

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

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. A key component of resource allocation is the traffic characterization of video sources that determines the resources required to support video connections. In this study, we propose a method for characterizing VBR video traffic with a fixed number of leaky buckets in networks with a deterministic service. We explore tradeoffs of network utilization in two directions: (1) the number of leaky buckets used for traffic characterization, and (2) the amount of information from a video sequence used to produce the characterization. We evaluate our method with a set of 30-minute long MPEG-compressed video traces.

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 require large amounts of bandwidth as well as minimum guarantees on their *quality-of-service* (QoS), for example, bounded delays, minimum throughput, and bounded jitter. 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 *policing mechanisms*. At connection establishment time, admission control functions decide whether traffic on a new connection can be supported 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, 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, 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 policing mechanisms that 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 for VBR video have been proposed in the literature. Several studies (see [7] for a recent survey) characterize VBR video traffic with stochastic processes that capture both the distribution of number of bits per frame as well as the autocorrelation function of VBR video sequences. 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 [3, 9, 10, 12] consider deterministic characterizations for video traffic. In [9], Knightly and Zhang present the Deterministic Binding Interval Dependent (D-BIND) traffic model that characterizes video traffic with a set of rate-interval pairs $\{(R_j, I_j) \mid 1 \leq j \leq n\}$ that define an n -segment piecewise linear function. The numerical examples presented in [9] suggest that the D-BIND model can characterize VBR video traffic accurately, but the problem of parameter-selection is not addressed.

An important class of traffic characterizations are those that can be policed by leaky bucket mechanisms [16]. Such traffic characterizations provide a good tradeoff between accuracy and simplicity; they can be efficiently implemented [14], and they are shown to accurately characterize VBR video in [10]. A leaky bucket mechanism has a burst parameter σ and a rate parameter ρ , and it can be im-

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

plemented with a single counter and timer. In [12] Low and Varaiya study the tradeoff between the burst σ and the rate ρ for resource allocation. They express the dependency of ρ on σ , and they define a class of leaky bucket parameters $\{(\sigma_j, \rho_j)\}$ that are dominant, i.e., they explicitly show the smallest possible value of ρ for any choice of σ . However, the selection of a fixed number m of these parameters for some number m of leaky buckets is not considered, and the QoS guarantees specified in this work and a more recent work by Chong [3] are specific to rate-proportional processor sharing (RPPS) multiplexers [13].

In this study, we consider two fundamental problems of video characterization. We first address the question of how much information about a video sequence is needed to accurately characterize the sequence. The traffic characterization is based on the so-called empirical envelope which defines the tightest upper bound on traffic generated by a particular video sequence. We present a method for determining an accurate traffic characterization based on a subset of the empirical envelope. The characterization algorithm requires some knowledge of the video sequence, and we investigate the impact of reducing the information used to produce a traffic characterization.

The second problem we consider is the selection of leaky bucket parameters for policing a video sequence. We present a heuristic algorithm that selects parameters for a fixed number of leaky buckets to characterize a VBR video sequence. Parameters are selected based on some subset of the empirical envelope as discussed above. We evaluate the effectiveness of our heuristic as compared to other approaches from the literature using a number of video segments coded with the MPEG compression algorithm [8].

The remainder of this paper is structured as follows. In §2 we review deterministic traffic characterizations of VBR video traffic and discuss the notion of the empirical envelope. In §3 we investigate the amount of information necessary to accurately characterize a VBR video sequence. In §4 we describe an algorithm that selects leaky bucket parameters. We evaluate the effectiveness of our characterization method in §5 with a number of MPEG-compressed video sequences. We conclude our study in §6.

2 Deterministic Characterization of VBR Video Traffic

2.1 Traffic Constraint Functions A^*

A worst-case characterization of a traffic source 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 A^* on A is called a *traffic constraint function* for A if for all times $\tau \geq 0$ and $t \geq 0$ the following holds [4]:

$$A[\tau, \tau + t] \leq A^*(t) \quad (1)$$

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 . For admission control tests of a deterministic service [11], we further require a

traffic constraint function to be *subadditive*, that is, A^* must satisfy the following equation:

$$A^*(t_1 + t_2) \leq A^*(t_1) + A^*(t_2) \quad \forall t_1, t_2 \geq 0 \quad (2)$$

Traffic constraint functions will be used in admission control tests that require subadditive functions [4, 11, 13].

