# Performance-driven Adaptive Admission Control for Multimedia Applications

Yin Bao
Lucent Technologies Inc.
ybao@dnrc.bell-labs.com

Adarshpal S. Sethi
University of Delaware
sethi@cis.udel.edu

## Abstract

*In this paper we study an adaptive measurement-based admission control scheme that is suitable for multimedia applications with diversified QoS requirements. The conservativeness of the proposed admission control scheme is self-adjusted based on the measured QoS performance of the existing multimedia flows. Extensive simulations performed on actual multimedia traffic with different combinations of traffic patterns, QoS requirements, and flow lifetimes have shown that the proposed scheme effectively balances the trade-off between high utilization and reliable QoS under different network conditions, hence enables a significant utilization increase compared with existing measurement-based admission control schemes.*

## I. Introduction

Admission control is essential for QoS (Quality of Service) control in the presence of real-time multimedia applications such as real-time audio and video transmissions because these applications tend to consume more network bandwidth and last on a longer time scale. It is also important for the network to evaluate the impact of a multimedia flow *beforehand* to prevent possible QoS deterioration of the existing flows due to the admission of the new flow, and to examine whether or not the requested QoS of the new flow may be supported under the current network condition.

Generally, there are two basic approaches to admission control: **parameter-based** and **measurement-based** [1]. The parameter-based approach seeks the worst-case or statistical analysis of the traffic based on *a priori* traffic characteristics and achieves admission decision by formal analysis [2] — [6]. It can be used to provide *guaranteed QoS*, however generally with the price of low network utilization [1].

The measurement-based approach focuses on achieving higher network utilization for real-time applications that are tolerant of QoS deterioration. It relies on the measurement of actual traffic load and QoS performance in making admission control decisions [1], [7] — [10]. When combined with conservative enough measurement methods, this approach provides decent network utilization as well as decent *predicted QoS*—in which the network will try to satisfy the requested QoS but do not provide any guarantee or promise [8]. It is believed that most of the real-time applications have sufficient adaptiveness to survive occasional service deterioration, especially if the service deterioration is controlled to happen infrequently [7], [8], [11], [12]. Many multimedia applications should thus benefit from measurement-based admission control because of the high utilization it provides.

In this paper, we propose a measurement-based admission control scheme that is "QoS conscious" — the conservativeness of the measurement is based on the packet loss performance of the participating flows. The measurement scheme is adaptive to the changing traffic, different mixtures of QoS requirements present in the network, and other network dynamics, therefore optimizing the admission ratio[1] while keeping the requested QoS.

The rest of this paper is organized in the following way. Section II presents the scheme to adaptively adjust the conservativeness of the admission control based on the packet loss performances of the existing flows. Section III presents an extensive simulation study for the proposed admission control scheme. In Section IV, we present the conclusion of our research and the future work.

## II. Adaptive Measurement for Admission Control

In measurement-based admission control, the admission decision is mainly made based on the comparison of measured residue resources[2] and the requested resources of the new flow. For example, a new flow may be modeled by a token bucket with $\rho_\alpha$ as the token generation rate. If the measured residue bandwidth is $\hat{u}$, this flow may be admitted if $\hat{u} > \rho_\alpha$. Obviously, how to obtain $\hat{u}$ directly affects the performance of the admission control and resulted network utilization.

We have done an experimental study on measuring the residue bandwidth, based on an algorithm similar to the one presented in [8]. In this algorithm, assuming all packets are of the same length. A *unit* is defined to be a packet transmission time. A measured sample is obtained over an *averag-*

[1] *Admission ratio:* the fraction of flows that are admitted into the network.
[2] *Residue resources*: the resources that are not utilized by the existing flows inside the network.

