Video Traffic Characterization for Multimedia Networks with a Deterministic Service*

Dallas E. Wrege *Jörg Liebeherr*

Department of Computer Science
University of Virginia
Charlottesville, VA 22903
Email: {dallas | jorg}@cs.virginia.edu

Abstract

One of the most important traffic types in future packet-switched networks is high-bandwidth, variable-bit-rate (VBR) video. Since video is a delay-sensitive media, the network must allocate resources to maintain quality-of-service (QoS) guarantees on throughput, delay, and delay jitter to video connections. It is difficult to allocate network resources to video connections while achieving a high utilization due to the burstiness of VBR video. A key component of resource allocation is the *traffic characterization* of video sources that determines the resources required to support video connections. The traffic characterization should accurately specify the traffic so that a large number of connections can be supported, but it must also map directly into policing mechanisms that monitor traffic on the connection. In this study, we propose a method for characterizing VBR video traffic with a fixed number of leaky bucket mechanisms in networks with a deterministic service. We explore tradeoffs of network utilization in two directions: (1) the amount of information from a video sequence used in a traffic characterization, and (2) the number of leaky buckets used to police each connection. We evaluate our method with a set of 30-minute long MPEG-compressed video traces, comparing the performance of our approach to other characterization techniques from the literature as well as a characterization that is optimal in terms of number of admissible connections.

Key Words: Multimedia Networks, Traffic Characterization, Bounded Delay Service, Quality-of-Service, Deterministic Service.

*This work is supported in part by the National Science Foundation under Grant No. NCR-9309224.

1 Introduction

The design of a packet-switched network that can support variable-bit-rate (VBR) video applications is a difficult task that has recently received much attention in the research community. VBR video connections place significant demands on the network because such connections are bursty and high in bandwidth usage while requiring minimum guarantees on their *quality-of-service* (QoS), for example, bounded delays, minimum throughput, and bounded jitter. Packet-switched networks have the potential to support VBR video connections efficiently since they are able to statistically multiplex bursty connections. However, a significant challenge remains in maintaining high utilization while providing QoS guarantees to all VBR video connections.

Many high-quality video applications are sensitive to statistical fluctuations in the service and demand a *deterministic service* that provides *worst-case* QoS guarantees. A key element for the implementation of a deterministic service is the *traffic characterization* that specifies the traffic on a connection. The traffic characterization in a network with a deterministic service is used in *admission control functions* and the *policing mechanisms.* At connection establishment time, admission control functions decide whether traffic conforming to the characterization can be transmitted at its specified QoS without violating the QoS of existing connections. Once a connection is established, policing mechanisms ensure that all traffic submitted to the network on the connection conforms to its characterization.

A traffic characterization for a deterministic service must be both *accurate* and *policable.* If the traffic characterization is not accurate in specifying the traffic on a connection, i.e., if the characterization specifies more traffic than is actually transmitted on the connection, the admission control tests will overestimate the resource requirements of a connection, resulting in poor network utilization. Due to the complex timely correlations of VBR video sequences (for example, see the MPEG video trace shown in Figure 1), elaborate traffic characterizations are needed to achieve a high degree of accuracy. However, such an elaborate characterization scheme may come at the expense of increased complexity of the policing mechanisms that must enforce the traffic characterization in real time. The choice of traffic characterization is a tradeoff between the high accuracy preferred by admission control tests and the simplicity required for implementing policing mechanisms.

Many traffic characterizations have been proposed in the literature, but here we consider traffic characterizations that can be policed by leaky bucket mechanisms [24]. Such traffic characterizations provide a good tradeoff between accuracy and simplicity; they can be efficiently implemented [22] and are shown to accurately characterize VBR video in [16]. A leaky bucket mechanism has a burst parameter σ and a rate parameter ρ, and it can be easily implemented with a single counter and timer. Note that the ATM Forum has adopted a *dual leaky bucket* scheme where one leaky bucket enforces the long-term average rate of a connection and the second bucket enforces the peak rate [1].

Many researchers have investigated the characterization of VBR video traffic. One approach

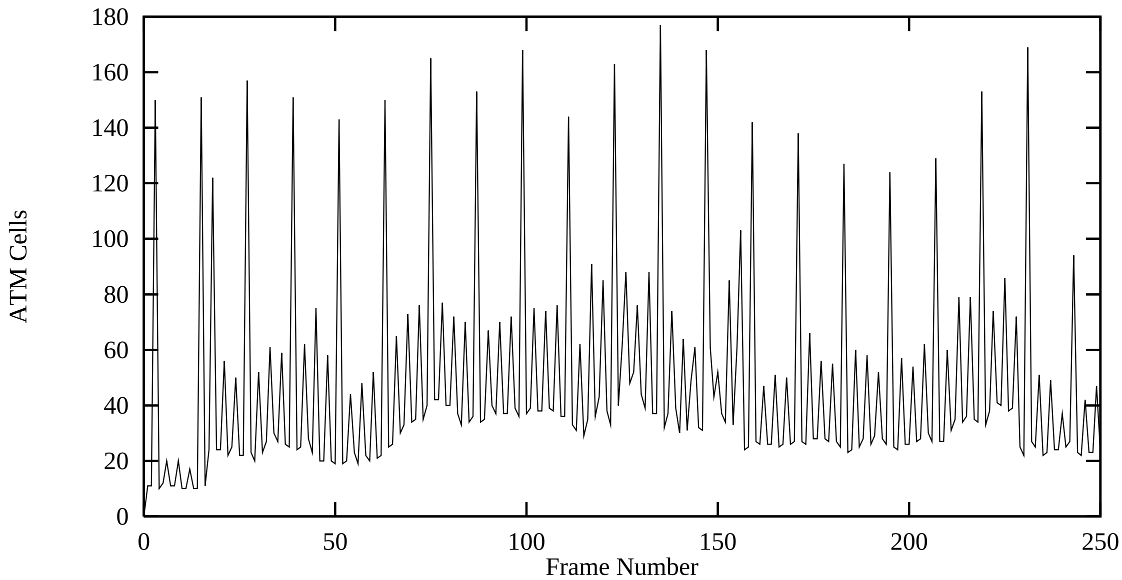


Figure 1: Traffic of an MPEG video stream.

(e.g., [5, 14]) allocates resources to connections based on their *equivalent bandwidths*, ensuring that the probability of buffer overflow or excessive delay is bounded. Since the equivalent bandwidth approach provides QoS guarantees only to the aggregate set of multiplexed connections, it does not provide guarantees to individual connections. Several studies (see [10] for a recent survey) characterize VBR video traffic with stochastic processes using models such as Markov-modulated, transform-expand-sample (TES) [20], and self-similar [12]. Since stochastic characterizations do not provide a worst-case bound on the traffic arrivals on a connection, they are inappropriate for use in a deterministic network.

Other works [4, 15, 18, 19] consider deterministic characterizations for video traffic. In [18, 19] Low and Varaiya investigate traffic that conforms to a leaky bucket mechanism. They quantify the tradeoff between buffer space and transmission bandwidth in network resource allocation with so-called *burstiness curves* that show how a smaller burst σ can be achieved by increasing the transmission rate ρ, and vice versa. Both [19] and a recent work by Chong [4] discuss how the burstiness curves can be used to select leaky bucket parameters, but their results are specific to rate-proportional processor sharing (RPPS) multiplexers [21] and to characterizations that use only a single leaky bucket.

In [15], Knightly and Zhang present the Deterministic Binding Interval Dependent (D-BIND) traffic characterization that specifies video traffic with a set of rate-interval pairs $\{(R_j, I_j) \,|\, 1 \leq j \leq n\}$. A D-BIND characterization defines an n-segment piecewise linear function. Such a characterization is a generalization of the n-piecewise linear *concave* function that corresponds to an n-level leaky bucket. The numerical examples presented in [15] suggest that the parameters of the D-BIND model can be selected to accurately characterize VBR video traffic, but the problem of parameter-selection is not addressed.

In this study, we investigate leaky bucket traffic characterizations of VBR video sequences that yield a high resource utilization in networks with a deterministic service. We consider two fundamental problems of video characterization. We first address the question of how much information about a video sequence is needed for an accurate characterization. As we will discuss in Section 2, the empirical envelope of a video sequence is the optimal characterization for use in deterministic networks in the sense that it is the tightest of all traffic constraint functions. Here we present a method for determining an accurate traffic characterization based on a subset of the empirical envelope. A characterization algorithm requires some knowledge of the video sequence, and we investigate the impact of reducing the information used in producing a characterization.

The second problem we consider is the selection of leaky bucket parameters. We present a heuristic algorithm that selects parameters for a fixed number of leaky buckets to characterize VBR video traffic. We evaluate the effectiveness of our heuristic with other approaches from the literature using a number of video segments coded with the MPEG compression algorithm [11].

The remainder of this paper is structured as follows. In Section 2 we review the empirical envelope $\mathcal{E}$ and investigate the amount of information necessary to accurately characterize a VBR video sequence. In Section 3 we describe an algorithm that selects leaky bucket parameters. We evaluate the effectiveness of our characterization method in Section 4 with a number of VBR video sequences compressed with the MPEG algorithm [11]. We conclude our study in Section 5.

