



Large impact of temporal/spatial correlations on per-session performance measures: single and multiple node cases

Marco Conti^{a,1}, Enrico Gregori^{a,1}, Ioannis Stavrakakis^{b,*,2}

^a CNR-Istituto CNUCE, Via Santa Maria 36, 56126 Pisa, Italy

^b Department of Informatics, University of Athens, 15784 Athens, Greece

Abstract

The problem of evaluating the end-to-end performance in an ATM environment is known to be difficult and largely open. Since applications may have very different quality of service (QoS) requirements, a performance measure associated with a *random* cell would not necessarily reveal the impact of multiplexing on the QoS of a specific application. For this reason, it is important that the traffic stream of interest be tagged and its distortion due to multiplexing be evaluated; for analysis purposes the co-existing traffic is aggregated and forms the background traffic. In the past studies the background traffic has been assumed to be uncorrelated and be diverted after a single multiplexing stage.

The objective in this work is to evaluate the impact of temporal and spatial correlations 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 queuing activity indicator (QAI) is proposed in this work to provide for a simple mechanism to capture these correlations. Results derived for various delay metrics of a tagged traffic stream associated with single and multiple nodes show the substantial impact of the spatial/temporal correlations. Furthermore, they suggest that the proposed QAI indicator is capable of capturing such correlations. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Traffic correlation; End-to-end performance; Queuing networks; Per session performance evaluation

1. Introduction

The most appealing characteristic of the asynchronous transfer mode (ATM) is its ability to achieve high resource utilization through statistical multiplexing of diversified applications [1]. Although maximum network utilization would be achieved by allowing for uncontrolled access to the network of all users, the delivered quality of service (QoS) would suffer substantially due to congestion. For this reason, an optimal compromise between network utilization and delivered QoS is desirable. The call admission

* Corresponding author. Tel.: +30-1-727-5343; fax: +30-1-727-5601.

E-mail address: istavrak@di.uoa.gr (I. Stavrakakis)

¹ Work supported by CNR in the framework of a bilateral project.

² Research 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).

control (CAC) and traffic regulation (TR) functions are being developed in an effort to achieve this optimal balance. The CAC function will determine whether more applications may be allowed to access the networking resources; this decision will be reached by deriving an estimate of the deliverable QoS after admission, based on limited and simple description of the traffic expected to be generated by the new application. The TR function will attempt to filter out traffic realizations which will potentially impact negatively on the QoS delivered by the network, by inducing network congestion. This filtering should be carried out in an effective manner so that it does not compromise by itself the QoS of this application.

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 queuing 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 [2–4], 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 such as the above may be of limited usefulness [5,6]. 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 [7–12]. The delay jitter has been adopted as a measure of the distortion and has been evaluated by developing analytic [8–10], and numerical [12] 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 [12] 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. The 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 background 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 are relaxed to a certain degree. At first, the background traffic is assumed to be a correlated process, which may be better reflecting a realistic environment. Correlated background traffic has also been considered in the recent work in [4,13,14].

Under certain assumptions, it is concluded in [4] that the spatial correlation may be insignificant and a nodal decomposition approach may be considered for the end-to-end performance evaluation. The work in [13] considers a single node environment and investigates the delay jitter process of an MMPP stream at the output of single node modeled as an MMPP/M/1/K queuing system served according to the FIFO policy; the autocorrelation function of this process is involved in this study. Finally, an analytic approximation for the first-order statistics of the delay jitter process is developed in [14] and then utilized in the study of the delay jitter induced by a network node under diverse system and traffic conditions. One of the objectives of the work presented in this paper is 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. In this work, a portion of this traffic is assumed to be forwarded to node $n + 1$ along with the tagged traffic. This “carried-on” traffic — which is allowed to be correlated — is added to the “fresh” background traffic at node n . The consideration of the carried-on traffic contributes to a coupling between the queuing 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 impact of the temporal/spatial correlation described above on the traffic profile of a tagged stream. For this reason, a system of consecutive multiplexing nodes is considered and measures of the distortion of the tagged traffic stream are derived by taking into consideration the temporal/spatial correlations in the stochastic processes of interest. One of the contributions of this work is to show that ignoring temporal/spatial correlations can lead to substantially inaccurate results. This is established by comparing results obtained by simulating the real system to those obtained by simulating or analytically evaluating a simplified, almost decoupled system, instead. To reduce complexity, correlation reducing approximations are employed in the simplified system.

Another major contribution of this work can be attributed to the development of a model for the generic multiplexing node which takes into consideration temporal/spatial correlations between processes associated with consecutive nodes. In addition to the insight gained in the process of developing and presenting the model, the results obtained under this model outperform those under a simplified, almost decoupled system and are found to be in good agreement with the simulative results of the real system in most cases. The latter suggests that the developed model delivers the most accurate results which are in general in good agreement with the simulative ones of the real system.

In the next section the proposed model for the generic multiplexing node is described. In Section 3, key stochastic processes are introduced and the queuing analysis is presented. In Section 4, the measures of the distortion of the tagged traffic are introduced and are calculated by employing processes derived from the queuing analysis of the generic node. Finally, a number of numerical results are presented and discussed in the last section.

2. Description and modeling of the generic multiplexing node

In this section a discussion is presented regarding the temporal/spatial correlation (coupling) associated with processes shaping the traffic profile of a tagged traffic stream. These correlations are captured approximately by a binary process — the queuing activity indicator (QAI) — introduced here and assumed to modulate key processes impacting on the traffic profile. Based on the QAI, a tractable model for the generic multiplexing node is proposed.

Let $\{1, 2, \dots, n-1, n, n+1, \dots, N\}$ denote a sequence of N consecutive nodes defining the path from the tagged traffic source to its destination. Let the background traffic at node n denote the (locally) generated traffic which is forwarded to node $n+1$ along with 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 the delay or interdeparture time) associated with tagged cell k and node n . Due to the memory present in the queuing process, $\{G_k^n\}_k$ will be a (time) correlated process. Since the queuing 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. Most of the past work has largely ignored such temporal and spatial correlations. Some results suggest that — at least under certain traffic conditions — these correlations may not be significant, [12].

In this work, the above assumptions (a) and (b) are relaxed. In addition to the memory in the queuing process, (temporal) correlation in the background traffic is assumed to 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 queuing behavior may be induced. Experience in queuing studies suggests that ignoring such correlations may result in very inaccurate performance calculations. By relaxing (b), the spatial correlation is expected to be increased through the contribution of the non-diverted background traffic which is shaped by the queuing 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 queuing activity at some node n from the queuing 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 QAI is adopted in this work, to provide for a simple mechanism to (approximately) capture these spatial and temporal correlations. The QAI is assumed to modulate the non-diverted background traffic as well as the tagged traffic stream, and provides for a limited coupling of queuing processes associated with consecutive cells and consecutive nodes.

The following definitions provide for a precise description of the traffic components at the input and output of the generic multiplexing node as depicted in Fig. 1.

The *tagged* traffic stream is defined to be the one of interest which is present in all multiplexing nodes. The tagged traffic stream at the output of a node will be, in general, a distorted version of the tagged traffic stream at the input (component c-1, Fig. 1).

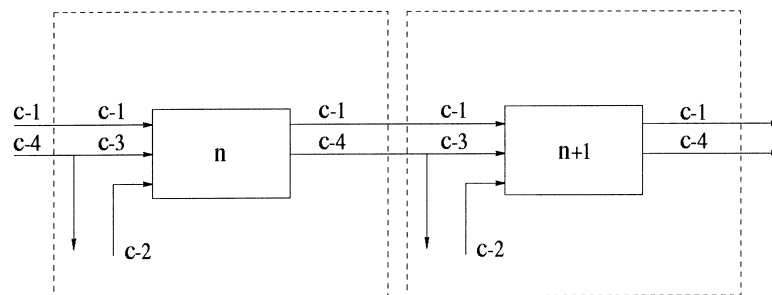


Fig. 1. Traffic components associated with the generic multiplexing node.

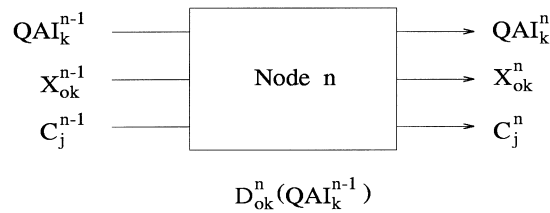


Fig. 2. Input–output quantities for a generic node.

The *background* traffic is defined to be the non-tagged traffic at the input of a multiplexing node which is *not* coming from the previous node of the tagged traffic path. This traffic is assumed to be independent from any other process in the system (component c-2, Fig. 1).

The *background traffic at the output* of a multiplexing node is defined to be the non-tagged traffic at the output of a node. Notice that this traffic component will contain the background traffic associated with this node as well as the other non-tagged traffic coming from previous nodes along the tagged traffic path, as explained next (component c-4, Fig. 1).

The *carried-on* traffic at a multiplexing node is defined to be the background traffic at the output of the previous node which is not diverted. That is, the carried-on traffic will be identical to the background traffic at the output of a previous node when the *splitting process* is off and it will be zero when the splitting process is on; the splitting process — which will be allowed to be correlated — is defined precisely later (component c-3, Fig. 1).

In view of the traffic components present at a generic multiplexing node, it is easy to identify the ones which contribute to the coupling of processes associated with consecutive multiplexing nodes, namely traffic components c-1 and c-3. For this reason, the input/output model for the generic node in isolation described below (Fig. 2) considers only these traffic components.

In this paper time is assumed to be slotted; the length of the slot (time unit) is equal to the cell transmission time. A subscript j denotes a quantity associated with time slot j . A subscript k denotes a quantity associated with the k th tagged cell arriving in the associated node. A superscript n denotes a quantity associated with node n . Consider the following quantities which are employed in the description of the input/output processes of the generic node (Fig. 2):

- QAI_k^{n-1} The QAI (defined precisely below) associated with node $n - 1$ and tagged cell k .
- X_k^{n-1} The tagged cell interdeparture time associated with node $n - 1$ and tagged cell k (component c-1, Fig. 1).
- C_j^{n-1} The background traffic (a 0–1 process) at the output of node $n - 1$ at time slot j (component c-4, Fig. 1).

The analysis of the generic node n will be based on the consideration of the input triplet $\{QAI_k^{n-1}, X_k^{n-1}, C_j^{n-1}\}$. The study will determine the impact of each of the multiplexing nodes on the tagged traffic as well as the output triplet which will feed the next node and provide for the coupling of the associated processes of interest.

Unlike previous studies where consecutive nodes were (weakly) coupled through a renewal tagged cell interdeparture process $\{X_k^{n-1}\}_k$ associated with the previous node, the present study allows for stronger coupling of the queuing processes associated with consecutive nodes and tagged cells. This is due to the consideration of: (a) the output process of the background traffic at node $n - 1$, $\{C_j^{n-1}\}_j$ – in addition to $\{X_k^{n-1}\}_k$; (b) the QAI associated with node $n - 1$ which is modeled as a correlated process. By allowing

for $\{X_k^{n-1}\}_k$ and $\{C_j^{n-1}\}_j$ to be modulated by the (correlated) process $\{QAI_k^{n-1}\}_j$, a potentially significant coupling in time and space of the queuing processes of interest may be achieved.

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_k^{n-1} as well as C_j^{n-1} ; i.e. the input traffic to node n will be shaped by the queuing activity at node $n - 1$, through the consideration of QAI_k^{n-1} . Given the current state of QAI_k^{n-1} , the conditional probabilities of X_k^{n-1} given QAI_k^{n-1} and C_j^{n-1} given QAI_k^{n-1} will be derived. Finally, the queuing analysis of node n will determine the output triplet $\{QAI_k^n, X_k^n, C_j^n\}$ which will be considered in the study of node $n + 1$.

3. Queuing analysis of the generic multiplexing node

The following definitions will be used in the precise description and analysis of the generic node. To facilitate the presentation, 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 Fig. 3:

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 $\bar{f}_a^n(k) = \Pr\{A_j^n > k / A_j^n > k - 1\} = 1 - f_a^n(k)$; let $E\{A_j^n\} = 1/\lambda^n$, where $E\{\cdot\}$ denotes the expectation operator.

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 as outlined in Appendix A, by utilizing the Markov chain $\{W_j\}_j$ defined below. I_j^n ($\bar{I}_j^n, \tilde{I}_j^n$) is the QAI associated with node n which modulates the output process of node n and is 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

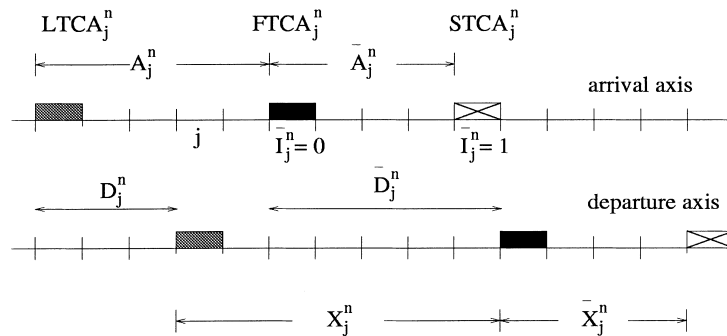


Fig. 3. Arrival and departure time axes.

identical to I_j^{n-1} , since the former is associated with $LTCA_j^n$ and the latter with $LTCA_j^{n-1}$, 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 $S_b = \{0, 1\}$, $\pi_b^n(i)$ and $p_b^n(i, j)$, $i, j \in S_b^n$ 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 $f_b^n(k, l) = \Pr\{\hat{B}_j^n(l) = k\}$, $0 \leq k \leq M_b, l = 0, 1$.

