Computer Communications 22 (1999) 1382–1391

# On the time scales in video traffic characterization for queueing behavior

H. Ahn[a,*], A. Baiocchi[b], J.-K. Kim[a]

[a]Visual Communications Laboratory, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Taejon, 305-701, South Korea

[b]Department of INFOCOM, University of Roma La Sapienza, Via Eudossiana 18 00184 Rome, Italy

Received 4 January 1999; received in revised form 13 May 1999; accepted 14 May 1999

**Abstract**

To guarantee quality of service (QoS) in future integrated service networks, traffic sources must be characterized to capture the traffic characteristics relevant to network performance. Recent studies reveal that multimedia traffic shows burstiness over multiple time scales and long range dependence (LRD). While researchers agree on the importance of traffic correlation, there is no agreement on how much correlation should be incorporated into a traffic model for performance estimation and dimensioning of networks.

In this article, we present an approach for defining a relevant time scale for the characterization of VBR video traffic in the sense of queueing delay. We first consider the Reich formula and characterize traffic by the *Piecewise Linear Arrival Envelope Function* (PLAEF). We then define the *cutoff interval* above which the correlation does not affect the queue buildup. The cutoff interval is the upper bound of the time scale which is required for the estimation of queue size and thus the characterization of VBR video traffic. We also give a procedure to approximate the empirical PLAEF with a concave function; this significantly simplifies the calculation in the estimation of the cutoff interval and delay bound with little estimation loss.

We quantify the relationship between the time scale in the correlation of video traffic and the queue buildup using a set of experiments with traces of MPEG/JPEG-compressed video. We show that the critical interval, i.e. the range for the correlation relevant to the queueing delay, depends on the traffic load: as the traffic load increases, the range of the time scale required for estimation for queueing delay also increases. These results offer further insights into the implication of LRD in VBR video traffic. 

*Keywords:* VBR video traffic; Traffic characterization; Cutoff interval; Deterministic bound; Multiple time scales; Long-range dependence; Queueing delay

## 1. Introduction

Resource allocation is necessary for a network that guarantees quality of service (QoS). To achieve this QoS guarantee, we must characterize traffic sources so that we can predict the QoS and allocate network resources efficiently. In general, the focus of traffic characterization and modeling lies not in precise modeling of traffic statistics, but rather in capturing the relevant characteristics so as to allow prediction and efficient management.

Variable bit-rate (VBR) compressed video traffic is expected to be a significant component of the traffic mix in integrated service networks. However, the burstiness and delay sensitivity of VBR video poses a severe challenge.

Recent research has demonstrated that a compressed video source shows burstiness over multiple time scales [1–3,35]. VBR video traffic shows burstiness not only over a period of milliseconds to seconds, corresponding to variations within a scene, but also over a period of tens of seconds to minutes, corresponding to scenes with differing information contents.

Recently, it has been reported that VBR video traffic exhibits long range dependence or asymptotically second-order self-similarity [1,4,5], in the sense that the autocorrelation function $\rho(k)\,(k = 1, 2, \ldots)$ has an infinite sum $\sum_k \rho(k) = \infty$. The relevance of LRD properties with cell loss and delay was studied in Refs. [6,7], and it was argued that there is a critical time scale that has a significant impact on queueing behavior.

A rich set of literature exists on characterizing VBR video traffic with a stochastic process, for example Poisson, Markov modulated, autoregressive, TES, and self-similar (see Ref. [8], and references therein; more recently, see Refs. [2,9]). Such stochastic models of sources have the advantage that they may be used to achieve potentially higher network utilization by exploiting the statistical

* Corresponding author. Tel.: + 82-42-869-5418; fax: + 82-42-869-8018.

*E-mail addresses:* cityboy@viscom.kaist.ac.kr (H. Ahn), andrea@infocom.ing.uniroma1.it (A. Baiocchi), kimjk@ee.kaist.ac.kr (J.-K. Kim)

0140-3664/99/$ - see front matter 
PII: S0140-3664(99)00119-X

properties of the sources. First- and second-order statistics, i.e. rate distribution and autocorrelation (equivalently, the power spectrum) are considered to be the most important characteristics [8,10,11]. However, for network performance estimation, approaches on two extreme conditions are now available. One is based on the bufferless multiplexor model, where the tail distribution of the aggregate rate of multiplexed input is crucial to cell loss probability. The Chernoff bound is one of the most effective methods for call admission control (CAC) under this assumption [8,12]. The other approach, the effective bandwidth (EB) theory, assumes a very large buffer condition, although it considers the fluctuation of traffic in time scales [8]. However, traffic characterization and CAC for a moderate buffer size is not yet developed. Moreover, it is difficult to implement a policing mechanism based on stochastic characterization and give significant bounds of performance measures in the network.

For these reasons, in this work we consider *deterministic* traffic characterization and its relationship with network performance. Recently, traffic control based on regulated sources has begun to receive increased attention [8,13–24]. Thus, ITU-T and ATM Forum [25,26] define the rule-based and deterministic traffic contract descriptors: peak cell rate (PCR), cell delay variation tolerance (CDVT), sustainable cell rate (SCR), and maximum burst size (MBS). Further, they define the generic cell rate algorithm (GCRA) for conformance testing of these parameters. An exhuastive survey and comparison of previous studies on guaranteed performance service disciplines with constrained burstiness sources is found in Ref. [27]. Especially in Refs. [19,22,23], traffic sources are regulated according to the leaky bucket constraint. Generalizations of this approach have recently been defined in Refs. [14,15,17,18]. By constraining the burstiness of input sources, the network can provide performance guarantees such as buffer backlog and delay bounds. In Refs. [28,29], a set of specific traffic characterization methods for VBR video sources are presented and compared with previous methods.

In this article, we concentrate on the time-scale problem in characterization of VBR video traffic: over what range does the traffic correlation have to be considered for estimation of queueing behavior? We first consider the Reich formula and characterize traffic by the piecewise linear arrival envelope function (PLAEF). We then introduce the concept of the *cutoff interval* above which the correlation of input traffic does not affect queue buildup. Since the results are based on deterministic characterization, the cutoff interval provides a fundamental bound of the time scale relevant to the estimation of queue size in the sense that it is free from the possible ambiguity in stochastic approaches for the same issue. The results can be used to determine the meaningful range of correlation for the characterization of VBR video traffic. We quantify the relationship using a set of experiments with compressed video traces by approximating the actual traffic profile with the PLAEF. From mathematical results and simulations, we show that the critical interval depends on the traffic load: as the traffic load increases, the range of interval required for estimation of queueing delay also increases.

The study on the time scale relevant to queueing behavior is motivated by the frequently asked question whether LRD in VBR video traffic is crucial for traffic engineering. Our results also offer another insight into the implication of LRD in VBR video traffic.

In addition, we propose an algorithm that approximates the empirical PLAEF to a concave PLAEF, which significantly simplifies the parameter measurement and estimation of cutoff interval and delay bound with little estimation loss.