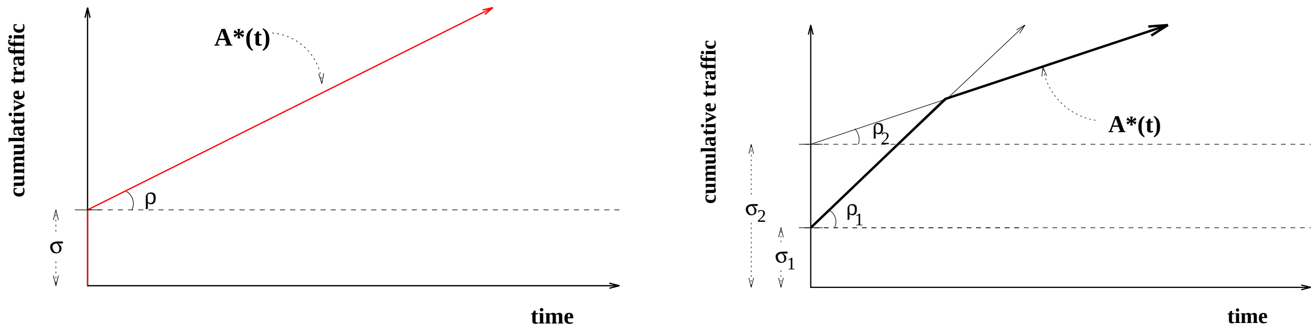
2 The Empirical Envelope and Approximations

In this section we present a class of traffic characterizations that is accurate in describing the traffic on a VBR video connection. We first review the so-called empirical envelope of a video sequence which provides a tight worst-case bound on the traffic on a connection [16]. We then investigate how much of this worst-case bound is needed to produce an accurate characterization. We show how a subset of the envelope can be used to characterize a video sequence, and we evaluate the impact of subset size k on the number of connections established by admission control tests.

2.1 The Empirical Envelope $\mathcal{E}$

A worst-case traffic characterization can be formally described as follows. Let A denote the actual traffic on a connection, where $A[\tau, \tau + t]$ denotes the traffic arrivals in the time interval $[\tau, \tau + t]$. An upper bound on A can be given by a function A^* if for all times $\tau \geq 0$ and $t \geq 0$ the following holds [3, 7, 17]:

$$A[\tau, \tau + t] \leq A^*(t) \tag{1}$$



(a) Leaky bucket (b) Multi-level leaky bucket

Figure 2: Traffic constraint functions.

Such a function A^* is called a *traffic constraint function* for A. Note that A^* is by construction a *time-invariant bound* of A, i.e., $A^*(t)$ provides a bound on the maximum traffic arrivals from any interval of length t. We further require a traffic constraint function to be a *subadditive* function, that is, A^* must satisfy the following equation:

$$A^*(t_1 + t_2) \leq A^*(t_1) + A^*(t_2) \qquad \forall t_1, t_2 \geq 0 \tag{2}$$

We insist that our traffic constraint functions be subadditive since they will be used in admission control tests that require subadditivity [7, 17, 21]. Thus, in this paper we consider only traffic constraint functions that satisfy both equations (1) and (2).

Traffic constraint functions can express the worst-case traffic admitted by a large class of policing mechanisms. For example, consider the traffic admitted to the network by a leaky bucket mechanism. A leaky bucket mechanism (σ, ρ) limits the rate and burstiness of a connection as follows: a credit counter is initialized to a positive value σ, and traffic may only enter the network if the counter is nonzero. The credit counter is decremented for each cell that enters the network, and the counter is incremented at rate ρ when its value is less than σ. The traffic constraint function A^* for traffic policed by a leaky bucket mechanism (σ, ρ) is then given by the linear constraint [7]:

$$A^*(t) = \sigma + \rho t \qquad \text{for all } t \geq 0 \tag{3}$$

Similarly, if multiple leaky buckets (σ_i, ρ_i) are connected in series (i.e., a multi-level leaky bucket [7, 16]), the amount of traffic admitted to the network is limited by each of the leaky buckets. The traffic constraint function for a m-level leaky bucket is given as follows [8, 15, 16]:

$$A^*(t) = \min_{1 \leq i \leq m} \{\sigma_i + \rho_i t\} \tag{4}$$

This function is an m-piecewise linear concave function. We illustrate curves A^* for both a single leaky bucket and a 2-level leaky bucket mechanism in Figure 2.

Many traffic constraint functions A^* can bound a given traffic source, but some are more appropriate for characterizing video than others. For example, a simple peak-rate characterization satisfies the definition of a traffic constraint function, but such a scheme overestimates the bandwidth requirements of a bursty connection. Since the network must reserve resources to support traffic on a connection according to its characterization, the traffic constraint function A^* for a connection should be selected as tight as possible (i.e., the characterization should be accurate).

In a previous study with Knightly and Zhang [16], we presented the tightest traffic constraint function for a given traffic source, the *empirical envelope* $\mathcal{E}$. The empirical envelope $\mathcal{E}$ of a video sequence is an optimal characterization in the sense that, for any subadditive traffic constraint function A^* that satisfies equation (1), $\mathcal{E}(t) \leq A^*(t)$ for all t. The empirical envelope $\mathcal{E}$ is given by the following equation [16]:

$$\mathcal{E}(t) = \max_{\tau > 0} A[\tau, \tau + t] \qquad \forall t > 0 \tag{5}$$

It follows from equation (5) that $\mathcal{E}$ is subadditive.

We next review a method from [16] for obtaining the empirical envelope for a given video sequence. Assume that a video sequence consists of N frames with fixed inter-frame time r. Further assume that frames are fragmented into 53-byte ATM cells with a payload of 48 bytes each; these cells are transmitted at equally-spaced intervals over the frame time r. If the sequence of frame sizes of a video sequence is given by $\{f_1, f_2, \ldots, f_N\}$, and if the transmission of this sequence starts at time 0, the arrival function A which describes the cumulative number of cell arrivals up to time t is given as follows [16]:

$$A[0,t] = \sum_{i=1}^{\lfloor \frac{t}{r} \rfloor} f_i + \left(\frac{t}{r} - \left\lfloor \frac{t}{r} \right\rfloor \right) f_{\lceil \frac{t}{r} \rceil} \qquad 0 \leq t \leq Nr \tag{6}$$

The empirical envelope $\mathcal{E}$ can then be constructed by combining equations (5) and (6). Alternatively, $\mathcal{E}$ can be constructed directly from the sequence of frames $\{f_1, f_2, \ldots, f_N\}$ by first calculating:

$$\mathcal{E}(i \cdot r) = \max_{0 < k < N-i} \sum_{j=k}^{k+i} f_j \qquad \text{for } i = 1, 2, \ldots N \tag{7}$$

The values of the empirical envelope at times that are not multiples of the frame time are then obtained by spacing the cells in $\mathcal{E}((i+1) \cdot r) - \mathcal{E}(i \cdot r)$ evenly over the frame time $[i \cdot r, (i+1) \cdot r]$. The empirical envelope for the video trace in Figure 1 is illustrated in Figure 3.

Once the empirical envelope $\mathcal{E}$ has been computed for a video sequence, it is straightforward to obtain a multi-level leaky bucket characterization if there is no restriction on the number of leaky bucket pairs available. Recall from equation (4) that a multi-level leaky bucket polices a concave, piecewise-linear function. Therefore, the concave hull of the empirical envelope $\mathcal{E}$ is the tightest traffic constraint function that can be policed by leaky buckets. We denote the concave hull of $\mathcal{E}$ by the function $\overline{\mathcal{E}}$. The function $\overline{\mathcal{E}}$ is clearly a traffic constraint function in the sense of equation (1), and the subadditivity of $\overline{\mathcal{E}}$ follows from its concavity. For a video sequence consisting of N frames, the function $\overline{\mathcal{E}}$ can be computed from $\mathcal{E}$ in $O(N^2)$ time with the straightforward

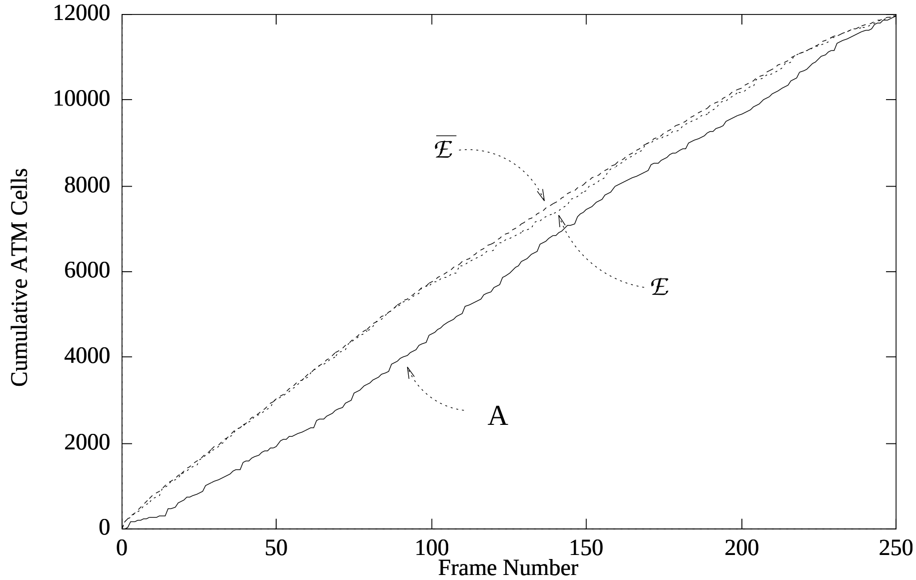


Figure 3: Accumulated traffic $A[0,t]$ of the MPEG-compressed sequence from Figure 1, the empirical envelope $\mathcal{E}(t)$, and its concave hull $\overline{\mathcal{E}}(t)$.

concave-hull algorithm described in [16]. More efficient algorithms, such as variants of Graham's scan (e.g., [6]), can compute the concave hull in $O(N \lg N)$ time. We illustrate the function $\overline{\mathcal{E}}$ in Figure 3.

2.2 Approximations of the Envelope

The empirical envelope $\mathcal{E}$ is the optimal traffic constraint function in terms of accuracy, but it has a significant drawback in that its production is computationally expensive. For example, the number of operations required to compute the empirical envelope $\mathcal{E}$ for a video sequence consisting of N frames is $O(N^2)$. Since N exceeds 200,000 for most entertainment motion pictures, it may not be possible to calculate the entire empirical envelope in real time. The characterization $\overline{\mathcal{E}}$ requires knowledge of the envelope, and so its production is also computationally intensive. We therefore seek other traffic constraint functions that closely approximate the envelope from above but can be calculated with fewer operations.