$X_j^n (\bar{X}_j^n)$

Interdeparture time between $LTCA_j^n$ and $FTCA_j^n$ ($FTCA_j^n$ and $STCA_j^n$).

The conditional probability mass function of X_j^n given J_j^{n-1} is derived as outlined in Appendix B.

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 , $1 \leq T_j^n \leq A_{\max}^n$.

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

‘It is easy to establish that the conditional probability mass function 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.

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

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

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

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).

$C_j^n (J_j^{n-1}, T_j^n)$

Indicator function assuming the value 1 if the background traffic at the output of node $n - 1$ at time j contains a cell. Notice that

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

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

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

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 as outlined in Appendix C; it is assumed that J_j^{n-1} and T_j^n determine completely the (probabilistic) behavior of $C_j^n(J_j^{n-1}, T_j^n)$.

F_j^n

State of the Markov splitting process associated with the background traffic at the output of node n ; let $S_f^n = \{0, 1\}$, $\pi_f^n(i)$ and $p_f^n(i, j)$, $i, j \in S_f^n$, 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 queuing activity would be expected; when it is equal to one, increased instantaneous rate of the tagged cell and/or background plus carried-on streams would be expected potentially leading to increased queuing intensity. 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 (Appendix B). By considering the correlation in the QAI process and capturing the dependence from the QAI of the input processes and induced tagged cell delay at node n , an approximate end-to-end performance based on limited nodal-coupling can be obtained. In addition, 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).

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 carried-on traffic from node $n - 1$ to node n which is forwarded to node $n + 1$. The last term describes a tagged cell arrival. In view of Eq. (8) and the definitions of $\{B_j^n\}_j$ and $\{F_j^n\}_j$, the following multi-dimensional process 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 introduced earlier:

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

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, and they are derived in Appendix D. The procedure leading to the derivation of the stationary probabilities of this Markov process, along with some complexity reducing techniques, may also be found in this appendix.

As it will be established in the next section, the measures of the distortion of the tagged traffic stream can be derived by employing the transition and stationary probabilities of $\{W_j\}_j$. To represent both the phase and the level of $\{W_j\}_j$, its stationary probabilities will be denoted by $\pi_w(\dots)$ in the rest of the paper.

4. Derivation of tagged traffic distortion measures

The impact of temporal/spatial correlations on the tagged traffic stream will be determined by considering various delay and delay jitter metrics. These metrics are derived in this section by employing the Markov chain $\{W_j\}_j$ which by construction takes into consideration such correlations. Numerical results will be presented in the next section.

As it will be seen below, the derivation of some performance measures requires convolutions as well as approximations for computational tractability. The definitions presented below will facilitate the description of the derivations that follow and the numerical results presented in the next section.

A pure convolution is defined to be an operation utilizing a convolution of the probability mass functions of the quantities of interest. It should be noted that the underlying assumption under a pure convolution is that the involved quantities are mutually independent, as well as independent from anything else in the system. Some of the temporal/spatial correlation will be captured under this operation since the employed probability mass functions are obtained from the system Markov chain $\{W_j\}_j$.

To evaluate the impact of the carried-on traffic on the performance metrics of interest, “no carried-on sim.” results will be presented as well. These will be simulative results obtained by replacing the carried-on traffic by a Bernoulli one of the same rate which is added to the background traffic of the next node. It should be noted that these results would be more accurate than the analytical results obtained through a simpler model based on nodal decomposition and assuming no carried-on traffic as well, since the latter would be derived under a renewal tagged cell interarrival process while the former would be derived under the true tagged cell interarrival process.

The QAI-based convolution is defined to be an operation that utilizes: (a) conditional probability mass functions of the quantities of interest given the associated QAI (s); (b) a Markov evolution of the QAI (s) (where relevant); (c) averaging over all possible (starting) states of the QAI (s).

Finally, the results derived under the “QAI approach” will be based on the proposed model and will utilize QAI-based convolutions if relevant. It should be noted that these results — although not uniformly accurate for all metrics considered — are expected to be the most accurate non-simulative results, since correlations are taken into consideration to greater extent than under models appeared in past studies or under other approximate calculations presented in this paper. Good accuracy of the results obtained under the “QAI approach” — as established by comparing them to simulative results of the real system (denoted by “real system”) — would suggest that the approximations involved in describing the generic node in terms of the Markov chain $\{W_j\}_j$, as well as in the Markovian evolution of the QAI and the conditional independence of the metric (s) from anything else in the system given the QAI, do not compromise the accuracy in the calculation of the specific metric (s). When performance measures are evaluated by incorporating the QAI (s) but assuming a specific state of the associated QAI (s), the results will be referred to as “conditional, QAI=(\dots)”.

4.1. Cell delay probability distribution

Let j be the time of arrival of tagged cell k to node n . Then

$$\{T_j^n, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\} \equiv \{1, J_k^{n-1}, F_k^n, B_k^n, D_k^n\}. \quad (12)$$

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

$$\{\Psi_k^n\}_k \equiv \{J_k^{n-1}, F_k^n, B_k^n, D_k^n\}_k \quad (13)$$

with stationary probabilities given by $\pi_\psi(i_1, i_2, i_3, i_4) = (1/\lambda^n)\pi_w(1, i_1, i_2, i_3, i_4)$. Clearly, the cell delay probability mass function, $P\{D_k^n\}$, can be derived from the stationary probabilities of $\{\Psi_k^n\}_k$.

4.2. Conditional cell delay probability distribution

As it should be clear from Section 2, the proposed model is constructed by introducing and utilizing the QAI as the process which will allow for the model to capture temporal/spatial correlations. By modeling the QAI process as a 2-state Markov process it is expected that the observed burstiness of this process — which is important in capturing correlations — will be preserved and represented in an analytically tractable way. The most meaningful way to assess the effectiveness of the proposed model — which is heavily dependent on the effectiveness of the QAI — would be by investigating, (a) the extent to which the QAI modulates key processes and (b) the accuracy of performance measures of interest derived by considering conditional independence of key quantities given the state of the QAI.

In view of the fact that the tagged cell delay is an important measure of the distortion to the tagged cell stream — which is also utilized in the derivation of other performance measures of interest, as discussed later — an outline of the derivation of the conditional delay of tagged cell k at the generic node n given the state of the QAI is presented.

The induced delay associated with tagged cell k and node n depends on: (a) the “recent past” queuing activity at node n which impacts on both tagged cell $k - 1$ and k ; (b) the “recent” cell arrival activity following the arrival of cell $k - 1$ and before the arrival of cell k . The former indicates that some dependence of D_k^n on QAI_k^n should be expected ($QAI_k^n \equiv I_j^n$ if cell k is the $LTCA_j^n$, Section 3). The latter indicates that the “instantaneous” rate of tagged cell arrivals (represented by X_k^{n-1}) and the “recent” carried-on traffic (represented by C_j^n over X_k^{n-1}) will impact on the induced value of D_k^n ; it should be noted that the background traffic of node n will impact as well, but this traffic is independent from anything else in the system. The argued dependence of D_k^n on X_k^{n-1} and C_j^n indicates the existence of a dependence of D_k^n on QAI_k^{n-1} ($\equiv J_k^{n-1}$, as defined in Section 3), since the latter will shape processes X_k^{n-1} and C_j^n .

The above discussion suggests that it is reasonable to consider that D_k^n depends on the QAI associated with cell k and nodes n and $n - 1$. In this paper it will be assumed that D_k^n is independent from anything else given (QAI_k^n, QAI_k^{n-1}) and will be calculated from

$$P\{D_k^n / (QAI_k^n, QAI_k^{n-1})\} = \frac{P\{D_k^n, QAI_k^n, QAI_k^{n-1}\}}{P\{QAI_k^n, QAI_k^{n-1}\}}, \quad (14)$$

where $P\{D_k^n, QAI_k^n, QAI_k^{n-1}\}$ and $P\{QAI_k^n, QAI_k^{n-1}\}$ are calculated as outlined in the Sections E.1 and E.2 of Appendix E, respectively. From the joint probability mass of the QAI associated with one cell and two consecutive nodes, $P\{QAI_k^n, QAI_k^{n-1}\}$, the conditional and joint probabilities shown below can be derived:

$$P\{QAI_k^n / QAI_k^{n-1}\} = \frac{P\{QAI_k^n, QAI_k^{n-1}\}}{P\{QAI_k^{n-1}\}}, \quad (15)$$

$$P\{QAI_k^1, \dots, QAI_k^n\} = P\{QAI_k^n / QAI_k^{n-1}\} \dots P\{QAI_k^2 / QAI_k^1\} P\{QAI_k^1\}, \quad (16)$$

where conditional independence of a QAI at node n from processes associated with nodes $\{n - 2, n - 3, \dots, 1\}$ is assumed given the QAI at node $n - 1$. This assumption is in line with the major assumptions intrinsic to the proposed model for the generic node.

4.3. End-to-end cell delay probability mass function

The probability mass function of the end-to-end delay of tagged cell k , $\hat{D}_k = \sum_{n=1}^N D_k^n$, is expected to be affected by the spatial correlations introduced primarily by the bursty carried-on traffic. Calculation of this performance metric, by taking into consideration the QAI (s) and evaluating its accuracy, would indicate the effectiveness of the developed model in capturing such correlations.

The end-to-end tagged cell delay is evaluated in terms of the conditional joint probability of $\{D_k^1, \dots, D_k^N\}$ given that the tagged cell is associated with a specific QAI pattern $\overline{QAI}_k = \{QAI_k^1, \dots, QAI_k^N\}$ as it traverses the N -node path. In view of the assumption of conditional independence of D_k^n given $\{QAI_k^n, QAI_k^{n-1}\}$, the above conditional probability can be calculated as follows:

$$\begin{aligned} & P\{(D_k^1, \dots, D_k^N)/\overline{QAI}_k\} \\ &= P\{(D_k^1, \dots, D_k^N)/(QAI_k^1 \dots QAI_k^N)\} \\ &= P\{D_k^N/(QAI_k^N, QAI_k^{N-1})\} P\{D_k^{N-1}/(QAI_k^{N-1}, QAI_k^{N-2})\} \dots P\{D_k^1/QAI_k^1\}, \end{aligned} \quad (17)$$

where the right-hand side terms are given by (14). The end-to-end delay can be derived as follows:

$$\begin{aligned} P\{\hat{D}_k = m\} &= \sum_{\overline{QAI}_k} P \left\{ \sum_{n=1}^N D_k^n = m / \overline{QAI}_k \right\} P\{\overline{QAI}_k\} \\ &= \sum_{\overline{QAI}_k} \sum_{m_1+m_2+\dots+m_N=m} P\{(D_k^1 = m_1, \dots, D_k^N = m_N)/\overline{QAI}_k\} P\{\overline{QAI}_k\}, \end{aligned} \quad (18)$$

where the right-hand side terms are given by (16) and (17).

4.4. Probability of M consecutive end-to-end deadline violations

An important performance metric associated with a time-constrained application is that of the probability that tagged cells experience an end-to-end delay exceeding a threshold. The “QAI approach” is applied for the evaluation of the probability that M consecutive tagged cells violate an end-to-end delay deadline. Under the “QAI approach” this probability is evaluated as described below, by adopting the previous assumptions of conditional independence.

$$\begin{aligned} & P\{\hat{D}_1 > T, \dots, \hat{D}_M > T\} \\ &= \sum_{\overline{QAI}_M} \dots \sum_{\overline{QAI}_1} P\{(\hat{D}_1 > T, \dots, \hat{D}_M > T)/(\overline{QAI}_1, \dots, \overline{QAI}_M)\} P\{\overline{QAI}_1, \dots, \overline{QAI}_M\} \\ &= \sum_{\overline{QAI}_M} \dots \sum_{\overline{QAI}_1} P\{\hat{D}_1 > T/\overline{QAI}_1\} \dots P\{\hat{D}_M > T/\overline{QAI}_M\} \\ & \quad \times P\{\overline{QAI}_M/\overline{QAI}_{M-1}\} \dots P\{\overline{QAI}_2/\overline{QAI}_1\} P\{\overline{QAI}_1\}. \end{aligned} \quad (19)$$

An alternative derivation based on complete independence of $\{\hat{D}_k\}_{k=1}^M$ is given by

$$P\{\hat{D}_1 > T, \dots, \hat{D}_M > T\} = P\{\hat{D}_1 > T\} \dots P\{\hat{D}_M > T\}, \quad (20)$$

where the end-to-end delays $\{\hat{D}_k\}_{k=1}^M$ are evaluated under the “QAI approach” (Section 4.1) and, thus, spatial correlation is not entirely ignored; this approach is referred to as “only spatial” approach, indicating that the coupling between consecutive cells is not taken into consideration, due to the assumed independence in (20).

