n$; let $\bar{f}_a^n(k) = Pr\{\bar{A}_j^n > k/\bar{A}_j^n > k-1\} = 1 - f_a^n(k)$; let $E\{A_j^n\} = \frac{1}{\lambda^n}$, where $E\{\cdot\}$ denotes the expectation operator.

$X_j^n(\bar{X}_j^n)$: Interdeparture time between $LTCA_j^n$ and $FTCA_j^n$ ($FTCA_j^n$ and $STCA_j^n$).

$D_j^n(\bar{D}_j^n, \tilde{D}_j^n)$: Delay of $LTCA_j^n$ ($FTCA_j^n, STCA_j^n$).

$I_j^n(\bar{I}_j^n, \tilde{I}_j^n)$: Indicator function assuming the value 1 if $LTCA_j^n$ ($FTCA_j^n, STCA_j^n$) finds the queue non-empty of tagged cells upon arrival to node n ; for instance, $\bar{I}_j^n = 1_{\{A_j^n \leq D_j^n\}}$. Process $\{I_j^n\}_j$ will be modeled in terms of a 2-state Markov process embedded at times of tagged cell arrivals. The parameters of this approximate process are derived by matching its transition probabilities to those of the true process; the latter are determined by considering the evolution of process $\{W_j^n\}_j$ (described in section 3). $I_j^n(\bar{I}_j^n, \tilde{I}_j^n)$ is the QAI associated with node n modulating the output process of node n and employed in the study of node $n+1$.

J_j^{n-1} : Indicator function assuming the value 1 if $LTCA_j^n$ finds the queue of node $n-1$ non-empty of tagged cells upon arrival to node $n-1$. Notice that J_j^{n-1} is not

Figure 3: Arrival and departure time axes.

identical to I_j^{n-1} , since the former refers to $LTCA_j^n$ and the latter to $LTCA_j^{n-1}$, which are different cells. J_j^{n-1} is the QAI associated with node $n-1$, modulating the output process of node $n-1$ and employed in the study of node n .

Q_j^n : Queue occupancy.

B_j^n : State of the Markov background traffic; let $S_b = \{0, 1\}$, $\pi_b^i(i)$ and $p_b^i(i, j)$, $i, j \in S_b^n$ denote the state space and the stationary and transition probabilities, respectively.

$\hat{B}_j^n(B_j^n)$: Number of background cells generated when the state of the background process is (B_j^n) ; let $f_b^n(k, l) = Pr\{\hat{B}_j^n(l) = k\}$, $0 \leq k \leq M_b$, $l = 0, 1$.

T_j^n : Time (in slots) that has elapsed since the arrival time of $LTCA_j^n$; $T_j^n = 1$ if a tagged cell arrives at slot j , as it is the case in Figure 3; $1 \leq T_j^n \leq A_{max}^n$.

$$T_{j+1}^n = \begin{cases} T_j^n + 1 & \text{if } \{A_j^n > T_j^n\} \\ 1, & \text{if } \{A_j^n = T_j^n\} \end{cases} \quad (1)$$

It is easy to establish that the conditional probability distribution for T_j^n – denoted by $p_t^n(l, k) = Pr\{T_j^n = k / T_{j-1}^n = l\}$ – is given in terms of that of A_j^n as shown below.

$$p_t^n(k-1, k) = Pr\{A_j^n > k-1 / A_j^n > k-2\} = \bar{f}_a^n(k-1), \quad 2 \leq k \leq A_{max}^n \quad (2)$$

$$p_t^n(k, 1) = Pr\{A_j^n = k / A_j^n > k-1\} = f_a^n(k), \quad 1 \leq k \leq A_{max}^n \quad (3)$$

$$p_t^n(k, l) = 0 \text{ elsewhere} \quad (4)$$

Determination of the evolution of this process requires knowledge of the conditional interarrival time A_j^n given that $A_j^n > T_j^n - 1$ or, equivalently, knowledge of X_j^{n-1} , which depends on J_j^{n-1} (the QAI). The conditional probability distribution of X_j^n given J_j^{n-1} is derived from the joint probability $\{X_j^n, J_j^{n-1}\}$ which is derived by considering the evolution of process $\{W_j^n\}_j$ (section 3).

