A Note on Statistical Multiplexing and Scheduling in Video Networks at High Data Rates *

Jörg Liebeherr
Department of Computer Science
University of Virginia
Charlottesville, VA 22904

January 2002

Abstract

This paper comments on the impact of link scheduling and statistical multiplexing on the multiplexing gain at high data rates. The multiplexing gain is evaluated using the number of MPEG video traces that can be provisioned with delay guarantees on a network link. The presented data indicates that, at high transmission rates, the multiplexing gain is substantial and is dominated by the effects of statistical multiplexing.

Key Words: Statistical Multiplexing, Statistical Service, Scheduling, Quality-of-Service.

1 Introduction

A distinguishing property of packet switching networks is that they achieve multiplexing gain by sharing resources. We speak of multiplexing gain when providing a certain grade of service to a group of traffic flows requires less network resources per flow than providing the same grade of service individually to each flow.

Research on Quality-of-Service (QoS) networks in the 1990s showed that multiplexing gain can be considerable even when service requirements of traffic are stringent and traffic is bursty, in the sense that the rate of traffic varies greatly over time. An example of a network service with stringent requirements is a bounded delay service, which guarantees that all traffic from a flow satisfies worst-case delay bounds and that no packets are dropped [4]. An example of bursty traffic is MPEG-1 video traffic, which has correlations of traffic over multiple time scales [5]. The burstiness of MPEG traffic is illustrated in Figure 1 for a sequence from an MPEG-1 compressed motion picture. In [13], it was shown that packet networks with a bounded delay service for MPEG-1 video traffic can capture multiplexing gain.

There are several approaches to improve multiplexing gain in a packet switching network with service guarantees. For example, buffering traffic which exceeds a given burstiness constraint at the network entrance effectively smoothes the traffic rate [14]. However, due to possibly significant

*This work is supported in part by the National Science Foundation through grant and ANI-0085955.

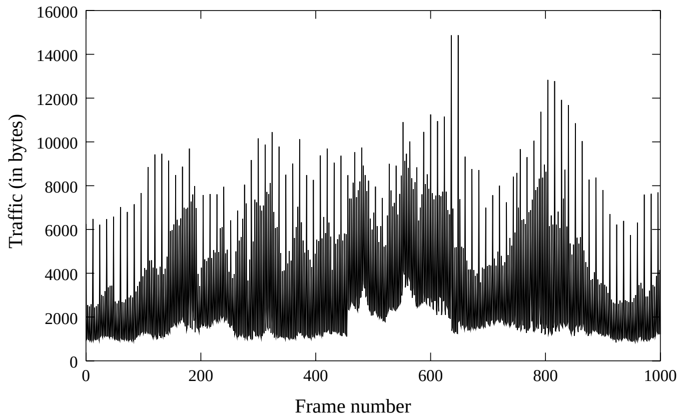


Figure 1: MPEG video traffic. Size of video frames for 1000 frames from the movie "Silence of the Lambs". The frame rate is 24 frames per second.

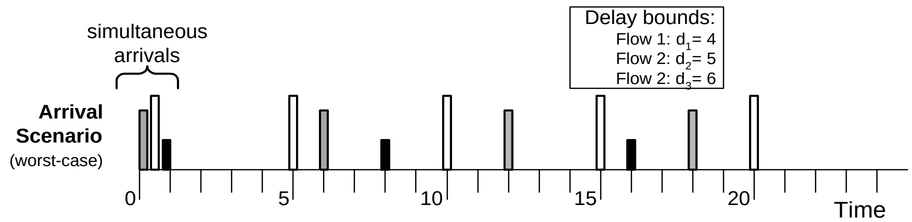
delays, smoothing is generally reserved for applications with a high delay tolerance. Scheduling algorithms at the output links of packet switches can improve multiplexing gain by transmitting backlogged traffic in the order which best satisfies the service requirements. Also, one can extract multiplexing gain by exploiting statistical properties of traffic. This is referred to as statistical multiplexing gain. By allowing a fraction of traffic to violate its service guarantees, for example, by making probabilistic service guarantees of the form: $Pr[Delay > X] < \varepsilon$, where ε is small, it is feasible to exclude worst-case traffic arrival scenarios where sharing of resources is difficult to achieve.

We refer to Figure 2 for a simple arrival scenario of three traffic flows. The arrival scenario in Figure 2(a) depicts a "worst-case" scenario with simultaneous arrivals from all flows at time $t = 0$. In Figures 2(b) and (c), two transmission scenarios are presented. In the first scenario in Figure 2(b), packets are transmitted in the order of arrivals (and in some arbitrary order for simultaneous arrivals). The figure shows that this scheduling scheme results in a violation of a delay bound for Flow 1 at time $t = 4$. In Figure 2(c), the transmission schedule gives highest priority to the packet with the smallest delay bound. This scheduler does not have delay bound violations. Thus, the scheduling algorithm in Figure 2(c) yields a better multiplexing gain.

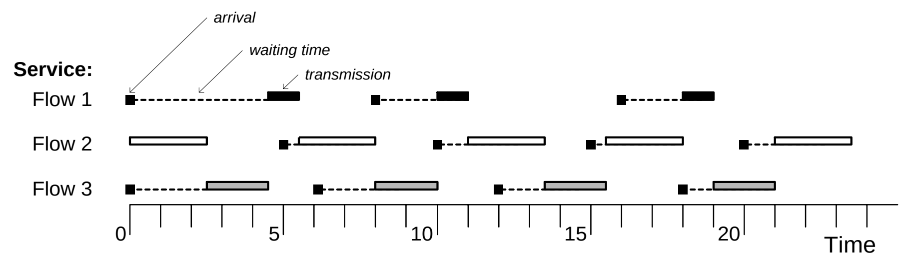
In Figure 3 we show an example which intends to demonstrate the benefits of statistical multiplexing. Figure 3(a) shows a "typical" arrival scenario where simultaneous arrivals of packets are relatively rare. By excluding these rare events, a schedule that transmits packets in the order of arrivals, which resulted in delay bound violations in Figure 3, does not result in a delay bound violation.

