

An Analytic Model of Correlations Induced in a Packet Stream by Background Traffic in IP Access Networks *

J. Kumaran^a, K. Mitchell^a, and A. van de Liefvoort^a

^aDivision of Computer Science & Electrical Engineering,
School of Interdisciplinary Computing and Engineering,
University of Missouri–Kansas City,
Kansas City, MO 64110 USA.

In this paper we present an analytic model that can quantify the impact of individual streams in an IP access network on Quality of Service (QoS). It differs from other models in that it incorporates the service characteristics of each per-stream flow as well as any general or correlated arrival process. The model shows that the service requirements at a server for background classes of traffic in an IP access network can have a significant effect on a tagged class, impacting the variance and introducing slowly decaying correlations in the tagged stream.

1. Introduction

The problem of ensuring QoS for different service classes in IP networks has sparked renewed interest in the design and analysis of mechanisms originally conceived for flow based networks. IP networks are relatively cheap to build, simple to manage, scale very well, and have been extremely successful in carrying delay insensitive data. Recently proposed applications in the fields of sensing, remote control, robotics, telemedicine, and telesurgery, require highly reliable data transport over IP networks with strict guarantees as to delay and jitter. This has sparked much debate about the ability of IP networks to provide the needed quality of service (QoS) over the Internet.

Ensuring QoS for a service class in IP networks poses the same issues and considerations as in flow based networks such as ATM except that the traffic characteristics can vary more widely. In flow based networks, traffic streams share bandwidth with their own classes only, and hence it is reasonable to assume a certain homogeneity in terms of the arriving streams and their service requirements. In IP networks, bandwidth is usually shared between all arriving streams of all classes, and assumptions of homogeneity may not be realistic. In fact, we will show that common assumptions regarding departure processes may not be valid.

Recent studies on dimensioning and admission control for access networks [1, 2, 20], address performance at the session level, where the multiplexing of individual packets from multiple applications can be modeled using a generalized processor sharing discipline. In these models, streaming traffic is assumed to occupy a fixed share of bandwidth. A packet level analysis may reveal distortion in the multiplexed stream. Present multimedia applications are relatively elastic and simple buffering at both the source and destination can provide acceptable QoS. However, for newly proposed applications, both bandwidth and especially latency requirements are of great concern.

Studies have shown that as the number of service classes becomes arbitrarily large in a stable system, a departing tagged stream will not suffer distortion, and the effect of multiplexing at the upstream node can be ignored [5]. However, in an IP environment running over an OSPF network, the maximum number of service classes is fairly small ([18]). In such a situation, particularly in access networks the assumption that the impact is negligible needs to be re-examined.

The study of traffic distortion (jitter) in flow based networks has been addressed in the literature [9, 15, 17, 19, 23–25]. De Veciana, Courcoubetis, and Walrand use the effective bandwidth of multiple

*This work has been supported in part by US NSF under grant No. ANI 0106640.

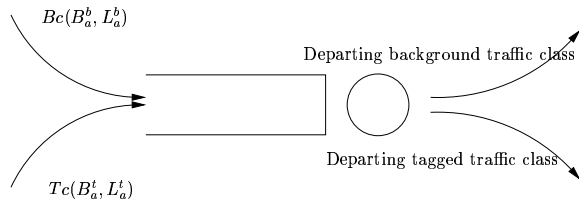


Fig. 1. Diagram for the shared buffer

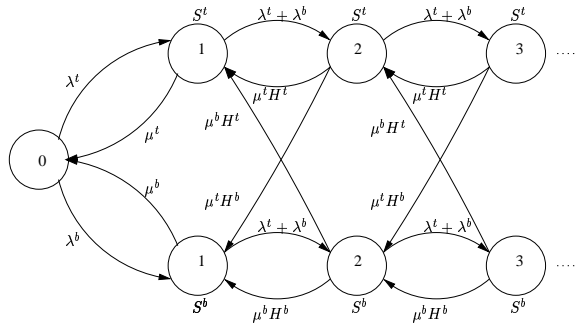


Fig. 2. Markov state diagram for the all exponential model

sources to determine the service rate needed (decoupling bandwidth) such that the asymptotics of the network reduce to the single buffer case [4]. Ohba, Murata, and Miyahara present a discrete time model with deterministic service times that captures the first-order characteristics of the interdeparture times [19]. Lau and Li present a frequency domain representation of the effects of statistical multiplexing [9]. Matraji, Bisdikian, and Sohraby present a heavy traffic model in discrete time using functional equations to determine the jitter of a tagged stream [13, 14]. Conti, Gregori, and Stavarakakis capture the effects of background correlations on the delay of the tagged arrivals in the shared queue [3]. These studies all assume that the service process is the same for each class of arriving traffic. The second-order per-stream effects of multiplexing using FIFO scheduling are, to our knowledge, not known for multiple classes with different service demands under general loads. We have found a few papers where such scenarios are studied under heavy load assumption, [13, 14, 22].

The model we introduce here is able to quantify the impact on the QoS of individual streams in an IP access network. It differs from other models in that it incorporates the differing service characteristics of each per-stream flow as well as any general or correlated arrival process. The model presented here is considerably different from earlier studies in that it captures the tagged stream departure characteristics for any service load.