4.5. Probability of M consecutive “overload” observations

Since the QAI can potentially capture an overload buffer state and it is not uncommon that such states tend to last, it seems reasonable to investigate the effectiveness of the “QAI approach” in evaluating the probability that M consecutive tagged cells find — upon arrival to a node — a buffer occupancy exceeding some value C . The results derived under the “QAI approach” will be compared to those obtained by assuming independent delays for consecutive tagged cells (“decoupled” approach) as calculated from the Markov chain $\{W_j\}_j$. The following derivation will be utilized under the “QAI approach”.

$$\begin{aligned} & P\{D_1^n > C, \dots, D_M^n > C\} \\ &= \sum_{\overline{QAI}^n} \sum_{\overline{QAI}^{n-1}} P\{(D_1^n > C, \dots, D_M^n > C) / (\overline{QAI}^n, \overline{QAI}^{n-1})\} P\{(\overline{QAI}^n, \overline{QAI}^{n-1})\} \\ &= \sum_{\overline{QAI}^n} \sum_{\overline{QAI}^{n-1}} P\{D_1^n > C / (QAI_1^n, QAI_1^{n-1})\} \dots P\{D_M^n > C / (QAI_M^n, QAI_M^{n-1})\} \\ & \quad \times P\{(\overline{QAI}^n, \overline{QAI}^{n-1})\}, \end{aligned} \quad (21)$$

where $\overline{QAI}^n = \{QAI_1^n, \dots, QAI_M^n\}$ and conditional independence of D_k^n from anything else given QAI_k^n and QAI_k^{n-1} is assumed. $P\{D_k^n > C / (QAI_k^n, QAI_k^{n-1})\}$, $1 \leq k \leq M$, is obtained from the corresponding joint probability as described in Section E.1 of Appendix E. Finally,

$$\begin{aligned} P\{(\overline{QAI}^n, \overline{QAI}^{n-1})\} &= P\{(QAI_M^n, QAI_M^{n-1}) / (QAI_{M-1}^n, QAI_{M-1}^{n-1})\} \\ & \quad \dots P\{(QAI_2^n, QAI_2^{n-1}) / (QAI_1^n, QAI_1^{n-1})\} P\{(QAI_1^n, QAI_1^{n-1})\} \end{aligned} \quad (22)$$

and $P\{(QAI_k^n, QAI_k^{n-1}) / (QAI_{k-1}^n, QAI_{k-1}^{n-1})\}$, $2 \leq k \leq M$, is derived from the corresponding joint probability as described in Section E.2 of Appendix E.

4.6. End-to-end m -cell block delay jitter probability distribution

An important performance metric, which is expected to be affected by temporal correlations between the associated metrics of consecutive cells, is the delay jitter associated with a block of M consecutive cells.

Let R^n denote the delay jitter associated with node n and a block of M consecutive cells defined as follows:

$$R^n = D_{oM}^n - D_{o1}^n. \quad (23)$$

Let \hat{R} denote the end-to-end delay jitter associated with a block of M consecutive cells defined as follows:

$$\hat{R} = \hat{D}_M - \hat{D}_1 = \sum_{n=1}^N D_{oM}^n - \sum_{n=1}^N D_{o1}^n = \sum_{n=1}^N (D_{oM}^n - D_{o1}^n) = \sum_{n=1}^N R^n, \quad (24)$$

where \hat{D}_M and \hat{D}_1 are the end-to-end delays of the 1st and M th cell of the block, as introduced in Section 4.3.

From (24) it is concluded that (temporal) correlations between the end-to-end delay \hat{D}_M and \hat{D}_1 will affect the delay jitter metric \hat{R} and that the smaller the size of the block M , the stronger the impact on the delay jitter will be due to temporal correlations in the delay.

By ignoring the (temporal) correlation between \hat{D}_M and \hat{D}_1 , the probability mass function of \hat{R} can be calculated as a pure convolution of the probability mass functions of \hat{D}_M and \hat{D}_1 . The analysis approach which employs the pure convolution will be referred to as the “decoupled” or “only spatial” approach, depending on the utilized probability mass functions for \hat{D}_M and \hat{D}_1 . Under the former approach, the probability mass function of \hat{D}_M and \hat{D}_1 is obtained as the pure convolution of the induced delays at various nodes. Under the latter approach, the end-to-end delay is calculated under the “QAI approach”, as outlined in Section 4.3, taking into consideration spatial correlations.

By allowing \hat{D}_M and \hat{D}_1 to be dependent on the QAI (s) associated with the various nodes and cells M and 1, respectively, and considering the M -step Markovian evolution of the QAI (s), the temporal coupling between the end-to-end delays of tagged cells M and 1 will be approximately captured to some extent. By adopting similar assumptions and the “QAI approach” followed in Section 4.3, the end-to-end delay jitter for a block of M cells will be derived under the “QAI approach” as follows:

$$\begin{aligned} P\{\hat{R} = m\} &= P\{\hat{D}_M - \hat{D}_1 = m\} \\ &= \sum_{\overline{QAI}_M} \sum_{\overline{QAI}_1} P\{\hat{D}_M - \hat{D}_1 = m / (\overline{QAI}_M, \overline{QAI}_1)\} P\{\overline{QAI}_M, \overline{QAI}_1\}, \end{aligned} \quad (25)$$

where $\overline{QAI}_M = \{QAI_M^1 \cdots QAI_M^N\}$ and $\overline{QAI}_1 = \{QAI_1^1 \cdots QAI_1^N\}$ represent the states of the N -dimensional process $\{QAI_k\}_k$ — modeled as a Markov chain — associated with cells M and 1, respectively. The stationary probability of $\{QAI_k\}_k$ is calculated as in (16); the M -step transition probability $P\{\overline{QAI}_M / \overline{QAI}_1\}$ is then obtained by considering the M -step transition of the QAI (s) employing the derivations in Section E.2 of Appendix E. Eq. (25) can then be written as follows:

$$\begin{aligned} P\{\hat{R} = m\} &= \sum_{\overline{QAI}_1} \sum_{\overline{QAI}_M} \sum_{m_2 - m_1 = m} P\{(\hat{D}_M = m_2, \hat{D}_1 = m_1) / (\overline{QAI}_M, \overline{QAI}_1)\} P\{\overline{QAI}_M / \overline{QAI}_1\} P\{\overline{QAI}_1\} \\ &= \sum_{\overline{QAI}_1} \sum_{\overline{QAI}_M} \sum_{m_2 - m_1 = m} P\{\hat{D}_M = m_2 / \overline{QAI}_M\} P\{\hat{D}_1 = m_1 / \overline{QAI}_1\} P\{\overline{QAI}_M / \overline{QAI}_1\} P\{\overline{QAI}_1\}, \end{aligned} \quad (26)$$

where conditional independence of \hat{D}_M given \overline{QAI}_M (and \hat{D}_1 given \overline{QAI}_1) from anything else in the system is assumed. The calculation in (26) is an example of “QAI-based” convolution.

5. Numerical results and discussion

As indicated in the introduction, the main objectives in this work are to investigate the impact of temporal/spatial correlations on the induced distortion to a tagged traffic stream and develop an analytical model capable of capturing these correlations to greater extent than under simplified models based on almost complete nodal decomposition. The results presented below have been obtained for a system of two (Section 5.1) or three (Section 5.2) consecutive multiplexing nodes. The results under the former system will help evaluate the impact of temporal/spatial correlations as well as the accuracy of the developed model, before additional approximations associated with a larger system come into the picture and lead to potentially more noisy results. In addition, simulation results are presented for a system of two nodes when the background process is described in terms of a multi-state Markov process developed in [17] in order to model the traffic behavior of an MPEG video stream. These results help evaluate the effectiveness of the proposed approach in the presence of more complex background traffic.

5.1. Results for a system of two nodes

The system considered in this case is depicted in Fig. 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 1 and 2. A portion of the background traffic at the output of 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 $\{W_j^1\}_j$ are described as follows:

$$\{\Phi_j^1\}_j \equiv \{T_j^1, B_j^1\}_j, \quad (27)$$

$$Q_j^1 = [Q_{j-1}^1 - 1]^+ + \hat{B}_j^1(B_j^1) + 1_{\{T_j^1=1\}}. \quad (28)$$

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.

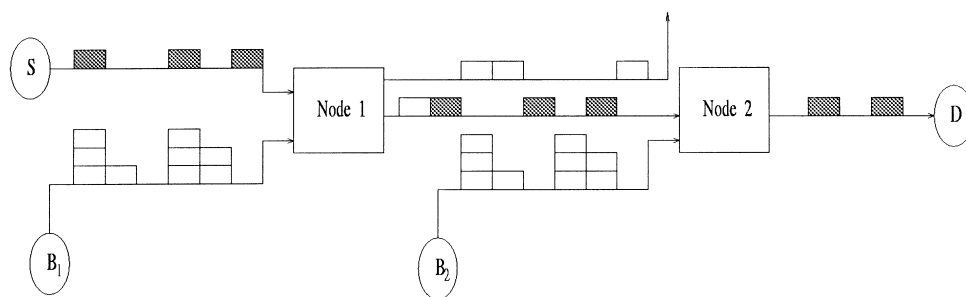


Fig. 4. A system of two consecutive nodes.

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 $\{W_j^1\}_j$, as described in Appendix A. $\{X_j^1(J_j^1)\}_j$ is derived from the joint probability $\{X_j^1, J_j^1\}_j$ as described in Appendix B. Finally, the background traffic at the output of node 1 is described as outlined in Appendix C. 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 $p_b^1(0, 0) = p_b^1(1, 1) = 0.99$ (and thus, $\pi_b^1(0) = \pi_b^1(1) = 0.5$); $f_b^1(k, l) = b(k, 6; p_l)$, where $b(k, 6; p_l)$ is the binomial probability with parameters 6 and $p_l, l = 0, 1$; $\lambda_b^1 = (\pi_b^1(0)p_0 + \pi_b^1(1)p_1)6 = 0.7$; $p_1/p_0 = 4$ (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, as well as determine the effectiveness of the QAI in characterizing this environment.

The traffic at node 2 consists of the tagged traffic ($\lambda^1 = \lambda^2 = 0.125$), the associated 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: $p_1/p_0 = 1$ and $\lambda_b^2 = 0.1$. That is, this process is uncorrelated. The parameters of the splitting process $\{F_j^1\}_j$ — determining the carried-on traffic — are: $\pi_f^1(0) = \pi_f^1(1) = 0.5, p_f^1(0, 0) = p_f^1(1, 1) = p_f, p_f \in \{0.50, 0.90, 0.99\}$. 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.

Figs. 5–7 present the tail of the probability mass function of the tagged cell delay at node 2 for different values of the burstiness p_f . The results under “QAI approach”, “no carried-on sim.” and “real system” are derived as described in Section 4. Notice that the simulative results have been obtained with a 90%

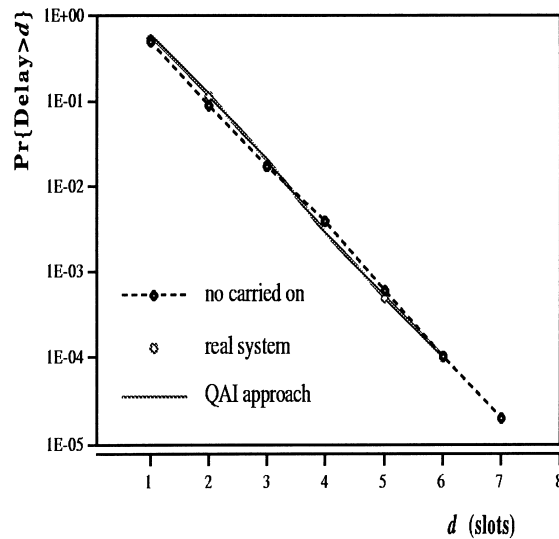


Fig. 5. Tail of node 2 tagged cell delay probability mass function for $p_f = 0.50$.

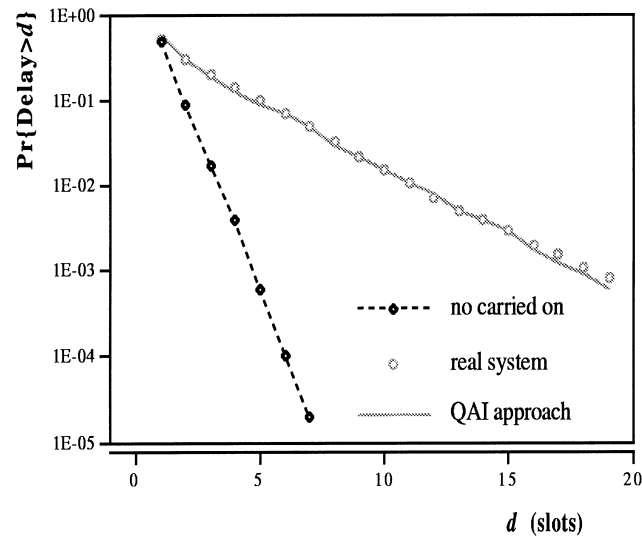


Fig. 6. Tail of node 2 tagged cell delay probability mass function for $p_f = 0.90$.

confidence level; the width of the confidence interval is always lower than 2% and for this reason it never appears in the figures.