The rest of this article is organized as follows. In Section 2, we introduce the PLAEF. In Section 3, we investigate the calculation of delay bound and cutoff interval for a source characterized by the PLAEF. In Section 4, we present numerical results of the application of the cutoff interval concept, with reference to VBR video traffic sequence. In Section 5, we present an algorithm that approximates the empirical PLAEF to a corresponding concave function. In Section 6, we conclude our work and suggest further studies.

## 2. The PLAEF: a deterministic traffic characterization

Consider a multiplexer with an output link capacity $C$ and an infinite buffer, loaded by $N$ virtual connections (VC). Let $v_i(s,t)$ denote the number of cell arrivals of $i$th VC in time $[s,t), i = 1, 2, \ldots, N$ and $v(s,t)$ the overall number of cells arrived at the multiplexer input in the same time interval. By the Reich formula [30]

$$Q(t) = \sup_{s \le t} \{v(s,t) - C(t-s)\}, \tag{1}$$

where $Q(t)$ denotes the multiplexor buffer occupancy at time $t$. If there exists a bounding function $f(u)$ such that for any $t$ and any $s \le t$, it can be written $v(s,t) \le f(t-s)$, then it follows that

$$Q(t) \le \sup_{s \le t} \{f(t-s) - C(t-s)\} = \sup_{u \ge 0} \{f(u) - Cu\}. \tag{2}$$

Inequality 2 implies that the delay bound is determined by the worst case arrival profile, which may be either for a single traffic $f_i(u)$ or a multiplexed input traffic $f(u)$. The bounding function $f(u)$ can be derived from either constraints imposed on the traffic source at the network ingress or real traffic traces. We resort to the second approach and choose $f(u)$ in the family of piecewise linear functions (i.e. first order splines), hence the name the Piecewise Linear Arrival Envelope Function (PLAEF). The aim in the construction of the PLAEF is using Eq. (2) to understand the relationship between queueing performance of video traffic multiplexing and the relevant time scale characterization of video traffic source models.

A few deterministic characterizations of VBR video

Table 1
Summary of the VBR traces used in the study

| Name | Video source | Length (frames) | Compression algorithm | Peak bit rate (bits/frame) | Mean bit rate (bits/frame) |
|---|---|---|---|---|---|
| JPEG | Star Wars movie [1] | 174,000 | Intra-frame | 627,672 | 27,791 |
| MPEG | Star Wars movie [1] | 174,000 | MPEG-1 | 185,267 | 15,601 |
| ATP | Tennis game [31] | 40,000 | MPEG-1 | 190,856 | 21,890 |
| MTV2 | Video clips [31] | 40,000 | MPEG-1 | 251,408 | 19,780 |
| Talk1 | Talk show [31] | 40,000 | MPEG-1 | 106,768 | 14,537 |
| Simpsons | Cartoon [31] | 40,000 | MPEG-1 | 240,376 | 18,576 |
| Settop | Video phone type [31] | 5000 | MPEG-1 | 46,200 | 6031 |

traffic were proposed in Ref. [28]. Some of them are equivalent to each other and each model has its advantages in accuracy, policing, easiness for parametrization, or simplicity. The PLAEF directly models the rate profile curves so that it is very accurate in delay estimation and easy to obtain from the real traces. Furthermore, it has explicit parameters for the time scale of burstiness and correlation of input traffic and thus it is suitable for explaining the effect of the time scale. In the PLAEF, $f(u)$ is obtained as follows. First, we define $P$ interval-arrival pairs (IAPs)

$$\{(t_i, a_i) | t_i < t_{i+1},\ a_i < a_{i+1},\ (t_0, a_0) = (0,0),\ i = 0,1,2,P-1\}, \tag{3}$$

where $a_i$ is the maximum amount of data over any time interval of duration $t_i$. The PLAEF is defined by the IAPs as the arrival envelope function $f(u)$ such that

$$v(s,t) \le f(t-s) = a_i + \frac{a_{i+1} - a_i}{t_{i+1} - t_i}(t - s - t_i),\ t_i \le t - s < t_{i+1},\ i = 0,1,2,\ldots,P-1. \tag{4}$$

For numerical results, we use two differently coded *Star Wars* traces [1] and 5 MPEG-coded traces [31]. We

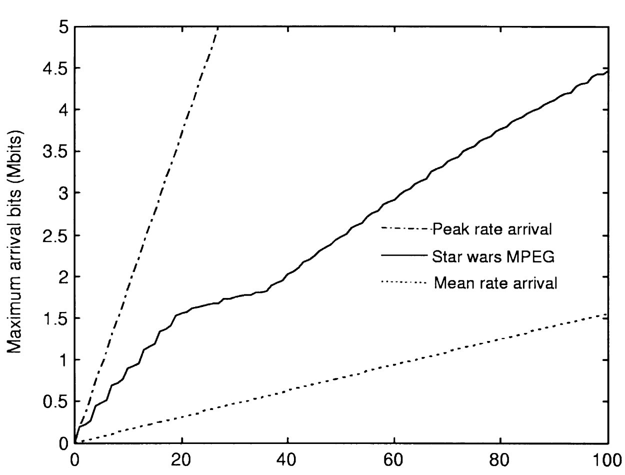


Fig. 1. The PLEAF, mean arrival, peak arrival functions for the MPEG trace.

summarize only the salient characteristics of the traces in Table 1.

We call the intraframe coded *Star Wars* trace the *JPEG trace*, even though its coding is not compatible with the JPEG standards. The JPEG algorithm also uses only the intraframe redundancy, and thus we expect that the JPEG trace show characteristics and results similar to a JPEG-coded trace. The other traces are coded according to the MPEG-1 video compression standards with the GOP structure IBBPBBPBBPBBP, where: (i) I denotes an intra-frame coded frame (i.e. without motion compensation); (ii) P denotes a predictive frame, where only the prediction error is coded and prediction is based on adjacent I or P frames; (iii) B denotes Interpolative frames, whose coding is based on interpolation between adjacent I and P frames.

Let $X(i)$, $i = 1,2,\ldots$ be the amount of bits generated in the $i$th frame. Because we have the data in units of bit per frame, we assume that the sources generate the data evenly over the frame interval $T$.[1] For numerical results, we assume all $T = 1/25$ s. The arrival function for a video trace $v(s,t)$ is given by

$$v(s,t) = \left(\left\lceil \frac{s}{T} \right\rceil - \frac{s}{T}\right) X\left(\left\lceil \frac{s}{T} \right\rceil\right) + \sum_{i=\lceil s/T \rceil + 1}^{\lfloor t/T \rfloor} X(i) + \left(\frac{t}{T} - \left\lfloor \frac{t}{T} \right\rfloor\right) X\left(\left\lceil \frac{t}{T} \right\rceil\right) \tag{5}$$

where $\lceil x \rceil$ denotes the ceiling of $x$, i.e. the least integer not less than $x$ and $\lfloor x \rfloor$ stands for the floor function of $x$, i.e. the largest integer not greater than $x$. The values of the IAPs for the PLAEF are obtained by

$$t_i = iT,\ a_i = \max_{s \ge 0} v(s, s + t_i),\ \ i = 1,2,3,\ldots. \tag{6}$$

This is a first order spline approximation of the minimal arrival curve defined in Ref. [14], i.e. the function

$$\gamma(t) = \sup_{s \ge 0} v(s, s+t). \tag{7}$$

In Ref. [14] it is shown that $\gamma(t)$ is the least function that upper bounds $v(s,t)$ for any $s$ belonging to the interval [0,$t$] and for any $t \ge 0$.