In section 2 the basic model is described and in section 3 the model development steps are given. Section 4 presents numerical results to support our claim that the common assumptions regarding departure processes may not be valid. Section 5 concludes this paper.

2. Basic Model Description

Without loss of generality, we assume only two classes of traffic: Tagged and Background, each with its own arrival and service requirement descriptions. In order to provide a tractable analysis, an infinite length buffer is assumed. Fig. 1 represents two classes of traffic multiplexed at a server serving according to a first come first serve (FCFS) scheduling policy. The departing flows are routed according to the path of each flow.

An exact Markov space representation for all states with k customers present, two classes of traffic, and $k \geq 0$, would require 2^k global states to keep track of the actual order of the two traffic types present in the buffer. This makes an exact analysis intractable unless k is small. We present an analytic model that allows us to study the characteristics of the departing tagged stream while having only a small number of global states, even when arrivals are correlated. The approximation is based on the observation that knowledge of the relative ordering of arriving types is needed only when the packets are taken out of the queue to start their service.

In our approximation we model the arrivals as a superimposed arrival process, where arrivals from either class are placed on a queue. Class distinctions are ignored at this transition and are ignored while in the queue. The service distribution for the packet taken out of the queue depends on the class to which it belongs and the class is determined by the probability of that class of arrival having occurred next.

This approximation is exact (and commonly applied) if both classes follow the same arrival process. In this case, the service is a mixture of the service requirements for the different traffic classes. But if the

arrivals adhere to different processes with differing marginal distributions and/or correlation structure, then this becomes an approximation, where we incorporate a second and a separate arrival component into the service space, whose function is to determine the class of a queued customer at the time that this customer is scheduled for service and to preserve the correlation structure. Thus two sample paths are generated : one from the timed arrival process (which helps to keep track of the number of customers present, for queueing purposes) and the other for the embedded arrival points at the departure epochs (which helps to decide which service process to activate). Both sample paths are drawn from the same process and are probabilistically equal. At the end of each idle period, both the time and type of arrival that occurs are known exactly. Thus the two sample paths are synchronized at the end of an idle period.

This synchronization operation copies the internal state of the timed arrival process into the internal state of the arrival point process that has become part of the service process. The timed arrival process is used to generate the event times at which either class generates an arrival to be placed on the queue as a generic customer. The arrival point process incorporated in the service time is used to determine the class of a generic customer when this generic customer is taken off the queue and placed in service.

This approximation is exact in the simple cases where the arrival distributions of the tagged and the background classes are identical or where they are both Poisson with different rates λ^t and λ^b respectively. In the latter case, it is commonly used as the service time becomes a mixture. The probability of the tagged arrival event happening before the background arrival event is $H_t = (\lambda^t + \lambda^b)^{-1}\lambda^t$ and the probability of the background arrival event happening before the tagged arrival event is $H_b = (\lambda^t + \lambda^b)^{-1}\lambda^b$. These probabilities decide the service type of the next packet as long as there is a packet waiting in the queue for service. Fig. 2 shows the state transition diagram for such an all exponential situation approximation discussed above. In the next section this model is described for non-exponential processes.

3. Model Description

3.1. Matrix Exponential Process

We use Linear Algebraic Queueing Theory (LAQT) to study the second order statistics of an isolated departure stream [12]. The needed material is briefly reviewed. A matrix exponential (ME) distribution [11] is defined as a probability distribution whose density can be written as

$$f(t) = \mathbf{p} \exp(-\mathbf{B}t) \mathbf{B} \mathbf{e}', \quad t \geq 0, \quad (1)$$

where \mathbf{p} is the starting operator for the process, \mathbf{B} is the process rate operator, and \mathbf{e}' is a summing operator consisting of a vector of all 1's. The n^{th} moment of the matrix exponential distribution is given by $E[X^n] = n! \mathbf{p} \mathbf{V}^n \mathbf{e}'$, where \mathbf{V} is the inverse of \mathbf{B} . The class of matrix exponential distributions is identical to the class of distributions that possess a rational Laplace-Stieltjes transform. As such, it is more general than continuous phase type distributions which have a similar appearance.

The joint density function of the first k -successive intervals between events describes a matrix exponential process (MEP):

$$f_k(x_1, \dots, x_k) = \mathbf{p} \exp(-x_1 \mathbf{B}) \mathbf{L} \dots \exp(-x_k \mathbf{B}) \mathbf{L} \mathbf{e}'. \quad (2)$$

The matrix \mathbf{L} is the event generator matrix. Examples for such processes are a Poisson process ($\mathbf{B}=[\lambda]$, $\mathbf{L}=[\lambda]$), a renewal process ($\mathbf{L} = \mathbf{B} \mathbf{e}' \mathbf{p}$), and a Markov Arrival Processes (MAP) ($\mathbf{B} = -\mathbf{D}_0$, $\mathbf{L} = \mathbf{D}_1$). Note that \mathbf{B} and \mathbf{L} are not limited to being Markovian rate matrices, so every MAP is an MEP, but not vice versa (see also [10]). We assume the process to be covariance stationary, so that \mathbf{p} is the stationary vector for the process at embedded event points (i.e., $\mathbf{p} \mathbf{V} \mathbf{L} = \mathbf{p}$). The expression for the lag- k covariance, the covariance between the first interval and the k th is

$$\text{cov}[X_1, X_k] = \mathbf{p} \mathbf{V} (\mathbf{V} \mathbf{L})^k \mathbf{V} \mathbf{e}' - (\mathbf{p} \mathbf{V} \mathbf{e}')^2. \quad (3)$$

Hence, the auto-correlation at lag- k , $r[k]$ can be found by dividing $\text{cov}[X_1, X_k]$ by the variance $\text{var}[X] = 2 \mathbf{p} \mathbf{V}^2 \mathbf{e}' - (\mathbf{p} \mathbf{V} \mathbf{e}')^2$.