0-7803-5287-4/99/$10.00 (c) 1999 IEEE

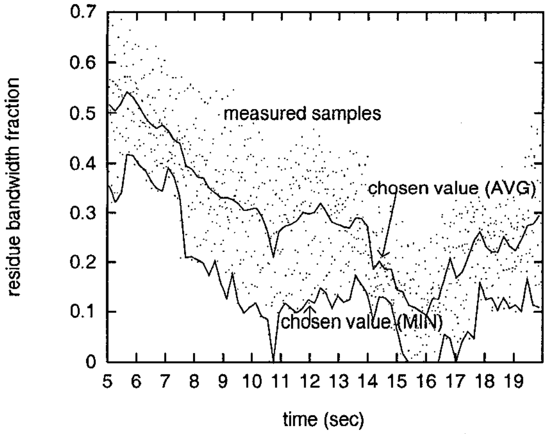


Fig. 1. **Different ways of residue bandwidth measurement**

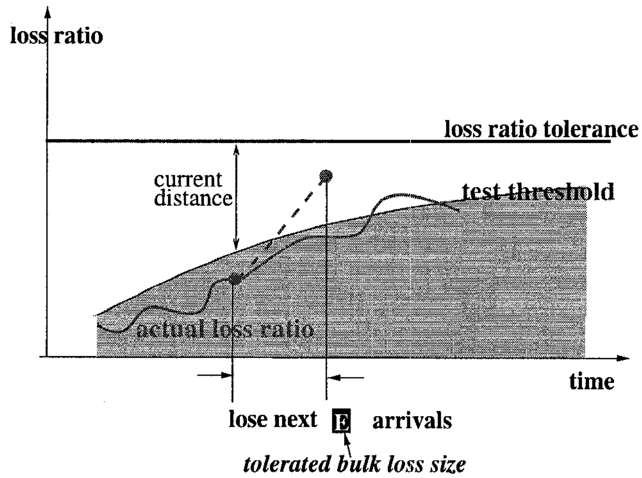


Fig. 2. **Pro-active performance test: the basic concept**

*ing period* whose size is $S$ units. Each *measurement block* contains $T$ packet transmission units, in which the following relationship holds: $T = nS$ ($n$ is an integer and $n \geq 1$). Each measured sample is obtained at the end of an averaging period of size $S$ units by counting the fraction of unused units among the $S$ transmission units. A measurement block of size $T$ $(= nS)$ units contains $n$ measured samples denoted by $S_1, S_2, ...S_n$. At the end of a measurement block, The *residue bandwidth fraction* $\hat{u}$ is chosen based on all the measured samples in the current block. A conservative measurement scheme would choose the *minimum* sample among $S_1, S_2, ...S_n$ to be $\hat{u}$.

Figure 1 depicts the measured residue bandwidth fraction samples and the residue bandwidth fraction $\hat{u}$ when two different schemes are used. One curve depicts the values of $\hat{u}$ obtained by taking the average of the samples $S_1, S_2, ...S_n$ in each measurement block (marked as AVG). The other curve plots the values of $\hat{u}$ obtained by taking the minimum of $S_1, S_2, ...S_n$ in each measurement block (marked as MIN). If we use the AVG curve in admission control, there will be more flows admitted to yield higher network utilization because the node will be perceived as having more available bandwidth resources. But doing this risks a higher possibility of QoS deterioration.

Simulations are done using the two measurement methods (MIN and AVG) and we obtained the following observation: when the requested QoS is loose (e.g. if the participating flows can tolerate higher loss ratios), it is possible to achieve higher network utilization while keeping reliable QoS by using a less conservative way to measure the residue resource. This observation poses the following question: if the network can choose different levels of conservativeness of residue resource measuring, how should it be done so that high network utilization can be achieved without jeopardizing the QoS performance?

In this section, we study an adaptive residue bandwidth measurement scheme that adjusts the conservativeness of admission control by toggling between using $\hat{u}_{MIN}$ (choose the minimum among all samples) or $\hat{u}_{AVG}$ (choose the average among all samples) as the estimated residue bandwidth fraction $\hat{u}$. The choice between these two measured values is based on the result of a *QoS performance test* on the existing flows. We adopt a *pro-active* approach to detect the performance deterioration **before** the QoS starts to suffer. In particular, we consider that the performance test fails if the loss ratio performance is within a certain *distance* of the requested loss ratio but still smaller than the requested loss ratio.

Figure 2 depicts the idea of the pro-active performance test. The convex line that is the upper boundary of the shaded area is considered to be the *performance test threshold*[3]. If the actual loss ratio performance exceeds this threshold line, the performance test fails. The shaded area is considered to be the *safe performance region*, where the loss ratio performances in this area result in successful performance tests. The definition of the shaded area is that if the loss ratio performance of a flow $i$ is inside this area, flow $i$ can sustain losing next $E$ arrivals consecutively without having the resulting loss ratio exceed its loss ratio tolerance. Intuitively, a "good" performance—i.e., a "successful" performance test—means that this flow can tolerate at least a bulk loss of size $E$ before reaching its loss ratio tolerance.