[1] We have the amount of bits in the unit of the slice for the JPEG trace, but not for the MPEG traces.

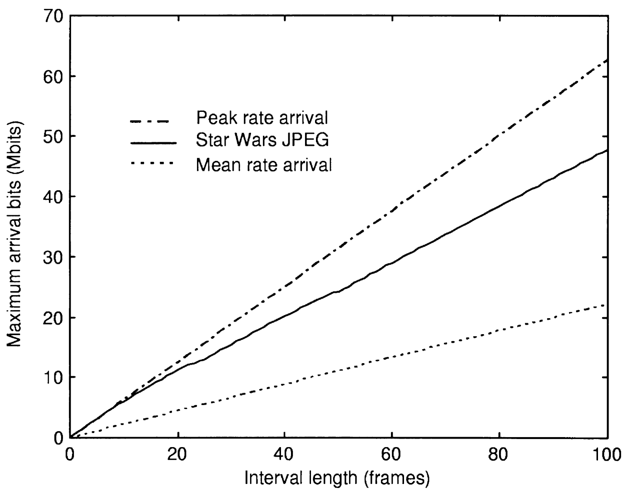


Fig. 2. The PLAEF, mean arrival, peak arrival functions for the JPEG trace.

The fact that the maximum value in Eq. (6) is obtained at points of the type $s = jT$ for the video traffic can make the computation easier. If we let $m(i)$ be the index such that the maximum of $v(s, s + t_i)$ is achieved at $s = m(i)T$, then

$$a_i = \sum_{k=m(i)+1}^{m(i)+i} X(k), \quad i = 1, 2, \ldots, P. \tag{8}$$

Figs. 1 and 2 show the PLAEFs for the MPEG trace and the JPEG trace, respectively. The PLAEF for the MPEG trace without any smoothing algorithm shows a periodic pattern due to the GOP structure of its coding algorithm and thus the PLAEF is not a concave function. In contrast, the PLAEF for the JPEG trace looks like a concave function (in fact, it is not a concave function, either). The concavity of the PLAEF makes the calculation of the delay bound and cutoff interval easy. We will consider this in Section 5.

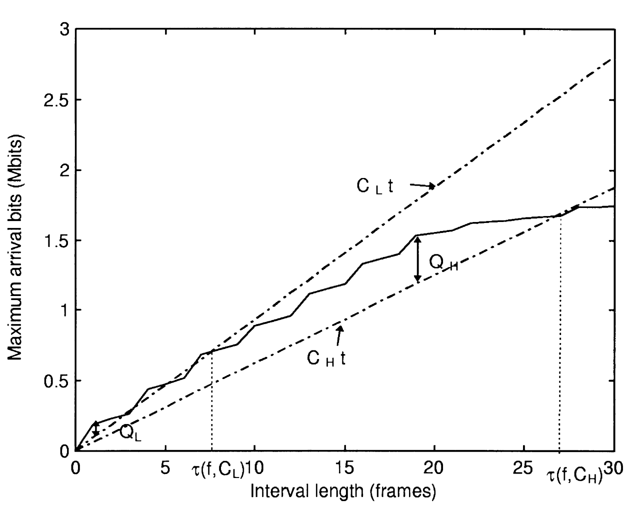


Fig. 3. The illustration of the cut off time, comparing heavy load and light load condition ($C_H < C_L$).

## 3. Range of correlation relevant to delay bound estimation

We now consider the relationship between the PLAEF and delay bound estimation. We here assume first-come-first-served (FCFS).

Fig. 3 shows clearly how the PLAEF can be used to assess the worst-case backlog of a multiplexer fed by the corresponding arrival pattern over any given time scale. Let $C_H$ and $C_L$ denote the output link rates ($C_H < C_L$). The subscript H and L mean heavy load and light load, respectively. Although in a practical situation, the link rate is fixed, here we change the link rate $C$ equivalently for simplicity of explanation. In Fig. 3, we use the MPEG coded Star Wars trace, and one third and a half of the peak rate of the trace for $C_H$, and $C_L$, respectively. This also confirms the intuition that a backlog can be found only up to a critical time scale beyond which no further overload is to be expected. This leads to the following definition of the cutoff interval.

**Definition.** Let $f(t)$ be an arrival envelope curve of the traffic flow feeding a multiplexer with output capacity $C$; the cutoff interval $\tau = \tau(f, C)$ is given by

$$\tau(f, C) = \sup\{u | u \geq 0 \text{ and } f(u) \geq Cu\}. \tag{9}$$

The existence of a finite value of $\tau$ can be proved as follows. It can be assumed that $f(t)$ is subadditive and $f(0) = 0$. If this were not the case, $f(t)$ could be replaced with another arrival envelope curve, $g(t)$, namely the subadditive closure of $f(t)$, such that $g(t) \leq f(t)$ for any $t \geq 0$ and such that $g(t)$ is subadditive and $g(0) = 0$ [14,17]. In Ref. [16] it is shown that for the stability of a queue whose input is bounded by $f(t)$, it must be

$$\lim_{t\to\infty} \frac{f(t)}{t} = a < C, \tag{10}$$

provided that $f(t)$ is subadditive and wide-sense increasing. By the definition of limit, it follows that there exists a finite $t_0$ such that $f(t) < Ct$ for any $t > t_0$. This implies that the cutoff interval $\tau(f, C)$ is finite, as long as the considered multiplexer is stable. In the special case of the PLAEF as defined in Section 2, $f(t)$ is actually wide-sense increasing and subadditive and there exists a finite value for the *cutoff interval* or equivalently a *cutoff index* $r$ such that

$$t_{r-1} < \tau(\text{PLAEF}, C) \leq t_r. \tag{11}$$

The cutoff index $r$ can also be defined as the maximum index $r \geq 1$ such that

$$\frac{a_r}{r} \leq CT < \frac{a_{r-1}}{r-1}. \tag{12}$$

We give a few properties of the cutoff interval here:

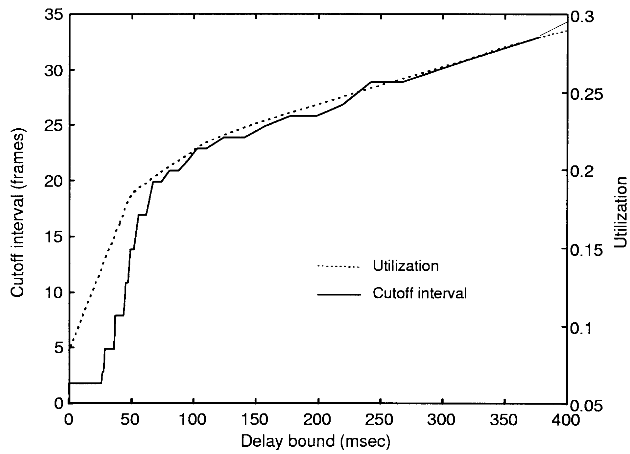


Fig. 4. The cutoff interval variation for the MPEG trace against the maximum delay bound.