We consider approximations based on a subset of the empirical envelope. In particular, we present in this section several approximations that can be derived from the first k frames of the envelope, $\mathcal{E}[0, k \cdot r]$, where r is the frame rate of the video sequence. We call $\mathcal{E}[0, k \cdot r]$ the "prefix" of size k frames, and we consider different methods for extrapolating $\mathcal{E}[0, k \cdot r]$ to obtain traffic constraint functions that are defined for all times t. We start with the prefix and extend, i.e., extrapolate, subject to the constraint that the extrapolation is subadditive and satisfies equation (1).

Feasible extrapolations of $\mathcal{E}[0, k \cdot r]$ must be at least as large as the empirical envelope $\mathcal{E}$ for all times t. Since we only know that $\mathcal{E}$ is a subadditive function with prefix $\mathcal{E}[0, k \cdot r]$, any extrapolation A^* must be at least as large as the *largest subadditive extension of* $\mathcal{E}[0, k \cdot r]$. We define the function E_k to be the largest such subadditive extension, and E_k is obtained by calculating:

$$E_k(i \cdot r) = \begin{cases} \mathcal{E}(i \cdot r) & \text{for all } i = 0, 1, \ldots, k. \\ \min_{1 \leq j < i} \{E_k(j \cdot r) + E_k((i-j) \cdot r)\} & \text{for all } i > k. \end{cases} \tag{8}$$

Note that equation (8) only defines E_k for times that are multiples of the frame time r. The values for intermediate values of E_k are then determined by spacing the cells evenly over each frame. Note that $E_k(t)$ is defined to be at least as small as $E_k(t_0) + E_k(t - t_0)$ for all t and $t_0 < t$, and so E_k is a subadditive function extension of $\mathcal{E}$.

Although the function E_k is the tightest traffic constraint function that can be obtained directly from a subset of k frames from the envelope, the production of E_k requires a large number of computations. In particular, we see from its definition in equation (8) that the number of computations required to construct E_k from the first k frames of the envelope is $O(N(N-k))$. Combining this complexity with the cost of obtaining k frames of the envelope yields a total computational complexity of $O(N^2)$, the same number of operations that are required for computing the empirical envelope itself. Since we seek an approximation that can be calculated with fewer computations, we turn to other approximation schemes.

Consider a function $\mathcal{E}_k$ obtained by a simple repetition of the frames from $\mathcal{E}[0, k \cdot r]$ over all times t. The relationship between the empirical envelope $\mathcal{E}$ and such a function $\mathcal{E}_k$ is illustrated in Figure 4. In the figure, the two functions are identical over the first k frame times, and the subsequent repeating of the first k frames for $\mathcal{E}_k$ is clearly shown. The function $\mathcal{E}_k$ can be given as follows:

$$\mathcal{E}_k(t) = \lfloor \frac{t}{k \cdot r} \rfloor \mathcal{E}(k \cdot r) + \mathcal{E}(t - \lfloor \frac{t}{k \cdot r} \rfloor (k \cdot r)) \qquad \text{for all } t \geq 0 \tag{9}$$

Observe that $\mathcal{E}_k$ is obtained in constant time from the first k frames of the envelope $\mathcal{E}[0, k \cdot r]$.

Recall that a feasible traffic constraint function must be as large as the empirical envelope and a subadditive function. We see that the function $\mathcal{E}_k$ satisfies the first of these conditions from the subadditivity of $\mathcal{E}$: since $\mathcal{E}(t)$ is less than the right-hand side of equation (9) for all times t, it follows that $\mathcal{E}(t) \leq \mathcal{E}_k(t)$ for all t.

Although $\mathcal{E}_k$ satisfies equation (1), the function is not necessarily subadditive. Consider for example a function $\mathcal{E}$ in which the second frame is smaller than the kth frame, that is, $\mathcal{E}[r, 2 \cdot r] < \mathcal{E}[(k-1) \cdot r, k \cdot r]$. Observe the following sequence of equations:

$$\begin{aligned} \mathcal{E}_k[(k-1) \cdot r, (k+1) \cdot r] &> \mathcal{E}_k[0, 2r] && (10) \\ \mathcal{E}_k[(k-1) \cdot r, (k+1) \cdot r] + \mathcal{E}_k((k-1) \cdot r) &> \mathcal{E}_k(2r) + \mathcal{E}_k((k-1) \cdot r) && (11) \\ \mathcal{E}_k(2r + (k-1)r) &> \mathcal{E}_k(2r) + \mathcal{E}_k((k-1) \cdot r) && (12) \end{aligned}$$

Equation (12) shows that the function $\mathcal{E}_k$ in this example violates the definition of subadditivity given in equation (2). Since $\mathcal{E}_k$ is not subadditive, it is not useful as a traffic constraint function

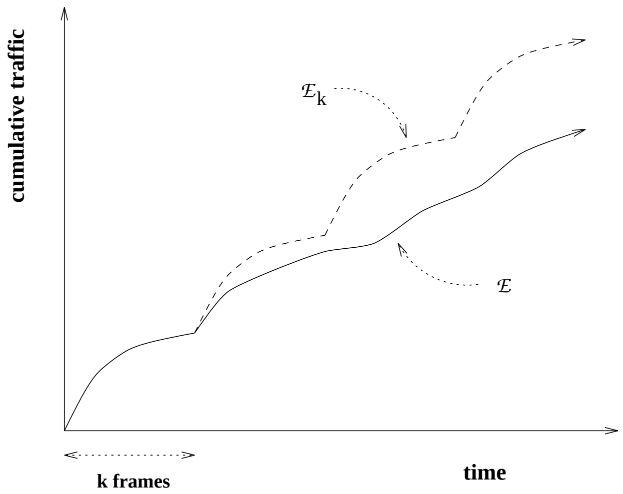


Figure 4: The empirical envelope $\mathcal{E}$ and $\mathcal{E}_k$.

for use in admission control tests. We next describe a subadditive function bounding $\mathcal{E}_k$ that can be used as a traffic characterization in deterministic networks.

We define the function $\overline{\mathcal{E}}_k$ to be be the concave hull of $\mathcal{E}_k$. In Figure 5 we depict functions $\mathcal{E}$, $\mathcal{E}_{25}$, and $\overline{\mathcal{E}}_{25}$ for the first 100 frames of the video sequence in Figure 1. We assume without loss of generality that $\overline{\mathcal{E}}_k$ is comprised of n piecewise-linear seqments, and we write the function $\overline{\mathcal{E}}_k$ as follows:

$$\overline{\mathcal{E}}_k(t) = \min_{1 \leq i \leq n} \{\overline{\sigma}_i + \overline{\rho}_i t\} \tag{13}$$

Using arguments similar to those in Section 2.1, we note that $\overline{\mathcal{E}}_k$ is a subadditive function and is at least as large as $\mathcal{E}_k$; $\overline{\mathcal{E}}_k$ is thus a feasible traffic constraint function. Also, the overall computational complexity of obtaining the traffic constraint function $\overline{\mathcal{E}}_k$ is then $O(k \cdot N)$.

2.3 Evaluation of the Envelope Approximation

In this section, we use traces of MPEG-compressed video to empirically evaluate the accuracy of the traffic constraint functions $\mathcal{E}$ and $\overline{\mathcal{E}}_k$ previously presented. We investigate the impact of the choice of k on the accuracy of $\overline{\mathcal{E}}_k$. These experiments provide insight into the tradeoff between network utilization and the amount of information from a video sequence used in the characterization.

We use two MPEG traces in the evaluation: one from the entertainment film *Jurassic Park* ("*Dino*"), and the second from a news broadcast ("*News*"). These sequences were captured on UNIX workstations and encoded with the Berkeley MPEG-encoder [23]. Both *Dino* and *News* are 30-minute video sequences with frame sizes of 384x288 and frame pattern IBBPBBPBBPBB. We note that *News* is much burstier than *Dino*; the ratio of the peak rate to the average rate for *News* and *Dino* are 5.9 and 3.1, respectively.