The choice of $E$—the *tolerated bulk loss size*—provides us a margin of safety in the performance test. The larger the $E$, the higher this margin of safety, thus causing the safe performance region to be smaller. Therefore with a larger $E$, the flow is more likely to fail the performance test. This causes the stricter residue bandwidth resource measurement (MIN)

[3]The reason this line is convex is that the packet loss ratio performance is evaluated cumulatively from the start of the transmission.

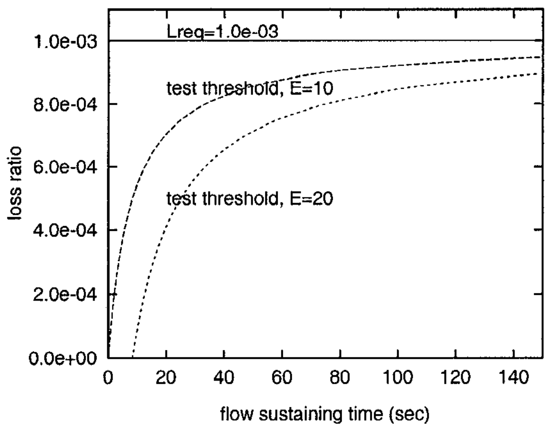


Fig. 3. **Performance test thresholds under different tolerated bulk loss size** $E$

to be used, and results in more conservative admission control. On the other hand, if $E = 0$, the test threshold line merges with the loss ratio tolerance. In that case, the performance test simply compares the current loss ratio performance with the loss ratio tolerance.

We derived the following formula to calculate the performance test threshold $L_t^T$ (details in [13]), given $L_{req}$ be the requested loss ratio; $t$ be the flow sustaining time; $A$ be the average rate of the flow; $E$ be the tolerated bulk size; $S_L$ be the initial number of lost packets[4]:

$$L_t^T = L_{req}(1 - \frac{E(1 - L_{req})}{tAL_{req} + S_L}) \quad (1)$$

With the result of Equation (1), assuming the actual loss ratio performance of a flow at time $t$ is denoted by $L_t$, if $L_t > L_t^T$, then the performance test fails. On the other hand, if $L_t \leq L_t^T$, the performance test succeeds.

Figure 3 depicts the performance test thresholds $L_t^T$ for the chosen tolerated bulk loss size $E = 10$ and $E = 20$. In both cases, we assume that the packet loss tolerance is $L_{req} = 1.0e - 3$ and the average traffic rate $A = 1200$ pkts/sec. From this figure, we can see that by choosing a different $E$, the system may have a different performance evaluation. The larger the $E$, the more careful the performance test and as a result, the more conservative the admission control.

The performance tests of multiple flows need to be converted on the same scale because the loss tolerances of different flows may be different by orders of magnitude. This can be done by taking the *relative distance* to $L_{req}$ and then averaging the performance test results among all flows. Interested readers may refer to [13] for details.

[4] When loss ratio is measured accumulatively, in order to avoid drastic loss ratio performance measure at the start of the transmission, we assume there are already initial numbers of lost and arrived packets [13].

## III. Simulation Study of the Dynamic Conservativeness

In this section, we present our simulation study of the proposed admission control scheme with self-adjustable performance-driven conservativeness. We have studied the proposed admission control scheme under a variety of real-time sources [14], [15] with different combinations of QoS requirements. The simulations are long enough so that the achieved network utilization stabilizes at $\pm 5\%$ tolerance with 90% confidence interval. The lengths of the simulation time are in the vicinity of tens of minutes with more than $2.0e7$ packets transmitted in each simulation. Inside the simulated network we assume that a scheduling scheme that provides *predicted* packet loss QoS control is deployed [16], even though the proposed admission control scheme is independent of what scheduling algorithm is used internally.