When $p_f = 0.50$, 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 “no carried-on sim.” case and, thus, the “real system” and “no carried-on sim.” results should be very close. This can be observed in Fig. 5 where $p_f = 0.50$. 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.

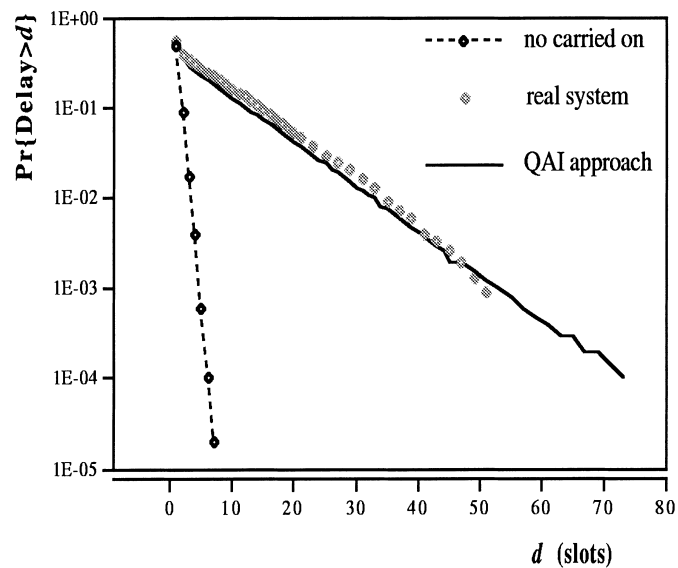


Fig. 7. Tail of node 2 tagged cell delay probability mass function for $p_f = 0.99$.

As expected, the queuing activity at node 2 increases as p_f increases. This is observed in Figs. 6 and 7 for $p_f = 0.90$ and 0.99 , respectively. Notice the increasing inaccuracy (as p_f increases) of the “no carried-on sim.” results and the consistent accuracy of the “QAI approach” results. These results suggest that

1. Destination correlation — indicated here by a large value of p_f — can have a significant impact on the queuing behavior. Ignoring such correlation may result in very inaccurate performance evaluation.
2. The modulation of the output processes (tagged and carried-on traffic) by employing the QAI results in an accurate evaluation of the queuing behavior in the next node, as determined by the tail of the probability mass function of the tagged cell delay. Thus, it may be argued that the QAI approach presented here seems to be capable of capturing the correlation among the queuing 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 queuing activity at node 1 induces increased queuing activity at node 2, as the results in Figs. 5–7 indicate.

Since a (time) correlated background traffic is expected to induce increased queuing 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 $P\{QAI_k^1 = 1/QAI_{k-1}^1 = 1\} = 0.96$ and $P\{QAI_k^1 = 0/QAI_{k-1}^1 = 0\} = 0.94$ under correlated background traffic at node 1 and $p_f = 0.99$ (Fig. 7). 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 queuing 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.

Fig. 8 shows the tail of the end-to-end tagged cell delay probability mass function, given that QAI (associated with node 1) is equal to 0 or 1. It can be observed that the observed state of the QAI affects significantly this performance measure. Given that process QAI remains unchanged for a long time (associated probabilities of change are 0.06 or 0.04), it may be concluded that a large number of consecutive tagged cells may all experience either small or large delays (temporal correlation). The remarkable agreement of the “real system” and the “conditional, $QAI=(\cdot)$ ” results can be clearly observed.

5.2. Results for a system of three nodes

A 3-node system is considered in this section to allow for the investigation of end-to-end results which may represent more accurately the interactions of processes in larger scale systems. The results are derived under the following system parameters.

The tagged cell interarrival time at node 1 is assumed to be constant and equal to 8; that is, $A_j^1 = 8$. The background traffic at all nodes is an uncorrelated process with rates equal to 0.7, 0.25 and 0.3 for nodes 1, 2 and 3, respectively; the number of background cells at each node follows the binomial distribution with maximum number equal to 6 and the proper success probability. The parameters of the correlated splitting processes at the output of nodes 1 and 2 are given by $\pi_f^1(1) = 0.5$, $p_f^1(1, 1) = 0.99$ and $\pi_f^2(1) = 0.5$, $p_f^2(1, 1) = 0.99$. The input traffic load to node 1 is equal to 0.825 (the tagged traffic rate equal to 0.125 and the background traffic rate equal to 0.7). The input traffic load to node 2 is equal to

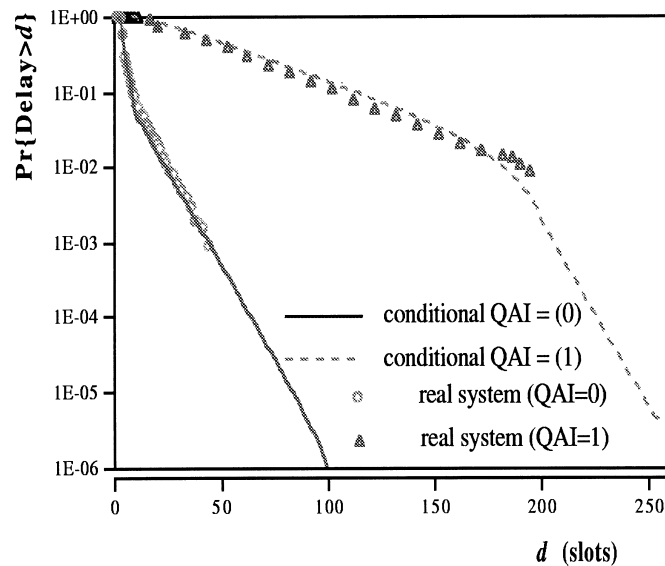


Fig. 8. Tail of the end-to-end (2-node) tagged cell delay probability mass function given the QAI state at the first node.

0.725 (the tagged traffic rate equal to 0.125, the background traffic rate equal to 0.25 and the carried-on traffic rate equal to 0.35). Notice that when the splitting process at the output of node 1 is off — and thus the entire output traffic of node 1 is forwarded to node 2 — the input traffic rate to node 2 is equal to 1.075 (the tagged traffic rate equal to 0.125, the background traffic rate equal to 0.25 and the carried-on traffic rate equal to 0.7). The input traffic load to node 3 is equal to 0.725 (the tagged traffic rate equal to 0.125, the background traffic rate equal to 0.3 and the carried-on traffic rate equal to 0.3). Notice that when the splitting process at the output of node 2 is off — and thus the entire output traffic of node 2 is forwarded to node 3 — the input traffic rate to node 3 is equal to 1.025 (the tagged traffic rate equal to 0.125, the background traffic rate equal to 0.3 and the carried-on traffic rate equal to 0.6). In view of the fact that the splitting process is very bursty ($p_f^1(1, 1) = 0.99$, $p_f^2(1, 1) = 0.99$) it is expected that the queue instability periods (when the traffic load exceeds 1) will be significant inducing severe queuing intensities which will not be observable under an evaluation assuming uncorrelated splitting. Finally, it should be noted that the correlation in the traffic that interferes with the tagged traffic stream is introduced primarily by the splitting process and it is not present in the background traffic.

The end-to-end delay probability mass function of a tagged cell is shown in Fig. 9. The result under the “QAI approach” is obtained as outlined in Section 4.3. The results under “conditional, QAI=(\dots)” are derived by assuming the indicated state of the QAI at the various nodes. Since larger delay is an indicator of higher queuing activity (congestion) and larger delay is observed when the QAI is set (Fig. 9), it may be concluded that the QAI is a reasonably good measure of congestion. Although there are realizations under which a node may not be congested although the previous tagged cell is found in the node upon arrival of the next one (QAI=1), the above results suggest that such realizations are not frequent and/or that the corresponding tagged cell arrival realization would represent one most likely shaped by a congested previous node.

Fig. 10 depicts the tail of the end-to-end tagged cell delay probability mass function. It should be noted again that the “no carried-on sim” results (Section 4) are expected to be more accurate than the

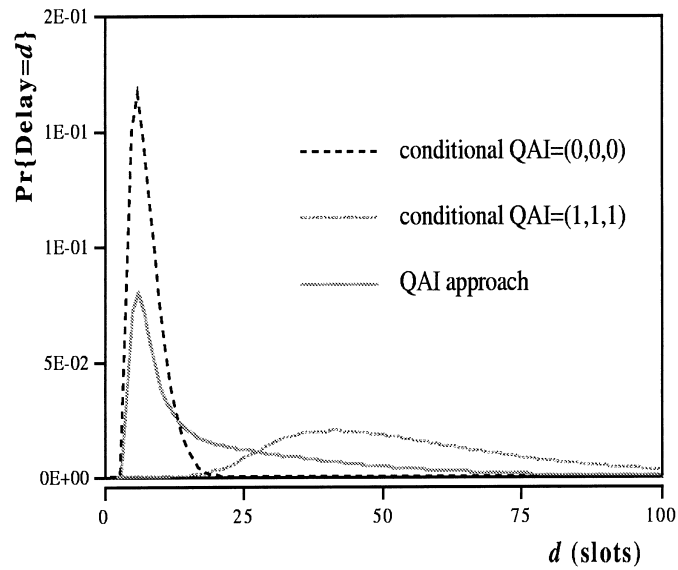


Fig. 9. End-to-end (3-node) conditional and unconditional tagged cell delay probability mass function.

analytical results derived by employing models proposed in the past which cannot accommodate the carried-on traffic or a non renewal tagged cell interarrival/interdeparture process. The results in Fig. 10 clearly indicate that ignoring the carried-on traffic can lead to substantial underestimation of the tail of the tagged cell delay probability mass function. Furthermore, the results obtained under the “QAI approach” (as outlined in Section 4.3) are seen to be reasonably close to the “real system” ones, suggesting again that the QAI is capable of capturing the spatial correlations associated with an end-to-end performance measure.

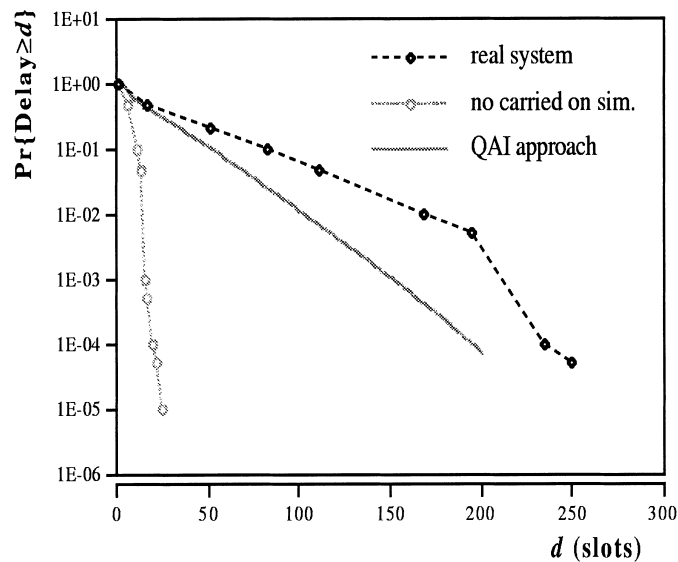


Fig. 10. End-to-end (3-node) tagged cell delay probability mass function.

Temporal correlations are expected to be better observed by considering a metric involving a sequence of consecutive cells, such as the delay jitter associated with an M -cell block. The probability mass function of the M -cell block end-to-end delay jitter is shown in Fig. 11. The impact of temporal and/or spatial correlations on the delay jitter results will be better understood by considering the fact that the delay jitter metric is equal to the difference of the end-to-end (spatial) delays of the 1st and M th (temporal) cells (Section 4.6). Results marked by “QAI, $M = x$ ” are obtained under the “QAI approach” (Section 4.6) for a sequence of x consecutive cells. The results under “only spatial, $M = x$ ” maintain the spatial correlation in the single cell end-to-end delay (as captured by the “QAI approach” associated with this computation) but assumes no temporal correlation between the 1st and M th cells which are involved in the computation of this metric (Section 4.3). The results under “decoupled, $M = x$ ” do not consider spatial and temporal correlation in the sense that the end-to-end cell delay is calculated as a pure convolution of the delays associated with each node and the end-to-end delays of the 1st and M th cells are assumed to be independent (Section 4.6).

The results shown in Fig. 11 suggest that the QAI is capable of capturing the temporal correlations associated with an M -consecutive cell metric. Specifically, as M increases — and the temporal correlation between the 1st and M th cells decreases — the support of the probability mass function increases as expected ($M = 5, 10, 20$). This indicates that results obtained under the QAI approach are shaped by (and, thus, capture) the temporal correlation. Results obtained under the QAI approach and $M > 20$ almost coincide with the results marked “QAI, $M = 20$ ” suggesting that the temporal correlations between the 1st and M th cells is insignificant for such valued of M . As a result, the “QAI, $M = 20$ ” result is expected to be very close to the one marked “only spatial, $M = 20$ ”, as it is the case in Fig. 11. Finally, by not considering spatial correlation as well — result marked “decoupled” — the delay jitter is expected to increase as it is observed in Fig. 11.