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 [16]. A leaky bucket mechanism (σ, ρ) with rate parameter ρ and burstiness parameter σ limits the traffic on a connection. The traffic constraint function A^* for traffic policed by a leaky bucket mechanism (σ, ρ) is given by the linear constraint [4]:

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

Similarly, if multiple leaky buckets (σ_i, ρ_i) are connected in series (i.e., a multi-level leaky bucket [5]), 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 an m -piecewise linear concave function given as follows [5, 10]:

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

2.2 The Empirical Envelope E^*

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.

We refer to the tightest traffic constraint function for a given traffic source as its *empirical envelope*, denoted by E^* [2, 10]. The empirical envelope E^* of a video sequence is optimal in the sense that, for any subadditive traffic constraint function A^* that satisfies equation (1), $E^*(t) \leq A^*(t)$ for all t . The empirical envelope E^* is given by the following equation [2, 10]:

$$E^*(t) = \sup_{\tau > 0} A[\tau, \tau + t] \quad \forall t > 0 \quad (5)$$

It follows from equation (5) that E^* is subadditive.

We next review a method from [10] 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 $\{f_1, f_2, \dots, f_N\}$, the empirical envelope E^* can then be constructed by first calculating [10]:

$$E^*(ir) = \max_{0 < k < N-i} \sum_{j=k}^{k+i} f_j \quad \text{for } i = 1, 2, \dots, N \quad (6)$$

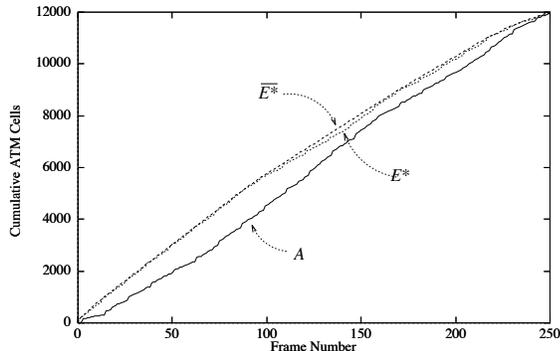


Figure 1: Relationship between $A[0, t]$, $E^*(t)$, and $\overline{E^*}(t)$.

The values of the empirical envelope at times that are not multiples of the frame time are then obtained by spacing the cells in $E^*((i+1)r) - E^*(ir)$ evenly over the frame time $[ir, (i+1)r]$.

Once the empirical envelope 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 E^* is the tightest traffic constraint function that can be policed by leaky buckets. We use $\overline{E^*}$ to denote the concave hull of the empirical envelope E^* . The function $\overline{E^*}$ is clearly a traffic constraint function in the sense of equation (1), and the subadditivity of $\overline{E^*}$ follows from its concavity. We illustrate functions E^* and $\overline{E^*}$ in Figure 1.

3 Approximations of the Envelope

The empirical envelope E^* is the optimal traffic constraint function in terms of accuracy, but it has a significant drawback in that its production is computationally expensive. The number of operations required to compute the empirical envelope 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 $\overline{E^*}$ 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. Next we show how a subset of the envelope can be used to characterize a video sequence in §3.1, and we evaluate the impact of subset size k on the number of connections established by admission control tests in §3.2.

3.1 Approximations S_k and \overline{R}_k

We consider approximations based on a subset of the empirical envelope. In particular, we present in this section approximations that are derived from the first k frames of the envelope, $E^*[0, kr]$, where r is the frame rate of the video sequence. We call $E^*[0, kr]$ the “prefix” of size k , and we consider different methods for extrapolating $E^*[0, kr]$

to obtain subadditive traffic constraint functions that are defined for all times t .