**Property 1.** The queueing delay does not depend on the values of the PLAEF for $t_i > \tau(\text{PLAEF}, C)$.

**Proof.** Property 1 comes directly from the definition of the cutoff interval. □

**Property 2.** The cutoff interval is a decreasing function of $C$.

**Proof.** Property 2 comes from the fact that the PLAEF is a non-decreasing function:

$$\{t|f(t) \geq C_H t\} \supset \{t|f(t) \geq C_L t\} \text{ for } C_H < C_L. \tag{13}$$

□

**Property 3.** The queueing delay is a decreasing function of $C$.

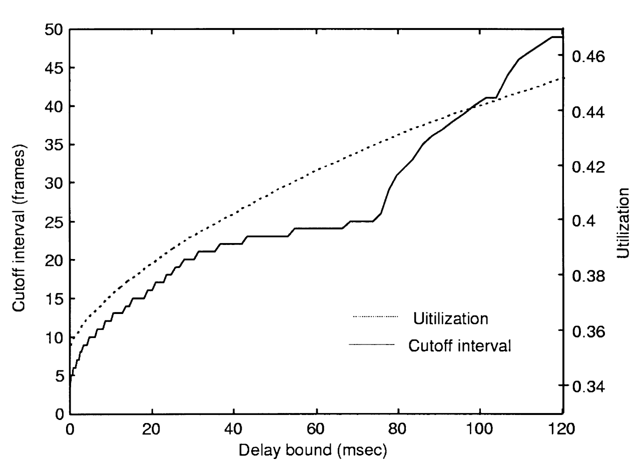


Fig. 5. The cutoff interval variation for the JPEG trace against the maximum delay bound.

**Proof.** The proof is also straightforward from the non-increasing property of the cutoff interval with respect to $C$:

$$\sup_{0 \leq t \leq \tau(f, C_H)} (f(t) - C_H t) \geq \sup_{0 \leq t \leq \tau(f, C_H)} (f(t) - C_L t)$$

$$\geq \sup_{0 \leq t \leq \tau(f, C_L)} (f(t) - C_L t). \tag{14}$$

□

Property 1 means that the range that is related to the delay of the queue is hard limited by the critical point $\tau(f, C)$ and thus a correlation structure longer than this time interval is of no importance at least in the estimation of queueing performance. Property 2 means the range of important correlation decreases as the link utilization increases.

The cutoff interval is in essence the largest value of the time scale for which the backlog in a queue receiving an input flow bounded by a function $f(u)$ and emptied with a capacity of $C$ can be non-negative. Then, if one considers such a queue and assumes that it is empty at time $t$, the worst case arrival pattern is such that the queue is empty again no later than $t + \tau(f, C)$. Hence, the cutoff interval $\tau(f, C)$ is an *upper bound* for the *duration of the queue busy period*. As such, it also represents the upper bound of delay experienced by any packet fed into the queue, regardless of the serving discipline. Chang [16] noticed such a thing: Here we develop a consistent framework that gives new insight to the remark in Ref. [16], by defining the concept of the cutoff interval and investigating its properties and applications.

## 4. Application of cutoff interval to VBR video traffic

This Section presents numerical results of the application of the cufoff interval concept, with reference to VBR video traffic sequences. Figs. 4 and 5 show cutoff intervals and link utilizations against the allowed maximum delay bound for a connection, $D$. The cutoff interval and the link utilization are obtained by calculating the minimum link capacity $C_{\min}$ that guarantees the maximum delay bound $D$. The minimum link capacity $C_{\min}$ is given by

$$C_{\min} = \sup_{u \geq 0} \left\{ \frac{f(u)}{u + D} \right\}, \tag{15}$$

and again the supreme value for the PLAEF is achieved at points of the type $u = jT$, $j = 0, 1, 2, \ldots$. The simulation results show that the possible link utilizations are rather low for the deterministic QoS guarantee: utilizations turn out to be 0.2 for the MPEG trace and 0.4 for the JPEG trace for a delay bound of 100 ms. The reason the utilization for the MPEG trace is lower than that for the JPEG trace is that the MPEG is more bursty than the JPEG. However, noting that the link rate is scaled by the input average, the number of MPEG connections admissible for the same link rate is larger than that of the JPEG connections. In addition, the cutoff interval curve shows a step-wise increase for a low

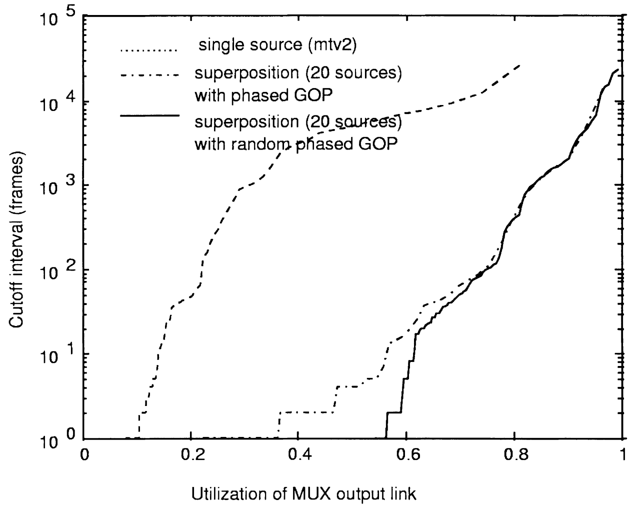


Fig. 6. The cutoff intervals versus the mean load of the multiplexer, for various input traffic mixes.

delay bound value. This is also due to the periodic burstiness of the GOP structure. The cutoff interval increases as the traffic intensity increases. Even with deterministic QoS, the cutoff interval is less than 50 frames for traffic loads of 0.5 (the JPEG trace) and 0.3 (the MPEG trace).

Fig. 6 shows the cutoff intervals derived from the PLAEF as a function of mean load for a multiplexer loaded with: (i) an MPEG-1 coded video trace representing video clips (MTV2), (ii) the superposition of 20 different MPEG-1 coded video traces with phased GOP, (iii) the superposition of the same 20 MPEG-1 coded video traces as in (ii) with random phasing. The 20 traces employed here are available and described in Ref. [31]. This figure shows clearly that the cutoff interval beyond which correlations do not matter for queueing performance can be very small (less than 10 frames, i.e. 400 ms) even for mean loads of 0.6, if the superposition of independently phased MPEG-coded video streams is considered.

Recently, LRD has been accepted as an inherent characteristic in VBR video traffic, and it has been reported that the LRD characteristic may impact queueing behavior significantly [32,33]. However, several studies have argued that LRD is not crucial in determining the queueing behavior of VBR video sources [5,7,12].

Our observations reflect these seemingly contradictory arguments: the LRD is also a characteristic in the large lag region, and the correlation in the large region is crucial for queueing performances under heavy loads but not for those under a light or moderate load condition.