There are many studies that have evaluated the impact of scheduling for video transmissions with a bounded delay service, e.g., [13], and many studies have evaluated the statistical multiplexing gain of video transmissions in a packet network, e.g., [6].[1] A direct comparison of the impact of both scheduling and statistical multiplexing has not been done, due to the lack of analytical tools that enable such a comparison. With a recently presented methods to determine the statistical multiplexing gain with various scheduling algorithms [2], such a comparison has become feasible.

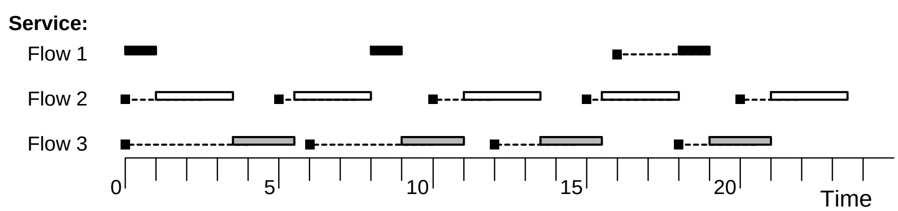
[1] The related work on these subjects is substantial, and a discussion is beyond the scope of this note. Specifically, the list of references of this note is incomplete, and highlights works co-authored by the author.



(a) "Worst-case" arrival scenario.

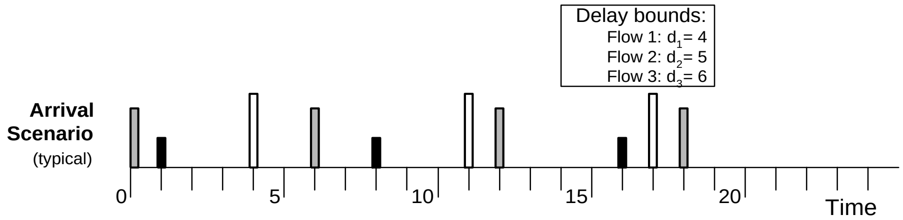


(b) Transmission Scenario 1.

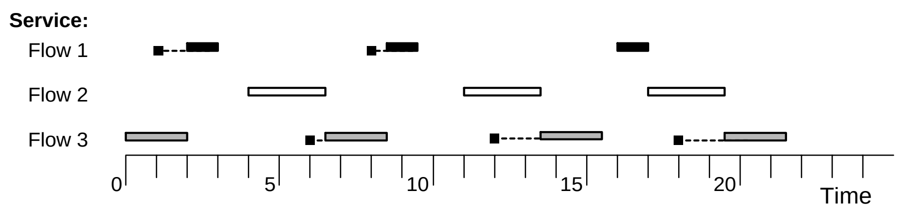


(c) Transmission Scenario 2.

Figure 2: Multiplexing gain through scheduling. In Figure 2(a), packet arrivals are indicated as boxes, where the color of a box indicates the flow type, and the length of the box indicates the transmission time of a packet. The transmission times for Flows 1, 2, and 3 are given by 1, 2.5, and 2 time units, respectively. The delay bounds for packets are given by $d_1 = 4$ for Flow 1, $d_2 = 5$ for Flow 2, and $d_3 = 6$ for Flow 3. We assume that packets from Flow 1, Flow 2, and Flow 3 arrive at most every eight, five, and six time units, respectively. In the depicted 'worst-case' arrival scenario, packet arrivals from all flows coincide at time $t = 0$. In (b) and (c), two transmission scenarios are presented. Dotted lines present waiting times and the horizontal boxes indicate packet transmissions. In the transmission scenario in (b), packets are transmitted in the order of their arrival. This results in a violation of the delay bound of Flow 1 at time $t = 4$. The scenario in (c) gives higher priority to packets with shorter delay bounds. Here, no delay bound violation occurs.



(a) "Typical" arrival scenario.



(b) Transmission Scenario.

Figure 3: Statistical multiplexing. In a 'typical' arrival scenario, packet arrivals do not coincide. A schedule which transmits packets in the order of arrivals does not result in a violation of delay bounds.

This paper discusses a set of numerical examples from a recent technical report [1] which use the approach in [2] to evaluate the number of video flows that can be provisioned with delay guarantees on a network link. The examples show that at high data rates, statistical multiplexing gain dominates the multiplexing gain, and, in comparison, the multiplexing gain due to scheduling is modest. Since the statistical multiplexing gain is due to the nature of traffic and not part of the network design, the results indicate that a relatively simple QoS network design may achieve a high multiplexing gain. We emphasize, however, that the conclusions of this paper are specific to the shown examples. Also, we emphasize that this paper makes assumptions, such as stationarity of video flows, which may not hold in practice.

The remaining sections of this paper are structured as follows. In Section 2 we specify our assumptions on the traffic and present schedulability conditions for a general class of packet schedulers. In Section 3 we discuss numerical examples using MPEG video traces, and compare the multiplexing gain attainable through scheduling and through statistical multiplexing.

2 Analysis Framework

The evaluation of scheduling and statistical multiplexing in this paper is based on analytical methods and not on experimental measurements. Specifically, we use schedulability conditions for a service with delay guarantees, which verify whether a network link can satisfy delay guarantees for a given set of traffic flows and a given scheduling algorithm. In this section, we discuss the schedulability conditions used in this paper. We first present functions that describe arrivals of single and aggregate traffic flows, and then use these functions in the schedulability conditions.

2.1 Traffic Arrivals and Envelope Functions

We consider arrivals of groups of video flows to an output link of a packet switch with transmission rate C. At the link, a scheduler determines the order in which backlogged traffic is transmitted. Consider a set $\mathcal{C}$ of flows which are partitioned into Q flow types, where $\mathcal{C}_q$ denotes the subset of type-q flows. Delay guarantees for a video flow $j \in \mathcal{C}_q$ are specified in terms of a delay bound d_q for type-j video flows. A delay bound violation occurs if traffic from flow j experiences a delay exceeding d_q.