Results for the probability mass function of the end-to-end delay jitter for $M = 10$ consecutive cells are shown in Fig. 12 for the real system, under the “QAI approach” (as in Fig. 11) and for a system with no carried-on traffic (simulative). The “real system” and “no carried-on sim.” results are in accordance

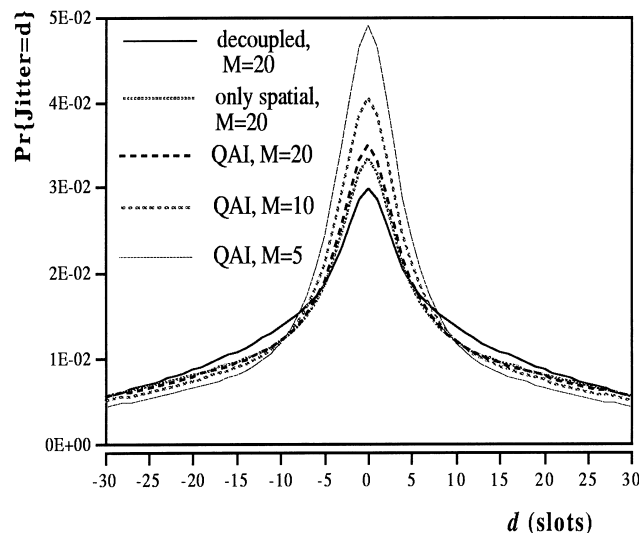


Fig. 11. Probability mass function for the M -cell block, end-to-end (3-node) delay jitter.

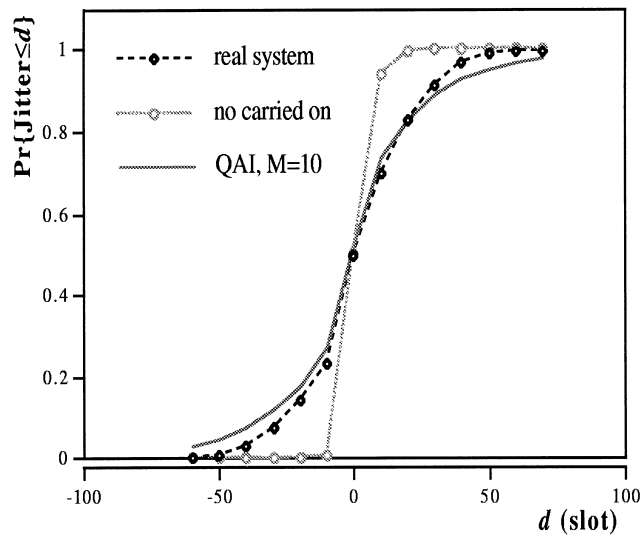


Fig. 12. Probability distribution function for the M -cell block, end-to-end (3-node) delay jitter.

with the expectation that the increased queuing activity due to the presence of the correlated carried-on traffic (Fig. 10) will induce higher delay jitter. The results under the QAI approach (QAI, $M = 10$) seem to be reasonably accurate.

Results for the probability that M consecutive cells miss a deadline $T = 100$ are shown in Fig. 13. The results are derived as described in Section 4.4. The “only spatial” results consider the end-to-end cell delay derived by applying the QAI approach but assume that these quantities are independent for consecutive cells. The difference between the “QAI approach” and the “only spatial” results is due to the

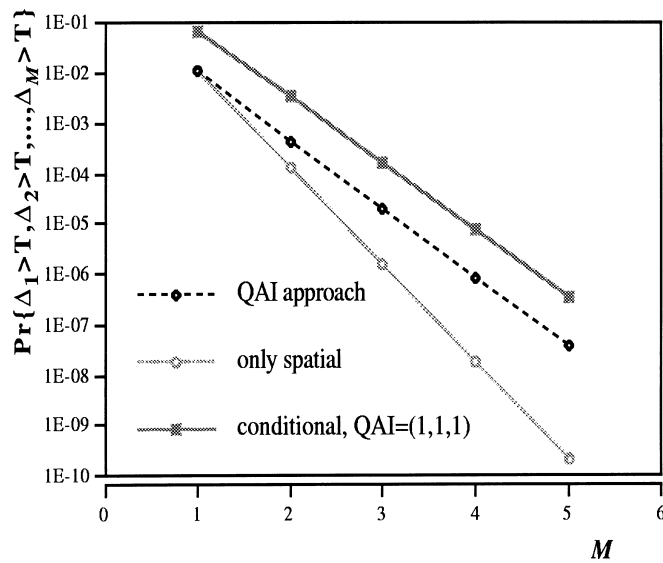


Fig. 13. Probability that M consecutive tagged cells miss a deadline of $T = 100$.

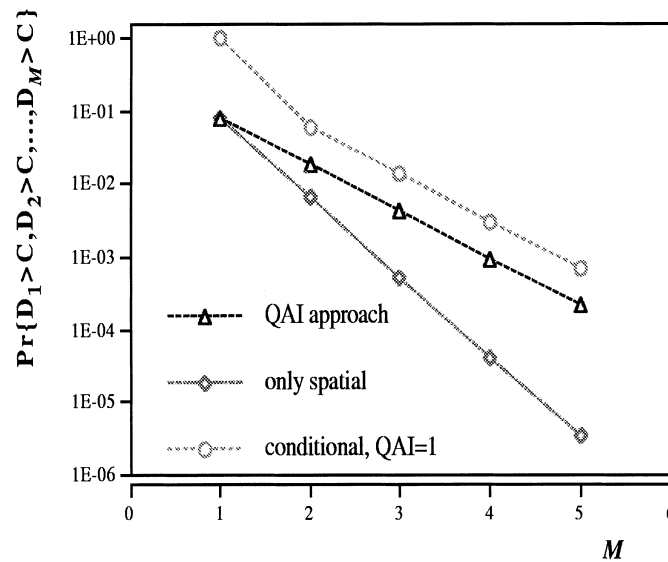


Fig. 14. Probability that M consecutive tagged cells observe a queue length exceeding $C = 35$ upon arrival to node 3.

positive temporal correlation between consecutive cells captured to greater extent by the QAI approach. Finally, the “conditional, $\text{QAI}=(1,1,1)$ ” result presents the probability that M consecutive cells miss their end-to-end deadline T given that the first cell observed a congested node (that is, QAI is equal to 1 in all three nodes).

Finally, the probability that M consecutive cells observe a queue length greater than $C = 35$ at node 3 is presented in Fig. 14. This probability may serve as a measure of the cell loss probability when the buffer capacity of node 3 is equal to C . The results are derived as described in Section 4.5. Similar results are shown and conclusions can be drawn as in Fig. 13.

5.3. Results for a system of two nodes under a complex background process

The results presented above suggest that the (simple) 2-state Markov QAI model can capture accurately the impact of temporal/spatial correlations present in consecutive nodes of tandem queuing systems. Since both the QAI model and the background traffic model have the same structure (2-state Markov), it has been questioned to what extent the effectiveness of the 2-state QAI model is due to the employment of a 2-state model for the background traffic. The results presented in this section suggest that the effectiveness of the 2-state Markov QAI approach is not limited to systems with 2-state Markov background traffic, suggesting that the developed QAI can capture the *intrinsic* characteristics of the temporal/spatial correlations processes with respect to their impact on the key metrics presented in Section 4.

In this section the system considered in Section 5.1 is modified in order to evaluate the effectiveness of the developed QAI approach in the presence of more complex background traffic. Specifically, the multi-state (two-dimensional) Markov model developed in [17] — in order to characterize an MPEG variable bit rate stream — is employed for the characterization of the background traffic. This type of traffic has a very complex behavior as it exhibits both a high degree of burstiness (for instance, the peak to average rate ratio is greater than 5) and strong temporal dependencies.

The video traffic model is described in terms of the two-dimensional process $\{L_k, H_k\}_{k \geq 0}$, where H_k describes the bit rate of the video source and L_k describes an underlying (low frequency) process which modulates the bit rate. To avoid unnecessary state space complexity, 8 states and 8 levels have been considered for the processes L_k and H_k , respectively; hence, the Markov chain has 64 states. The transition probabilities of this Markov chain may be found in [18] and have been obtained by applying the procedure developed in [17] and matching them to those associated with the MPEG-2 movie “The Sheltering Sky”. The resulting average quantized rate has been normalized in the simulations presented below, in order to obtain exactly the same average load in the nodes as in Section 5.1.

The accuracy of the QAI approach under the background traffic described above is evaluated through simulations, since the formulae derived in the previous sections would require significant modifications, which is beyond the scope of this paper. As in Section 5.1, the effectiveness of the QAI approach is determined by comparing the delay distribution obtained: (a) with the QAI approach, (b) for the real system, and (c) under “no carried-on” traffic; in case (c), the background traffic in node 2 has an average rate equal to the sum of the average rate of the carried-on traffic (coming from node 1) and the average rate of the local background traffic.

As in Section 5.1, the traffic at node 2 consists of the tagged traffic (with rate 0.125 cells per time unit), the carried-on traffic from node 1 and the node 2 local background traffic. In the real system (and also in the QAI approach simulation), 50% of the carried-on traffic (i.e. the node 1 background traffic) is diverted, as specified by the 2-state splitting process with parameter p_f . In the no carried-on simulations all the carried-on is diverted. The model considered for the background traffic for node 2 is identical to that used for node 1 (i.e. $\{L_k, H_k\}_{k \geq 0}$) but with a different arrival rate. The rate of the background traffic at node 1 is equal to 0.7 cells per time unit. In node 2 the background traffic has an arrival rate of 0.25 cells per time unit in the real system (and also in the QAI approach simulation) and 0.60 cells per time unit in the no carried-on case.

Figs. 15–17 present the tail of the probability mass function of the tagged cell delay for different values of the splitting process probability. Due to the computational cost required to accurately estimate

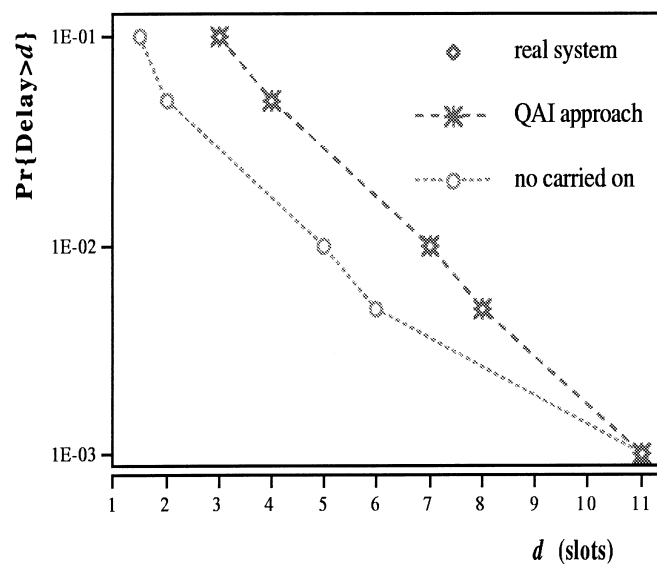


Fig. 15. Complex background traffic case: tail of node 2 tagged cell delay probability mass function for $p_f = 0.50$.

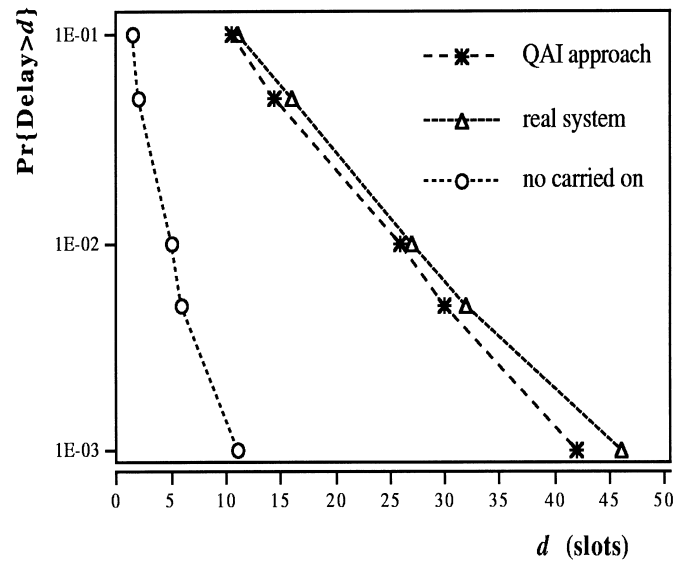


Fig. 16. Complex background traffic case: tail of node 2 tagged cell delay probability mass function for $p_f = 0.90$.

rare events with simulative techniques [19], reliable estimates up to the 99.5 percentile of the tail of the delay distribution have been obtained. The estimates related to the 99.9 percentile have large confidence intervals (for instance, the confidence interval is about 25 slots for $p_f = 0.99$).