Feasible extrapolations of $E^*[0, kr]$ must be at least as large as the empirical envelope E^* for all times t . Since we only know that E^* is a subadditive function with prefix $E^*[0, kr]$, any extrapolation must be at least as large as the *largest subadditive extension of $E^*[0, kr]$* . We denote this largest subadditive extension by S_k , where S_k is obtained by calculating:

$$S_k(ir) = \begin{cases} E^*(ir) & \text{for } i \leq k. \\ \min_{1 \leq j < i} \{S_k(jr) + S_k((i-j)r)\} & \text{for } i > k. \end{cases} \quad (7)$$

Note that equation (7) only defines S_k for times that are multiples of the frame time r . The values for intermediate values of S_k are then determined by spacing the cells evenly over each frame. The relationship between S_k and the empirical envelope E^* is shown in Figure 2(a).

Although the function S_k is the tightest traffic constraint function that can be obtained directly from a subset of k frames from the envelope, the production of S_k requires a large number of computations. In particular, we see from its definition in equation (7) that the number of computations required to construct S_k is $O(N^2)$, the same number required for computing the empirical envelope itself. Since our goal is an approximation that can be calculated with fewer computations, we turn to other approximation schemes.

Consider a function obtained by a simple repetition of the frames from $E^*[0, kr]$ over all times t . We denote such a function by R_k , where R_k is given by:

$$R_k(t) = \lfloor \frac{t}{kr} \rfloor E^*(kr) + E^*(t - \lfloor \frac{t}{kr} \rfloor (kr)) \quad \text{for } t \geq 0. \quad (8)$$

Observe that R_k can be immediately obtained from the prefix $E^*[0, kr]$, so the computational complexity of computing R_k is $O(kN)$. The relationship between R_k and the empirical envelope E^* is illustrated in Figure 2(b). In the figure, note that the two functions are identical over the first k frame times and that the first k frames of R_k are subsequently repeated.

Although the function R_k satisfies equation (1), it is not necessarily subadditive, and hence does not satisfy our definition of a traffic constraint function. We thus define the function \overline{R}_k to be the concave hull of R_k . The function \overline{R}_k , depicted in Figure 2(c), is clearly a subadditive traffic constraint function. \overline{R}_k can be expressed in terms of the multi-level leaky bucket model as shown in equation (4) as follows:

$$\overline{R}_k(t) = \min_{1 \leq i \leq n} \{\overline{\sigma}_i + \overline{\rho}_i t\} \quad (9)$$

where parameters $\overline{\sigma}_i$ and $\overline{\rho}_i$ are determined by some appropriate algorithm.

3.2 Evaluation

We now empirically evaluate the accuracy of the traffic constraint functions E^* , S_k , and \overline{R}_k with traces of MPEG-compressed video. We investigate the impact of the choice of k on the accuracy of both S_k and \overline{R}_k .