Finally, the marginal process is matrix exponential with density given in equation (1).

3.2. Model Development

The arrival and the service process for both the tagged and the background processes are assumed to be matrix exponential processes. Let the dimensions of the tagged arrival process, background arrival process, tagged service process, background service process be m_a^t , m_a^b , m_s^t , and m_s^b respectively. The

subscripts “ a ”, (“ ap ”, “ s ”, and, “ ta ”) are used to indicate arrival space, arrival point space, service space, and timed arrival space respectively. The superscripts are class indicators.

For the analytic development of the model, we define the following matrix and vector operators:

$\mathbf{B}_a^t, \mathbf{L}_a^t$: The MEP rate operator and event operator for the tagged arrivals.

$\mathbf{B}_a^b, \mathbf{L}_a^b$: The MEP rate operator and event operator for the background arrivals.

$\mathbf{B}_{ta}, \mathbf{L}_{ta}$: The MEP rate and the event operator for the superposition of tagged and background arrivals.

$\mathbf{B}_{ta} = \mathbf{B}_a^t \oplus \mathbf{B}_a^b, \mathbf{L}_{ta} = \mathbf{L}_a^t \oplus \mathbf{L}_a^b$, where \oplus denotes the Kronecker sum.

e_{ta} : The summing operator for the timed arrival space.

\mathbf{H}_{ap}^t : The MEP operator that transforms the arrival components of the system to immediately after the occurrence of a tagged arrival.

$$\mathbf{H}_{ap}^b = \int_0^\infty \exp(-\mathbf{B}_a^t \otimes \mathbf{I}_a^t t) \exp(-\mathbf{I}_a^t \otimes \mathbf{B}_a^b t) (\mathbf{L}_a^t \otimes \mathbf{I}_a^b) dt = \mathbf{B}_{ta}^{-1} (\mathbf{L}_a^t \otimes \mathbf{I}_a^b), \quad (5)$$

where \mathbf{I}_a^b is the identity operator of dimension m_a^b .

\mathbf{H}_{ap}^b : The MEP operator that transforms the arrival components of the system to immediately after the occurrence of a background arrival.

$$\mathbf{H}_{ap}^b = \int_0^\infty \exp(-\mathbf{I}_a^t \otimes \mathbf{B}_a^b t) \exp(-\mathbf{B}_a^t \otimes \mathbf{I}_a^t t) (\mathbf{I}_a^t \otimes \mathbf{L}_a^b) dt = \mathbf{B}_{ta}^{-1} (\mathbf{I}_a^t \otimes \mathbf{L}_a^b) \quad (6)$$

The \mathbf{H}_{ap}^t (or \mathbf{H}_{ap}^b) operators transform the arrival components of the system to immediately after the occurrence of an arrival, conditioned on the current arrival being tagged (or background). They maintain the internal state information of the tagged(background) arrival processes. Hence, these operators effectively transfer the arrival components of the system from one arrival to another without incorporating their inter-event time, see [11].

The service space includes the arrival points and the individual service spaces of both the tagged and the background demands although only one process will be active at any point of time. For the service process, we define the following matrix and vector operators:

$\mathbf{B}_s^t, \mathbf{L}_s^t$: The MEP rate operator and event operator for the service process of tagged arrivals.

$\mathbf{B}_s^b, \mathbf{L}_s^b$: The MEP rate operator and event operator for the service process of background arrivals.

$\mathbf{B}_s, \mathbf{L}_s$: The MEP rate operator and the event operator describing the service process.

$$\mathbf{B}_s = \begin{bmatrix} \mathbf{I}_{ap} \otimes \mathbf{B}_s^t \otimes \mathbf{I}_s^b & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{ap} \otimes \mathbf{I}_s^t \otimes \mathbf{B}_s^b \end{bmatrix}, \mathbf{L}_s = \begin{bmatrix} \mathbf{H}_{ap}^t \otimes \mathbf{L}_s^t \otimes \mathbf{I}_s^b & \mathbf{H}_{ap}^b \otimes \mathbf{L}_s^t \otimes \mathbf{I}_s^b \\ \mathbf{H}_{ap}^t \otimes \mathbf{I}_s^t \otimes \mathbf{L}_s^b & \mathbf{H}_{ap}^b \otimes \mathbf{I}_s^t \otimes \mathbf{L}_s^b \end{bmatrix} \quad (7)$$

where \mathbf{I}_{ap} is defined by $\mathbf{I}_{ap} = \mathbf{I}_a^t \otimes \mathbf{I}_a^b$.

e'_s : The summation operator in the service space.