The above results suggest that the QAI approach captures effectively spatial/temporal correlations when the background traffic is described not only by a 2-state Markov model but more complex ones. As the spatial/temporal correlation increases the no carried-on approach provides QoS estimates which become

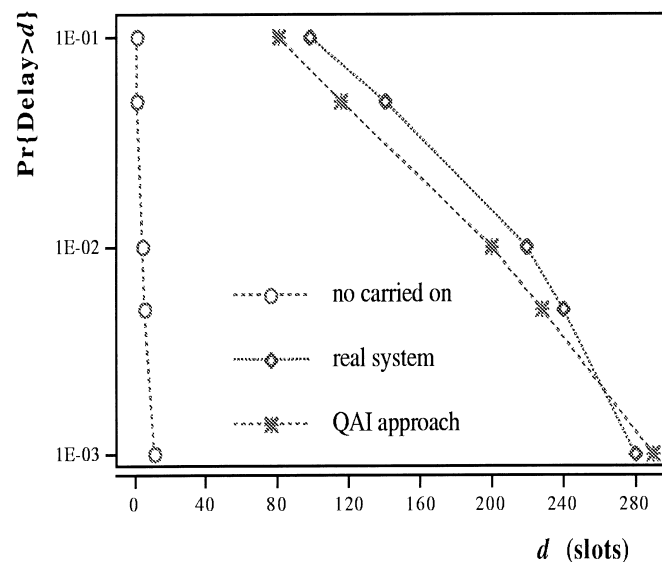


Fig. 17. Complex background traffic case: tail of node 2 tagged cell delay probability mass function for $p_f = 0.99$.

more and more optimistic. On the other hand, the estimates under the QAI approach seem to be in good agreement with the values obtained for the real system.

Appendix A

The 2-state, first-order Markov approximation to the binary process $\{I_j^n\}_j$ (or $\{J_j^{n-1}\}_j$) is developed by matching the transition probabilities of the approximating process to those of the true process.

Let $p_i^n(k, l)$ and $\pi_i^n(k)$, $k, l \in \{0, 1\}$ denote the transition and steady state probabilities of the 2-state Markov approximation to $\{I_j^n\}_j$. These probabilities are given by

$$p_i^n(k, l) = \frac{\Pr\{\bar{I}_j^n = k, \tilde{I}_j^n = l\}}{\Pr\{\bar{I}_j^n = k\}}, \quad \pi_i^n(k) = \Pr\{\bar{I}_j^n = k\}, \quad k, l \in \{0, 1\}, \quad (\text{A.1})$$

where \bar{I}_j^n and \tilde{I}_j^n denote two consecutive states of $\{I_j^n\}_j$ as shown in Fig. 3. The above joint and marginal probabilities can easily be computed in terms of the known probabilistic description of $\{W_j^n\}_j$ as described below.

Let j mark the arrival time of $LTCA_j^n$. At this time instant, $\{W_j^n\}_j = \{1, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\}$. If $Q_j^n \geq 2A_{\max}^n$ then $(\bar{I}_j^n, \tilde{I}_j^n) = (1, 1)$ since the queue cannot become empty of tagged cells before the next two tagged cell arrivals occur. If $Q_j^n < 2A_{\max}^n$ then $\bar{I}_j^n = 1_{\{A_j^n \leq Q_j^n\}}$; i.e. $\bar{I}_j^n = 1$ as long as the departure of $LTCA_j^n$ — which occurs Q_j^n time slots following its arrival instant — does not occur before the next tagged cell arrival. Similarly, if $FTCA_j^n$ occurs at slot i , $i = j + A_j^n$, then $\tilde{I}_j^n = 1_{\{\bar{A}_j^n \leq Q_j^n\}}$. More details are given below.

As described earlier, $P\{\bar{I}_j^n = k, \tilde{I}_j^n = l\}$ is completely determined by the following: the number of packets in the queue at the $LTCA_j^n$ arrival instant; the interarrival time of the $FTCA_j^n$ packet; the number of packets in the queue at the $FTCA_j^n$ arrival instant; the interarrival time of the $STCA_j^n$ packet. To facilitate the derivations, consider the auxiliary Markov chain $\{W_m^{n,T=1}\}_m = \{\hat{Q}_m^n, \phi_m^n, A_m^n\}$, embedded at the tagged packet arrival instants; \hat{Q}_m^n denotes the queue occupancy at the m th tagged packet arrival instant, $\phi_m^n = \{J_m^{n-1}, F_m^n, B_m^n\}$ and A_m^n denotes the interarrival time between the m th and $(m+1)$ th tagged packet arrival. From the steady state and transition probabilities of the Markov chain $\{W_m^{n,T=1}\}_m$, the joint probabilities of two consecutive states of the QAI, $\{\bar{I}_j^n, \tilde{I}_j^n\}$, can be easily determined. For instance,

$$\Pr\{\bar{I}_j^n = 0, \tilde{I}_j^n = 0\} = \sum_{\phi_m^n} \sum_{\phi_{m+1}^n} \sum_{\hat{Q}_m^n < A_m^n} \sum_{\hat{Q}_{m+1}^n < A_{m+1}^n} \Pr\{W_{m+1}^{n,T=1} / W_m^{n,T=1}\} \Pr\{W_{m+1}^{n,T=1}\}. \quad (\text{A.2})$$

The steady state and transition probabilities of the Markov chain $\{W_m^{n,T=1}\}_m$ can be derived from those of the Markov chain $\{W_j^n\}_j$, embedded at each time slot. The steady state probabilities are obtained by conditioning on the event $\{T_j^n = 1\}$; i.e.

$$\begin{aligned} \hat{\pi}^{n,T=1}\{l, \phi_0, a\} &= \lim_{m \rightarrow \infty} \Pr\{\hat{Q}_m^n = l, \phi_m^n = \phi_0, A_m^n = a\} \\ &= \frac{\pi_w^n\{l, T = 1, \phi_0\}}{\sum_{l, \phi_0} \pi_w^n\{l, T = 1, \phi_0\}} \Pr\{A^n = a / \phi^n = \phi_0\}, \end{aligned} \quad (\text{A.3})$$

where $\pi_w^n\{l, T = 1, \phi_0\}$ are the steady state probabilities $\{W_j^n\}_j$ and $\Pr\{A^n = a/\phi^n = \phi_0\} = \Pr\{A^n = a/J^{n-1}\}$, where A^n , ϕ^n and J^{n-1} denote the generic random variables for $\{A_j^n\}_j$, ϕ_m^n and $\{J_j^{n-1}\}_j$, respectively.

The computation of the transition probabilities $\Pr\{W_{m+1}^{n,T=1} / W_m^{n,T=1}\}$ of the auxiliary Markov chain will be facilitated by rewriting them as follows:

$$\Pr\{\hat{Q}_{m+1}^n, \phi_{m+1}^n, A_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n\} = \Pr\{\hat{Q}_{m+1}^n, \phi_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n\} \Pr\{A_{m+1}^n / \phi_{m+1}^n\}, \quad (\text{A.4})$$

where, as before, $\Pr\{A_{m+1}^n = a/\phi_{m+1}^n = \phi_0\} = \Pr\{A_m^n = a/J^{n-1}\}$. The only unknown is therefore $\Pr\{\hat{Q}_{m+1}^n, \phi_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n\}$. Let j denote the slot at which the m th tagged cell arrives, then

$$\Pr\{\hat{Q}_{m+1}^n, \phi_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n\} = \Pr\{Q_{j+A_j^n}^n, T_{j+A_j^n}^n = 1, J_{j+A_j^n}^{n-1}, F_{j+A_j^n}^n, B_{j+A_j^n}^n / Q_j^n, T_j^n = 1, J_j^{n-1}, F_j^n, B_j^n, A_j^n\}. \quad (\text{A.5})$$

The above quantity can be computed starting from state W_j^n and making A_j^n transitions in which no tagged cell arrives in the first $A_j^n - 1$ transitions (with probability 1) and one tagged cell arrives at the last transition (with probability 1). This computation is facilitated by slightly modifying the submatrices $\mathbf{A}_i(h, h+1)$ and $\mathbf{A}_i(h, 1)$ of the transition matrix \mathbf{P} , by taking into consideration that the interarrival time is deterministic (given) and equal to A_j^n . The following hold true for the modified submatrices $\tilde{\mathbf{A}}_i(h, h+1)$ and $\tilde{\mathbf{A}}_i(h, 1)$

$$\tilde{\mathbf{A}}_i(h, h+1) = \frac{1}{p_i^n(h, h+1)} \mathbf{A}_i(h, h+1), \quad \tilde{\mathbf{A}}_i(h, 1) = \frac{1}{p_i^n(h+1, 1)} \mathbf{A}_i(h, 1). \quad (\text{A.6})$$

By using the above submatrices, the probability in the right-hand side of (A.5) is derived by constructing the sequence of vectors $\{\mathbf{r}_t\}_j$, $1 \leq t \leq A_j^n$, where $r_t(h, \mathbf{y})$ denotes the probability $\Pr\{Q_{j+t} = h, \phi_{j+t} = \mathbf{y}\} / Q_j, \phi_j, A_j$, and

$$r_1(h, \mathbf{y}) = \begin{cases} 1 & \text{if } \hat{Q}_j = h \text{ and } \phi_j = \mathbf{y}, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.7})$$

\mathbf{r}_t is evaluated using the following relationships:

$$\mathbf{r}_{t+1}(h) = \mathbf{r}_t(0) \tilde{\mathbf{A}}_h(t, t+1) + \sum_{j=1}^{h+1} \mathbf{r}_t(j) \tilde{\mathbf{A}}_{h-j+1}(t, t+1), \quad t < A_j^n. \quad (\text{A.8})$$

From the last vector $\mathbf{r}_{A_j^n}$ of the sequence, the steady state probability vector just after the $FTCA_j^n$ arrival, conditioned on the state of the system just after the $LTCA_j^n$ arrival, is computed as follows:

$$\mathbf{r}_{FTCA_j^n}(h) = \mathbf{r}_{A_j^n}(0) \tilde{\mathbf{A}}_h(A_j^n, 1) + \sum_{j=1}^{h+1} \mathbf{r}_{A_j^n}(j) \tilde{\mathbf{A}}_{h-j+1}(A_j^n, 1). \quad (\text{A.9})$$

From $\mathbf{r}_{FTCA_j^n}(h)$ the probability in the right-hand side of (A.5) is directly obtained:

$$\Pr\{Q_{j+A_j^n}^n, T_{j+A_j^n}^n = 1, J_{j+A_j^n}^{n-1}, F_{j+A_j^n}^n, B_{j+A_j^n}^n / Q_j^n, T_j^n = 1, J_j^{n-1}, F_j^n, B_j^n, A_j^n\} = \mathbf{r}_{FTCA_j^n}(Q_{j+A_j^n}^n, \mathbf{x}) \quad (\text{A.10})$$

where $\mathbf{x} = \{J_{j+A_j^n}^{n-1}, F_{j+A_j^n}^n, B_{j+A_j^n}^n\}$.

Following the above approach, all joint probabilities $\Pr\{\bar{I}_j^n = k, \tilde{I}_j^n = l\}$ can be computed. Alternatively, all probabilities can be computed from $\Pr\{\bar{I}_j^n = 0, \tilde{I}_j^n = 0\}$ and the probability $\Pr\{\bar{I}_j^n = 1, \tilde{I}_j^n = 0\}$ (computed as above); i.e. by computing only the probabilities of the joint events with $\tilde{I}_j^n = 0$ and, as a result, considering only the cases in which the $STCA_j^n$ arrival finds no tagged cell in the queue. Thus, not every probability in each vector $\{\mathbf{r}_t\}$, $1 \leq t \leq A_j^n$, needs to be computed but only $r_t(h, \mathbf{y})$ for $h < A_{\max}^n + A_j^n - t$. Finally, the remaining probabilities are computed from the following expressions:

$$\Pr\{\bar{I}_j^n = 0\} = \sum_{l=0}^{A_{\max}^n-1} \sum_{a=l+1}^{A_{\max}^n} \sum_{\phi_0} \hat{\pi}^{n,T=1}\{l, \phi_0, a\}, \quad \Pr\{\bar{I}_j^n = 1\} = 1 - \Pr\{\bar{I}_j^n = 0\}, \quad (\text{A.11})$$

$$\Pr\{\bar{I}_j^n = 0, \tilde{I}_j^n = 0\} + \Pr\{\bar{I}_j^n = 0, \tilde{I}_j^n = 1\} = \Pr\{\bar{I}_j^n = 0\}, \quad (\text{A.12})$$

$$\Pr\{\bar{I}_j^n = 1, \tilde{I}_j^n = 0\} + \Pr\{\bar{I}_j^n = 1, \tilde{I}_j^n = 1\} = \Pr\{\bar{I}_j^n = 1\}. \quad (\text{A.13})$$

Appendix B

One of the innovations of this work is due to the allowed dependence of $\{A_j^n\}_j$ (or X_j^{n-1}) from J_j^{n-1} (the QAI associated with node $n - 1$). That is, the time of the next tagged arrival to node n depends on the QAI associated with node $n - 1$. The conditional probability mass function of X_j^{n-1} given J_j^{n-1} can be derived from the (joint) probability of event $\{X_j^{n-1}, J_j^{n-1}\}$. The corresponding probability for node n is derived below by referring to quantities shown on Fig. 3.