The traffic arrivals from flow j in the interval $[t_1, t_2)$ are denoted by a function $A_j(t_1, t_2)$ with the following properties:

- **Additivity.** For any $t_1 < t_2 < t_3$, we have $A_j(t_1, t_2) + A_j(t_2, t_3) = A_j(t_1, t_3)$.
- **Subadditive Bound.** A_j is bounded by a deterministic subadditive envelope A_j^* as $A_j(t, t+\tau) \leq A_j^*(\tau)$ for all $t \geq 0$ and for all $\tau \geq 0$.[2]

The selection of "subadditive" bounds is motivated by the result that a bound for a traffic flow, which is not subadditive, can be improved by replacing it with a tighter subadditive bound [3]. Given the traffic arrivals of a video flow, a deterministic envelope for that video flow can be constructed by

$$\mathcal{E}_j^*(\tau) = \sup_{t \geq 0} A_j[t, t+\tau] \qquad \forall \tau \geq 0 \ . \tag{1}$$

In [13], this function is referred to as "empirical envelope", and shown to be the smallest subadditive envelope for a traffic flow. To reduce the number of parameters of the empirical envelope, we apply a method from [13], which approximates the concave hull of the empirical envelope by a piecewise linear function with K segments. For a flow j, the k-th segment of this function is characterized by a burst parameter σ_{jk} and a rate parameter ρ_{jk}, resulting in a subadditive envelope of the form

$$A_j^*(\tau) = \min_{k=1,\dots,K} \{\sigma_{jk} + \rho_{jk}\tau\} \ , \tag{2}$$

where $\{\sigma_{jk}, \rho_{jk}\}_{k=1}^K$ are the parameters of the piecewise linear segments. An algorithm to reduce the number of piecewise linear segments can be found in [11].

To exploit statistical multiplexing, we view the arrivals $A_j(t_1, t_2)$ as a family of random variables, which, in addition to the assumptions above, satisfy the following:

- **Stationarity.** The A_j are *stationary* so that for all $t, t' \geq 0$ we have $Pr[A_j(t, t+\tau) \leq x] = Pr[A_j(t', t'+\tau) \leq x]$.
- **Independence.** The A_i and A_j are stochastically independent for all $i \neq j$.

Within the constraints of these assumptions, we consider arrival scenarios where each video flow exhibits its worst possible ('adversarial') behavior. However, even if flows individually behave in a worst-case fashion, as allowed by their subadditive bounds, the independence assumption prevents the flows from 'conspiring' to yield a joint worst-case behavior. These assumptions effectively exclude scenarios as shown in Figure 2, where arrival bursts from multiple flows coincide.

[2] A function f is subadditive if $f(t_1 + t_2) \leq f(t_1) + f(t_2)$, for all $t_1, t_2 \geq 0$.

For the calculation of statistical multiplexing gain we will take advantage of the notion of *effective envelopes* [2]. Effective envelopes are functions that are, with high probability, upper bounds on multiplexed traffic from a set of flows satisfying the given assumptions. Effective envelopes have been shown to be a useful tool for calculating the statistical multiplexing gain at a network node [9, 8].[3]

Consider the set $\mathcal{C}_q$ of type-q flows. We use $A_{\mathcal{C}_q}$ to denote the aggregate arrivals of all type-q flows, that is, $A_{\mathcal{C}_q}(t, t+\tau) = \sum_{j \in \mathcal{C}_q} A_j(t, t+\tau)$. Let N_q denote the number of flows in set $\mathcal{C}_q$. All flows of the same type have the same subadditive bound. Thus, we use A_q^* to denote the bound of a type-q flow with $A_j^*(\tau) = A_q^*(\tau)$ for all $j \in \mathcal{C}_q$.

An effective envelope for $A_{\mathcal{C}_q}(t, t+\tau)$ is a function $\mathcal{G}_{\mathcal{C}_q}$ with:

$$Pr\Big[A_{\mathcal{C}_q}(t, t+\tau) \le \mathcal{G}_{\mathcal{C}_q}(\tau; \varepsilon) \Big] \ge 1 - \varepsilon, \quad \forall\, t,\ \tau \ge 0 \ . \tag{3}$$

Thus, an effective envelope provides a bound for the aggregate arrivals $A_{\mathcal{C}_q}$ for each time interval of length τ, which is violated with probability at most ε.

Explicit expressions for effective envelopes can be obtained with large deviation techniques. The construction of effective envelopes $\mathcal{G}_{\mathcal{C}_q}$ for a set $\mathcal{C}_q$ of type-q flows uses the moment generating function of A_j, denoted as $M_j(s,t) = E[e^{A_j(0,t)s}]$, where $E[.]$ denotes the expected value. As shown in [2], with the above assumptions, it can be shown that, for a flow $j \in \mathcal{C}_q$, $M_j(s,t) \le \overline{M}_q(s,t)$, where

$$\overline{M}_q(s,t) \;=\; 1 + \frac{\rho_q\, t}{A_q^*(t)} \big(e^{s A_q^*(t)} - 1 \big) \ , \tag{4}$$

and where $\rho_q := \lim_{t \to \infty} A_q^*(t)/t$ is assumed to exist. With the independence of flows we obtain from the Chernoff bound that

$$Pr\{ A_{\mathcal{C}_q}(t) \ge x \} \le e^{-xs} \overline{M}_q(s,t)^{N_q} \ . \tag{5}$$

From here, we can obtain an effective envelope as follows

$$\mathcal{G}_{\mathcal{C}_q}(t) \;=\; \inf_{s>0} \frac{1}{s} \big(N_q \log \overline{M}_q(s,t) + \log \varepsilon^{-1} \big) \ . \tag{6}$$

Note that the effective envelope can also be expressed in terms of the *effective bandwidth* given by $\alpha(s,t) := \frac{1}{st} \log M_j(s,t)$, which is widely used in the literature on statistical multiplexing [7].

We will use the effective envelope given by Eqn. (6) in the numerical examples in Section 3. We will see that effective envelopes capture the statistical multiplexing gain well. If N_q is large, generally, we have that $\mathcal{G}_{\mathcal{C}_q}(t) \ll N_q \cdot A_q^*(t)$.

2.2 Schedulability

In a packet-switched network, packets from a particular flow traverse the network on a path of packet switches and links. Figure 4 shows a sketch of a typical packet switch. The shown switch

[3] In [2] two notions of effective envelopes are introduced, called local effective envelope and global effective envelope. In this paper, we only use local effective envelopes and refer to them as effective envelopes.