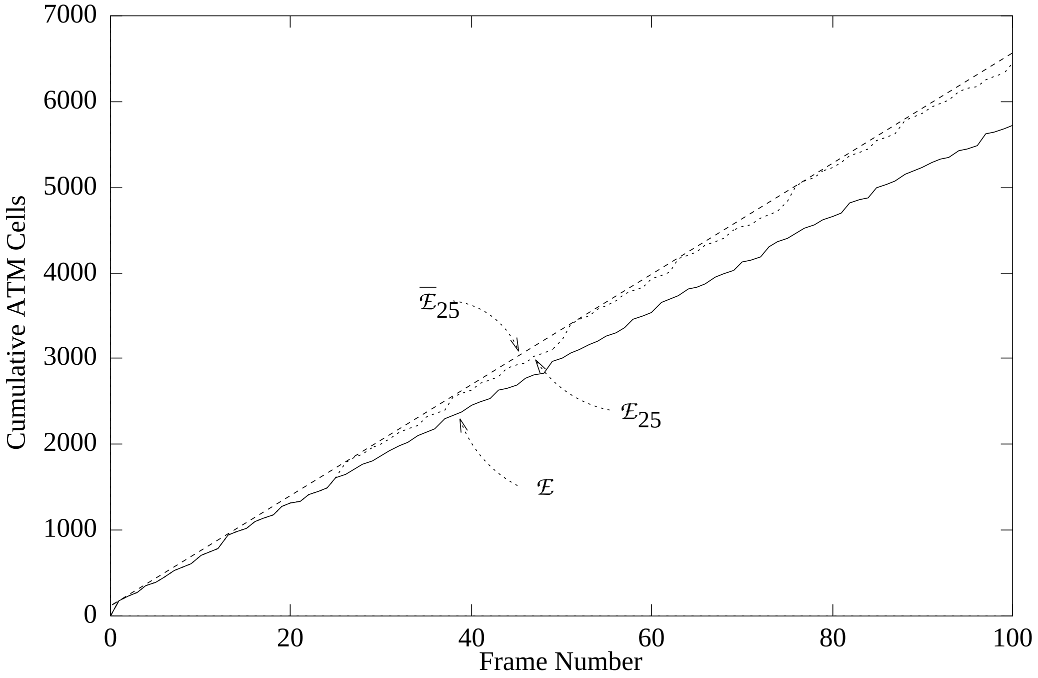


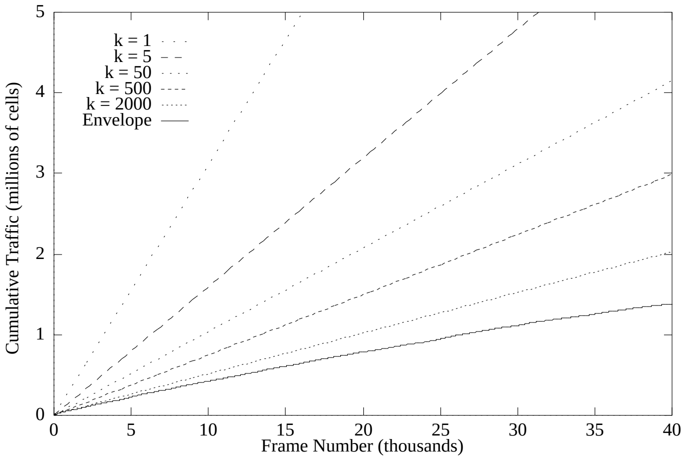
Figure 5: The first 100 frames of the empirical envelope $\mathcal{E}$ and its approximations $\mathcal{E}_{25}$ and $\overline{\mathcal{E}}_{25}$.

In these experiments, we consider a single multiplexer that operates at 155.52 Mbps, a data rate that corresponds to OC-3. We assume that the switch transmits its packets with a First-Come-First-Served (FCFS) discipline.[1]

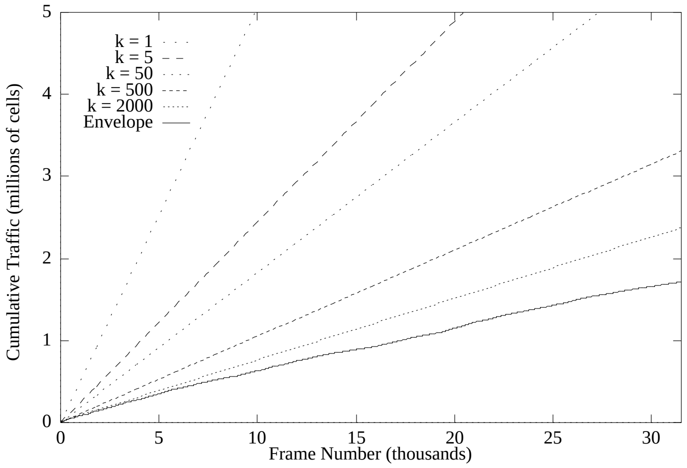
Figures 6(a) and 6(b) illustrate various traffic constraint functions for the *Dino* and *News* traffic traces, respectively. For each trace, we generate the empirical envelope $\mathcal{E}$ and the function $\overline{\mathcal{E}}_k$ for various values of k. In the graphs, the empirical envelope $\mathcal{E}$ is shown as the lowest curve, while the functions $\overline{\mathcal{E}}_k$ are the higher curves. As expected, the functions $\overline{\mathcal{E}}_k$ approximate the empirical envelope $\mathcal{E}$ more closely for larger values of k.

Based on the traffic constraint functions shown in Figure 6, one may get the impression that a large value of k is necessary to accurately characterize a video sequence. In both Figures 6(a) and 6(b), each function $\overline{\mathcal{E}}_k$ diverges from the empirical envelope $\mathcal{E}$ for larger values of t; the long-term average rate of each function $\overline{\mathcal{E}}_k$ is larger than that of the empirical envelope $\mathcal{E}$ for all values of k studied. However, the number of connections that can be admitted to the network for each of these traffic constraint functions cannot be determined directly from Figure 6. To better evaluate the accuracy of each characterization, we next compare the utilizations that can be achieved at a network switch using each traffic constraint function.

[1]The exact admission control test for FCFS multiplexers is given by $d \geq \sum_{j \in \mathcal{N}} A_j^*(t) - t + \max_{k \in \mathcal{N}} s_k$ for all $t \geq 0$ [7]. In this admission control test, $\mathcal{N}$ denotes the set of all connections at a multiplexer, d denotes the single delay bound the multiplexer supports, and $\max s_k$ denotes the cell size.



(a) *Dino*

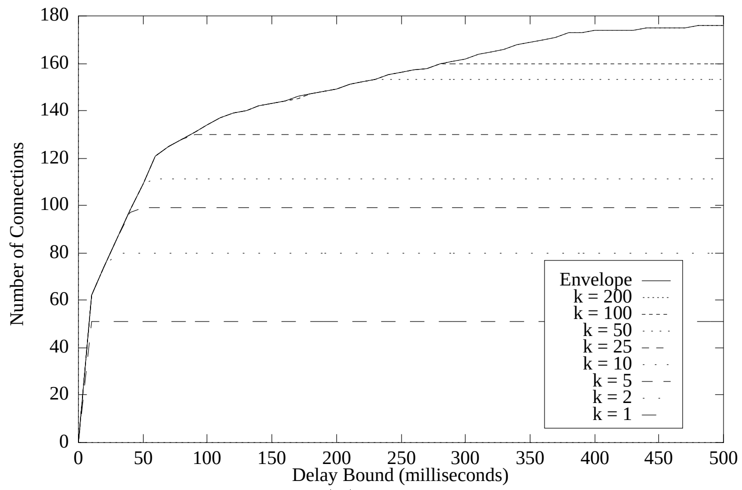


(b) *News*

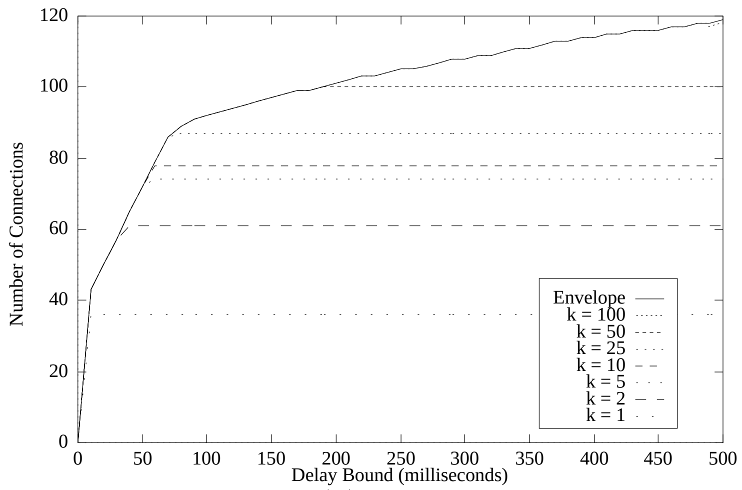
Figure 6: A comparison of $\mathcal{E}$ with $\overline{\mathcal{E}}_k$ for various values of k.

In Figure 7, we illustrate the network utilization that can be obtained at a multiplexer for traffic characterizations such as those in Figure 6. All connections at a multiplexer are assumed to be of the same type and have identical delay bounds. For each characterization, we plot the number of connections that can be admitted simultaneously against the delay bounds of those connections. For example, we observe in Figure 7(b) that the traffic constraint function $\overline{\mathcal{E}}_2$ can be used to support 61 *News* connections for delay bounds larger than 35 milliseconds.

Notice in both graphs that as the number of frames k is increased, more connections can be accepted. A characterization $\overline{\mathcal{E}}_k$ admits the same number of connections as the empirical envelope for some delay bound range $[0, d]$. For example, observe in Figure 7(a) that $\overline{\mathcal{E}}_{25}$ will admit as many connections as the envelope if the delay bound is less than 90ms. For larger values of k, higher delay bounds can be supported precisely; note in Figure 7(a) that $\overline{\mathcal{E}}_{50}$ admits the same number of



(a) *Dino*



(b) *News*

Figure 7: A utilization comparison of $\mathcal{E}$ with $\overline{\mathcal{E}}_k$ for various values of k.

connections as the envelope for delay bounds up to 230ms. Contrast the delay bound range for $\overline{\mathcal{E}}_{50}$ with the range for $\overline{\mathcal{E}}_{25}$ described previously.

This experiment provides insight into how much of the empirical envelope $\mathcal{E}$ is necessary to characterize a video sequence. For example, note that for both video sequences, the function $\overline{\mathcal{E}}_{200}$ admits the same number of connections as the empirical envelope $\mathcal{E}$ for delay bounds up to 500 milliseconds. Thus, the first 200 frames of the envelope (i.e., $\overline{\mathcal{E}}_k$ where $k = 200$) are sufficient to characterize both video sequences for the delay bound range considered.

3 A Heuristic for Leaky Bucket Parameter Selection

In the previous section, we showed how much of the envelope was needed to produce an accurate traffic characterization, and we produced such a traffic characterization $\overline{\mathcal{E}}_k$ that can be policed by a number of leaky buckets. Note however that the number of leaky buckets required for policing $\overline{\mathcal{E}}_k$ can be large. For example, 12 leaky bucket mechanisms are needed to police $\overline{\mathcal{E}}_{200}$ for the *News* sequence described in Section 2.3. In an actual network, the number of leaky buckets available to police a connection may be limited to only two or three; recall that the ATM Forum specifies that connections are policed by only two leaky buckets [1]. While the goal of the previous section was to reduce the information needed for video characterization, the goal here is to reduce the number of leaky buckets used to police the characterization. Given a characterization with a large number of leaky buckets, we approximate it with a characterization that requires fewer leaky buckets. To minimize the difference between the original characterization and the approximation, we require: (1) a cost function that expresses the difference between the two characterizations, and (2) an algorithm to minimize the cost function.