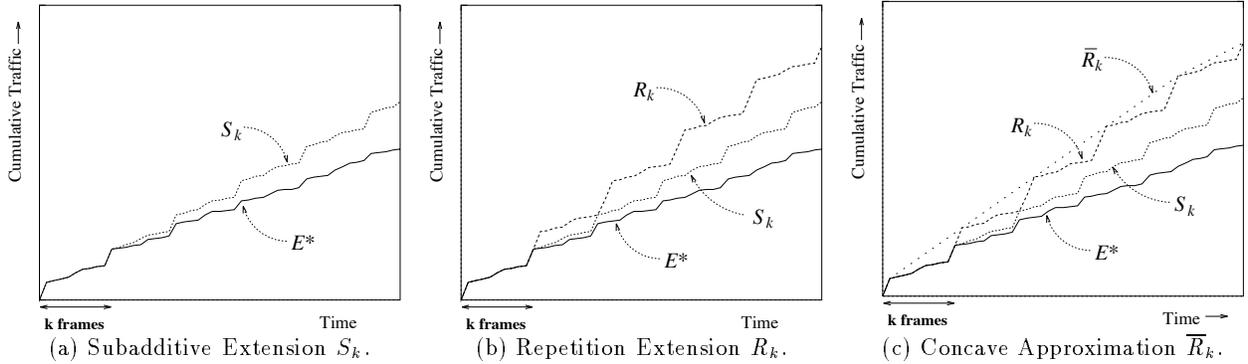


Figure 2: Approximations of the Empirical Envelope

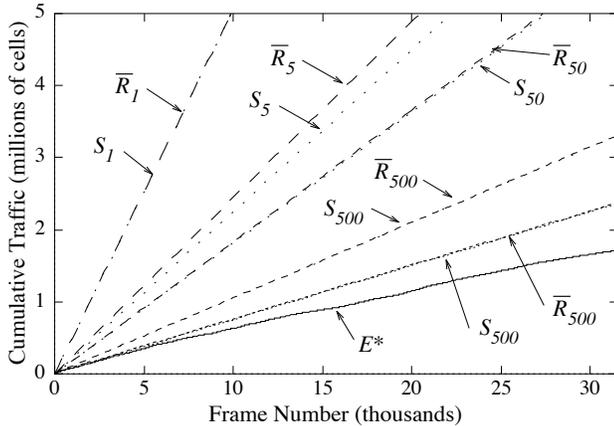


Figure 3: Traffic constraint functions for *News*.

We use two MPEG traces in the evaluation: one from the entertainment film *Jurassic Park* (“*Park*”), and the second from a news broadcast (“*News*”). These sequences were captured on UNIX workstations and encoded with the Berkeley MPEG-encoder [15]. Both *Park* and *News* are 30-minute video sequences with frame sizes of 384x288 and frame pattern IBBPBBPBBPBB.

In these experiments, we consider a single multiplexer that operates at 155 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

Figure 3 illustrates traffic constraint functions for the *News* trace. We show the empirical envelope E^* as well as functions S_k and \bar{R}_k for various values of k . In the figure, the empirical envelope E^* is shown as the lowest curve, while the functions S_k and \bar{R}_k are depicted by dotted and dashed curves, respectively. As expected, the approximation functions estimate the empirical envelope E^* more accurately for larger values of k .

A key observation from Figure 3 is that the function \bar{R}_k approximates S_k closely for all values of k considered; only \bar{R}_5 and S_5 differ considerably. Since S_k is the tightest

¹The exact admission control test for FCFS multiplexers is given by $d \geq \sum_{j \in \mathcal{N}} A_j^*(t) - t + s$ for all $t \geq 0$ [4]. 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 s denotes the maximum cell transmission time.

traffic characterization that can be produced from k frames of the empirical envelope, we note that \bar{R}_k provides an accurate characterization.²

To better evaluate the accuracy of each characterization, we next consider the utilizations that can be achieved at a network switch using traffic constraint functions \bar{R}_k . In Figure 4, we illustrate the network utilization obtained at a multiplexer using E^* as well as \bar{R}_k for various values of k . 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 4(b) that the traffic constraint function \bar{R}_2 can be used to support 61 *News* connections for delay bounds larger than 35ms.

The general trend in both graphs is that the number of connections accepted using \bar{R}_k as the traffic constraint function is increasing in k . For example, note in Figure 4(a) that \bar{R}_{25} and \bar{R}_{50} admit the same number of connections as the empirical envelope for delay bounds up to 90ms and 230ms, respectively.

Note that for both video sequences, the function \bar{R}_{200} admits the same number of connections as the empirical envelope E^* for delay bounds up to 500 milliseconds. Thus, the first 200 frames of the envelope (i.e., \bar{R}_k where $k = 200$) are sufficient to characterize both video sequences for the delay bound range considered.

4 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 \bar{R}_k that can be policed by a number of leaky buckets $(\bar{\sigma}_i, \bar{\rho}_i)$. However, the number of leaky buckets required for policing \bar{R}_k can be large. For example, 12 leaky bucket mechanisms are needed to police \bar{R}_{200} for the *News* sequence described in §3.2. In an actual network, the number of leaky buckets available to police a connection may be limited to only two or three.

In this section, we show how to reduce the number of

²Results for the *Park* sequence, omitted due to space limitations, are similar to those for the *News* sequence shown.