Table I depicts the simulations when the flows participating in admission are homogeneous flows with the same loss tolerance. In all these simulations, the tolerated bulk loss size $E$ is chosen to be 15. Table I lists three sets of simulation comparisons, performed under three video clips — advertisement clip, Star Wars movie clip, and lecture clip [14], respectively. For each set, in general, when the packet loss tolerance gets loose, the admission ratio as well as the network utilization get higher. The reason for this is that successful performance tests are more likely when the packet loss tolerance is loose. This results in higher fraction of AVG sampling being used in the admission decision. For example, for the advertisement source simulation sets in Table I, when the loss tolerance increases as $1.0e-4 \rightarrow 1.0e-3 \rightarrow 1.0e-2 \rightarrow 1.0e-1$, the AVG fraction used in admission

TABLE I

**Admission control: homogeneous sources with the same QoS**

$L_{req}$: loss tolerance
ADM%: admission ratio
UT%: bandwidth utilization
AVG%: the fraction of time in which the admission control uses the AVG residue bandwidth measuring scheme
QoS vio%: the fraction of flows that have violated QoS among all finished flows throughout the simulation

| source clip | $L_{req}$ | ADM % | UT % | AVG % | QoS vio% |
|---|---|---|---|---|---|
| ad | $L = 1.0e-4$ | 25.5 | 67.88 | 0.0 | 0.0 |
| ad | $L = 1.0e-3$ | 28.2 | 72.30 | 25.9 | 0.0 |
| ad | $L = 1.0e-2$ | 35.2 | 84.46 | 92.3 | 4.5 |
| ad | $L = 1.0e-1$ | 39.7 | 90.63 | 100.0 | 0.0 |
| movie | $L = 1.0e-4$ | 43.6 | 66.34 | 0.0 | 0.0 |
| movie | $L = 1.0e-3$ | 43.6 | 66.34 | 0.0 | 0.0 |
| movie | $L = 1.0e-2$ | 57.0 | 81.53 | 91.8 | 2.77 |
| movie | $L = 1.0e-1$ | 63.1 | 88.69 | 100.0 | 0.0 |
| lecture | $L = 1.0e-4$ | 38.5 | 68.68 | 0.0 | 0.0 |
| lecture | $L = 1.0e-3$ | 39.1 | 68.76 | 1.0 | 0.0 |
| lecture | $L = 1.0e-2$ | 52.7 | 84.83 | 98.6 | 0.0 |
| lecture | $L = 1.0e-1$ | 58.0 | 90.66 | 100.0 | 0.0 |

decision increases as 0% → 25.9% → 92.3% → 100%, respectively. Thus supports higher admission ratio (25.5% → 28.2% → 35.2% → 39.7%) and higher network utilization (67.88% → 72.30% → 84.46% → 90.63%). This series of simulations shows that the performance tests have effectively adjusted the conservativeness of the admission control, thereby supporting higher network utilization when the requested QoS is loose. Furthermore, despite the relaxation of admission control, the resulted QoS is reasonably reliable — in most cases the QoS violation ratio is 0%. Even though occasionally QoS violation happens, the ratio has been kept reasonably low.

Similar simulations have been conducted on heterogeneous sources with different QoS requirements. Table II depicts some of these simulation results. In all these simulations, a newly generated flow can choose to transmit from one of the following four video clips randomly: advertisement clip, lecture clip, video conferencing clip, and Star Wars movie clip. These four video clips vary in their traffic rates and characteristics [14]. Also, a newly generated flow randomly chooses a loss ratio tolerance from a particular pool of different loss ratio tolerances. We picked 3 pools of loss tolerances to use in the simulation:

- *strict range* = $\{0.0, 1.0e-4, 1.0e-3, 1.0e-2\}$
- *wide range* = $\{0.0, 1.0e-3, 1.0e-2, 1.0e-1\}$
- *relaxed range* = $\{1.0e-3, 1.0e-2, 5.0e-2, 1.0e-1\}$

The simulation results in Table II presents similar observations to those in the homogeneous source case. When the mixture of QoS requirements gets loose, the performance test enables less conservative admission control, hence supports higher network utilization without jeopardizing the reliability of the QoS. Figure 4 illustrates the actual loss ratio performance distribution of all admitted flows for the simulation conducted in Table II when the requested loss tolerances being the wide range pool (Figure 4 does not include the loss ratio performances of the flows requesting no packet loss because for all of these flows, the actual loss ratio performances are 0% packet loss). As we can see, almost all the flows' actual loss performances are better than requested and are distributed based on their packet loss tolerances.