$$e'_s = \begin{bmatrix} e'_a{}^t \otimes e'_a{}^b \otimes e'_s{}^t \otimes e'_s{}^b \\ e'_a{}^t \otimes e'_a{}^b \otimes e'_s{}^t \otimes e'_s{}^b \end{bmatrix} \quad (8)$$

The timed arrival and service processes for each class are concurrently active and are statistically independent of one another. These spaces are combined into the system space by using the *Kronecker product*, whose operator is denoted as “ \otimes ”. The matrices and the vectors operators for the processes involved are defined in the the system space as $\widehat{\mathbf{B}}_{ta} = \mathbf{B}_{ta} \otimes \mathbf{I}_s$, $\widehat{\mathbf{L}}_{ta} = \mathbf{L}_{ta} \otimes \mathbf{I}_s$, $\widehat{\mathbf{B}}_s = \mathbf{I}_{ta} \otimes \mathbf{B}_s$, $\widehat{\mathbf{L}}_s = \mathbf{I}_{ta} \otimes \mathbf{L}_s$, $\widehat{e}'_s = \mathbf{I}_{ta} \otimes e'_s$.

Where the matrices \mathbf{I}_{ta} and \mathbf{I}_s are identity matrices of the dimensions of the timed arrival and the service spaces respectively.

We now introduce a sync operator \otimes , that synchronizes the arrival points embedded in the system space with the timed arrivals. This synchronization operation is performed at the end of each idle period. The sync operator is defined by

$$\otimes(M, N) \triangleq [M_{c1}|N_{c1}|M_{c2}|N_{c2}\dots|M_{cm}|N_{cm}],$$

where the matrices M, N are of dimensions $(m \times m)$ and the operation M_{ci} is the matrix obtained by replacing all the columns of the matrix M with zeroes except the i^{th} column.

The following example explains this sync operator. For the matrices M and N , let

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad N = \begin{bmatrix} e & f \\ g & h \end{bmatrix}, \quad \text{then } \otimes(M, N) = \begin{bmatrix} a & 0 & e & 0 & 0 & b & 0 & f \\ c & 0 & g & 0 & 0 & d & 0 & h \end{bmatrix}$$

For brevity, we introduce L_{\otimes} as

$$L_{\otimes} = \otimes(L_a^t \otimes I_a^b, I_a^t \otimes L_a^b) \otimes (p_s^t \otimes p_s^b). \quad (9)$$

The intuitive notion for L_{\otimes} is: an arrival event which finds the server idle activates the class-appropriate service ($p_s^t \otimes p_s^b$). At the same time, a cloning of the arrival space takes place, one to generate the next generic arrival in time, the other to generate the next type of arrival regardless of when it happens. The steady state balance equations can now be written and solved as a standard QBD process as depicted in Fig. 3

$$\pi(0)B_{ta} = \pi(1)\widehat{L}_s e'_s \quad (10)$$

$$\pi(1)(\widehat{B}_{ta} + \widehat{B}_s) = \pi(0)L_{\otimes} + \pi(2)\widehat{L}_s \quad (11)$$

$$\pi(n)(\widehat{B}_{ta} + \widehat{B}_s) = \pi(n-1)\widehat{L}_{ta} + \pi(n+1)\widehat{L}_s, \quad n > 1. \quad (12)$$

3.3. Approximation for the Departure Process of the Tagged Stream

The departure process of the tagged stream is characterized by mapping the transitions that result in the departure of the tagged customers into the L matrix and all the other transitions are mapped into the B matrix. The B and the L descriptors for the departing tagged stream have an infinite state representation and are shown below

$$B_d = \begin{bmatrix} B_{ta} & L_{\otimes} & 0 & 0 & \cdots \\ \widehat{L}_s^b e'_s & (\widehat{B}_{ta} + \widehat{B}_s) & \widehat{L}_{ta} & 0 & \cdots \\ 0 & \widehat{L}_s^b & (\widehat{B}_{ta} + \widehat{B}_s) & \widehat{L}_{ta} & \cdots \\ \vdots & \ddots & \ddots & \ddots & \cdots \end{bmatrix}, \quad L_d = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ \widehat{L}_s^t e'_s & 0 & 0 & \cdots \\ 0 & \widehat{L}_s^t & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

where

$$\widehat{L}_s^t = I_{ta} \otimes \begin{bmatrix} H_{ap}^t \otimes L_s^t \otimes I_s^b & H_{ap}^b \otimes L_s^t \otimes I_s^b \\ 0 & 0 \end{bmatrix}, \quad \widehat{L}_s^b = I_{ta} \otimes \begin{bmatrix} 0 & 0 \\ H_{ap}^b \otimes I_s^t \otimes L_s^b & H_{ap}^b \otimes I_s^t \otimes L_s^b \end{bmatrix}$$