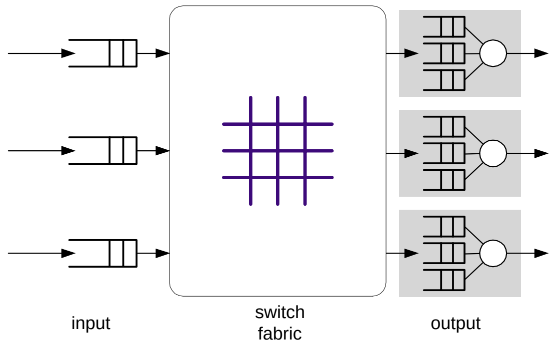


Figure 4: Packet switch.

performs buffering at the input and the output of a switch fabric. For each output link, a scheduler determines the order in which backlogged packets are transmitted. The selection of a particular scheduling discipline for an output link involves a tradeoff between the need to support a large number of flows with diverse delay requirements and the need for simplicity of the scheduling operations. Here we consider well-known scheduling disciplines First-Come-First-Served (FCFS), Static Priority (SP), and Earliest-Deadline-First (EDF).

- *First-Come-First-Served (FCFS):* A FCFS scheduler transmits packets in the order of their arrival. The main advantage of FCFS is its simplicity. However, since a FCFS scheduler treats all traffic in the same way, it is not well suited to support different delay guarantees.
- *Static Priority (SP):* An SP scheduler assigns to each flow type a priority level and a separate FCFS queue. Traffic is always transmitted from the highest priority FCFS queue. By convention, a lower index indicates a higher priority level.
- *Earliest-Deadline-First (EDF):* An EDF scheduler tags traffic with a deadline which is set to the arrival time plus the delay bound d_q, and transmits traffic in the order of deadlines. It has been shown that the EDF scheduling algorithm is optimal for a service with delay guarantees, in the sense that, among all scheduling algorithm, it can support the most flows with deterministic delay guarantees [10].

Given a scheduling algorithm and a set of delay bounds, a *schedulability condition* verifies that, for all flows, the delay of each packet is less than its required delay bound. In the following, we will not take into consideration that packet transmissions on a link cannot be preempted. This assumption is reasonable when packet transmission times are short. We assume that the transmission rate of the link is normalized, that is $C = 1$.

In addition to the scheduling algorithm and statistical multiplexing, the number of flows that are admitted by a schedulability condition depends on the method in which traffic is characterized in the schedulability condition, and on the tightness of the schedulability condition itself. If a traffic characterization method overestimates the actual traffic of a flow, the schedulability condition will underestimate the achievable multiplexing gain. Since schedulability conditions are expressed in terms of bounds, loose bounds have the same effect. The traffic characterization used in this paper, which is based on the empirical envelope of an MPEG trace, has been demonstrated to be very accurate and cannot be easily improved [11]. Likewise, the schedulability conditions presented here, from [2], are tight in many cases.

A schedulability condition for a set of flows with deterministic delay guarantees is given as follows. If a set of flows which satisfies the assumptions on additivity and subadditive bounds, no delay violation occurs if the d_q are selected such that, for all $\tau \geq 0$, we have

$$\sup_{\tau} \left\{ \sum_{p} \sum_{j \in \mathcal{C}_p} A_j^*(x_p^{q,\tau} + \tau) - \tau \right\} \leq d_q \ , \tag{7}$$

where $x_p^{q,\tau}$ is given by

$$\begin{array}{lll} \text{FCFS :} & x_p^{q,\tau} = 0 & \\ \text{SP:} & x_p^{q,\tau} = \begin{cases} -\tau & ,p > q \\ 0 & ,p = q \\ d_q & ,p < q \end{cases} & \\ \text{EDF:} & x_p^{q,\tau} = \max\{-\tau, d_q - d_p\} & . \end{array}$$

Next we describe a schedulability condition for a statistical service which exploits statistical multiplexing gain. The condition requires stationarity and independence of flows. For the purposes of this note, we make the convenient assumption that

$$Pr\left[\sup_{\tau} \left\{ \sum_{p} A_{\mathcal{C}_p}(t-\tau, t + x_p^{q,\tau}) - \tau \right\} \leq d_q \right] \approx \inf_{\tau} Pr\left[\sum_{p} A_{\mathcal{C}_p}(t-\tau, t + x_p^{q,\tau}) - \tau \leq d_q \right] \ . \tag{8}$$

Assuming that Eqn. (8) holds with equality, we have that an arbitrary type-q arrival has a deadline violation with probability $< \varepsilon$ if d_q is selected such that

$$\sup_{\tau} \left\{ \sum_{p} \mathcal{G}_{\mathcal{C}_p}(x_p^{q,\tau} + \tau, \varepsilon/Q) - \tau \right\} \leq d_q \ . \tag{9}$$

Remark: The drawback of the condition in Eqn. (9) is the dependence on the assumption in Eqn. (8). Since the assumption does not hold in general, the resulting schedulability condition may be overly optimistic. In [2], it was shown that a more conservative effective envelope, called 'global effective envelope' can result in a rigorous bound which does not require to make the assumption of Eqn. (8). Also, it should be noted that the assumption of stationarity for MPEG traffic sources may be too strong.

3 Evaluation of Multiplexing Gain

We will now evaluate the multiplexing of MPEG video sources using the schedulability conditions from Subsection 2.2. The performance measure for the evaluation is the number of video flows that can be provisioned on a link with delay guarantees. The following allocation methods will be considered in the evaluation:

- **Peak Rate Allocation:** A peak rate allocation reserves bandwidth at the peak rate of a traffic flow. While a peak rate allocation yields deterministic delay bound guarantees, it does not exploit any multiplexing gain, and, therefore, is an inefficient method for achieving delay guarantees. The number of flows that can be supported with a peak rate allocation serves as a lower bound for any method for provisioning delay guarantees.

- **Deterministic Allocation:** Here, we use the schedulability condition from Eqn. (7) and obtain a service with deterministic delay guarantees. A deterministic allocation captures multiplexing gain achievable through scheduling, but does not exploit statistical multiplexing gain.