We formulate the problem to be addressed in this section as follows: Given a function $\overline{\mathcal{E}}_k$ that is written as in equation (13):

$$\overline{\mathcal{E}}_k(t) = \min_{1 \le i \le n} \{\overline{\sigma}_i + \overline{\rho}_i t\}$$

We want to find a set of m leaky bucket parameters $\{(\sigma_i, \rho_i) \,|\, 1 \le i \le m\}$ that determine a traffic constraint function B^*:

$$B^*(t) = \min_{1 \le i \le m} \{\sigma_i + \rho_i t\} \tag{14}$$

Assume that we are given a cost function $C(B^*, \overline{\mathcal{E}}_k)$ that expresses the difference between B^* and $\overline{\mathcal{E}}_k$. Note that this cost function C is dependent on the choices of (σ_i, ρ_i) that determine B^*. We select parameters (σ_i, ρ_i) to solve the following optimization problem:

minimize a cost function $C(B^*, \overline{\mathcal{E}}_k)$

subject to the constraint $\overline{\mathcal{E}}_k(t) \le B^*(t)$ for all times $t \ge 0$.

The constraint guarantees that B^* can be used as a traffic constraint function for a deterministic service. In the remainder of this section we describe a general form for the cost function $C(B^*, \overline{\mathcal{E}}_k)$, present a heuristic algorithm to solve the optimization problem, and evaluate particular choices of cost functions.

3.1 Cost Function $C(B^*, \overline{\mathcal{E}}_k)$

In this section, we describe a general cost function $C(B^*, \overline{\mathcal{E}}_k)$ that expresses the difference between the two functions B^* and $\overline{\mathcal{E}}_k$. The minimization of C subject to the constraint $\overline{\mathcal{E}}_k(t) \le B^*(t)$ for all times $t \ge 0$ should result in a traffic constraint function B^* that allows the admittance of a large number of connections. While it is clear that the function B^* should be as tight as possible, the choice of cost function C is not obvious. It is not clear that a straightforward minimization of the absolute difference between the two functions is an appropriate cost function. For example, since the large burstiness of VBR video limits the number of admitted connections at small delay bounds, it is important that the function B^* approximates $\overline{\mathcal{E}}_k$ more closely for small values of t.

We present a general parameterized cost function C, given as follows:

$$C(B^*, \overline{\mathcal{E}}_k) = \int_0^{k \cdot r} \frac{(B^*(t) - \overline{\mathcal{E}}_k(t))^\alpha}{(t+1)^\beta (\overline{\mathcal{E}}_k(t))^\gamma} \, dt, \tag{15}$$

where the integer exponents α, β, and γ in equation (15) determine the shape of the cost function.

Many different cost metrics can be determined with appropriate selections for these parameters. For example, a selection of $(2, 0, 0)$ for the (α, β, γ)-tuple results in an approximation where the square of the difference between B^* and $\overline{\mathcal{E}}_k$ is minimized. However, a least-squares model may not be appropriate since the function B^* is required to be larger than $\overline{\mathcal{E}}_k$. If the parameters $(1, 0, 1)$ are selected, then the cost function measures the overestimation of B^* relative to the function $\overline{\mathcal{E}}_k$. Selecting positive values for β results in a cost function that gives more weight to smaller times t. We evaluate choices for α, β, and γ in Section 3.3.

3.2 A Heuristic Algorithm

Assume that we are given a cost function that conforms to equation (15). We next present a heuristic algorithm that selects leaky bucket parameters to produce a traffic constraint function B^* of low cost. We evaluate the effectiveness of our heuristic with other schemes from the literature in Section 4.

The heuristic algorithm takes as input the function $\overline{\mathcal{E}}_k$, the number m of leaky buckets (σ_i, ρ_i) available to police B^*, a cost function $C(B^*, \overline{\mathcal{E}}_k)$, and a sensitivity parameter $\epsilon > 0$. The algorithm produces as output a set of m leaky bucket pairs (σ_i, ρ_i). The approach of the heuristic algorithm is to select initial values for all leaky bucket pairs (σ_i, ρ_i) and then iteratively modify these values to reduce the cost $C(B^*, \overline{\mathcal{E}}_k)$ of the function B^* determined by the (σ_i, ρ_i) pairs. The heuristic algorithm is presented in Figure 8.

The selection of initial values for the pairs (σ_i, ρ_i) is shown in steps 2 through 5 of Figure 8. These pairs are selected from the set of pairs $\{(\overline{\sigma}_j, \overline{\rho}_j)\}$ that determine the function $\overline{\mathcal{E}}_k$. The heuristic improves this initial selection by altering the (σ_i, ρ_i) pairs as shown in steps 6 through 11. The heuristic algorithm iteratively considers each (σ_i, ρ_i) pair and modifies its values to minimize the cost function C until the cost cannot be significantly improved.

The critical step of the algorithm is step 9, where a pair (σ_i, ρ_i) is selected that minimizes the cost function. We describe here how such a selection can be determined quickly. Consider first the

Input: A set of n leaky bucket parameters $\{(\overline{\sigma}_j, \overline{\rho}_j)\}$ that define the function $\overline{\mathcal{E}}_k$, the number m of available leaky buckets, a cost function $C(B^*, \overline{\mathcal{E}}_k)$, and the sensitivity parameter ϵ.

Output: A set of m leaky bucket parameters (σ_i, ρ_i) to determine the traffic constraint function B^*.

Procedure Find_Parameters ($\overline{\mathcal{E}}_k$, m, $C(B^*, \overline{\mathcal{E}}_k)$, ϵ)
    For $i = 1$ **To** m /* Select initial values for (σ_i, ρ_i) pairs */
        $\sigma_i \leftarrow \overline{\sigma}_{\lfloor \frac{i \cdot n}{m} \rfloor}$
        $\rho_i \leftarrow \overline{\rho}_{\lfloor \frac{i \cdot n}{m} \rfloor}$
    End For
    Do /* Modify selections, making greedy improvements */
        Cost $\leftarrow C(B^*, \overline{\mathcal{E}}_k)$
        For $i = m$ **Down To** 1
            Select (σ_i, ρ_i) such that $C(B^*, \overline{\mathcal{E}}_k)$ is minimized
        End For
    While ((Cost $- C(B^*, \overline{\mathcal{E}}_k)$) $> \epsilon$)
    Output $B^* \leftarrow \min_{1 \le i \le m} \{\sigma_i + \rho_i t\}$
End Procedure

Figure 8: A heuristic to determine leaky bucket parameters.

function $B^*(t) = \min_i \{\sigma_i + \rho_i t\}$. Observe that we can rewrite B^* as follows:

$$B^*(t) = \begin{cases} \sigma_1 + \rho_1 t & \text{for } 0 \le t < \frac{\sigma_2 - \sigma_1}{\rho_1 - \rho_2} \\ \sigma_2 + \rho_2 t & \text{for } \frac{\sigma_2 - \sigma_1}{\rho_1 - \rho 2} \le t < \frac{\sigma_3 - \sigma_2}{\rho_2 - \rho_3} \\ \vdots & \\ \sigma_{m-1} + \rho_{m-1} t & \text{for } \frac{\sigma_{m-1} - \sigma_{m-2}}{\rho_{m-2} - \rho m-1} \le t < \frac{\sigma_m - \sigma_{m-1}}{\rho_{m-1} - \rho_m} \\ \sigma_m + \rho_m t & \text{for } t \ge \frac{\sigma_m - \sigma_{m-1}}{\rho_{m-1} - \rho m} \end{cases} \tag{16}$$

Equation (16) identifies the m intervals for which the function B^* is linear. By decomposing $\overline{\mathcal{E}}_k$ in a similar fashion, we can rewrite the integral on the right hand side of equation (15) as follows:

$$C(B^*, \overline{\mathcal{E}}_k) = \sum_{i=1}^{m+n} \int_{t_i}^{t_{i+1}} \frac{((\sigma_{\eta(i)} + \rho_{\eta(i)}) - (\overline{\sigma}_{\nu(i)} + \overline{\rho}_{\nu(i)}))^\alpha}{(t+1)^\beta (\overline{\sigma}_{\nu(i)} + \overline{\rho}_{\nu(i)})^\gamma} \, dt, \tag{17}$$

where the $[t_i, t_{i+1}]$ are time intervals for which both B^* and $\overline{\mathcal{E}}_k$ use a single linear constraint as described above, and the functions η and ν identify the relevant linear segments of B^* and $\overline{\mathcal{E}}_k$, respectively, for each time interval. Note that the times t_i are dependent on the choices for parameters (σ_j, ρ_j). For many fixed values of α, β, and γ, the integral in equation (17) can be solved straightforwardly.[2]

In the heuristic, each (σ_j, ρ_j) pair is altered iteratively while keeping all other parameters constant, and it is possible in this case to write the cost function $C(B^*, \overline{\mathcal{E}}_k)$ as an expression where σ_j is the only independent variable. We observe that the variable ρ_j is dependent on σ_j with the help of the graph shown in Figure 9. Assume that the burst value σ_j is fixed; the rate parameter ρ_j will be selected as small as possible, i.e., the line $\sigma_j + \rho_j t$ will be tangential to the function $\overline{\mathcal{E}}_k(t)$ at some time τ. In Figure 9, notice that the (σ_j, ρ_j) line is tangential to $\overline{\mathcal{E}}_k(t)$ at a point τ where two linear segments of $\overline{\mathcal{E}}_k(t)$ intersect. Since $\overline{\mathcal{E}}_k(t)$ is concave and piecewise linear, this tangent will always occur at such a point τ, and the intersection point will necessarily occur at either time $t = 0$ or one of the $n-1$ times $\dfrac{\overline{\sigma}_i - \overline{\sigma}_{i-1}}{\overline{\rho}_{i-1} - \overline{\rho}_i}$. We obtain the following equation at the point of tangency τ:

$$\sigma_j + \rho_j \tau = \overline{\sigma}_i + \overline{\rho}_i \tau \qquad \text{for some } 1 \le i \le n \tag{18}$$

Solving for ρ_j in equation (18) , we obtain:

$$\rho_j = \frac{\overline{\sigma}_i + \overline{\rho}_i \tau - \sigma_j}{\tau} \tag{19}$$

In the empirical results presented in Sections 3.3 and 4, we select the (σ_j, ρ_j) pair of minimum cost in step 9 by an exhaustive search through all possible values of σ_j. However, with ρ_j expressed in terms of σ_j, such a selection can be determined by setting the derivative of $C(B^*, \overline{\mathcal{E}}_k)$ in equation (17) to zero.