3.4. Finite Approximation

A simple finite approximation for this departure process can be constructed by truncating to a finite buffer. The problem with such a finite buffer is that the buffer size must be sufficiently large in order to accurately capture both the marginal distribution and the correlation structure of the departure process of the infinite queue under high load. Green provides an approximation for the departure process of the MAP/PH/1 queue is introduced by letting the sojourn time of the last state be equal to the busy period of the tail of the infinite queue [6]. We modified this algorithm to include MEPs and to be able to separate the tagged and background departures. The MEP descriptors $\langle B_d, L_d \rangle$ for the departing tagged stream of the finite approximation are shown below

$$B_d = \begin{bmatrix} \ddots & & & & & \\ \ddots & \ddots & & & & \\ \ddots & \ddots & \widehat{L}_{ta} & & & \\ \widehat{L}_s^b & (\widehat{B}_{ta} + \widehat{B}_s) & E_0 & & & \\ & E_2^b & E_1 - E_3^b & & & \end{bmatrix}, \quad L_d = \begin{bmatrix} \ddots & \ddots & \ddots & \cdots & 0 \\ 0 & \cdots & \widehat{L}_s^t & 0 & 0 \\ 0 & 0 & \cdots & E_2^t & E_3^t \end{bmatrix}$$

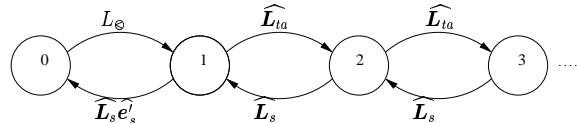


Fig. 3. Markov state diagram for the approximation

where $E_0 = \widehat{L}_{ta}(e'_{ta} \otimes I_s)$, $E_1 = B_s$, $E_2^t = L_s^t e'_s \mathbf{y}_{k-1}$, $E_2^b = L_s^b e'_s \mathbf{y}_{k-1}$, $E_3^t = L_s^t (1 - \mathbf{y}_{k-1} e')$ and $E_3^b = L_s^b (1 - \mathbf{y}_{k-1} e')$.

The vector e' is a summing vector of the appropriate dimension. The vector \mathbf{y}_{k-1} is defined as the unconditional distribution vector of the return to level $(k - 1)$, Appendix(A). Note that the expression for B_d in equation (13) is the truncated version of equation (13). The resulting model is one that approximates an infinite queue where correlated superimposed arrival streams share a server in a FCFS fashion. The approximation combines essentially both heavy traffic and light traffic models, and does surprisingly well when the operating range is set to medium ($\rho = 0.4$ through 0.7). Extensive numerical experimentation shows that this approximation works well.

4. Numerical Analysis and Discussion

The model developed in the previous section allows us to study first- and second-order characteristics of an isolated stream in a network sharing a single server with a background service class. In this section, we present numerical results for packets of a tagged service class multiplexed with packets of a background service class having correlated arrival times. In order to quantify the impact caused by the background classes of traffic, we also present results for the queueing behavior of the departing tagged class when passing through a second queue. We perform extensive numerical experiments to explore various combinations and present the interesting and insightful results, while discussing trends and special cases.

4.1. Experimental Setup

In order to perform parametric studies and clearly understand the distortions that the tagged traffic experiences, we use Poisson distributed arrivals and exponential service times for the tagged class. The mean service rate ratio, defined as α , is $\alpha = \frac{\mu_t}{\mu_b}$. The parameter k [6], representing the size of the finite approximation for the first queue is fixed at $k = 15$. The mean arrival rate of the tagged and background classes is the same ($\lambda_t = \lambda_b = 5.0$) and the utilization at the second queue is fixed at $\rho_2 = 0.8$. For the first queue, the tagged class service rate and the background class service rate are determined based on the value of α and ρ_1 , where ρ_1 is the utilization of the first queue. We have exercised the model extensively and present here the numerical results only for interesting and typical cases. In particular, we want to study the impact of the arrival time distributions of the background traffic at a multiplexer. In order to better evaluate the distortions in the tagged service class departing from a multiplexer, we also pass the departing stream into an exponential server with an infinite queue. The CSIM simulation package was used to validate our model.

4.2. Impact of the Background Stream with short range correlations

We characterize the background stream distribution using their first- and second-order characteristics. We specify the mean rate, λ , and squared coefficient of variation, c^2 , for marginals and the correlation, $r[1]$, as the second-order parameter in the construction of the arrival processes. We construct the process in such a manner that we have independent control over c^2 and r . We use the hyper exponentials with balanced means to characterize the first two moments of the background stream. Such a distribution can be represented in LAQT as

$$\mathbf{p} = [p \quad 1 - p], \quad \mathbf{B} = \lambda \begin{bmatrix} 2p & 0 \\ 0 & 2(1 - p) \end{bmatrix}, \quad \mathbf{L} = \mathbf{B}e'\mathbf{p} \quad (13)$$

where p is defined by $p = \frac{1}{2} + \frac{1}{2} \sqrt{\frac{c^2 - 1}{c^2 + 1}}$. This process is an uncorrelated sequence. In order to construct correlated processes with geometrically decaying covariances that share the same marginals, we use the approach presented in [16]. Define $\mathbf{L}^{(\gamma)}$ for $-1 < \gamma < 1$ as

$$\mathbf{L}^{(\gamma)} = (1 - \gamma)(\mathbf{B}e'\mathbf{p} - \mathbf{B}) + \mathbf{B}. \quad (14)$$

The $\mathbf{L}^{(\gamma)}$ constructed introduces geometrically decaying correlations in the process, while leaving the marginals invariant. Now, we have

$$r[k] = \frac{\mathbf{p}V^2e' - (\mathbf{p}Ve')^2}{2\mathbf{p}V^2e' - (\mathbf{p}Ve')^2} \gamma^k. \quad (15)$$