The probability of event $\{X_j^{n-1}, J_j^{n-1}\}$ can be derived by proceeding as in part (a). Let j mark the arrival time of $LTCA_j^n$; at this time instant, $\{W_j^n\}_j = \{1, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\}$. It is easy to establish that $D_j^n = Q_j^n$, $\bar{D}_j^n = Q_{j+A_j^n}^1$, by noting that the delay of a tagged (or any) cell is equal to the queue occupancy upon arrival to the queue. The probability of the joint event $\{X_j^n = \bar{D}_j^n - D_j^n + A_j^n, \bar{I}_j^n\}$ can be determined in terms of the stationary probabilities that $\{W_j^n\}_j$ is in state $\{1, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\}$ and considering the A_j^n transition probabilities, $1 \leq A_j^n \leq A_{\max}^n$. More details are given below.

From the above discussion it is evident that $\Pr\{X_j^n, \bar{I}_j^n\}$ can be expressed in terms of the joint probabilities of the following events: the number of packets queued at the $LTCA_j^n$ arrival instant; the interarrival time of the $LTCA_j^n$ packet; the number of packets queued at the $FTCA_j^n$ arrival instant. Specifically,

$$\Pr\{X_m^n = k, \bar{I}_m^n = 0\} = \sum_{h=0}^{A_{\max}^n-1} \sum_{a=h+1}^{A_{\max}^n} \Pr\{\hat{Q}_m^n = h, A_m^n = a, \hat{Q}_{m+1}^n = k + h - a\}, \quad (\text{B.1})$$

$$\Pr\{X_m^n = k, \bar{I}_m^n = 1\} = \sum_{h=1}^{\infty} \sum_{a=1}^h \Pr\{\hat{Q}_m^n = h, A_m^n = a, \hat{Q}_{m+1}^n = k + h - a\}, \quad (\text{B.2})$$

$\Pr\{\hat{Q}_m^n = h, A_m^n = a, \hat{Q}_{m+1}^n = k + h - a\}$ can be derived from the following:

$$\Pr\{\hat{Q}_m^n = h, A_m^n = a, \hat{Q}_{m+1}^n = k + h - a\} \\ \sum_{\phi_m^n} \Pr\{\hat{Q}_{m+1}^n = k + h - a / \hat{Q}_m^n = h, \phi_m^n, A_m^n = a\} \Pr\{\hat{Q}_m^n = h, \phi_m^n, A_m^n = a\},$$

where the probabilities $\Pr\{\hat{Q}_m^n = h, \phi_m^n, A_m^n = a\}$ are the steady state probabilities of the auxiliary Markov chain $\{W_m^{n,T=1}\}_m$ (see Appendix A), and the probabilities $\Pr\{\hat{Q}_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n\}$ can be derived by applying the same procedure as in Appendix A.

In principle — as shown by (B.1) and (B.2) — the computation of $\Pr\{X_m^n = k, \bar{I}_m^n = 1\}$ requires that for each value of the interarrival $A_m^n = a$ the probability calculation procedure involved in (B.1) and (B.2), needs to be carried out for each $Q_m^n \geq a$, that is infinite times. This computation can be simplified though by noting that since the queue never empties for this case (i.e. $\bar{I}_m^n = 1$), the event $\{\hat{Q}_{m+1}^n / \hat{Q}_m^n, \phi_m^n, A_m^n = a\}$ depends only on the number of arrivals in A_m^n steps; the latter depends only on the evolution of the process $\{\phi_j^n\}_j$. Thus,

$$\Pr\{\hat{Q}_{m+1}^n = h / \hat{Q}_m^n = l, \phi_m^n, A_m^n\} = \delta_{\phi_m^n, A_m^n}(h - l + A_m^n),$$

where $\delta_{\phi_m^n, A_m^n}(k)$ denotes the probability of k arrivals in A_m^n steps with process $\{\phi_j^n\}_j$ starting from state ϕ_m^n . The function $\delta_{\phi_m^n, A_m^n}(k)$ can be computed by constructing the sequence of vectors $\{\mathbf{r}_t\}$, $1 \leq t \leq A_m^n$, for each A_m^n , where

$$r_1(h, \mathbf{y}) = \begin{cases} 1 & \text{if } \hat{Q}_m^n = A_m^n = h \text{ and } \phi_j^n = \mathbf{y}, \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.3})$$

(see Appendix A). In this case, it is easy to observe that $\delta_{\phi_m^n, A_m^n}(k) = \sum_{\mathbf{y}} \mathbf{r}_{FTCA_j^n}(k, \mathbf{y})$.

Appendix C

The presence or absence of a background cell at the output of node n at time j is assumed to be (probabilistically) determined by I_j^n (the QAI) and T_j^{n+1} . If $T_j^{n+1} > 1$ it is implied that $X_j^n > T_j^{n+1}$ and these are the cases that need to be considered (Eq. (7)). Thus, $P\{C_j^{n+1}(0, T_j^{n+1}) = 1 / X_j^n > T_j^{n+1}\}$ needs to be derived for all $1 < T_j^{n+1} \leq A_{\max}^n$. Or, equivalently, the joint probability $P\{C_j^{n+1}(0, T_j^{n+1}) = 1, X_j^n > T_j^{n+1}\}$ needs to be derived.

Let j mark the arrival time of $LTCA_j^n$; at that time $\{W_j^n\}_j = \{1, J_j^{n-1}, F_j^n, B_j^n, Q_j^n\}$. Only the cases for $Q_j^n < A_{\max}^n$ (so that $\bar{I}_j^n = 0$) need to be considered. Consider the Q_j^n step evolution of $\{W_j^n\}_j$. If $X_j^n \leq Q_j^n$ then $\bar{I}_j^n = 1$ and such paths are not considered. If $X_j^n > Q_j^n$ then $\bar{I}_j^n = 0$. Let $\{Q_j^n, \dots, Q_{j+Q_j^n}^n\}$ be the state of $\{W_j^n\}_j$ reached at time $j + Q_j^n$. If $Q_{j+Q_j^n}^n > A_{\max}^n - Q_j^n$ then $P\{C_j^{n+1}(0, Q_{j+l}^n) = 1, X_j^n > Q_{j+l}^n\} = \bar{f}_a^n(Q_{j+l}^n)$, $1 \leq l < A_j^n - Q_j^n$ (until the next tagged cell arrival). If $Q_{j+Q_j^n}^n \leq A_{\max}^n - Q_j^n$ then $\{X_j^n > Q_{j+l}^n\}$ with probability $\bar{f}_a^n(Q_{j+l}^n)$ and $\{C_j^{n+1}(0, Q_{j+l}^n) = 1_{\{Q_{j+l}^n \neq 0\}}\}$ for $1 \leq l < A_j^n - Q_j^n$ (until the next tagged cell arrival). More details are provided below, where the QAI associated with the previous node, J_j^n , is employed in the derivations for quantities at the input of node $n + 1$.

To compute $\Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 1, X_j^n(J_j^n) > T_j^{n+1} / J_j^n = 0\}$ it is convenient to compute its complement $\Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1} / J_j^n = 0\}$. The following can be obtained by employing basic probabilistic arguments:

$$\begin{aligned} & \Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1} / J_j^n = 0\} \\ &= \frac{\sum_{l=0}^{A_{\max}^n-1} \sum_{a=l+1}^{A_{\max}^n} \sum_{\phi_j^n} \Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1}, \hat{Q}_j^n = l, \phi_j^n, A_j^n = a\}}{\sum_{l=0}^{A_{\max}^n-1} \sum_{a=l+1}^{A_{\max}^n} \sum_{\phi_j^n} \Pr\{\hat{Q}_j^n = l, \phi_j^n, A_j^n = a\}}. \end{aligned} \quad (C.1)$$

The denominator is easily derived by noting that $\Pr\{\hat{Q}_j^n = l, \phi_j^n, A_j^n = a\}$ are the steady state probabilities of the process $\{W_m^{n,T=1}\}_m = \{\hat{Q}_m^n, \phi_m^n, A_m^n\}_m$ (see Appendix A). The probability $\Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1}, \hat{Q}_j^n = l, \phi_j^n, A_j^n = a\}$ is written as $\Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1} / \hat{Q}_j^n = l, \phi_j^n, A_j^n = a\} \Pr\{\hat{Q}_j^n = l, \phi_j^n, A_j^n = a\}$, where the conditional probability is computed from the following:

$$\begin{aligned} & \Pr\{C_j^{n+1}(J_j^n, T_j^{n+1}) = 0, X_j^n(J_j^n) > T_j^{n+1} / \hat{Q}_j^n = l, \phi_j^n, A_j^n = a\} \\ &= \Pr\{\hat{Q}_{j+l+T_j^{n+1}} = 0, X_j^n(0) > T_j^{n+1} / \hat{Q}_j = l, \phi_j^n, A_j^n = a\}, \end{aligned} \quad (C.2)$$

$$\begin{aligned} & \Pr\{\hat{Q}_{j+l+T_j^{n+1}} = 0, X_j^n(0) > T_j^{n+1} / \hat{Q}_j = l, \phi_j^n, A_j^n = a\} \\ &= \begin{cases} \Pr\{\hat{Q}_{j+l+T_j^{n+1}} = 0 / \hat{Q}_j = l, \phi_j^n, A_j^n = a\}, & \text{if } l + T_j^{n+1} < A_j^n, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \quad (C.3)$$

where $\Pr\{\hat{Q}_{j+l+T_j^{n+1}} = 0 / \hat{Q}_j = l, \phi_j^n, A_j^n = a\} = \sum_{\mathbf{y}} r_{l+T_j^{n+1}}(0, \mathbf{y})$ and the sentence of vectors $\{\mathbf{r}_t\}$, $1 \leq t \leq A_j^n$ is obtained as explained in the Appendix A.

Appendix D

In this appendix the Markov chain $W_j^n \equiv \{\Phi_j^n, Q_j^n\}$ is described and the derivation of its stationary probabilities is outlined.

$$\mathbf{P} = \begin{matrix} Q_j^n = 0 \\ Q_j^n = 1 \\ Q_j^n = 2 \\ Q_j^n = 3 \\ \vdots \end{matrix} \begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{A}_4 & \cdots \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{A}_4 & \cdots \\ \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

By taking into consideration that in each transition the time T_j^n increases by one (time unit) until an arrival occurs, it can be easily established that submatrices \mathbf{A}_i have the following structure:

$$\mathbf{A}_i = \begin{matrix} T_j^n = 1 \\ T_j^n = 2 \\ T_j^n = 3 \\ \vdots \\ T_j^n = A_{\max-1} \\ T_j^n = A_{\max} \end{matrix} \left[\begin{array}{cccccc} \mathbf{A}_i(1, 1) & \mathbf{A}_i(1, 2) & \mathbf{0} & \dots & & \\ \mathbf{A}_i(2, 1) & \mathbf{0} & \mathbf{A}_i(2, 3) & \mathbf{0} & \dots & \\ \mathbf{A}_i(3, 1) & \mathbf{0} & \mathbf{0} & \mathbf{A}_i(3, 4) & \mathbf{0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{A}_i(A_{\max-1}, 1) & \mathbf{0} & \dots & \dots & \dots & \mathbf{A}_i(A_{\max-1}, A_{\max}) \\ \mathbf{A}_i(A_{\max}, 1) & \mathbf{0} & \dots & \dots & \dots & \dots \end{array} \right]$$

Submatrices $\mathbf{A}_i(h, h+1)$ represent transitions in which no tagged arrival occurs at time $T_j^n = h$ and i packets (carried-on plus background) arrive at slot j . These submatrices are given by

$$\begin{aligned} & \mathbf{A}_i(h, h+1) \\ & \begin{matrix} J_h^{n-1} = 0, F_h^n = 0 \\ J_h^{n-1} = 0, F_h^n = 1 \\ J_h^{n-1} = 1, F_h^n = 0 \\ J_h^{n-1} = 1, F_h^n = 1 \end{matrix} \\ & = \left[\begin{array}{cccc} \mathbf{B}_i\{c(h+1)p_f(0, 0) \\ + (1 - c(h+1))\} & \mathbf{B}_{i-1}c(h+1)p_f(0, 1) & \mathbf{0} & \mathbf{0} \\ \mathbf{B}_i c(h+1)p_f(1, 0) & \mathbf{B}_{i-1}c(h+1)p_f(1, 1) \\ & + \mathbf{B}_i(1 - c(h+1)) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_i p_f(0, 0) & \mathbf{B}_{i-1} p_f(0, 1) \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_i p_f(1, 0) & \mathbf{B}_{i-1} p_f(1, 1) \end{array} \right] p_i^n(h, h+1), \end{aligned}$$

where $c(h+1)$ is the probability that a carried-on packet arrives at time $T_j^n = h+1$ and \mathbf{B}_i is a 2×2 matrix which contains the product of the transition probabilities of the background process and the probability that i background packets arrive given the status of the background process; $p_f(l, m)$ is the probability that the splitting process moves from state l to state m ; $p_i^n(h, h+1)$ is the transition probability for T_j^k (defined earlier). Submatrices $\mathbf{A}_i(h, 1)$ represent transitions in which a tagged arrival occurs at time $T_j^n = h$ and hence: (a) no transition occurs in the splitting process; (b) $i-1$ new background packets arrive in slot j .