We comment briefly on the computational complexity of the heuristic algorithm. While we cannot make guarantees on the running time of the algorithm, the examples that we have run converge quickly. For all examples, no more than six iterations were required.

In this section, we have presented a heuristic algorithm to select parameters for a fixed number of leaky buckets that determine a traffic constraint function B^*. A general cost function shown in equation (15) is used in the algorithm; we will evaluate different cost functions with traces of MPEG video in the next section.

3.3 Cost Function Evaluation

In the previous sections we presented a heuristic algorithm for selecting leaky bucket parameters according to a general cost function. Here we evaluate different choices for the parameters (α, β, γ) to determine a specific cost function that yields an accurate selection of leaky bucket parameters. We show data based on the same MPEG sequences described in Section 2.3, and we again consider a single multiplexer at a switch that operates at 155.52 Mbps.

We evaluate the performance of a cost function as follows. For a given characterization B^*, we plot the ratio of the number of connections admitted using B^* to the number admitted using the

[2] We have obtained closed-form solutions for all choices of (α, β, γ) considered in Section 3.3. These choices include $(1,0,1)$, $(1,1,0)$, $(2,0,0)$, $(2,1,1)$, $(2,0,2)$, and $(2,2,0)$.

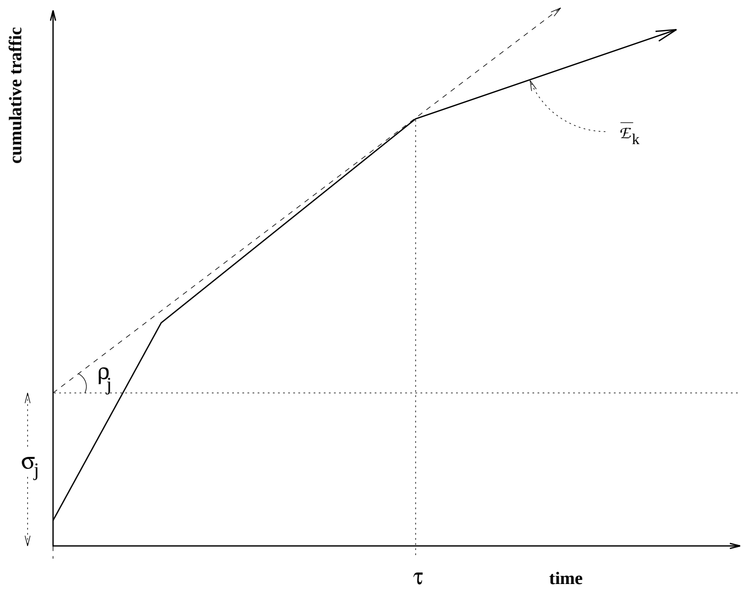


Figure 9: Illustration of dependency between σ_j and ρ_j.

optimal characterization (i.e., the empirical envelope $\mathcal{E}$) as a function of the delay bound of those connections, that is, for a given function B^* we plot:

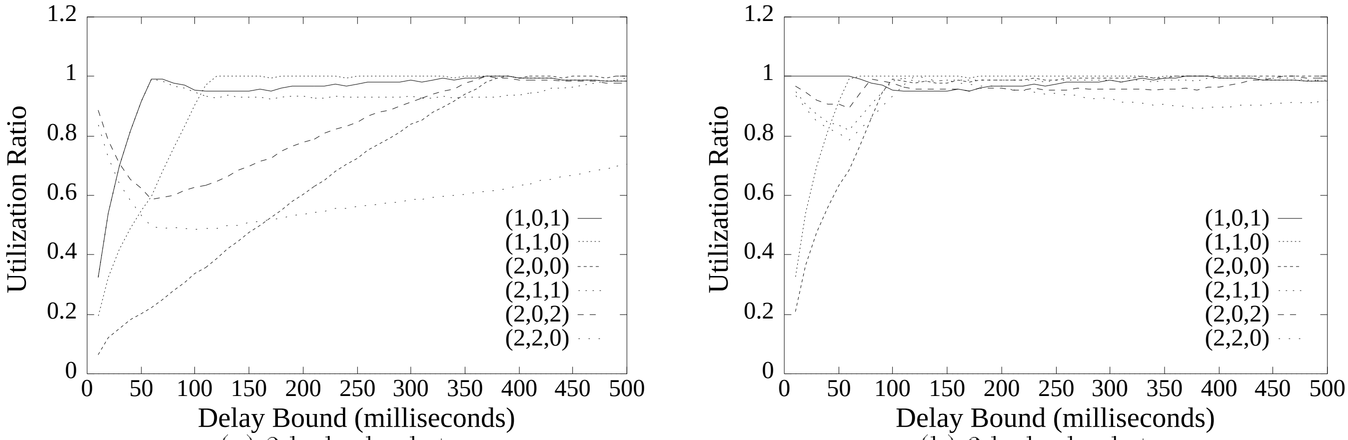
$$\text{Utilization Ratio}(B^*) = \frac{\#\text{ Connections with } B^*}{\#\text{ Connections with } \mathcal{E}} \tag{20}$$

For example, a traffic characterization that admits an optimal number of connections would correspond to the constant curve Utilization Ratio$(B^*) = 1$.

Figures 10 and 11 depict the utilization ratios for characterizations produced using the heuristic with various cost functions. In Figures 10 and 11 we plot utilization ratios for *Dino* and *News* characterizations, respectively. The number of leaky buckets m is fixed in each graph. We plot the utilization ratios corresponding to six different cost functions (α, β, γ): $(1,0,1)$, $(1,1,0)$, $(2,0,0)$, $(2,1,1)$, $(2,0,2)$, and $(2,2,0)$. All of these characterizations are produced with the heuristic in Figure 8 using $\overline{\mathcal{E}}_{200}$ as the input. Recall from Figure 7 that 200 frames from the empirical envelope were sufficient to characterize both of these video sequences.

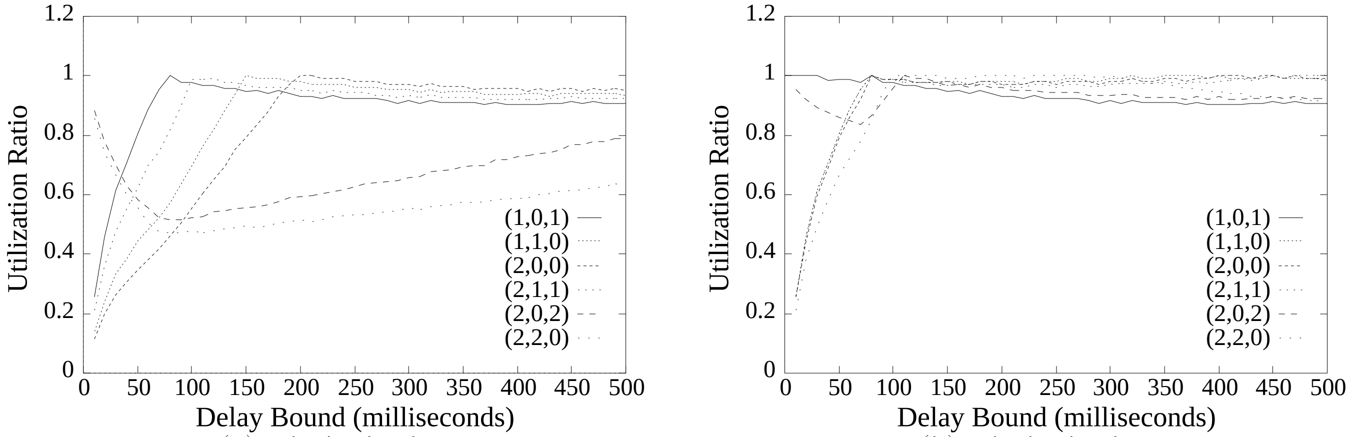
The most striking feature of these graphs is the accuracy of all six approximation algorithms. In particular, notice from Figures 10(b) and 11(b) that four of the characterizations have a utilization ratio of higher than 80% over the entire delay bound range [0, 500ms] using three leaky buckets, and the $(1,0,1)$ characterization has a utilization ratio of better than 90% over the same delay bound range. Even with only two leaky buckets as in Figures 10(a) and 11(a), the $(1,0,1)$ and $(2,1,1)$ characterizations have a high utilization ratio for delay bounds greater than 50ms.

Based on these experiments, the $(1,0,1)$ cost function (appearing in all graphs as a solid line) appears to be the best choice. In all four graphs, observe that the utilization ratio for the $(1,0,1)$



(a) 2 leaky buckets. (b) 3 leaky buckets.

Figure 10: Evaluation of cost functions for *Dino*.



(a) 2 leaky buckets. (b) 3 leaky buckets.