- **Statistical Allocation:** We use Eqn. (9) to determine admissibility of flows with the effective envelope from Eqn. (6). The service guarantees of the statistical allocation are probabilistic delay guarantees. The statistical allocation exploits statistical multiplexing gain, as well as the multiplexing gain due to scheduling.

- **Average Rate Allocation:** An average rate allocation merely guarantees average throughput and finiteness of delays, but does not support delay guarantees. Since the number of flows admitted with an average rate allocation is always close to 100% of the link capacity, the average rate allocation provides an upper bound for the number of flows admitted by an allocation method.

Movie Trace	Average frame size (bits/frame)	Mean Rate (Mbps)	Peak Rate (Mbps)
Terminator	10,904	0.261	1.90
Lambs	7,312	0.171	3.22

Table 1: Parameters of the movie traces.

We use statistics of MPEG-compressed video as traffic sources. The evaluation with MPEG streams is analogous to that in [13], which explored the multiplexing gain of a service with deterministic delay guarantees. In our examples, a number of MPEG-compressed video sequences are multiplexed on 622-Mbps links. We assume that the video sequences are played continuously with a randomly shifted starting time chosen uniformly over the length of the trace. We consider two traces of MPEG-compressed video from [12]. The first trace is taken from the movie "Terminator 2" (*Terminator*), and the second trace is obtained from the movie "Silence of the Lambs" (*Lambs*). Both traces are digitized to 384 by 288 pixels with 12 bit color information and compressed at 24 frames per second with frame pattern IBBPBBPBBPBB (12 frames). Each sequence consists of a total of 40,000 video frames, corresponding to approximately 30 minutes of video. The data of these traces is given in terms of frame sizes. In Table 1, we show some of the parameters of the traces. We assume that the arrival of a frame is spread evenly over an interframe interval (of length 1/24 s); Hence, a (normally instantaneous) frame arrival occurs at a constant rate.

For each of the MPEG traces, we assume that a deterministic regulator is obtained using the method described in [13]: (1) empirical envelopes are obtained from the MPEG traces using

Silence of the Lambs (*Lambs*)	
Rate parameter (Bits per second)	Burst parameter (Bits)
$\rho_1 = 3,221,376.0$	$\sigma_1 = 0.0$
$\rho_2 = 867,008.0$	$\sigma_2 = 98,098.7$
$\rho_3 = 759,628.8$	$\sigma_3 = 156,262.4$
$\rho_4 = 694,336.0$	$\sigma_4 = 246,149.3$
$\rho_5 = 656,472.0$	$\sigma_5 = 321,122.0$
$\rho_6 = 647,850.7$	$\sigma_6 = 372,131.6$
$\rho_7 = 563,438.9$	$\sigma_7 = 1,126,242.3$
$\rho_8 = 502,912.0$	$\sigma_8 = 2,042,261.3$
$\rho_9 = 448,013.1$	$\sigma_9 = 2,911,892.3$
$\rho_{10} = 208,800.0$	$\sigma_{10} = 3,157,800.0$

Terminator 2 (*Terminator*)	
Rate parameter (Bits per second)	Burst parameter (Bits)
$\rho_1 = 1,909,440.0$	$\sigma_1 = 0.0$
$\rho_2 = 869,056.0$	$\sigma_2 = 43,349.3$
$\rho_3 = 791,680.0$	$\sigma_3 = 75,589.3$
$\rho_4 = 624,776.3$	$\sigma_4 = 165,995.4$
$\rho_5 = 592,576.0$	$\sigma_5 = 214,296.0$
$\rho_6 = 425,421.1$	$\sigma_6 = 485,922.6$
$\rho_7 = 361,641.5$	$\sigma_7 = 679,919.0$
$\rho_8 = 346,464.0$	$\sigma_8 = 961,968.0$
$\rho_9 = 317,920.00$	$\sigma_9 = 1,563,770.7$
$\rho_{10} = 304,514.7$	$\sigma_{10} = 1,853,100.7$

Table 2: Rate and burst parameters of the movie traces using the algorithm from [13].

Eqn. (1), (2) the concave hull of the empirical envelopes is approximated by a piecewise linear function, (3) the segments of the resulting functions yield a set of rate and burst parameters, which determines a deterministic envelope function as in Eqn. (2). In Table 2 we present the parameters which are obtained from the two MPEG traces with this method.

3.1 Example 1: Comparison of Envelope Functions for MPEG Traces

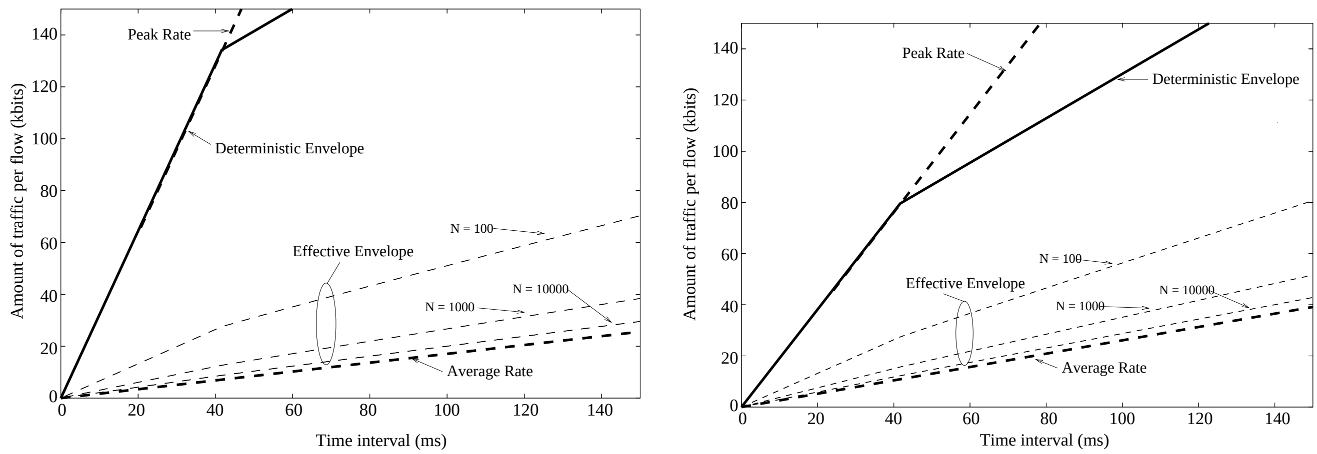
We first compare envelope functions for MPEG traces. A deterministic envelope of a flow j is given by $A_j^*(\tau) = \min_k\{\sigma_{jk} + \rho_{jk}\tau\}$, where the parameters $\{\sigma_{jk}, \rho_{jk}\}_{k=1,\dots,K}$, given in Table 2, are obtained from a concave hull of the empirical envelope of the movie traces [13]. The effective envelope for a group of flows is obtained from the deterministic envelopes using Eqn. (6).