The interpretation of the cutoff interval as an upper bound of the multiplexer busy period explains why the significance of LRD depends on the input traffic load. Since video traffic, as a component of multimedia stream within an interactive service, has real time constraints, the maximum delay allowed through the network (hence through any multiplexer of the network) must not exceed a quite limited quantity $D$ (i.e. it might be a few video frames, at most a few hundred ms). Since the cutoff interval is one (possibly rough) bound of the maximum delay through a queue, it follows that the cutoff interval cannot be very large. In contrast, this means that the correlations that matter for queueing performance under the cited delay constraint are rather limited and hence LRD could be of no importance, depending on the delay constraints.

We can infer that with probabilistic QoS in low and moderate utilization (i.e. a mean load of 0.4 to 0.5 or even 0.6 in the case of superposition of many independent video sources), Markovian models, such as the discrete autoregressive DAR [12,34] and the circulant Markovian modulated Poisson process (CMMPP) [6,10], will be enough for the estimation of buffer occupancy if they can model the correlation up to about 100 frames lag.

## 5. Approximation of PLAEF by a concave function

In this section, we consider a concave approximation algorithm of the original PLAEF and its advantages in calculation of performance estimation. Let

$$m_i = \frac{a_{i+1} - a_i}{t_{i+1} - t_i}, \quad i = 0, 1, 2, \dots, P - 1. \tag{16}$$

Then the PLAEF, defined by a set of IAPs $(t_i, a_i)$, $i = 0, 1, 2, \dots, P$, is a concave function if the sequence of the slope coefficients of the linear pieces $m_i$, $i = 0, 1, 2, \dots, P - 1$, is a non-increasing sequence. If each source has a rate profile that is a concave function, we can simplify the calculation significantly as follows.

**Property 4.** If the rate profile function $f(u)$ is a concave function, (i) the largest backlog $Q_{\max}$ is obtained by $f(t_k) - Ct_k = a_k - Ct_k$, where $k$ is the unique index such that

$$m_{k-1} \geq C \text{ and } m_k < C. \tag{17}$$

And (ii) the cutoff interval $\tau(f, C)$ and the cutoff index $r$ are uniquely determined by

$$\tau(f, C) = \frac{a_{r-1}t_r - a_r t_{r-1}}{C(t_r - t_{r-1}) - (a_r - a_{r-1})}, \tag{18}$$

and

$$a_{r-1} - Ct_{r-1} > 0 \text{ and } a_r - Ct_r \leq 0. \tag{19}$$

| |
|---|
| STEP 1: set $j = 0$, $k(j) = 0$, and $(\bar{t}_0, \bar{a}_0) = (0, 0)$. |
| STEP 2: $j = j + 1$. |
| STEP 3: $k(j) = \arg\max_{i > k(j-1)} (a_i - \bar{a}_{j-1})/(t_i - \bar{t}_{j-1})$ |
| STEP 4: $(\bar{t}_j, \bar{a}_j) = (t_{k(j)}, a_{k(j)})$. |
| STEP 5: if $k(j) < P$, go to STEP 2. |

Fig. 7. The approximation algorithm of an original PLAEF to a concave one.

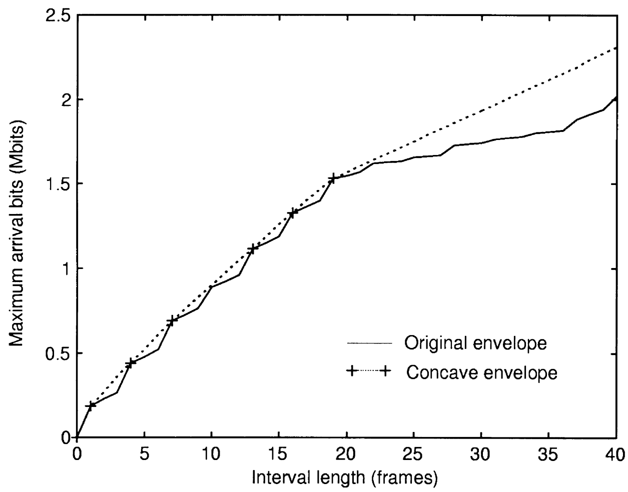


Fig. 8. The concave PLAEF and original PLAEF for the MPEG trace.

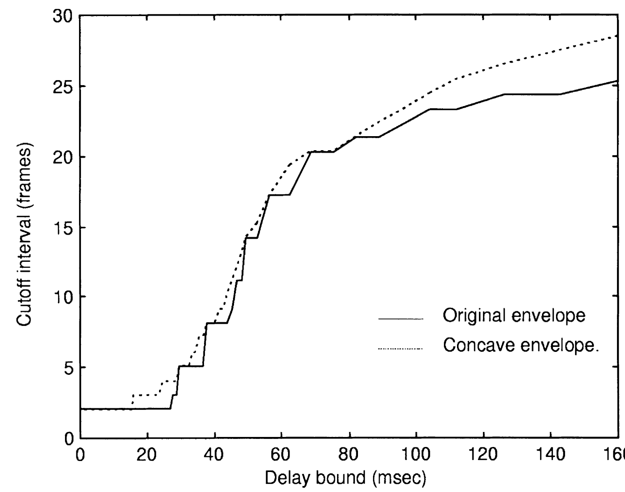


Fig. 10. The changes of cutoff intervals due to the concave approximation for the MPEG trace against the maximum delay bound.

**Proof.** The location of the maximum backlog is easily obtained if we consider that $f(u)–Cu$ is also piecewise linear so that it attains its maximum value at the break point $t_k$ from which the function begins to decrease. The cutoff interval is uniquely obtained by solving $f(t) = Ct$ because of the concavity of $f(u)$. In other words, the critical interval is where $f(t) - Ct$ changes the sign from positive to negative. Noting $t_{r-1} < \tau \leq t_r$ and the linearity in $(t_{r-1}, t_r]$, the equation for the cutoff interval is obtained. □

Property 4 means that for sources with a concave PLAEF, to obtain the largest backlog, we need only the arrival information $a_r$ up to values less than $Ct_r$ and the link rate $C$.

When the PLAEF is not a concave function, the backlogs at all time intervals $t_i$ have to be compared to obtain the maximum backlog. However, as shown in Section 3, the experimental PLAEF is not necessarily a concave function but just a non-decreasing function, especially in the MPEG trace. Therefore, in this section we define an algorithm to derive a concave PLAEF that approximates the empirical PLAEF and examine the difference of the cutoff interval estimation and the utilization resulting from the two functions. The algorithm is described in Fig. 7.