In the previously presented simulation results, a constant tolerated bulk loss size $E$ is used for all simulations being compared. In the analysis of Section II, it was shown that different values of $E$ represent different levels of *carefulness* of the admission control. Now we present our study on the impact of this parameter.

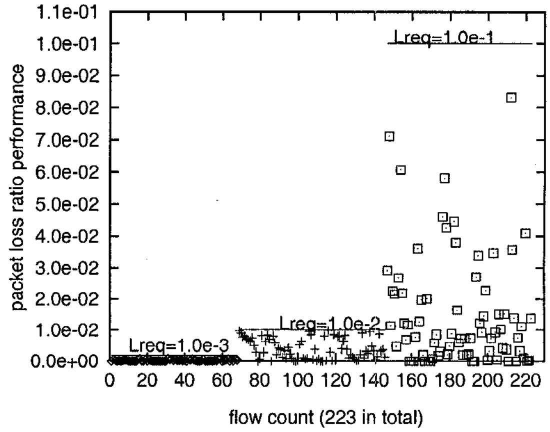


Fig. 4. **Distribution of the loss ratio performances** (lifetime 30..90sec, $L_{req} = \{0.0, 1.0e-3, 1.0e-2, 1.0e-1\}$)

Table III depicts a set of simulations that were performed under the same simulation configurations. The only different parameter between the simulations in comparison is the tolerated bulk loss size $E$. We can see that when $E$ is small ($E = 10, 20$), there is a considerable amount of QoS violation. But when $E$ is increased to 30 or 40, there is little or no QoS violation anymore. This observation tells us that finding a minimum tolerated bulk loss size $E$ is crucial. If $E$ is too small, QoS cannot be achieved reliably. On the other hand, if the tolerated bulk loss size $E$ actually used is larger than necessary, the drawback is that the network utilization is lower.

From this study we conclude that the performance of the proposed admission control depends on the appropriate choices of performance parameters. Similar conclusions have also been reached by other measurement-based admission control schemes [17], [18]. Finding the optimal control parameter value is not an easy task and is still considered to be an open

TABLE II
**Admission control: heterogeneous sources (flow lifetime 30..90 sec)**

| loss tolerance | ADM % | UT % | AVG % | QoS vio% |
|---|---|---|---|---|
| strict range | 30.9 | 69.50 | 0.8 | 0.0 |
| wide range | 37.3 | 79.98 | 41.4 | 0.95 |
| relaxed range | 44.9 | 90.91 | 99.4 | 0.0 |

TABLE III
**Impact of different tolerated bulk size $E$ (flow lifetime 30..90sec)**

| $E$ | AVG fraction (%) | bandwidth utilization (%) | QoS violation ratio (%) |
|---|---|---|---|
| 10 | 50.97 | 83.42 | 9.59 |
| 20 | 42.37 | 79.72 | 3.25 |
| 30 | 34.59 | 79.91 | 0.77 |
| 40 | 26.22 | 77.43 | 0.0 |
| 60 | 16.46 | 74.65 | 0.0 |

issue in measurement-based admission control. Admission control is expected to run over a very long time scale, over which period of time is it very likely that various network conditions—mixture of traffic, requested QoS, and flow lifetimes, etc.—will change. It is necessary to have an adaptive algorithm to adjust the control parameter $E$ to maintain a stable long term admission control performance.

In our further study we have introduced an algorithm to adjust the value of $E$ based on the long term QoS performance in the network. This algorithm effectively adjusts the conservativeness of admission control on a longer time scale to reflect network dynamics. Due to limitation of space, we do not present our study on this issue here. Interested readers can find details of the algorithm in [13].

## IV. Conclusion and Future Work

In this paper, we have proposed a new measurement-based admission control scheme whose conservativeness is self-adjusted over time to balance the trade-off between QoS reliability and network utilization. The major contribution of our research is to address the adaptiveness of admission control. The proposed admission controller toggles between using either strict or relaxed residue bandwidth estimates based on the QoS performance, thereby trying to achieve the highest network utilization with reliable QoS. The self-adaptiveness makes the proposed method a viable approach to handle admission control under different network conditions and different mixtures of traffic patterns and QoS demands. Our further study has shown that another level of adaptiveness can be used to maintain a satisfactory long term QoS performance [13].