Figures 5(a) and 5(b), respectively, show the results for N multiplexed *Lambs* and *Terminator* traces, where N is set to $N = 100, 1000$, and 10000 flows. In the graphs, we plot the size of the envelopes normalized by the number of flows as functions of time. We use $\varepsilon = 10^{-6}$ for all envelopes.

We observe that the effective envelopes are much smaller than the deterministic envelope or the peak rate. Note that increasing the number of flows N increases the statistical multiplexing gain, leading to a lower traffic rate for each flow.

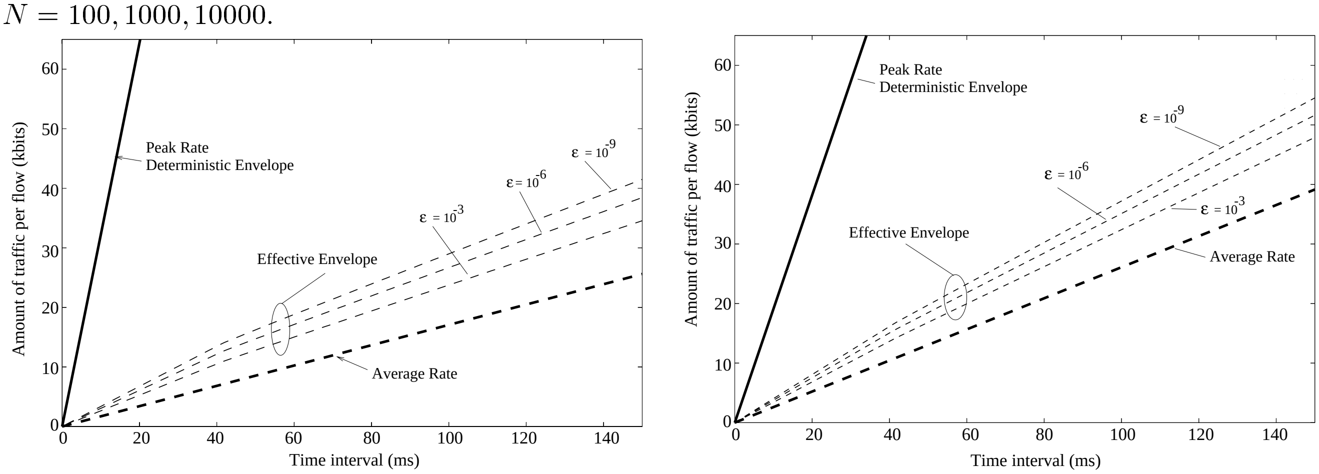
In Figure 6 we show the shape of the envelopes for a fixed number of flows, $N = 1000$, and different values of ε, namely $\varepsilon = 10^{-3}, 10^{-6}$ and 10^{-9}. Figure 6 shows that the effective envelopes are not very sensitive to variations of the parameter ε.

In Figure 7 we show how the effective envelopes vary if the number of flows N is increased. We consider the values of the envelopes at the (fixed) time interval $\tau = 50$ ms. For comparison, we include the peak and average rates into the graph. When N is large, the effective envelopes are close to the average traffic rate, indicating a significant statistical multiplexing gain.



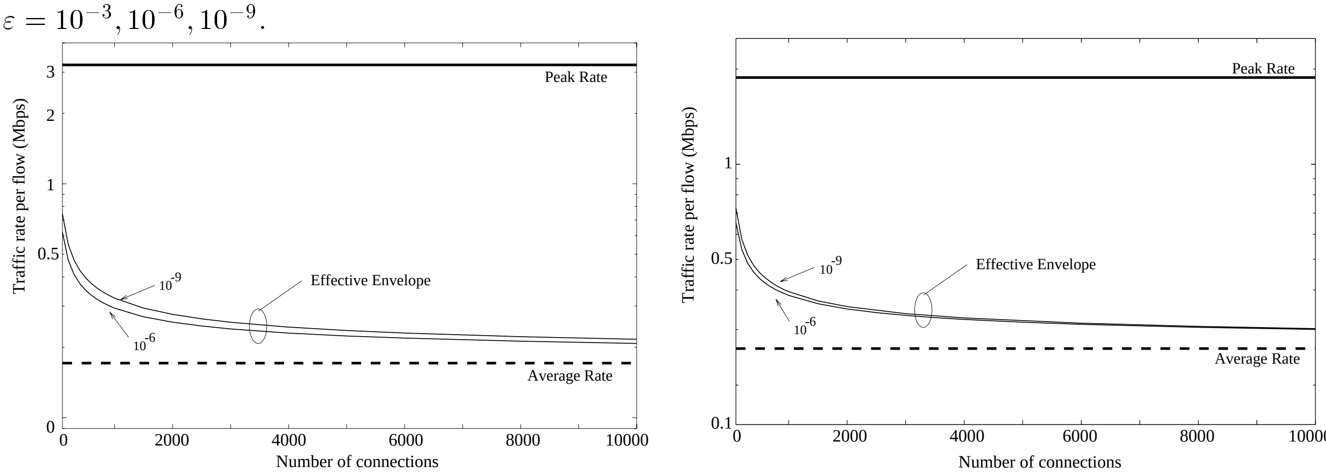
(a) *Lambs*. (b) *Terminator*.

Figure 5: Example 1: Comparison of envelopes for $\tau \leq 150\ ms$, $\varepsilon = 10^{-6}$, and for number of flows $N = 100, 1000, 10000$.



(a) *Lambs*. (b) *Terminator*.