$C_j^n(J_j^{n-1}, T_j^n)$: Indicator function assuming the value 1 if a background cell of node $n - 1$ arrives at node n at time j . Notice that

$$C_j^n(1, T_j^n) = 1_{\{T_j^n \neq 1\}} \quad (5)$$

$$C_j^n(0, 1) = 0 \quad (6)$$

$$C_j^n(0, T_j^n) = l, \quad l = 0, 1, \quad T_j^n \neq 1 \quad (\text{to be evaluated}) \quad (7)$$

Notice that $J_j^{n-1} = 1$ implies that the arrival of a tagged cell found the previous tagged cell in the queue and, thus, the interval between their transmission instants will be filled up with background traffic. The probabilistic description of $C_j^n(0, T_j^n)$ is derived assuming that J_j^{n-1} and T_j^n determine completely the (probabilistic) behavior of $C_j^n(J_j^{n-1}, T_j^n)$ and considering the evolution of the process $\{W_j^n\}_j$ (section 3).

F_j^n : State of the Markov splitting process associated with the carried-on traffic; let $S_j^n = \{0, 1\}$, $\pi_j^n(i)$ and $p_j^n(i, j)$, $i, j \in S_j^n$, denote its state space, and its stationary and transition probabilities, respectively. If a background cell from node $n - 1$ arrives at node n at slot j ($C_j^n(J_j^{n-1}, T_j^n) = 1$), then this cell is forwarded to node $n + 1$ (along the path of the tagged stream) if $F_j^n = 1$ – and this is called *carried-on* cell – and it is diverted otherwise. Notice that transitions of this Markov chain are assumed to occur at slots containing background traffic ($C_j^n(J_j^{n-1}, T_j^n) = 1$) from node $n - 1$.

It should be noted that when the QAI J_j^{n-1} (or, I_j^n , $\bar{I}_j^n$, $\tilde{I}_j^n$) is equal to zero, light to moderate queueing activity may be assumed; when it is equal to one, moderate to serious queueing activity may be assumed. It is expected that this QAI biases the delay and interdeparture distributions. For this reason, families of these distributions will be obtained by considering the different values of the QAI, as indicated earlier in the presentation of Figure 1. By considering the correlation in the QAI process and capturing the dependence from the QAI associated with node $n - 1$ of the input processes and the induced tagged cell delay at node n , an end-to-end performance based on limited nodal-coupling can be obtained. In addition to the previous, the particular selection of the QAI facilitates the description of the background traffic at the output of the corresponding node. When $I_j^n = 1$, every slot in the output link over the interdeparture interval X_j^n must contain a background cell from node n , as indicated in (5). When $I_j^n = 0$, the previous is not necessarily true, (7).

3 OUTLINE OF THE QUEUEING ANALYSIS

In this section, the queueing measures of interest are derived for node n and its output process is characterized. Details regarding the derivations are omitted due to space limitations.

Assuming that the departures occur before arrivals over the same slot, the evolution of the queue occupancy process is given by

$$Q_j^n = [Q_{j-1}^n - 1]^+ + \hat{B}_j^n(B_j^n) + 1_{\{C_j^n(J_j^{n-1}, T_j^n)=1, F_j^n=1\}} + 1_{\{T_j^n=1\}} \quad (8)$$

where the first term describes departures from the queue provided that it is non-empty and the second term describes fresh background arrivals. The third term describes background traffic from node $n - 1$ to node n which is forwarded to node $n + 1$ (carried-on traffic). The last term describes a tagged cell arrival. In view of equation (8) and the definitions of $\{B_j^n\}_j$ and $\{F_j^n\}_j$, the following multi-dimensional process:

$$\{W_j^n\}_j \equiv \{T_j^n, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\}_j, \quad (9)$$

becomes a Markov process under the approximations associated with processes $\{J_j^{n-1}\}_j$, $\{T_j^n\}_j$ and $\{C_j^n(J_j^{n-1}, T_j^n)\}_j$, as outlined above. Let

$$\Phi_j^n \equiv \{T_j^n, J_j^{n-1}, F_j^n, B_j^n\}. \quad (10)$$

Then $\{W_j^n\}_j$ can be written as

$$W_j^n \equiv \{\Phi_j^n, Q_j^n\} \quad (11)$$

The transition probability matrix of (11) has the M/G/1 structure with Φ_j^n and Q_j^n corresponding to the phase and level processes, respectively. The numerical complexity in deriving the stationary probabilities $\pi_w(\cdot, \cdot, \cdot, \cdot, \cdot)$ of (11) is determined by the dimensionality of the space of the phase process (Neuts, 1989) which is equal to $2^3 A_{max}^n \times 2^3 A_{max}^n$. Due to the particular evolution of process T_j^n (it increases by 1 or returns to 1), the number of possible states of the phase process is significantly smaller than $2^3 A_{max}^n \times 2^3 A_{max}^n$.

Let j be the time of arrival of the k th tagged cell to node n (Figure 3). Then

$$\{T_j^n, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\} \equiv \{1, J_{ok}^{n-1}, F_{ok}^n, B_{ok}^n, D_{ok}^n\} \quad (12)$$

Notice that the queue occupancy upon the k th tagged cell arrival is equal to the delay of this cell and that $T_j^n = 1$ when a tagged cell arrival occurs.

Let

$$\{\Psi_k^n\}_k \equiv \{J_{ok}^{n-1}, F_{ok}^n, B_{ok}^n, D_{ok}^n\}_k \quad (13)$$

with stationary probabilities given by $\pi_\psi(i_1, i_2, i_3, i_4) = \frac{1}{\lambda^n} \pi_w(1, i_1, i_2, i_3, i_4)$. Clearly, the joint probability of $\{J_{ok}^{n-1}, D_{ok}^n\}$ can be derived from the stationary probabilities of $\{\Psi_k^n\}_k$ and finally

$$P\{D_{ok}^n / J_{ok}^{n-1} = i\} = \frac{P\{J_{ok}^{n-1} = i, D_{ok}^n\}}{P\{J_{ok}^{n-1} = i\}}, \quad i = 0, 1. \quad (14)$$

4 NUMERICAL RESULTS AND DISCUSSION

In this work results have been derived for the system of two consecutive nodes depicted in Figure 4. The tagged source S generates a traffic stream which is multiplexed with the background traffic B_1 and B_2 – at nodes 1 and 2, respectively – before it reaches the destination D . The FIFO service discipline and infinite queue capacity are assumed at nodes

Figure 4: A system of two consecutive nodes.

1 and 2. A portion of the background traffic at node 1 is forwarded to node 2 while the remaining is diverted. The cumulative arrival process to node 2 contains tagged cells, fresh background cells from B_2 and carried-on traffic coming from B_1 .

By definition, carried-on traffic is the background traffic that is coming from the previous node and is forwarded to the next node, together with the tagged traffic. As a consequence, carried-on traffic is not present in the first node and the phase and level processes for $\{W_j^1\}_j$ will be described as follows.

$$\{\Phi_j^1\}_j \equiv \{T_j^1, B_j^1\}_j \quad (15)$$

$$Q_j^1 = [Q_{j-1}^1 - 1]^+ + \hat{B}_j^1(B_j^1) + 1_{\{T_j^1=1\}} \quad (16)$$

Under the assumption that $\{A_j^1\}_j$ is an independent process and $\{B_j^1\}_j$ is a Markov process, it is easy to establish that $\{W_j^1\}_j$ is a Markov process. The block matrices in the transition matrix of $\{W_j^1\}_j$ (M/G/1 structure) are easily determined.