Figure 11: Evaluation of cost functions for *News*.

characterization is above 90% for a large range of delay bounds. Thus, the cost function that we employ for the remainder of the paper is given as follows:

$$C(B^*, \overline{\mathcal{E}}_k) = \int_0^{kr} \frac{(B^*(t) - \overline{\mathcal{E}}_k(t))}{\overline{\mathcal{E}}_k(t)}\, dt, \tag{21}$$

4 Empirical Evaluation

We are now ready to evaluate our characterization method by comparing it with other traffic characterization schemes from the literature. With the results from Sections 2 and 3, we can make recommendations for a characterization method. In Section 2, we found that 200 frames of the envelope are sufficient to produce an accurate VBR characterization, i.e., we use the function $\overline{\mathcal{E}}_{200}$. In Section 3, we found that the cost function $(1, 0, 1)$ is a good choice for use in the heuristic algorithm. In this section, we compare our characterization method using the above selections with other schemes using the MPEG video traces *Dino* and *News* described in Section 2.3.

The traffic characterizations from the literature that we use for comparison include a peak-rate allocation, a dual leaky bucket scheme that uses the peak and average traffic rates, and the so-called "concave hull approach" presented previously in [16]. In addition to the evaluation technique with a single traffic type used in Section 3.3, here we also consider scenarios with multiple traffic types and delay bounds.

We use the following three benchmarks:

(a) *Peak-rate*: The straightforward peak-rate allocation is the simplest of all traffic characterizations. A peak-rate characterization is determined by a single rate parameter ρ that corresponds with the peak rate of the connection. Such a characterization can be policed with a cell-spacer that ensures a minimum interarrival time between consecutive cells [2]. The traffic constraint function for peak-rate allocation is given as follows:

$$A^*_{\text{peak}}(t) = \rho t \tag{22}$$

(b) *Dual bucket*: The dual bucket scheme enforces both the peak rate ρ_0 and the average rate ρ_1 of a video connection as specified by the ATM Forum in [1]. The peak rate is policed with a cell-spacer that ensures a maximum rate ρ_0, while a leaky bucket mechanism (σ_1, ρ_1) enforces the average rate of the connection. The value of σ_1 is selected as small as possible subject to the constraint that $\sigma_1 + \rho_1 t \geq \mathcal{E}(t)$ for all times t. For the dual-bucket scheme, the traffic constraint function is given by:

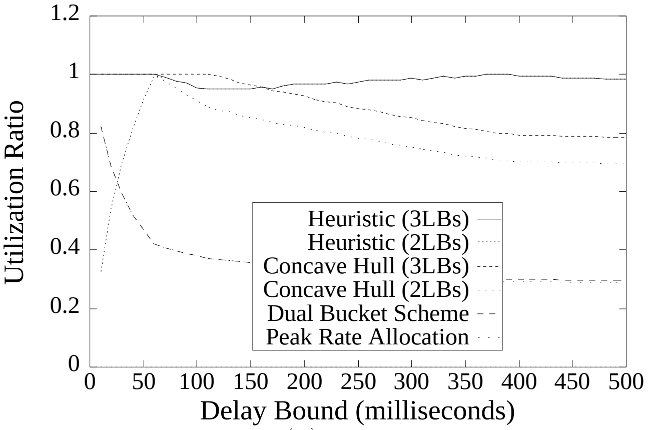
$$A^*_{\text{db}}(t) = \min\{\rho_0 t \, , \, \sigma_1 + \rho_1 t\} \tag{23}$$

(c) *Concave hull* [16]: Similar to the heuristic algorithm presented in this paper, the concave hull approach selects a number m of leaky bucket parameters to characterize the traffic on a video sequence. The parameters are taken directly from the concave hull of empirical envelope. If the concave hull of the empirical envelope is an l-piecewise linear function characterized by a set of l parameters $\{(\hat{\sigma}_i, \hat{\rho}_i) | 1 \leq i \leq l\}$, where $\hat{\sigma}_i < \hat{\sigma}_j$ for $i < j$, then the concave hull approach forms its traffic constraint function from the first m such parameters, that is:

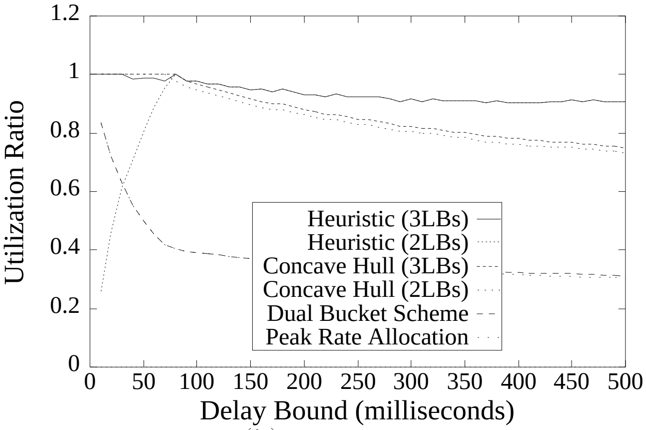
$$A^*_{\text{hull}}(t) = \min_{1 \leq i \leq m} \{\hat{\sigma}_i + \hat{\rho}_i t\} \tag{24}$$

4.1 Evaluation for a Single Traffic Type

Figures 12(a) and 12(b) show the utilization ratios of *Dino* and *News* connections, respectively, for the suite of traffic characterizations described above, namely, B^*, A^*_{peak}, A^*_{db}, and A^*_{hull}. Our



(a) *Dino.*



(b) *News.*

Figure 12: Utilization comparison of our heuristic with several approaches from the literature.

heuristic algorithm and the concave hull approach can determine values for a variable number of leaky buckets m, and so we depict characterizations B^* and A^*_{hull} for $m = 2$ and $m = 3$.[3]

Notice the poor performance of both A^*_{peak} and A^*_{db} in both graphs. The function A^*_{peak} achieves a utilization ratio of less than 40% for delay bounds greater than 50ms for these video sequences. The function A^*_{db} corresponding with the dual bucket strategy described by the ATM Forum performs only slightly better than A^*_{peak}. The peak-rate and dual-bucket schemes are clearly inferior to both the heuristic algorithm and the concave hull approach.

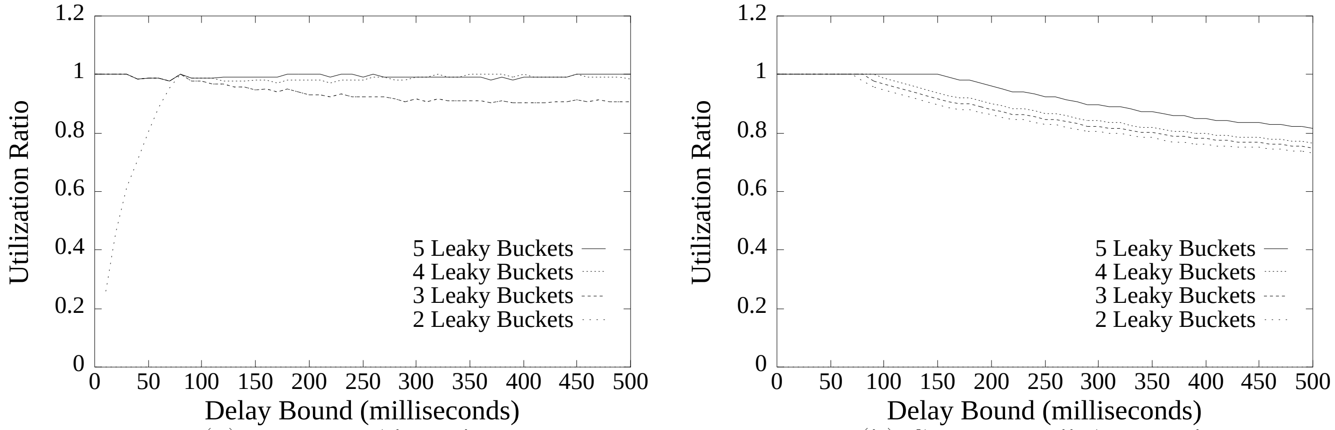
The utilization ratio for each curve A^*_{hull} is 1 for small delay bounds, and the ratio decreases gradually as the delay bound increases. The delay bound at which the utilization ratio begins decreasing is proportional to the number of leaky buckets m. Notice in Figure 12(a) that the point of decline begins at delay bound $d = 70$ms for $m = 2$ and at delay bound $d = 110$ms for $m = 3$. Observe that both characterizations produced by the heuristic in this paper are superior to those produced by the concave hull approach for delay bounds greater than 165ms in *Dino* and 80ms in *News*. However, the concave hull approach dominates for some small delay bounds.

4.2 Evaluation of the Number of Leaky Buckets

Here we examine the impact of the number of leaky buckets on network utilization. We consider parameter selections for $m = 2$, $m = 3$, $m = 4$, and $m = 5$ leaky buckets, and we show the functions B^* produced by our heuristic algorithm and A^*_{hull} produced by the concave hull scheme.

Figures 13(a) and 13(b) illustrate utilization ratios obtained for B^* and A^*_{hull}, respectively; these figures depict utilization ratios for characterizations of the *News* sequence. For two leaky buckets, we see that neither algorithm is dominant over the entire range of delay bounds. However, notice

[3]Note that for the case $m = 2$, i.e, two leaky buckets, the parameters selected by our algorithm are distinct from those selected by the dual bucket algorithm.



(a) Heuristic Algorithm.

(b) Concave Hull Approach.

Figure 13: Utilization comparison of two parameter-selection algorithms for the *News* segment.

in Figure 13(a) that the utilization ratio for B^* is close to 1 over the entire interval using only 4 leaky buckets. Conversely, additional leaky buckets do not provide as substantial a performance gain for A^*_{hull}; observe in Figure 13(b) that the utilization ratio for A^*_{hull} declines significantly for higher delay bounds even if 5 leaky buckets are used.