Due to the structure of submatrices $\mathbf{A}_i(h, h+1)$ and $\mathbf{A}_i(h, 1)$, matrix \mathbf{A}_0 has several zero columns: all the columns corresponding to submatrices $\mathbf{A}_0(h, 1)$, $1 \leq h \leq A_{\max}$, are zero columns (total of eight). In addition, each submatrix $\mathbf{A}_0(h, h+1)$ has its last two columns equal to zero. As a result, $2^3 + 2(A_{\max} - 1)$ columns of matrix \mathbf{A}_0 vanish. This implies that the corresponding columns of matrix \mathbf{G} defined by the unique stochastic solution to equation:

$$\mathbf{G} = \sum_{k=0}^{\infty} \mathbf{A}_k \mathbf{G}^k, \quad (\text{D.1})$$

— see Section 3.4 in [15] — also vanish and hence \mathbf{G} is reducible. In this case, the solution approach should be modified. By permuting the phases, submatrices \mathbf{A}_k can be partitioned as shown below:

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{A}_k(1) & \mathbf{A}_k(2) \\ \mathbf{A}_k(3) & \mathbf{A}_k(4) \end{bmatrix}.$$

$\mathbf{A}_0(1)$ and $\mathbf{A}_0(3)$ are empty and thus matrix \mathbf{G} has the following structure:

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{G}(2) \\ \mathbf{0} & \mathbf{G}(4) \end{bmatrix}.$$

By noting that

$$\mathbf{G}^k = \begin{bmatrix} \mathbf{0} & \mathbf{G}(2)\mathbf{G}^{k-1}(4) \\ \mathbf{0} & \mathbf{G}^k(4) \end{bmatrix},$$

$\mathbf{G}(2)$ and $\mathbf{G}(4)$ are obtained by instantiating (12) to this case. Thus,

$$\begin{aligned} \mathbf{G}(2) &= \sum_{k=0}^{\infty} \mathbf{A}_k(2)\mathbf{G}^k(4) + \sum_{k=1}^{\infty} \mathbf{A}_k(1)\mathbf{G}(2)\mathbf{G}^{k-1}(4), \\ \mathbf{G}(4) &= \sum_{k=0}^{\infty} \mathbf{A}_k(4)\mathbf{G}^k(4) + \sum_{k=1}^{\infty} \mathbf{A}_k(3)\mathbf{G}(2)\mathbf{G}^{k-1}(4). \end{aligned}$$

Substantial computational gains can be achieved by following the above calculation procedure and employing only the submatrices of \mathbf{G} which are not empty. The computational complexity is further reduced by exploiting the structure of matrices \mathbf{A}_k . As pointed out before, $2 \cdot 2^3$ elements in each row of matrix \mathbf{A}_k are non-zero, reducing the complexity of the product $\mathbf{A}_k\mathbf{G}^k$ to $2 \cdot 2^3$ products for each non-zero element in \mathbf{G} . It should also be noted that the resulting system exhibits a homogeneous behavior for all states; that is, rows corresponding to $Q_j^n = 0$ and $Q_j^n = 1$ are identical in \mathbf{P} , and hence $\mathbf{K} = \mathbf{G}$ (see [15]). By following the standard procedure described in [15] — employing the computational reduction as outlined above — the steady state probabilities \mathbf{x}_0 for the boundary states are calculated; the elements of \mathbf{x}_0 are the steady state probabilities associated with the level (occupancy) 0 and some phase.

Finally, the queue occupancy probability mass function is calculated by applying Ramaswami’s algorithm, [16]. Let \mathbf{x}_i denote the vector of the steady state probabilities associated with a level (occupancy) i and some phase. Then,

$$\mathbf{x}_i = \left[\sum_{k=0}^{i-1} \mathbf{x}_k \bar{\mathbf{A}}_{i+1-k} \right] (\mathbf{I} - \bar{\mathbf{A}}_1)^{-1}, \quad \bar{\mathbf{A}}_k = \sum_{i=k}^{\infty} \mathbf{A}_i \mathbf{G}^{i-k} \quad (k \geq 0). \tag{D.2}$$

Clearly, the complexity reduction considerations in the \mathbf{G} matrix computation (as discussed above) hold also for the computation of matrix $\bar{\mathbf{A}}_k$.

Appendix E

E.1. Derivation of $\Pr\{D_k^n, QAI_k^n, QAI_k^{n-1}\}$

The derivation of $\Pr\{D_k^n, QAI_k^n, QAI_k^{n-1}\}$ is presented in this part. This quantity is computed by conditioning on the status of the system observed by the previous tagged packet (packet $ok - 1$) and the interarrival between packets $ok - 1$ and ok :

$$\begin{aligned} & \Pr\{D_k^n, QAI_k^n, QAI_k^{n-1}\} \\ &= \sum_{\hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n} \Pr\{D_k^n, QAI_k^n, QAI_k^{n-1} / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\} \Pr\{\hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\}. \quad (E.1) \end{aligned}$$

$\Pr\{\hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\}$ is the steady state probability of the auxiliary process $\{W_m^{n,T=1}\}_m$ and $\Pr\{D_k^n, QAI_k^n, QAI_k^{n-1} / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\} = \Pr\{QAI_k^n / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\} \Pr\{D_k^n, QAI_k^{n-1} / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\}$, where $\Pr\{QAI_k^n = 1 / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\} = I_{\{\hat{Q}_{ok-1}^n \geq A_{ok-1}^n\}}$ and $\Pr\{D_k^n, QAI_k^{n-1} / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\} = \Pr\{\hat{Q}_{ok-1}^n, QAI_k^{n-1} / \hat{Q}_{ok-1}^n, \phi_{ok-1}^n, A_{ok-1}^n\}$ is obtained directly from the transition probabilities of the auxiliary process.

E.2. Derivation of $\Pr\{QAI_M, QAI_{M-1}\}$

The derivation of $\Pr\{\overline{QAI}_M, \overline{QAI}_{M-1}\}$ is presented in this part.

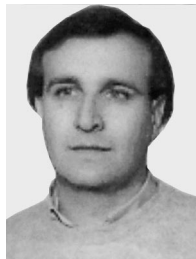
$$\begin{aligned} \Pr\{\overline{QAI}_M, \overline{QAI}_{M-1}\} &= \Pr\{QAI_M^1, QAI_{M-1}^1, QAI_M^2, QAI_{M-1}^2, \dots, QAI_M^N, QAI_{M-1}^N\} \\ &= \left(\prod_{i=2}^N \Pr\{QAI_M^i, QAI_{M-1}^i / QAI_M^{i-1}, QAI_{M-1}^{i-1}\} \right) \Pr\{QAI_M^1, QAI_{M-1}^1\}. \quad (E.2) \end{aligned}$$

Thus, the problem is reduced to the computation of probabilities $\Pr\{QAI_M^i, QAI_{M-1}^i / QAI_M^{i-1}, QAI_{M-1}^{i-1}\}$, which are obtained from the auxiliary process $\{W_m^{n,T=1}\}_m$. It is easy to observe that $\Pr\{QAI_M^i, QAI_{M-1}^i, QAI_M^{i-1}, QAI_{M-1}^{i-1}\}$ can be expressed in terms of $\Pr\{W_{M-1}^{n,T=1}, W_M^{n,T=1}\}$.

References

- [1] R. Onvural, Asynchronous Transfer Mode: Performance Issues, Artec House, Norwood, MA, 1993.
- [2] I. Stavrakakis, Queueing behavior of two interconnected buffers of a packet network with application to the evaluation of packet routing policies, Int. J. Digital Analog Commun. Systems 4 (1991) 249–260.
- [3] I. Stavrakakis, Efficient modeling of merging and splitting processes in large networking structures, IEEE J. Selected Areas Commun. 9 (8) (1991) 1336–1347.
- [4] W.-C. Lau, S.-Q. Li, Traffic analysis in large-scale high speed integrated networks: validation of nodal decomposition approach, in: Proceedings of the IEEE Infocom'93, San Francisco, CA, 1993.
- [5] J. Kurose, Open issues and challenges in providing quality of service guarantees in high speed networks, Comput. Commun. Rev. 23 (1) (1993).
- [6] R. Nagarajan, J. Kurose, On defining, computing and guaranteeing quality-of-service in high-speed networks, in: Proceedings of the IEEE Infocom'92, Florence, Italy, 1992.

- [7] F. Guillemin, W. Monin, Management of cell delay variation in ATM networks, in: Proceedings of the IEEE Globecom'92, Orlando, FL, 1992.
- [8] C. Bisdikian, W. Matragi, K. Sohraby, A study of the jitter in ATM multiplexers, in: Proceedings of the Fifth International Conference on Data Communication Systems and their Performance, Raleigh, NC, October 25–28, 1993.
- [9] W. Matragi, C. Bisdikian, K. Sohraby, Jitter calculus in ATM networks: single node case, in: Proceedings of the IEEE Infocom'94, Toronto, Canada, 1994
- [10] W. Matragi, C. Bisdikian, K. Sohraby, Jitter calculus in ATM networks: multiple node case, in: Proceedings of the IEEE Infocom'94, Toronto, Canada.
- [11] J. Roberts, F. Guillemin, Jitter in ATM networks and its impact on peak rate enforcement, *Perform. Eval.* 16 (1–3) (1992) 35–48.
- [12] R. Landry, I. Stavrakakis, A queueing study of peak-rate enforcement for jitter reduction in ATM networks, in: Proceedings of the IEEE GLOBECOM'94, San Francisco, November 27–December 1, 1994.
- [13] C. Fulton, S.-Q. Li, Delay jitter correlation analysis for traffic transmission on high speed networks, in: Proceedings of the IEEE Infocom'95, Boston, April 4–6, 1995.
- [14] C. Fulton, S.-Q. Li, Delay jitter first-order statistical function in high-speed multi-media networks, in: Proceedings of the ICC'96, Dallas, June 23–27, 1996.
- [15] M.F. Neuts, *Structured Stochastic Matrices of the M/G/1 Type and their Applications*, Marcel Dekker, New York, 1989.
- [16] V. Ramaswami, A stable recursion for the steady state vector in Markov chains of the M/G/1 type, *Commun. Statist. Stochast. Models* 4 (1) (1988) 183–188.
- [17] M. Conti, E. Gregori, A. Larsson, Analysis and modeling of MPEG-1 Video Sources, *IEEE J. Selected Areas Commun.* 14 (7) (1996) 1445–1471.
- [18] M. Conti, E. Gregori, Traffic characterization and bandwidth allocation for MPEG 2 VBR video traffic, in: Proceedings of the IEEE GLOBECOM'97, Phoenix, AZ, November 8–10, 1997.
- [19] P. Heidelberger, Fast simulation of rare events in queueing and reliability models, in: L. Donatiello, R. Nelson (Eds.), *Lecture Notes in Computer Science*, Vol. 729, Springer, Berlin, 1993 pp. 165–202.



Marco Conti received the Laurea degree in Computer Science from the University of Pisa, Italy, in 1987. In 1987 he joined the Networks and Distributed Systems Department of CNUCE, an institute of the Italian National Research Council (CNR) where he is currently a senior researcher. He has written a large number of research papers in the areas of design, modeling and performance evaluation of computer communication systems. He is co-author of the book “Metropolitan Area Networks” (Springer, London, 1997). His current research interests include Internet architecture and protocols, wireless networks, and QoS in packet switching networks.



Enrico Gregori received the “Laurea” in electronic engineering from the University of Pisa in 1980. He joined CNUCE, an institute of the Italian National Research Council (CNR) in 1981. He is currently a CNR Research Director. In 1986 he held a visiting position in the IBM Research Center in Zurich working on network software engineering and on heterogeneous networking. He has contributed to several national and international projects on computer networking. He was the Italian delegate in the OSI standardization committees initially working on the standardization of the transport layer and later in OSI management committees. He has authored a large number of papers in the area of computer networks and has published in international journals and conference proceedings. He is co-author of the book “Metropolitan Area Networks” (Springer, Berlin, 1997).



Ioannis Stavrakakis received his Diploma in Electrical Engineering, from Aristotelian University of Thessaloniki (Greece), in 1983 and Ph.D. in Electrical Engineering from University of Virginia (USA), in 1988; He worked as an Assistant Professor in CSEE, University of Vermont (USA) during 1988–1994; Associate Professor of ECE, Northeastern University, Boston (USA), 1994–1999; and currently he is an Associate Professor of Informatics, University of Athens (Greece). His teaching and research interests are focused on resource allocation protocols and traffic management for communication networks. His past research has been published in over 90 scientific journals and conference Proceedings. His research has been funded by NSF, DARPA, GTE, BBN and Motorola (USA) as well as Greek and European Union Funding agencies. He has served in NSF research proposal review panels and involved in the organization of numerous conferences sponsored by IEEE, ACM, ITC and IFIP societies. He is a senior member of IEEE, a member of the IEEE Technical Committee on Computer Communications (TCCC) and a member of IFIP WG6.3. He has served as an elected officer for TCCC, as a co-organizer of the 1996 International Teletraffic Congress (ITC) mini-seminar on “Performance Modeling and Design of Wireless/PCS Networks”, as the organizer of the 1999 IFIP WG6.3 Workshop and a technical program Co-chair for the IFIP Networking 2000 conference. He is an Associate Editor of the ACM/Baltzer Wireless Networks Journal.