The concavity of the new PLAEF comes from

$$\frac{\bar{a}_{j+1} - \bar{a}_j}{\bar{t}_{j+1} - \bar{t}_j} < \frac{\bar{a}_{j+1} - \bar{a}_{j-1}}{\bar{t}_{j+1} - \bar{t}_{j-1}} < \frac{\bar{a}_j - \bar{a}_{j-1}}{\bar{t}_j - \bar{t}_{j-1}}. \quad (20)$$

Note also that the new concave PLAEF is an upper bound of the corresponding PLAEF since

$$\frac{a_i - \bar{a}_{j-1}}{t_i - \bar{t}_{j-1}} \leq \frac{\bar{a}_j - \bar{a}_{j-1}}{\bar{t}_j - \bar{t}_{j-1}} \text{ for } \bar{t}_{j-1} < t_i \leq \bar{t}_j \quad (21)$$

and thus

$$a_i \leq \bar{a}_{j-1} + \frac{\bar{a}_j - \bar{a}_{j-1}}{\bar{t}_j - \bar{t}_{j-1}}(t_i - \bar{t}_{j-1}) \text{ for } \bar{t}_{j-1} < t_i \leq \bar{t}_j. \quad (22)$$

In Figs. 8 and 9, we compare the results obtained by the

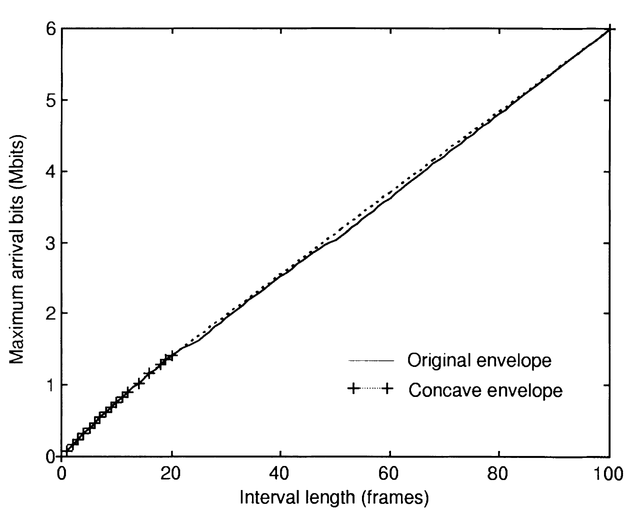


Fig. 9. The concave PLAEF and original PLAEF for the JPEG trace.

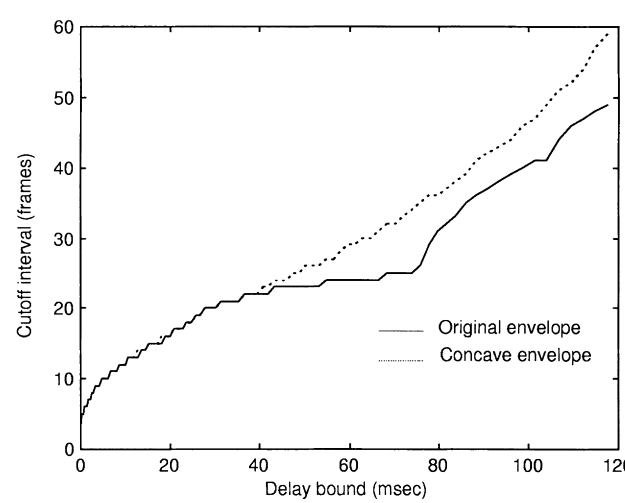


Fig. 11. The changes of cutoff intervals due to the concave approximation for the JPEG trace against the maximum delay bound.

Table 2
Simulation results for various video traces: cutoff intervals, $t_r, \bar{t}_r$ (frames), and link utilization

| Maximum delay (ms) | | JPEG | MPEG | MTV2 | ATP | Simpsons | Talk1 | Settop |
|---|---|---|---|---|---|---|---|---|
| | $t_r$ | 7 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1 | $\bar{t}_r$ | 7 | 2 | 2 | 2 | 2 | 2 | 2 |
| | G | 1.014 | 1.027 | 1.032 | 1.027 | 1.027 | 1.032 | 1.027 |
| | $t_r$ | 12 | 2 | 2 | 2 | 2 | 2 | 2 |
| 10 | $\bar{t}_r$ | 12 | 2 | 2 | 2 | 2 | 2 | 2 |
| | G | 1.052 | 1.260 | 1.258 | 1.252 | 1.253 | 1.253 | 1.260 |
| | $t_r$ | 16 | 2 | 3 | 2 | 3 | 2 | 2 |
| 20 | $\bar{t}_r$ | 17 | 3 | 4 | 3 | 3 | 3 | 2 |
| | G | 1.093 | 1.504 | 1.518 | 1.504 | 1.502 | 1.518 | 1.502 |
| | $t_r$ | 21 | 5 | 4 | 3 | 3 | 3 | 2 |
| 30 | $\bar{t}_r$ | 21 | 5 | 4 | 4 | 3 | 4 | 3 |
| | G | 1.093 | 1.751 | 1.553 | 1.770 | 1.758 | 1.767 | 1.776 |
| | $t_r$ | 22 | 8 | 15 | 5 | 3 | 5 | 3 |
| 40 | $\bar{t}_r$ | 23 | 10 | 19 | 6 | 4 | 5 | 4 |
| | G | 1.136 | 2.024 | 1.938 | 2.008 | 2.004 | 2.016 | 2.020 |
| | $t_r$ | 23 | 14 | 24 | 11 | 4 | 6 | 4 |
| 50 | $\bar{t}_r$ | 26 | 16 | 28 | 11 | 5 | 8 | 4 |
| | G | 1.166 | 2.217 | 2.028 | 2.273 | 2.242 | 2.262 | 2.252 |
| | $t_r$ | 24 | 20 | 27 | 13 | 4 | 17 | 5 |
| 60 | $\bar{t}_r$ | 31 | 20 | 32 | 15 | 6 | 19 | 6 |
| | G | 1.190 | 2.304 | 2.062 | 2.513 | 2.538 | 2.532 | 2.551 |
| | $t_r$ | 25 | 21 | 36 | 15 | 6 | 50 | 6 |
| 70 | $\bar{t}_r$ | 35 | 21 | 36 | 20 | 8 | 51 | 6 |
| | G | 1.209 | 2.398 | 2.092 | 2.681 | 2.725 | 2.747 | 2.793 |
| | $t_r$ | 35 | 21 | 37 | 15 | 6 | 57 | 8 |
| 80 | $\bar{t}_r$ | 41 | 22 | 38 | 22 | 9 | 62 | 10 |
| | G | 1.229 | 2.451 | 2.128 | 2.747 | 2.933 | 2.755 | 3.012 |
| | $t_r$ | 39 | 22 | 38 | 16 | 9 | 65 | 14 |
| 90 | $\bar{t}_r$ | 45 | 23 | 39 | 25 | 11 | 73 | 14 |
| | G | 1.2422 | 2.506 | 2.160 | 2.809 | 3.086 | 2.809 | 3.257 |
| | $t_r$ | 43 | 23 | 39 | 17 | 11 | 86 | 17 |
| 100 | $\bar{t}_r$ | 52 | 24 | 41 | 30 | 14 | 88 | 19 |
| | G | 1.256 | 2.558 | 2.232 | 2.874 | 3.268 | 2.865 | 3.559 |
| | $t_r$ | 202 | 27 | 41 | 163 | 67 | 93 | 51 |
| 200 | $\bar{t}_r$ | 202 | 32 | 51 | 194 | 72 | 116 | 53 |
| | G | 1.379 | 2.950 | 2.353 | 3.356 | 4.405 | 3.003 | 4.525 |
| | $t_r$ | 211 | 32 | 47 | 223 | 71 | 134 | 78 |
| 300 | $\bar{t}_r$ | 212 | 43 | 67 | 226 | 81 | 147 | 88 |
| | G | 1.397 | 3.279 | 2.532 | 3.460 | 4.587 | 3.077 | 4.902 |
| | $t_r$ | 220 | 35 | 56 | 229 | 77 | 179 | 203 |
| 400 | $\bar{t}_r$ | 229 | 56 | 87 | 231 | 92 | 179 | 267 |
| | G | 1.423 | 3.584 | 2.688 | 3.509 | 4.785 | 3.145 | 5.348 |
| | $t_r$ | 232 | 81 | 71 | 235 | 82 | 196 | 247 |
| 500 | $\bar{t}_r$ | 242 | 82 | 126 | 237 | 107 | 214 | 366 |
| | G | 1.441 | 3.937 | 2.865 | 3.559 | 4.975 | 3.226 | 5.618 |

concave PLAEF to those by the original PLAEF. For the JPEG trace, the original PLAEF is already nearly a concave function so that the concave PLAEF only slightly differs from the original one. However, for the MPEG trace the regions of the concave hull are interpolated by a new concave function.