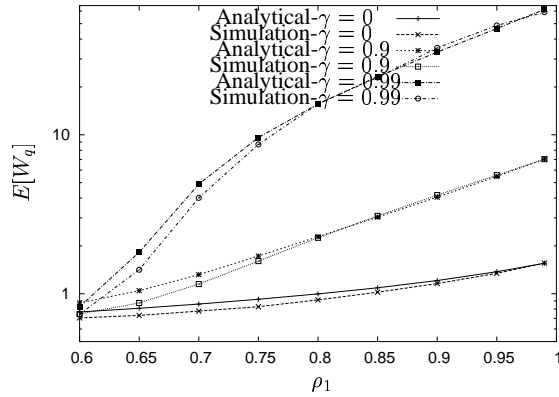


Fig. 4. $E[W_q]: c_b^2 = 4, \alpha = 4$

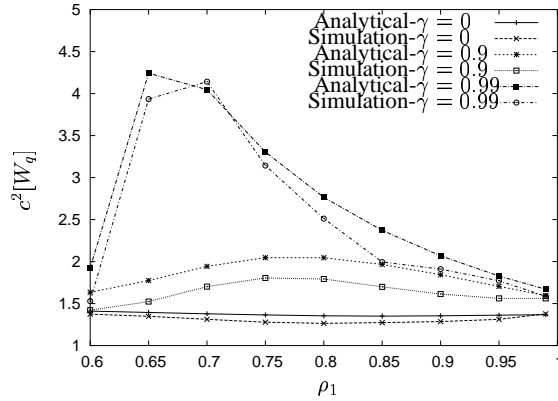


Fig. 5. $c^2[W_q]: c_b^2 = 4, \alpha = 4$

This formulation enables us to vary either the covariance decay or the c^2 of a process (or both). The arrival processes for the background class are constructed using equations (13) and (14). We use the results from [8] to solve for the first two moments of the waiting times $G/M/1$ system.

4.2.1. Impact of the Background Stream on Varying Loads

The characteristics of the tagged departing stream are described in terms of its marginal distribution and its correlation structure. In this section we consider a background class with $\gamma \in \{0, 0.9, 0.99\}$, the squared coefficient of variation $c^2 = 4$, and $\alpha = 4$, resulting in 4 tagged packets for every background packet. The service distribution of the background class was assumed to be Erlang-2 and we observe the impact of multiplexing by looking at the first two moments of the waiting time in the second queue for varying loads. In Fig. 4 we observe the mean waiting time in the second queue. It is interesting to note the significant role the correlations of the background class plays in affecting the mean waiting time as the utilization of the first queue moves toward heavy load.

In Fig. 5, we observe the squared coefficient of variation of the waiting time in the second queue for varying load. The greater the correlations in the background stream, the greater the squared coefficient of variation in the queueing time of the tagged stream.

4.2.2. Effect of Background Correlations on a Departing Tagged Stream

We now study the correlation structure of the departing tagged class. We consider a background class $\gamma \in \{0, 0.9, 0.99\}$ and $\alpha = 1$, thus resulting in equal number of tagged and the background packets. The utilization of the first queue was fixed at $\rho_1 = 0.9$.

In Fig. 6, we compare the waiting times to understand the impact of distortions introduced by the background stream on the tagged stream for varying squared coefficient of variation of the background arrival process. It is interesting to note the differences in waiting times experienced by a tagged stream