The phase process associated with node 2 is given by (10) and the evolution of the level process by (8). $\{B_j^2\}_j$ and $\{F_j^2\}_j$ are assumed to be Markov processes with given parameters (section 2). $\{J_j^1\}_j$ is approximated by a 2-state first-order Markov chain by utilizing the solution of $\{W_j^1\}_j$. Process $\{X_j^1(J_j^1)\}_j$ as well as the background traffic at the output of node 1 are calculated as indicated section 2. The results presented below have been derived for the system of 2 nodes under the following traffic parameters.

The tagged cell interarrival time at node 1 is constant and equal to 8; that is, $A_j^1 = 8$. The background traffic at node 1 has parameters $p_b^1(0,0) = p_b^1(1,1) = .99$, (and thus, $\pi_b^1(0) = \pi_b^1(1) = .5$); $f_b^1(k,l) = b(k,6;p_l)$, where $b(k,6;p_l)$ is the binomial probability with parameters 6 and p_l , $l = 0, 1$; $\lambda_b^1 = (\pi_b^1(0)p_0 + \pi_b^1(1)p_1)6 = .7$; $\frac{p_1}{p_0} = 4$ (burstiness measure). Notice that the background traffic is very bursty. On the average, it stays in state 1 for about 100 slots delivering background cells at a rate greater than 1. The objective in selecting such burstiness of the background traffic at node 1 is to investigate how congestion in node 1 affects the performance figures at node 2, and determine the effectiveness of the QAI in characterizing this environment.

The traffic at node 2 consists of the tagged traffic ($\lambda^1 = \lambda^2 = .125$), the fresh background traffic and the carried-on traffic from node 1. The model considered for the background traffic at node 2 is identical to that at node 1 with the following differences in the parameters: $\frac{p_1}{p_0} = 1$ and $\lambda_b^2 = .1$. That is, this process is uncorrelated. The parameters of the splitting process

$\{F_j^2\}_j$ – determining the carried-on traffic – are: $\pi_j^2(0) = \pi_j^2(1) = .5$, $p_j^2(0,0) = p_j^2(1,1) = p_f$, $p_f \in \{.50, .90, .95, .99\}$. That is, half of the background traffic from node 1 becomes carried-on traffic and competes with the tagged traffic and the background traffic of node 2 for the same resources; p_f determines the burstiness of the carried-on traffic.

Figure 5: Tail of node 2 tagged cell delay distribution for $p_f = .50$.

Figure 6: Tail of node 2 tagged cell delay distribution for $p_f = .90$.

Figure 7: Tail of node 2 tagged cell delay distribution for $p_f = .95$.

Figure 8: Tail of node 2 tagged cell delay distribution for $p_f = .99$.

Figures 5-8 present the tail distribution of the tagged cell delay at node 2 for different values of the burstiness p_f . The results under “QAI approach” are derived by applying the

approach developed in this work. The results under “Decoupled” refer to simulation results of the system where the carried-on traffic has been replaced by a Bernoulli fresh background traffic at node 2 of the same rate (.35). The results under “Simulation” refer to the simulation of the real system with the carried-on traffic present. Notice that the simulative results have been obtained with a 90% confidence level; the width of the confidence interval is always lower than 2% and for this reason it never appears in the figures.

When $p_f = .50$, the carried-on traffic is generated through an uncorrelated splitting process of the background traffic coming from node 1. This traffic should be very similar to the additional fresh background traffic considered in the “Decoupled” case and, thus, the “Simulation” and “Decoupled” results should be very close. This can be observed in Figure 5 where $p_f = .50$. In addition, the “QAI approach” results are very close to the other ones, indicating that the approximations involved in this approach do not compromise its accuracy.