We examine the change in cutoff intervals (see Figs. 10 and 11) caused by the concave approximation. As we can expect, in the JPEG trace the cutoff interval for the concave PLAEF is almost the same as the cutoff interval for the original PLEAF, especially in the small delay bound region. In the MPEG trace, the concave PLAEF is an upper bound of the original PLAEF, and thus the resulting cutoff interval is an upper bound of the original one as well. We observe that the cutoff interval does not show a stepwise increase which is due to the GOP structure of the MPEG video coding algorithm.

To examine the variation and efficiency of the proposed concave algorithm, we show extensive results in Table 2. First, we define the gain $G$ as the ratio of peak bandwidth to allocated bandwidth, i.e. $G = \max_i(X(i))/CT$, where $T$ is the video frame time. We find that the JPEG trace has a poorer multiplexing gain than all the MPEG-coded traces for the same delay bounds. Among the MPEG coded traces, Settop and Simpson show a little higher gain than others. Again, we observed that the PLAEF by the concave approximation

Table 3
The numbers of IAPs $r$ required to cover upto the cutoff interval: $r$ for the concave PLAEFs ($r$ for the original PLAEFs)

| Maximum delay (ms) | JPEG | MPEG | MTV2 | ATP | Simpsons | Talk1 | Settop |
|---|---|---|---|---|---|---|---|
| 10 ms | 7(12) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) | 2(2) |
| 50 ms | 18(23) | 5(14) | 7(24) | 3(11) | 3(4) | 2(6) | 2(4) |
| 100 ms | 18(43) | 7(23) | 8(38) | 3(17) | 3(11) | 8(86) | 4(17) |
| 500 ms | 54(232) | 8(81) | 16(71) | 16(235) | 8(82) | 19(196) | 19(247) |

estimates the cutoff interval accurately. For the delay bound of less than 100 ms, the cutoff interval is less than 50 frames and very accurate. Even for the delay bound of 500 ms, the error of the cutoff interval is less than 25%. Note that the cutoff index values for the original PLAEF and the concave PLAEF can be much different even though the corresponding cutoff intervals are almost same.

In Table 3, we show the size of the IAP sets $r$ for the original and concave PLAEFs required to cover the correlation range up to the cutoff intervals for a delay requirement of 10, 50, 100, 500 ms, respectively. We again observe that the concave approximation substantially reduces the number of IAPs for the cutoff interval. The IAP set sizes for the MPEG traces are significantly smaller than those of the JPEG traces. Especially, the number of IAPs required for characterizing the relevant time scale region is less than 10 for up to 100 ms but it increases steeply above that point. This is the effect of the GOP structure of MPEG traces. However, the number of 10 even for the concave PLAEF is still large for practical traffic control. Although the focus of this work lies in the charterization problem, in a future work we will examine the effect of matching the original PLEAF to two IAPs, which is equivalent to the present standards, (PCR, CDVT), and (SCR, MBS).

## 6. Conclusions

In traffic theory, rate distribution and time-correlation for input traffic have been considered of first importance for the estimation of network performance. Whether LRD in VBR video traffic is crucial for queueing behavior is also a frequently asked question.

In this article, to investigate the relevance between the correlation range and queueing performance, we characterized VBR video traffic by PLAEF function which is the worst arrival pattern for input traffic. The deterministic characterization, unlike in previous stochastic approaches, can give a fundamental bound of the correlation interval relevant to the network delay. We then defined cutoff interval above which the correlation does not affect the queue buildup. The cutoff interval is the clear bound of the time scale relevant to the estimation of queue size and thus of the meaningful range of correlation for the characterization of VBR video traffic. We quantified the relationship using a set of experiments with traces of MPEG/JPEG-compressed video. We showed that the range for important time scales depends on the traffic load; as traffic load increases, the range of time scale required for estimation of queueing delay also increases. These results offer another insight into the implication of LRD in VBR video traffic.

As found in this work and related literature, traffic utilization for deterministic QoS is rather low. For this drawback, a probabilistic approach is taken in Refs. [10,12]. Especially, Chong and Li [10] generalized the deterministic bounded sources to a probabilistic bounded curve (PBC). Thus, extension to the probabilistic bounded source is considered as a future work.

## Acknowledgements

This work was supported in part by Korea Science and Engineering Foundation (KOSEF) under Grant 98-0101-02-01-3. We thank Dr. O. Rose and Dr. M.W. Garrett for making their traces available. One of the authors, H. Ahn, would like to thank J. Scherer for her invaluable discussion on the manuscript.

## References

[1] M.W. Garrett, Contributions toward real-time services on packet networks, PhD Dissertation, Columbia Uni., New York, NY, May, 1993. ftp address and directory of the used video trace:bellcore.com/pub/vbr.video.trace/.
[2] M.M. Krunz, A.M. Makowski, Modeling video traffic using M/G/∞ input processes: a compromise between Markovian and LRD models, IEEE J. Select. Areas Commun. 16 (5) (1998) 733–748.
[3] D.N. Tse, R.G. Gallager, J.N. Tsitsiklis, Statistical multiplexing of multiple time-scale Markov streams, IEEE J. Select. Areas Commun. 13 (6) (1995) 1028–1038.
[4] J. Beran, R. Sherman, M. Taqqu, W. Willinger, Long-range dependence in variable-bit-rate video traffic, IEEE Trans. Commun. 43 (1995) 1566–1579.
[5] D.P. Heyman, T.V. Lakshman, What are the implications of long-range dependence for VBR video traffic engineering?, IEEE/ACM Trans. Networking 4 (3) (1996) 301–317.
[6] Y. Kim, S.Q. Li, Time-scale of interest in traffic measurement for link bandwidth allocation design, Proceedings of the INFOCOM'96, March 1996, pp. 738–746.
[7] M. Montgomery, G. De. Veciana, On the relevance of time scales in performance oriented traffic characterizations, Proceedings of the IEEE INFOCOM'96, March, 1996, pp. 520–913.
[8] J.W. Roberts, U. Mocci, J. Virtamo (Eds.), Broadband Network Teletraffic: Final Report of Action COST 242 Springer, Berlin, 1996.
[9] B. Melamed, D.E. Pendarakis, Modeling full-length VBR video using Markov-renewal-modulated TES models, IEEE J. Select. Areas Commun. 16 (5) (1998) 600–611.