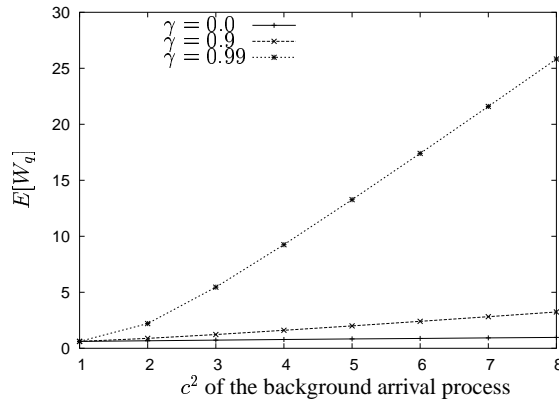


Fig. 6. $E[W_q]: \rho_1 = 0.9, \alpha = 1$

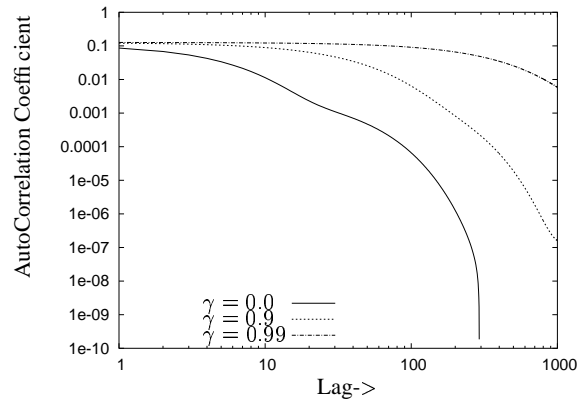


Fig. 7. Departure lag: $\rho_1 = 0.9, \alpha = 1, c_b^2 = 8$

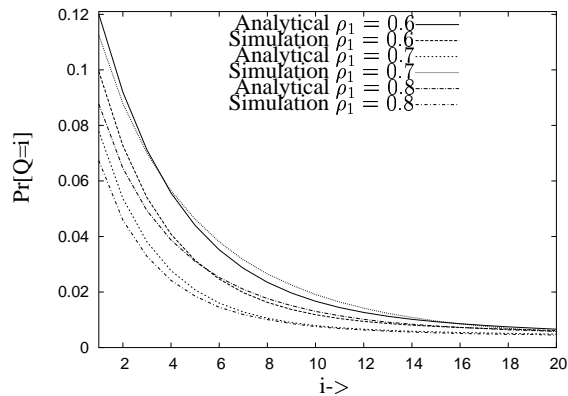


Fig. 8. $E[W_q]$: $\alpha = 4$

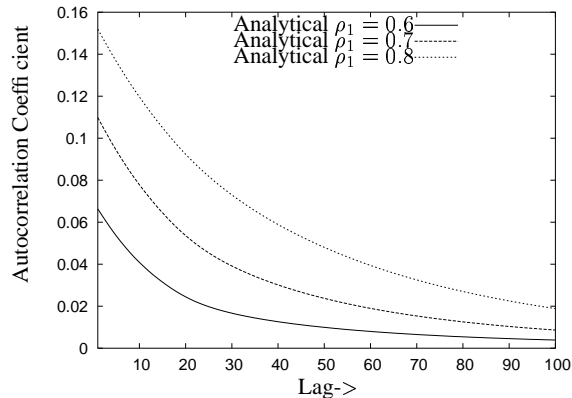


Fig. 9. Lag correlations : $\alpha = 4$

that is multiplexed with a background stream of $\gamma = 0.9$ and $\gamma = 0.99$ for higher c^2 , even when the difference between the lag1 correlations are almost negligible. This can be explained by observing Fig. 7, where we have plotted the first 1000 lags of the departure correlations of the tagged stream when the c^2 of the background class is fixed at 8 for $\gamma \in \{0, 0.9, 0.99\}$. The decay in the correlation structure is almost negligible in the case of $\gamma = 0.99$ when compared with $\gamma = 0$ or $\gamma = 0.9$ resulting in significant increase in the mean waiting time.

4.3. Impact of the Background Stream with long range correlations

We characterize the background stream using the N-Burst model [21] which is a superposition of N-independent ON/OFF sources with ON time distributions having ME representations and OFF periods distributed exponentially. Each source generates packets according to poisson distribution during the on periods. For our studies we used the truncated power tail distribution described in [7] for the ON periods which is defined by

$$R(x) = \frac{1 - \theta}{1 - \theta^M} \sum_{i=0}^{M-1} \theta^i \exp\left(-\frac{x}{\gamma^i}\right) \quad (16)$$

The parameters for the truncated power tail distribution were fixed at $\theta = 0.5$, $\gamma = 2$, $M = 2$. The number of sources for our experiment was set at 2. The mean of the background stream was fixed at $\lambda_b = 5$ and the ratio of the tagged to background packets was fixed at $\alpha = 4$. The service for both the background and the tagged streams were assumed to be exponential with service rates determined by α . Fig. 9 we compare the queuelength distribution at the second queue for utilization at the first queue $\rho \in \{0.6, 0.7, 0.8\}$. Fig. 8 we observe the first 100 lag correlations of the departing tagged stream for loads in the first queue $\rho \in \{0.6, 0.7, 0.8\}$. Observe the slowly decaying correlations induced by the background stream in to the tagged stream even when the region of operation of the sytem is not heavily loaded.

5. Conclusion and Future Work

In this paper we have captured the effect of multiplexing a tagged stream with a correlated background stream and the region of operation around where the coupling can have significant effect on the departing tagged stream. First we build a model based on Linear Algebraic Queueing principles where the arrival processes involved can be non-exponential and correlated. Using this model, we show the operational range where significant correlations are introduced by the background traffic when the arrival time distribution for the background class deviates from those of the tagged class. This can have a large impact on QoS in IP access networks. We also show how severe this distortion can be by looking at the first two moments of the waiting time in the second queue. Moreover, this model has the inherent ability for extending the analysis to N node networks. Further research is being conducted on extending this model where the service process of the individual streams can be correlated. Additionally further studies are also conducted on a finite approximation different from the one we have used in this paper.