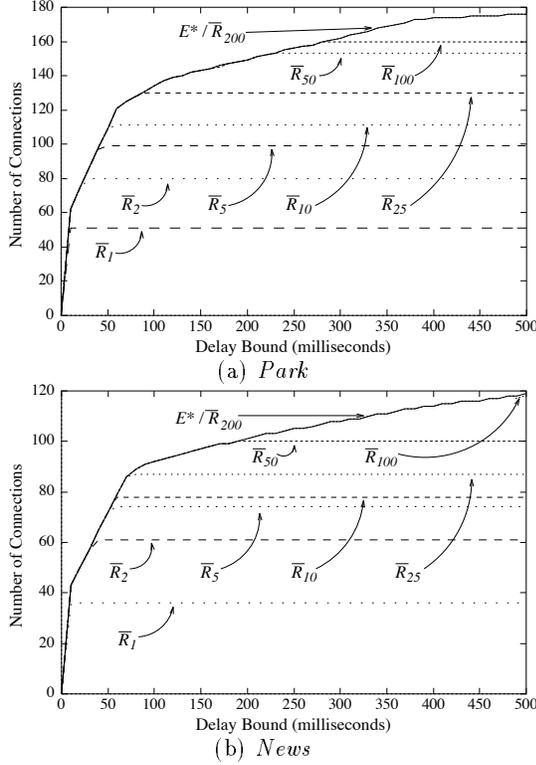


Figure 4: Utilization comparison.

leaky buckets used to police the characterization. To minimize the difference between the original characterization and the approximation, we require: (1) a cost function that expresses the difference between two traffic characterizations, and (2) an algorithm to minimize such a cost function.

We formulate the problem as follows: Given a function \bar{R}_k that can be policed by n leaky buckets (see equation (9)):

$$\bar{R}_k(t) = \min_{1 \leq i \leq n} \{\bar{\sigma}_i + \bar{\rho}_i t\}$$

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

$$B^*(t) = \min_{1 \leq i \leq m} \{\sigma_i + \rho_i t\} \quad (10)$$

Assume that we are given a cost function $C(B^*, \bar{R}_k)$ that expresses the difference between B^* and \bar{R}_k . We select parameters (σ_i, ρ_i) to solve the following optimization problem:

$$\begin{aligned} &\text{minimize a cost function } C(B^*, \bar{R}_k) \\ &\text{subject to the constraint } \bar{R}_k(t) \leq B^*(t) \quad \forall t \geq 0. \end{aligned}$$

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 $C(B^*, \bar{R}_k)$ and present a heuristic algorithm to solve the optimization problem.

4.1 Cost Function $C(B^*, \bar{R}_k)$

Here, we describe a cost function $C(B^*, \bar{R}_k)$ that expresses the difference between the two functions B^* and \bar{R}_k . The minimization of C subject to the constraint $\bar{R}_k(t) \leq B^*(t)$ for all times $t \geq 0$ 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 apparent 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 \bar{R}_k more closely for small values of t .

One can find a number of cost functions with different tradeoffs. Due to space limitations, we do not evaluate different choices for C in this paper. Here we choose the following cost function to be used in the remainder:

$$C(B^*, \bar{R}_k) = \int_0^{k \cdot r} \frac{(B^*(t) - \bar{R}_k(t))}{\bar{R}_k(t)} dt, \quad (11)$$

We considered several cost functions, and we selected the function shown in equation 11 because it produced good empirical results.

4.2 A Heuristic Algorithm

We next present a heuristic algorithm that selects parameters for a fixed number m of leaky buckets to produce a traffic constraint function B^* of low cost. The algorithm takes as input the function \bar{R}_k , the number m of leaky buckets (σ_i, ρ_i) available to police B^* , the cost function $C(B^*, \bar{R}_k)$, and a sensitivity parameter $\epsilon > 0$. The heuristic algorithm produces as output a set of m leaky bucket pairs (σ_i, ρ_i) . The approach of the 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 $C(B^*, \bar{R}_k)$. The algorithm is presented in Figure 5.

The initialization of the pairs (σ_i, ρ_i) is shown in steps 2 through 5 of Figure 5. Observe that these values are selected from the set of pairs $\{(\bar{\sigma}_j, \bar{\rho}_j)\}$ that determine the function \bar{R}_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 reduce the cost function C , and the process completes only when the cost cannot be significantly reduced. The critical step of the algorithm is step 9, where a single pair (σ_i, ρ_i) is modified to minimize the cost function. During this step, the values for pairs $\{(\sigma_j, \rho_j) | j \neq i\}$ are fixed, and the pair (σ_i, ρ_i) is selected subject to the constraint that $\sigma_{i-1} < \sigma_i < \sigma_{i+1}$.³ Note that the choice of ρ_i is dependent on σ_i according to the relationship described in [3, 12].