Our simulation studies have shown that the proposed measurement-based admission controller can effectively control the incoming traffic to support reliable QoS with high network utilization even for bursty MPEG encoded VBR traffic. Furthermore, our studies also demonstrate that the adaptiveness is an important feature of admission control to balance the trade-off between reliable QoS and high network utilization under different network conditions.

The future work includes studying the performance of the admission control when different internal QoS control and scheduling schemes are used. We are also interested in the performance of the proposed admission control scheme when packet loss ratio performance of the flow is evaluated over a window rather than cumulatively over the lifetime of the flow. The involved issues include answering the question of whether it is better to choose a universal window size for the whole network, or to choose different window sizes for different applications, the impact of different window sizes to the performance of admission control, etc.

## REFERENCES


[1] S. Jamin, S.J. Shenker, and P.B. Danzig. Comparison of measurement-based admission control algorithm for controlled-load service. In *Proc. INFOCOM '97*, volume 3, pages 973–980, Kobe, Japan, April 1997.

[2] 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):368–279, April 1990.

[3] P. Pancha and M.E. Zarki. Leaky bucket access control for VBR MPEG video. In *Proc. IEEE INFOCOM '95*, volume 2, pages 6d.3.1 – 6d.3.8, April 1995.

[4] E. Knightly and H. Zhang. Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model. In *Proc. IEEE INFOCOM '95*, volume 3, pages 1137–1145, Boston, MA, April 1995.

[5] H. Zhang and E. Knightly. Providing end-to-end statistical performance guarantee with bounding interval dependent stochastic models. In *Proc. ACM SIGMETRICS '94*, pages 211–220, May 1994.

[6] R. Guerin, H. Ahmadi, and M. Naghshineh. Equivalent capacity and its application to bandwidth allocation in high speed networks. *IEEE Journal on Selected Areas in Communications*, 9(7):968–981, September 1991.

[7] S. Floyd. Comments on measurement-based admissions control for controlled-load services. ftp://ftp/ee/lbl.gov/papers/admit.ps.Z, July 1996. Submitted to Computer Communication Review.

[8] S. Jamin, P.B. Danzig, S.J. Shenker, and L. Zhang. A measurement-based admission control algorithm for integrated services packet networks (extended version). *IEEE/ACM Transactions on Networking*, 5(1):56–70, February 1997

[9] D. Tse and M. Grossglauser. Measurement-based call admission control: analysis and simulation. In *Proc IEEE INFOCOM '97*, volume 3, pages 981–989, Kobe, Japan, April 1997.

[10] M. Degermark, T. Kohler, S. Pink, and O. Schelen. Advance reservations for predictive service. In *Proc. 5th Intl. Workshop on Network and Operating System Support for Digital Audio and Video*, pages 3–14, Durham, NH, April 1995.

[11] D.D. Clark, S. Shenker, and L. Zhang. Supporting real-time applications in an integrated service packet network: Architecture and mechanism. In *Proc. ACM SIGCOMM '92*, pages 14–26, Baltimore, MD, August 1992.

[12] H. Zhang and E. Knightly. A new approach to support delay-sensitive VBR video in packet-switched networks. In *Proc. 5th Intl. Workshop on Network and Operating System Support for Digital Audio and Video*, pages 275–286, Durham, NH, April 1995.

[13] Y. Bao. *Quality of Service Control for Real-Time Multimedia Applications in Packet-Switched Networks*. PhD thesis, University of Delaware, Newark, DE, February 1998.

[14] O. Rose and H. Zoelzer. MPEG-I traces. ftp://ftp-info3.informatik.uni-wuerzburg.de /pub/MPEG/traces, September 1995.

[15] H. Zhang. MPEG-I traces from live videos. ftp://tenet.cs.berkeley.edu/pub/dbind/traces.

[16] Y. Bao and A.S. Sethi. OCP_A: An efficient QoS control scheme for real time multimedia communications. In *Proc. GLOBECOM '97*, Pheonix, AZ, November 1997.

[17] C. Casetti, J. Kurose, and D. Towsley. A new algorithm for measurement-based admission control in integrated services packet networks. In *Proc. Protocols for High Speed Networks '96*, Sophia Antipolis, France, October 1996.

[18] S. Jamin and S. Shenker. Measurement-based admission control algorithms for controlled-load service: A structure examination. Technical Report CSE-TR-333-97, CSE Division, Department of EECS, University of Michigan, Ann Arbor, MI, April 1997.