REFERENCES

1. N. Benameur, S. B. Fredj, S. Oueslati-Boulahia, and J. W. Roberts, "Quality of service and flow level admission control in the Internet," *Computer Networks*, vol. 40, pp. 57–71, 2002.
2. T. Bonald, P. Olivier, and J. Roberts, "Dimensioning of high speed IP access networks," in *ITC 18*, (Berlin, Germany), Sept. 2003. to appear.
3. M. Conti, E. Gregori, and I. Stavrakakis, "Study of the impact of temporal/spatial correlations on a tagged traffic stream," in *Data Communications and their Performance (S. Fdida and R. Onvural, eds.)*, pp. 194–208, Chapman & Hall, 1996.
4. G. de Veciana, C. Courcoubetis, and J. Walrand, "Decoupling bandwidths for networks," *IEEE/ACM Transactions on Networking*, 1997.
5. D. Eun and N. Shroff, "Simplification of network analysis in large bandwidth systems," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, 2003.
6. D. Green, "Lag correlations of approximating departure processes of MAP/PH/1 queues," in *3rd Int. Conf. on Matrix-Analytic Methods in Stochastic Models*, pp. 135–151, 2000.
7. M. Greiner, M. Jobman, L. Lipsky, "The Importance of Power-Tail Distributions for Modeling Queueing Systems," *Operations Research*, vol. 47, no. 2, pp. 313–326, 1999.
8. A. Heindl, *Traffic-based Decomposition of General queueing Networks with Correlated Input Processes*. PhD thesis, Technical University Berlin, 2001.
9. W.-C. Lau and S.-Q. Li, "Traffic distortion and inter-source cross-correlation in a high speed integrated networks," *Computer Networks and ISDN Systems*, vol. 29, 1997.
10. Y. Lee, A. van de Liefvoort, and V. Wallace, "Modeling correlated traffic with generalized IPP," *Performance Evaluation*, vol. 40, pp. 99–114, 2000.
11. L. Lipsky, *Queueing theory: A linear algebraic approach*. New York: MacMillan, 1992.
12. L. Lipsky, P. Fiorini, W. Hsin, and A. van de Liefvoort, "Auto-correlation of lag-k for customers departing from semi-Markov processes," Tech. Rep. TUM-19506, Technical University Munich, 1995.
13. W. Matrangli, C. Bisdikian, and K. Sohraby, "Jitter calculus in ATM networks. multiple node," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, 1994.
14. W. Matrangli, C. Bisdikian, and K. Sohraby, "Jitter calculus in ATM networks. single node," in *Proceedings of the Conference on Computer Communications (IEEE Infocom)*, 1994.
15. M.D'Ambrosia, "Evaluating the limit behavior of the ATM traffic within a network," *IEEE/ACM Transactions on Networking*, vol. 3, no. 6, pp. 832–841, 1995.
16. K. Mitchell, "Constructing correlated sequence of matrix exponentials with invariant first-order properties," *Operations Research Letters*, vol. 28, no. 1, pp. 27–34, 2001.
17. K. Mitchell, A. van de Liefvoort, and J. Place, "Correlation properties of the token leaky bucket departure process," *Computer Communications*, vol. 21, 1998.
18. K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the differentiated services field (DS field in the IPv4 and IPv6 headers)." IETF rfc 2474.
19. Y. Ohba, M. Murata, and M. Miyahara, "Analysis of interdeparture processes for bursty traffic in ATM networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, pp. 468–476, 1991.
20. A. Riedl, M. Perske, T. Bauschert, and A. Probst, "Dimensioning of IP access networks with elastic traffic," in *9th International Telecommunication Network Planning Symposium*, (Toronto), Sept. 2000.
21. H. Schwefel, *Performance Analysis of Intermediate Systems Serving Aggregated ON/OFF Traffic with Long-Range Dependent properties*. PhD thesis, Technical University Munich, 2000.
22. S. Srivastava, K. Mitchell, and A. van de Liefvoort, "Correlations induced in a packet stream by background traffic in a multiplexed environment," in *ITC 18*, vol. 5a, (Berlin, Germany), Sept. 2003.
23. I. Stavrakakis, "Efficient modeling of merging and splitting processes in large network structure," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 8, pp. 1336–1347, 1991.
24. J. van den Berg, E. Diks, J. Resing, and J. van der Wal, "The change of traffic characteristics in ATM networks," in *Proceedings B-ISDN Teletraffic Modelling Symposium*, (Antwerp), Feb. 1995.
25. S. Wittevrongel and H. Bruneel, "Analytic study of the queueing performance and the departure process of a leaky bucket with burst input traffic," *AE*, vol. 50, no. 1, pp. 1–10, 1996.

A. Approximation for the departing tagged stream

We discuss in this section the process involved in the computation of the vector y_{k-1} for the finite approximation of the departure process. The vector y_{k-1} defined in [6] as the unconditional distribution of the return phase at level $(k-1)$ is given by

$$y_{k-1} = \frac{x_{k-1}}{\sum_{j=k-1}^{\infty} x_j e^j} \quad (17)$$

where x_{k-1} is defined as distribution of the QBD process (shown in Fig. 3) at level $k-1$ conditional on a departure having just occurred, and calculated from the stationary distribution of the QBD obtained by solving equations (10), (11), and (12). The unique stationary distribution of the QBD is given by

$$\Psi = \pi_0 [R_0, R_0 R, R_0 R^2, \dots] \quad (18)$$

$$x_{k-1} = \begin{cases} \pi_0 R_0 B_2 (p_a L_a e'_a)^{-1} & k = 1 \\ \pi_0 R_0 R^{k-1} A_2 (p_a L_a e'_a)^{-1} & k > 1 \end{cases} \quad (19)$$

where p_a is the stationary vector for the arrival process at embedded arrival epochs. The matrix R is the minimal non-negative solution to the matrix quadratic equation

$$\sum_{i=0}^2 R^i A_i = 0. \quad (20)$$

The matrices A_0, A_1, A_2 are given by

$$A_0 = -\widehat{L}_{ta}, \quad A_1 = (\widehat{B}_{ta} + \widehat{B}_s), \quad A_2 = -(\widehat{L}_s^t + \widehat{L}_s^b) \quad (21)$$

Numerous methods can be found for the efficient solution of this equation.