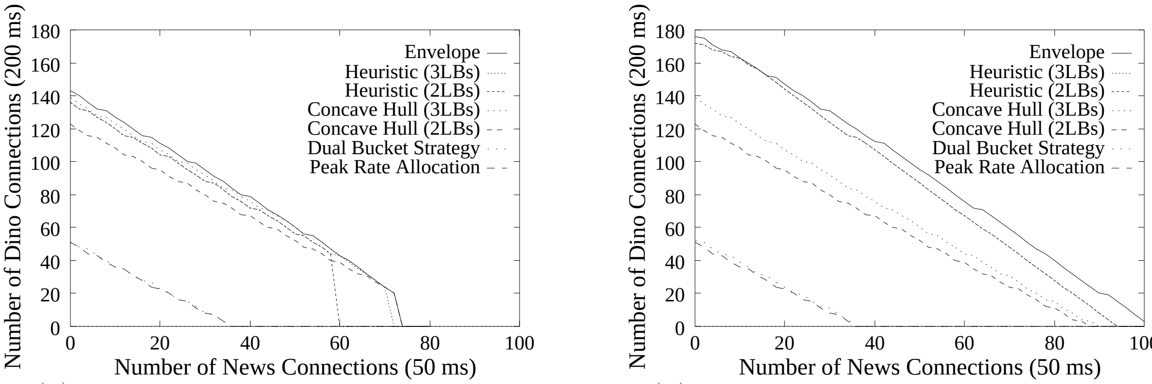
Based on these results, neither the heuristic algorithm nor the concave hull approach is dominant for two leaky buckets. Observe that the concave hull approach is superior for $m = 2$ if the delay bounds to be supported are small. However, the heuristic algorithm is superior to the concave hull approach if three or more leaky buckets are available.

4.3 Evaluation for Multiple Traffic Types

In the final example, we consider the support of two traffic types simultaneously. We consider two types of connections: those carrying the *Dino* sequence and those carrying the *News* sequence. All connections of the same type have identical delay bounds. Since there is more than one traffic type, the scheduling discipline at the switch will impact network utilization [17]. Since we are interested in the impact of the traffic constraint function on network utilization, we consider a switch that uses the optimal Earliest-Deadline-First (EDF) multiplexer [13]. The Earliest-Deadline-First (EDF) multiplexer orders cells for transmission according to their deadlines [9].[4]

Figure 14 illustrates the maximum number of *Dino* and *News* connections that can be simultaneously supported at an EDF multiplexer for the specified delay bounds. The delay bounds are $d_{Dino} = 200$ms and $d_{News} = 50$ms for Figure 14(a), while the delay bounds for Figure 14(b)

[4]The exact admission control test for EDF multiplexers is given as follows [17]: $t \geq \sum_{j \in \mathcal{N}} A^*_j(t - d_j) + \max_{d_k > t} s_k$ for all $t \geq d_1$. In this admission control test, $\mathcal{N}$ denotes the set of all connections at a multiplexer, d_j denotes the delay bound for a connection $j \in \mathcal{N}$, and $\max s_k$ is the cell size.



(a) $d_{News} = 50$ms, $d_{Dino} = 200$ms. (b) $d_{News} = 200$ms, $d_{Dino} = 500$ms.

Figure 14: A comparison of different traffic constraint functions for EDF multiplexing.

are $d_{Dino} = 500$ms and $d_{News} = 200$ms.

We first note in both graphs that the number of connections admitted using B^* is a factor of two or three greater than the number admitted using either A^*_{peak} or A^*_{db}. For example, in Figure 14(a), A^*_{peak} can support 30 *News* connections and 8 *Dino* connections, while B^* with two leaky buckets can support 92 *Dino* connections for the same number of 30 *News* connections.

Note in Figure 14(a), that B^* admits more connections than A^*_{hull} for $m = 2$ when more than 45 *Dino* connections are supported. Note also in both of these graphs that use of our heuristic algorithm with three leaky buckets yields a characterization that is close to optimal.

In Figure 14(b) we see the superiority of our heuristic over the concave hull approach for higher delay bounds. Notice that B^* with $m = 2$ admits more connections than A^*_{hull} for $m = 3$.

5 Concluding Remarks

The traffic characterization used for VBR video connections has a significant impact on the number of connections that can be established in a packet-switched network with a deterministic service. In this paper, explored two fundamental problems of video characterization: we determined how much of a video sequence is needed to produce an accurate characterization, and we presented a heuristic algorithm to select parameters for a number of leaky buckets. We combined these two techniques and presented a method for selecting leaky bucket parameters to characterize VBR video traffic. We gained the following insights from the paper:

- For the MPEG video sequences considered in the paper, we saw that as few as 200 frames of the empirical envelope are sufficient to yield a characterization that admits the same number of connections as the entire empirical envelope. This observation suggests that the important information regarding an MPEG sequence is contained in a small subset of the envelope.

- Using our heuristic algorithm, three leaky buckets were shown to be sufficient to admit nearly the same number of connections as the empirical envelope for both MPEG video sequences used in this paper. Based on the good performance of our heuristic algorithm, it may not be worthwhile to investigate more complex algorithms for video characterization.

- The dual leaky bucket scheme used by the ATM Forum was shown to result in poor network utilization in all experiments from Section 4. The performance gain of the dual bucket scheme over a simple peak-rate allocation was shown to be negligible. However, the poor performance of the dual bucket is not due to the fact that only two leaky buckets are used, but rather to a poor selection of leaky bucket parameters. Our heuristic algorithm with two leaky buckets was able to achieve a markedly higher performance than the dual bucket scheme.

- In [16], the numerical examples indicated that a large number of leaky bucket pairs were needed to approximate the empirical envelope. Here we argue that the number of leaky buckets needed to achieve performance similar to the envelope is small, i.e., three or four. This discrepancy can be resolved in two ways. First, we note that the examples in this paper only consider delay bounds up to 500ms, while the examples in [16] consider delay bounds up to 2000ms. For a larger delay bound range, additional leaky buckets will be necessary to approximate the empirical envelope. Second, the heuristic algorithm presented here is superior to the concave hull approach from [16]. We showed that the heuristic algorithm in this paper can closely approximate the empirical envelope with fewer leaky buckets.

Acknowledgment

We would like to thank Sudhir Srinivasan for his helpful discussions and suggestions.

References

[1] ATM Forum, ATM User-Network Interface Specification Version 3.0, Prentice-Hall, 1993.

[2] P.E. Boyer, F.M. Guillemin, M.J. Servel, and J.-P. Coudreuse. Spacing Cells Protects and Enhances Utilization of ATM Network Links. *IEEE Network*, 6(5):38–49, September 1992.

[3] C.-S. Chang. Stability, Queue Length, and Delay of Deterministic and Stochastic Queueing Networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

[4] S. Chong and S. Li. (σ, ρ)-Characterization Based Connection Control for Guaranteed Services in High Speed Networks. In *Proc. IEEE Infocom '95*, pages 835–844, 1995.

[5] S. Chong, S. Li, and J. Ghosh. Predictive Dynamic Bandwidth Allocation for Efficient Transport of Real-Time VBR Video over ATM. *IEEE Journal on Selected Areas in Communication*, 13(1):12–23, January 1995.

[6] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms.* MIT Press, 1992.

[7] R. L. Cruz. A Calculus for Network Delay, Part I: Network Elements in Isolation. *IEEE Transactions on Information Theory*, 37(1):114–131, January 1991.

[8] R. L. Cruz. A Calculus for Network Delay, Part II: Network Analysis. *IEEE Transactions on Information Theory*, 37(1):132–141, January 1991.

[9] D. Ferrari and D. C. Verma. A Scheme for Real-Time Channel Establishment in Wide-Area Networks. *IEEE Journal on Selected Areas in Communications*, 8(3):368–379, April 1990.

[10] V. Frost and B. Melamed. Traffic Modelling for Telecommunications Networks. *IEEE Communications Magazine*, 32(3):70–81, March 1994.

[11] D. Le Gall. MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, 34(4):46–58, April 1991.

[12] M. W. Garrett and W. Willinger. Analysis, Modeling and Generation of Self-Similar VBR Video Traffic. In *Proc. ACM Sigcomm '94*, pages 269–280, August 1994.

[13] L. Georgiadis, R. Guerin, and A. Parekh. Optimal multiplexing on a single link: Delay and buffer requirements. In *Proc. IEEE Infocom '94*, pages 524–532, 1994.

[14] G. Kesidis, J. Walrand, and C.-S. Chang. Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources. *IEEE/ACM Transactions on Networking*, 1(4):424–428, August 1993.

[15] E. Knightly and H. Zhang. Traffic characterization and switch utilization using deterministic bounding interval dependent traffic models. In *Proc. IEEE Infocom '95*, pages 1137–1145, April 1995.

[16] E. W. Knightly, D. E. Wrege, J. Liebeherr, and H. Zhang. Fundamental Limits and Tradeoffs of Providing Deterministic Guarantees to VBR Video Traffic. In *Proc. ACM Sigmetrics '95*, pages 98–107, May 1995.

[17] J. Liebeherr, D. E. Wrege, and Domenico Ferrari. Exact Admission Control in Networks with Bounded Delay Services. Technical Report CS-94-29, University of Virginia, July 1994.

[18] S. Low and P. Varaiya. A Simple Theory of Traffic and Resource Allocation in ATM. In *Proc. IEEE Globecom '91*, pages 1633–1637, 1991.

[19] S. Low and P. Varaiya. Burstiness Bounds for Some Burst Reducing Servers. In *Proc. IEEE Infocom '93*, pages 2–9, 1993.

[20] B. Melamed and B. Sengupta. TES Modeling of Video Traffic. *IEICE Transactions on Communications*, E75-B(12):1292–1300, 1992.

[21] A. K. Parekh and R. G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, June 1993.

[22] E. P. Rathgeb. Modeling and Performance Comparison of Policing Mechanisms for ATM Networks. *IEEE Journal on Selected Areas in Communications*, 9(4):325–334, April 1991.

[23] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. Technical Report 101, Institute of Computer Science, University of Wurzburg, February 1995. The traces used in this paper are available via anonymous ftp from the site ftp-info3.informatik.uni-wuerzburg.de in the directory /pub/MPEG/.

[24] J. S. Turner. New Directions in Communications (or Which Way to the Information Age?). *IEEE Communications Magazine*, 25(8):8–15, October 1986.