Figure 6: Example 1: Comparison of envelopes for $\tau \leq 150\ ms$, number of flows $N = 1000$ and $\varepsilon = 10^{-3}, 10^{-6}, 10^{-9}$.



(a) *Lambs*. (b) *Terminator*.

Figure 7: Example 1: Traffic rates $\mathcal{G}_{\mathcal{C}}(\tau;\varepsilon)/(N\tau)$ of *Lambs* and *Terminator* for $\tau = 50$ ms and $\varepsilon = 10^{-6}$ or 10^{-9} .

3.2 Example 2: Number of Admitted MPEG Flows

We consider a single link with a FCFS scheduler and compare how many flows can be admitted without violation of deterministic or probabilistic delay guarantees. We also include results from simulations of the statistical rate allocation. The traffic sources are either all flows from the *Lamb* MPEG trace or all flows from the *Terminator* MPEG trace. The results for this example are shown in Figure 8. In the example, we set $C = 622$ Mbps and $\varepsilon = 10^{-6}$. The figure shows the number of admitted flows as a function of the delay bound. The results in Figures 8 show that a deterministic allocation improves upon a peak rate allocation. However, the statistical allocation, which expresses the statistical multiplexing gain admits almost as many flows as an average rate allocation scheme.

In Figure 9, we show how the achievable average utilization of a link with a FCFS multiplexer increases as the capacity of the link is increased. We fix the delay bound of traffic to $d = 50$ ms and we set $\varepsilon = 10^{-6}$. Figure 9 illustrates the achievable average link utilization as a function of the link capacity. The average achievable link utilization is the sum of the average rates of flows which are admitted according to a chosen allocation method.

The results in Figures 9(a) and (b) show that for both MPEG traces, an average utilization of more than 80% is attainable if the link capacity is above 600 Mbps or more.

3.3 Example 3: Number of Admitted MPEG Flows with Different Types

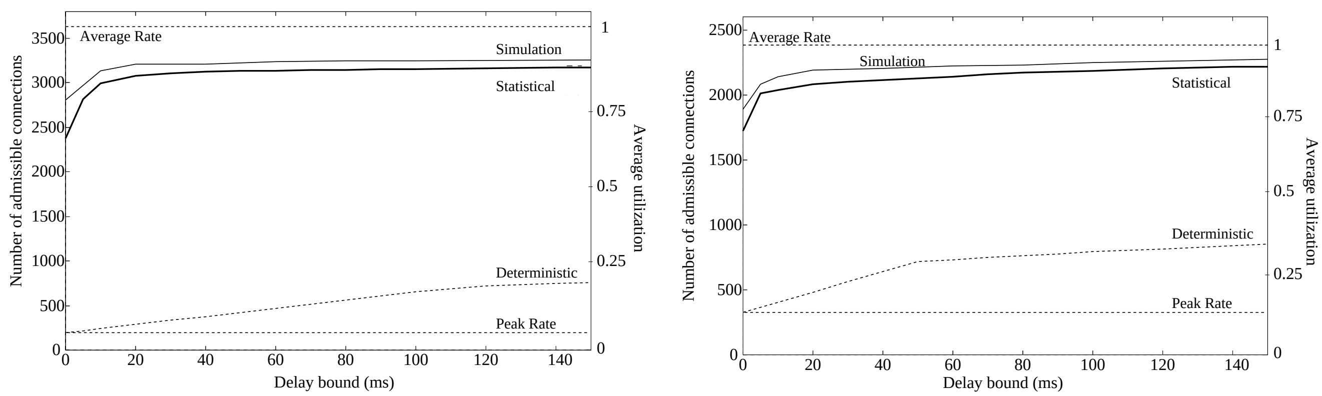
Finally, we explore the multiplexing gain at a link with two types of traffic, flows of type *Lambs* and flows of type *Terminator*. The link has a capacity of 622 Mbps. The delay bounds are set to $d_{Terminator} = 50$ ms for flows of type *Terminator*, and to $d_{Lambs} = 100$ ms for flows of type *Lambs*.

We consider two scheduling algorithms: Static Priority (SP) and Earliest-Deadline-First (EDF). For purposes of comparison, we include results for a peak rate allocation, average rate allocation, and deterministic delay guarantees.

In Figure 10, we show the maximum number of admitted MPEG flows for the various allocation methods. The figure illustrates that a statistical allocation method admits significantly more traffic than a deterministic allocation. Since the difference between the deterministic and statistical allocation method is the consideration of statistical multiplexing, we conclude that the multiplexing gain due to statistical multiplexing is significant. For both deterministic and statistical allocations, the difference between SP and EDF schedulers is modest. Hence, we conclude that the impact of the scheduling algorithm on the multiplexing gain is limited in this example. Finally, note that the results for the effective envelope are close to those attainable with an average rate allocation. This indicates that statistical delay guarantees can be provided without leaving many network resources unused.

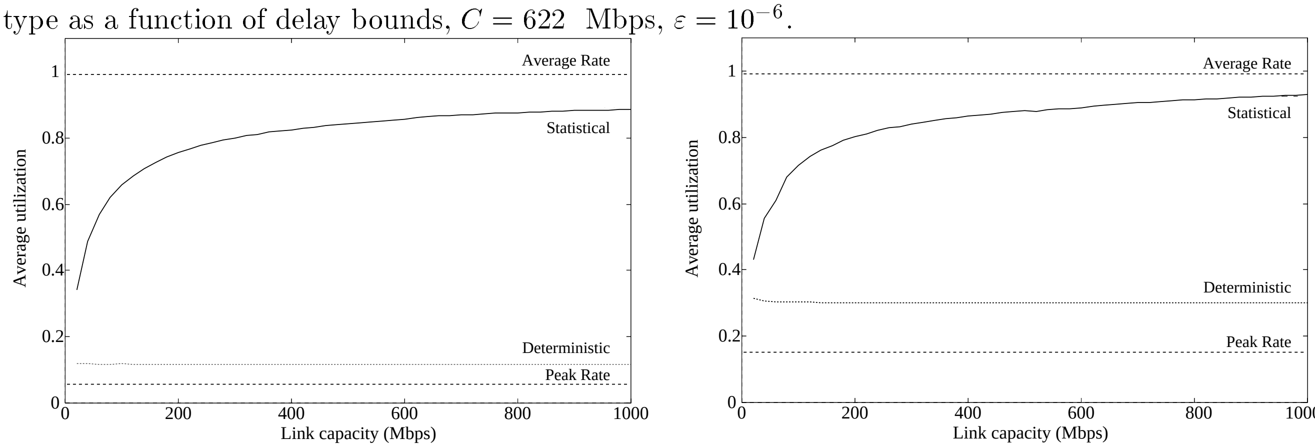
Acknowledgments

This note is based on a technical report [1], which was jointly written with Robert Boorstyn, Almut Burchard, and Chaiwat Oottamakorn. All numerical computations presented in this paper were performed by Chaiwat Oottamakorn.