In the empirical results presented in §5, we select the (σ_i, ρ_i) pair of minimum cost in step 9 through an

³Boundary conditions for this selection are $\sigma_1 \geq 0$ and $\sigma_m \leq \bar{\sigma}_n$.

<p>Input: A set of n leaky bucket parameters $\{(\bar{\sigma}_j, \bar{\rho}_j)\}$ that define the function E_k^*, the number m of available leaky buckets, a cost function $C(B^*, \bar{R}_k)$, and a sensitivity parameter ϵ.</p> <p>Output: A set of m leaky bucket parameters (σ_i, ρ_i) to determine the traffic constraint function B^*.</p>
<pre> 1. Procedure Parameterize ($\bar{R}_k, m, C(B^*, \bar{R}_k), \epsilon$) 2. For $i = 1$ To m /* Initialize (σ_i, ρ_i) */ 3. $\sigma_i \leftarrow \bar{\sigma}_{\lfloor \frac{i}{m} \rfloor}$ 4. $\rho_i \leftarrow \bar{\rho}_{\lfloor \frac{i}{m} \rfloor}$ 5. End For 6. Do /* Greedy modifications */ 7. Cost $\leftarrow C(B^*, \bar{R}_k)$ 8. For $i = m$ Down To 1 9. Select (σ_i, ρ_i) to minimize $C(B^*, \bar{R}_k)$ 10. End For 11. While (Cost - $C(B^*, \bar{R}_k) > \epsilon$) 12. Output $B^* \leftarrow \min_{1 \leq i \leq m} \{\sigma_i + \rho_i t\}$ 13. End Procedure </pre>

Figure 5: Parameterization algorithm.

exhaustive search through all possible values of σ_i . However, with ρ_i expressed in terms of σ_i , it is possible to write $C(B^*, \bar{R}_k)$ with σ_i as the only independent variable, and the selection can be determined analytically by setting $\frac{\partial C}{\partial \sigma_i}$ to zero. Also, while we do not make guarantees on the running time of the algorithm, the examples that we ran converged rapidly. In all examples using a sensitivity parameter $\epsilon = 0$, no more than six iterations were required.

5 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 §3 and §4, we can make recommendations for a characterization method as follows. In §3, we found that 200 frames of the envelope are sufficient to produce an accurate VBR characterization, i.e., we use the function \bar{R}_{200} . In §4, we chose a particular cost function C 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 *Park* and *News* described in §3.2 and a single multiplexer at a switch that operates at 155 Mbps.

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 [10]. These benchmarks are described in the following:

(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 [14]. The traffic constraint function for

peak-rate allocation is given as follows:

$$A_{\text{peak}}^*(t) = \rho t \quad (12)$$

(b) *Dual bucket*: The dual bucket scheme uses a 2-level leaky bucket that 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 ρ_1 of the connection. The value of σ_1 is selected as small as possible subject to the constraint that $\sigma_1 + \rho_1 t \geq E^*(t)$ for all times t . For the dual-bucket scheme, the traffic constraint function is given by:

$$A_{\text{db}}^*(t) = \min\{\rho_0 t, \sigma_1 + \rho_1 t\} \quad (13)$$

(c) *Concave hull* [10]: 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 the empirical envelope. If the concave hull of the empirical envelope is an n -piecewise linear function characterized by a set of n parameters $\{(\bar{\sigma}_j, \bar{\rho}_j) | 1 \leq j \leq n\}$, where $\bar{\sigma}_i < \bar{\sigma}_j$ for $i < j$, then the concave hull approach forms its traffic constraint function from the first m such parameters, that is:

$$A_{\text{hull}}^*(t) = \min_{1 \leq j \leq m} \{\bar{\sigma}_j + \bar{\rho}_j t\} \quad (14)$$

We evaluate the accuracy of a traffic constraint function as follows. For a given characterization A^* , we plot the ratio of the number of connections admitted using A^* to the number admitted using the optimal characterization with the empirical envelope E^* as a function of the delay bound of those connections. That is, for a given function A^* we plot:

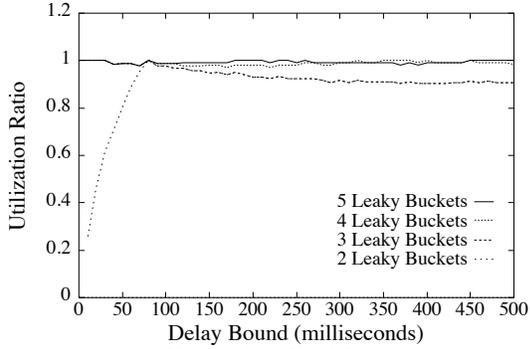
$$\text{Utilization Ratio}(A^*) = \frac{\# \text{ Connections with } A^*}{\# \text{ Connections with } E^*} \quad (15)$$

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

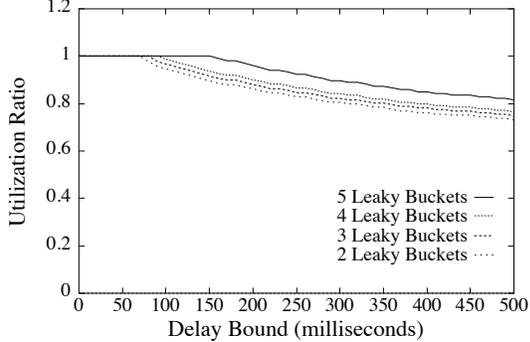
5.1 Number of Leaky Buckets

In this section, 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 consider the two traffic constraint functions that yield a variable number of leaky buckets: B^* as produced by our heuristic algorithm, and A_{hull}^* as produced by the concave hull scheme.

Figures 6(a) and 6(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 in Figure 6(a) that the utilization ratio for B^* is close to 1 over the entire interval if more than 2



(a) Heuristic Algorithm.



(b) Concave Hull Approach.

Figure 6: Leaky bucket evaluation.

leaky buckets are used. Observe in Figure 6(b) that the utilization ratio for A_{hull}^* declines significantly for higher delay bounds even if 5 leaky buckets are used.

In summary, 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.

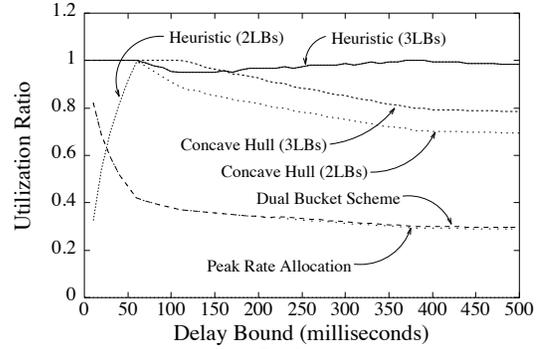
5.2 Single Traffic Type

Figures 7(a) and 7(b) show the utilization ratios of *Park* and *News* connections, respectively, for the entire suite of traffic characterizations described previously, namely, B^* , A_{peak}^* , A_{db}^* , and A_{hull}^* . We depict characterizations B^* and A_{hull}^* for $m = 2$ and $m = 3$.⁴

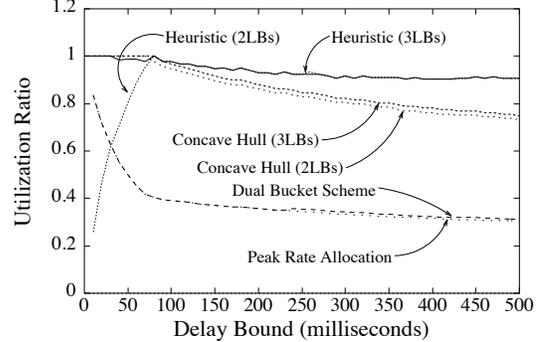
Notice the poor performance of both A_{peak}^* and A_{db}^* in both graphs. The function A_{peak}^* achieves a utilization ratio of less than 40% for delay bounds greater than 60ms for these video sequences. The function A_{db}^* performs only slightly better than A_{peak}^* . The peak-rate and dual-bucket 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 dependent on the number of leaky buckets m . Notice in Figure 7(a) that the point of decline begins at $d = 70\text{ms}$ for $m = 2$ and at $d = 110\text{ms}$

⁴Note 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) *Park*.