[10] S. Chong, S. Li, Probabilistic burstiness-curve-based connection control for real-time multimedia services in ATM networks, IEEE J. Select. Areas Commun. 15 (6) (1997) 1072–1086.
[11] B. Hajek, L. He, On variations of queue responses for inputs with identical means and autocorrelation functions, Proceedings of the 1996 Conference on Information Sciences and systems, Princeton University, Princeton, NJ, March, 1996, pp. 1195–1201.
[12] A. Elwalid, D. Heyman, T.V. Lakshman, D. Mitra, A. Wiess, Fundamental bounds and approximations for ATM multiplexor with applications to video teleconferencing, IEEE J. Select. Areas Commun. 13 (6) (1995) 1004–1016.
[13] A. Baiocchi, Minimum buffer size for overflow avoidance in ATM VBR/CBR connection multiplexing, IEEE Commun. Letters 2 (4) (1998) 192–194.
[14] J.-Y. Le Boudec, Application of network calculus for guaranteed service networks, IEEE Trans. Inform. Theory 44 (3) (1998) 1087–1096.
[15] J.-Y. Le Boudec, P. Thiran, Network calculus viewed as a min-plus system theory applied to communication networks, Technical Report SSC/1998/016 EPFL-DI, 1998.
[16] C.S. Chang, Stability, queue length, and delay of deterministic and stochastic queueing networks, IEEE Trans. Automat. Control 39 (5) (1994) 913–931.
[17] C.S. Chang, A filtering theory for deterministic traffic regulation, Proceedings of the IEEE INFOCOM '97, Kobe Japan, June, 1997, pp. 437–444.
[18] C.S. Chang, Matrix extensions of the filtering theory for deterministic traffic regulation and service guarantees, IEEE J. Select. Areas Commun. 16 (5) (1998) 708–718.
[19] R.L. Cruz, Caculus for network delay. Part I: network elements in isolation, IEEE Trans. Inform. Theory 37 (1) (1991) 114–131.
[20] S.J. Golestani, A self-clocked fair queueing scheme for broadband applications, Proceedings of the IEEE INFOCOM'94, Toronto, Canada, July, 1994, pp. 636–646.
[21] F. Lo Presti, Z.-L. Zhang, J. Kurose, D. Towsley, Source time scale and optimal buffer/bandwidth trade-off for regulated traffic in an ATM node, in Proceedings of the IEEE INFOCOM '97, Kobe (Japan), June, 1997, pp. 676–683.
[22] A.K. Parekh, R.G. Gallager, A Generalized processor sharing approach to flow control in integrated service networks: the single node case, IEEE/ACM Tran. Networking 1 (3) (1993) 344–357.
[23] A.K. Parekh, R.G. Gallager, A Generalized processor sharing approach to flow control in integrated service networks: the multiple node case, IEEE/ACM Tran. Networking 2 (2) (1994) 137–150.
[24] J.W. Roberts, Virtual spacing for flexible traffic control, Internat. J. Commun. Sys. 7 (1994) 307–318.
[25] ATM Forum, Traffic management specification, version 4.0, April 1996.
[26] ITU-T Study Group 13, Recommendation I.371, Traffic control and congestion control in B-ISDN, Geneva, Switzerland, 1996.
[27] H. Zhang, Service disciplines for guaranteed performance service in packet-switching networks, Proc. IEEE 83 (10) (1996) 1374–1396.
[28] E. Knightly, H. Zhang, An accurate traffic model for providing QoS guarantees to VBR traffic, IEEE/ACM Tran. Networking 5 (2) (1997) 219–231.
[29] D.E. Wrege, E. Knightly, H. Zhang, J. Liebeherr, Deterministic delay bounds for VBR video in packet-switching networks: fundamental limits and practical trade-offs, IEEE/ACM Trans. Networking 4 (3) (1996) 352–362.
[30] E. Reich, On the intergrodifferential equation of Takacs, Ann. Math. Statist. 29 (1958) 563–570.
[31] O. Rose, Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems, University of Wuerzburg, Institute of Computer Science Research Report Series, Report no. 101, February, 1995, ftp address and directory of the used video trace: ftp-info3.informatik.uni-wuerzburg.de/pub/MPEG/
[32] A. Erramilli, O. Narayan, W. Willinger, Experimental queueing analysis with long-range dependent packet traffic, IEEE/ACM Trans. Networking 4 (2) (1996) 209–223.
[33] I. Norros, On the use of fractional Brownian motion in the theory of connectionless networks, IEEE J. Select. Areas Commun. 13 (6) (1995) 953–962.
[34] D.P. Heyman, A. Tabatabai, T.V. Lakshman, Statistical analysis and simulation study of video teleconferencing traffic in ATM networks, IEEE Trans. Circuits Syst. Video Technol. 2 (1) (1992) 49–59.
[35] D.Y. Eun, H. Ahn, B.H. Roh, J.-K. Kim, Effects of long-range dependence of VBR video traffic on queueing performances, Proceedings of the GLOBECOM'97, Phoenix, USA, November, 1997, pp. 1440–1444.

**Heejune Ahn** received BS and MS degrees from KAIST in electrical engineering in 1993 and 1995, respectively. Currently he is working towards a PhD degree in electrical engineering at KAIST. He joined LGIC (Lucky-Goldstar Information and Communication) as a research engineer in 1996. His research interests are ATM traffic control, video traffic characterization and modeling, and multimedia communication systems. He is a member of the KICS of KOREA and a student member of IEEE. A paper jointly authored by Prof. Jaekyoon Kim won a prize in the fifth (1999) Samsung Electronics Humantech thesis contest.

**Andrea Baiocchi** received his 'Laurea' degree in Electronics Engineering in 1987 and his PhD degree in Information and Communications Engineering in 1992, both from the University of Roma 'La Sapienza', where he has been an Associate Professor since November 1998. From 1991 to 1992 he was a Researcher at the Department of Mathematical Methods and Models for Applied Sciences, where he held lectures in Numerical Analysis. From July 1992 to October 1998 he stayed with the INFOCOM Department as a Researcher in Communications. His current research interests are in the area of Communications Networks and specifically on traffic modeling and performance evaluation of packet networks and multiaccess protocols for fixed and mobile radio interfaces. Andrea Baiocchi is a member of IEEE Communications Society.

**Jae-kyoon Kim** received a BS degree from National Aviation College in 1961, a MS degree from Seoul National University, Seoul, Korea in 1967, and a PhD degree from the University of Southern California, Los Angeles in 1971 all in electrical engineering. From April 1972 to March 1973, he was a Research Associate at NASA/GSFC, Greenbelt, MD in USA. Since April 1973 he has been a faculty member of KAIST, where he is now a professor in the Department of Electrical Engineering. His current research interests are in the area of video data compression, visual and multimedia communication systems. From 1993 to 1994, he was the president of KICS (Korean Institute of Communications Science) of Korea. He is now an associate editor of Signal processing: Image communication (EURASIP). He is a senior member of IEEE, a member of the KITE, KICS of KOREA, and a member of SMPTE, SPIE, IEICE.