As expected, the queueing activity at node 2 increases as p_f increases. This is observed in Figures 6-8 for $p_f = .90, .95, .99$, respectively. Notice the increasing inaccuracy (as p_f increases) of the “Decoupled” results and the consistent accuracy of the “QAI approach” results. These results suggest that :

- (a) The modulation of the output processes (tagged and carried-on traffic) by the QAI results in an accurate evaluation of the queueing behavior in the next node, as determined by the tagged cell delay tail distribution.
- (b) Destination correlation – indicated here by a large value of p_f – can have a significant impact on the queueing behavior. Ignoring such correlation may result in very inaccurate performance evaluation.
- (c) The QAI approach presented here seems to be capable of capturing the correlation among the queueing processes associated with consecutive nodes (spatial correlation). When node 1 is temporarily overloaded by what will become highly correlated carried on traffic – when strong spatial correlation is present due to a high value of p_f –, the increased queueing activity at node 1 induces increased queueing activity at node 2, as the results in Figures 5-8 indicate.

Since a (time) correlated background traffic is expected to induce increased queueing activity at node 1 leading to a sustainable value of the QAI equal to 1, it is expected that the Markov approximation to the QAI process will also exhibit similar level of correlation. Indeed, it was found that $P\{QAI_k^1 = 1/QAI_{k-1}^1 = 1\} = .96$ and $P\{QAI_k^1 = 0/QAI_{k-1}^1 = 0\} = .94$ under correlated background traffic at node 1 and $p_f = .99$ (Figure 8). That is, a temporal correlation in the input process to node 1 seems to be well captured by the temporal correlation of the QAI process. Thus, the resulting increased temporal correlation in the queueing process due to the temporal correlation in the arrival process seems to be well captured by the temporal correlation in the QAI process. This may be important in accurately evaluating the temporal correlation in the end-to-end tagged cell delay process which may be useful in identifying potential starvation problems when a large number of consecutive cells are delayed excessively.

Figure 9: Tail of the end-to-end tagged cell delay distribution given the value of QAI in the first node.

Figure 10: Tail of the end-to-end tagged cell delay distribution under independent delays in the nodes (“convolution”) and considering the delay in node 2 as shaped by the QAI in node 1 (“QAI approach”).

Figure 9 shows the end-to-end tagged cell delay tail distribution given that QAI (associated with node 1) is equal to 0 or 1. It can be observed that the QAI affects significantly the performance measure. Given that process QAI remains unchanged for a long time (associated change probabilities are .06 or .04), it may be concluded that a large number of consecutive tagged cells may all experience either small or large delays (temporal correlation).

Finally, Fig. 10 presents the end-to-end tagged cell delay tail distribution by assuming independent delays at nodes 1 and 2 under "Convolution" and considering the delay at node 2 given the QAI at node 1 under the "QAI approach". The "Convolution" involves the distributions of $\{D_{ok}^1/QAI_k^1 = 0\}$ (derived from the Markov chain $\{W_j^1\}_j$) and $\{D_{ok}^2\}$ (derived from the Markov chain $\{W_j^2\}_j$). The "QAI approach" involves the conditional distributions associated with $\{D_{ok}^1/QAI_k^1 = 0\}$ and $\{D_{ok}^2/QAI_k^1 = 0\}$. Notice that the results under the "QAI approach" represent lighter queueing conditions at node 2, suggesting that the QAI process seems to effectively modulate the output processes from node 1 and induce lower queueing activity at node 2 when QAI=0 at node 1, as expected. Again, given that process QAI remains unchanged for a long time (associated change probabilities are .06 or .04), it may be concluded that a large number of consecutive tagged cells may all experience either small or large delays (temporal correlation).

Additional results to further substantiate the points made above are under derivation.