(b) *News*.

Figure 7: Comparison for single traffic type.

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 *Park* and 80ms in *News*.

5.3 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 *Park* sequence and those carrying the *News* sequence. All connections of the same type have identical delay bounds. We consider a switch that uses Earliest-Deadline-First (EDF) multiplexing, that is, cells are ordered for transmission according to their deadlines [6].⁵

Figure 8 illustrates the maximum number of *Park* and *News* connections that can be simultaneously supported at an EDF multiplexer for the specified delay bounds. The delay bounds are $d_{\text{Park}} = 200\text{ms}$ and $d_{\text{News}} = 50\text{ms}$ for Figure 8(a), while the delay bounds for Figure 8(b) are $d_{\text{Park}} = 500\text{ms}$ and $d_{\text{News}} = 200\text{ms}$.

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 8(a), A_{peak}^* can support 30 *News* con-

⁵The exact admission control test for EDF multiplexers is given by [11]: $t \geq \sum_{j \in \mathcal{N}} A_j^*(t - d_j) + s(t)$ for all $t \geq d_1$. Here, \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 $s(t)$ is the maximum cell transmission time from any connection k with $d_k > t$.

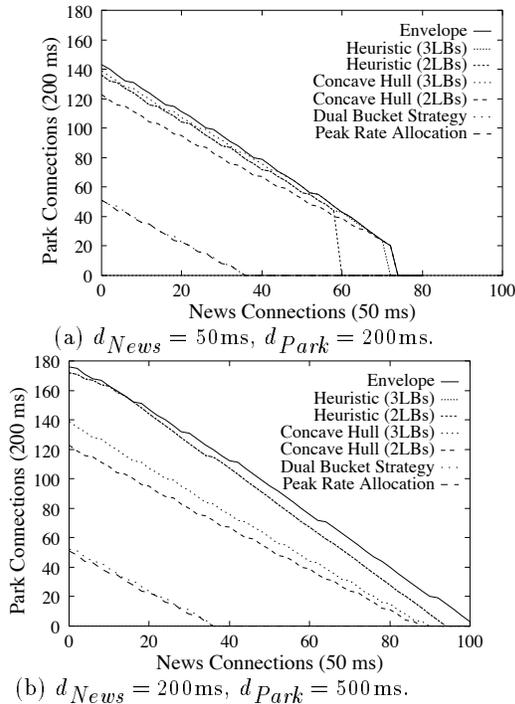


Figure 8: Comparison for multiple traffic types.

nections and 8 *Park* connections, while B^* with two leaky buckets can support 92 *Park* connections for the same number of 30 *News* connections.

In Figure 8(a), observe that B^* admits more connections than A_{hull}^* for $m = 2$ when more than 45 *Park* connections are supported. In both graphs the use of our heuristic algorithm with three leaky buckets yields a characterization that is close to optimal. In Figure 8(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 $m = 3$.

6 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, we 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.

References

- [1] ATM Forum, ATM Forum Traffic Management Specification Version 4.0, October, 1995.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] R. L. Cruz. A Calculus for Network Delay, Part II: Network Analysis. *IEEE Transactions on Information Theory*, 37(1):132–141, January 1991.
- [6] 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.
- [7] V. Frost and B. Melamed. Traffic Modelling for Telecommunications Networks. *IEEE Communications Magazine*, 32(3):70–81, March 1994.
- [8] D. Le Gall. MPEG: A Video Compression Standard for Multimedia Applications. *Communications of the ACM*, 34(4):46–58, April 1991.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] S. Low and P. Varaiya. A Simple Theory of Traffic and Resource Allocation in ATM. In *Proc. IEEE Globecom '91*, pages 1633–1637, 1991.
- [13] 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.
- [14] 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.
- [15] 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/.
- [16] J. S. Turner. New Directions in Communications (or Which Way to the Information Age?). *IEEE Communications Magazine*, 25(8):8–15, October 1986.