(a) *Lambs.* (b) *Terminator.*

Figure 8: Example 2: Admissible number of flows at a FCFS scheduler for flows from the same type as a function of delay bounds, $C = 622$ Mbps, $\varepsilon = 10^{-6}$.



(a) *Lambs.* (b) *Terminator.*

Figure 9: Example 2: Average utilization vs. link capacity, $\varepsilon = 10^{-6}$ and $d = 50$ ms.

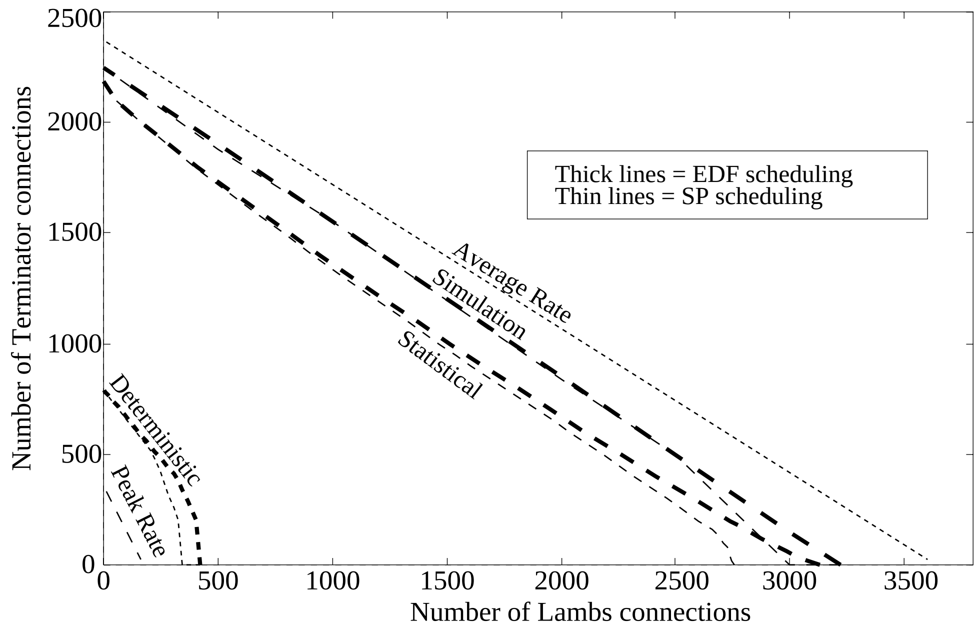


Figure 10: Example 3: Admissible region of multiplexing *Lambs* and *Terminator* flows with $\varepsilon = 10^{-6}$ and $d_{Terminator} = 50$ ms and $d_{Lambs} = 100$ ms.

References

[1] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Statistical multiplexing gain of link scheduling algorithms in QoS networks. Technical Report CS-99-21, University of Virginia, Computer Science Department, June 1999.

[2] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Oottamakorn. Statistical service assurances for traffic scheduling algorithms. *IEEE Journal on Selected Areas in Communications. Special Issue on Internet QoS*, 18(12):2651–2664, December 2000.

[3] C. Chang. Stability, queue length, and delay of deterministic and stochastic queueing networks. *IEEE Transactions on Automatic Control*, 39(5):913–931, May 1994.

[4] D. Ferrari and D. 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.

[5] 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.

[6] P. R. Jelenkovic, A. A. Lazar, and N. Semret. The effect of multiple time scales and subexponentiality in mpeg video streams on queueing behavior. *IEEE Journal on Selected Areas in Communications.Special Issue on Real-Time Video Services in Multimedia Networks*, 15(6):1052–1071, August 1997.

[7] Frank Kelly. Notes on effective bandwidths. In *Stochastic Networks: Theory and Applications (Editors F.P. Kelly, S. Zachary and I.B. Ziedins), Oxford University Press*, pages 141–168, 1996.

[8] J. Liebeherr, S. D. Patek, and A. Burchard. A calculus for end-to-end statistical service guarantees. Technical Report CS-01-19, University of Virginia, Computer Science Department, August 2001.

[9] J. Liebeherr, S. D. Patek, and E. Yilmaz. Tradeoffs in designing networks with end-to-end statistical qos guarantees. In *Proceedings of IEEE/IFIP Eighth International Workshop on Quality of Service (IWQoS '2000)*, pages 221–230, June 2000.

[10] J. Liebeherr, D. Wrege, and D. Ferrari. Exact admission control for networks with bounded delay services. *IEEE/ACM Transactions on Networking*, 4(6):885–901, December 1996.

[11] J. Liebeherr and D. E. Wrege. An Efficient Solution to Traffic Characterization of VBR Video in Quality-of-Service Networks. *ACM/Springer Multimedia Systems Journal*, 6(4):271–284, July 1998.

[12] 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 *Talk* trace used in this paper is available via anonymous ftp from the site ftp-info3.informatik.uni-wuerzburg.de in the directory /pub/MPEG/.

[13] D. Wrege, E. Knightly, H. Zhang, and J. Liebeherr. Deterministic delay bounds for VBR video in packet-switching networks: Fundamental limits and practical tradeoffs. *IEEE/ACM Transactions on Networking*, 4(3):352–362, June 1996.

[14] Z. Zhang, J. F. Kurose, J. D. Salehi, and D. F. Towsley. Smoothing, statistical multiplexing, and call admission control for stored video. *IEEE Journal on Selected Areas in Communications.Special Issue on Real-Time Video Services in Multimedia Networks*, 15(6):1148–1166, August 1997.