5 REFERENCES

- Onvural, R. (1993) *Asynchronous Transfer Mode: Performance Issues*. Artec House, Boston.
- Stavarakakis, I. (1991a) Queueing Behavior of Two Interconnected Buffers of a Packet Network with Application to the Evaluation of Packet Routeing Policies. *International Journal of Digital and Analog Communication Systems*, John Wiley and Sons, Vol. 4.
- Stavarakakis, I. (1991b) Efficient Modeling of Merging and Splitting Processes in Large Networking Structures. *IEEE Journal on Selected Areas in Communications*, Vol.9 No. 4.
- Lau, W.-C. and Li, S.-Q. (1993) Traffic Analysis in Large-Scale High Speed Integrated Networks: Validation of Nodal Decomposition Approach. in *IEEE Infocom'93*, San Francisco, CA.
- Kurose, J. (1993) Open Issues and Challenges in Providing Quality of Service Guarantees in High Speed Networks. *Computer Communications Review*, Vol. 23, No. 1.
- Nagarajan, R. and Kurose, J. (1992) On Defining, Computing and Guaranteeing Quality-of-Service in High-Speed Networks. in *IEEE Infocom'92*, Florence, Italy.
- Guillemin, F. and Monin, W. (1992) Management of Cell Delay Variation in ATM Networks. in *IEEE Globecom'92*, Orlando, FL.
- Bisdikian, C. and Matragi, W. and Sohraby, K. (1993) A Study of the Jitter in ATM Multiplexers. in *Fifth International Conference on Data Communication Systems and their Performance*, Raleigh, NC.
- Matragi, w. and Bisdikian, C. and Sohraby, K. (1994a) Jitter Calculus in ATM Networks: Single Node Case. ", in *IEEE Infocom'94*, Toronto, Canada.
- Matragi, W. and Bisdikian, C. and Sohraby, K. (1994b) Jitter Calculus in ATM Networks: Multiple Node Case. in *IEEE Infocom'94*, Toronto, Canada.

- Roberts, J. and Guillemin, F. (1992) Jitter in ATM Networks and its Impact on Peak Rate Enforcement. *Performance Evaluation*, Vol. 16, No. 1-3.
- Landry, R. and Stavrakakis, I. (1994) A Queueing Study of Peak-Rate Enforcement for Jitter Reduction in ATM Networks", *IEEE GLOBECOM'94*, San Francisco.
- Neuts, M.f. (1989) *Structured Stochastic Matrices of the M/G/1 Type and their Applications*, Marcel Dekker, New York.

6 BIOGRAPHY

Marco Conti received the Laurea degree in Computer Science from the University of Pisa, Pisa, Italy, in 1987. In 1987 he joined the Networks and Distributed Systems department of CNUCE, an institute of CNR (the Italian National Research Council). He has worked on modeling and performance evaluation of Metropolitan Area Network MAC protocols. His current research interest include ATM, Wireless Networks, design, modeling and performance evaluation of computer communication systems.

Enrico Gregori graduated in Electronic Engineering from the University of Pisa in 1980. He joined CNUCE, an institute of the Italian National Research Council (CNR), in 1981. In 1986 he spent one year as a visiting at the IBM Research Laboratory in Zurich. He has worked in several projects on network architectures and protocols. His current research interests include high-speed network design and performance evaluation. He is author or co-author of numerous publications in this field. From 1989 to 1991, he gave a course on Metropolitan Area Networks and their performance evaluation in the Computer Science department of the University of Pisa. He is now lecturing on computer network architectures, protocols and their performance evaluation in the Faculty of Engineering of the University of Siena, Italy.

Ioannis Stavrakakis received the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece, 1983, and the Ph.D. degree in Electrical Engineering from the University of Virginia, 1988. In 1998, he joined the faculty of Computer Science and Electrical Engineering of the University of Vermont as an assistant and then associate professor. Since 1994, he has been an associate professor of Electrical and Computer Engineering of the Northeastern University, Boston. His research has been focused on the design and performance evaluation of communication networks. He is an active member of the IEEE Communications Society (Computer Communications) and has been involved in the technical program of conferences sponsored by the IEEE, ACM and IFIP societies.