

Measurement-Based Admission Control with Aggregate Traffic Envelopes

Jingyu Qiu and Edward W. Knightly, *Member, IEEE*

Abstract—The goal of admission control is to support the quality-of-service demands of real-time applications via resource reservation. In this paper, we introduce a new approach to measurement-based admission control for multiclass networks with link sharing. We employ adaptive and measurement-based *maximal rate envelopes* of the aggregate traffic flow to provide a general and accurate traffic characterization that captures its temporal correlation as well as the available statistical multiplexing gain. In estimating applications' future performance, we introduce the notion of a schedulability confidence level which describes the uncertainty of the measurement-based "prediction" and reflects temporal variations in the measured envelope. We then devise techniques to control loss probability for a buffered multiplexer servicing heterogeneous and bursty traffic flows, even in the regime of a moderate number of traffic flows, which is important in link-sharing environments. Finally, we have developed an implementation of the scheme on a prototype router and performed a testbed measurement study, which together with extensive trace-driven simulations illustrates the effectiveness of the approach in practical scenarios.

Index Terms—Admission control, quality of service, traffic envelopes, real-time flows.

I. INTRODUCTION

ENSURING minimum quality-of-service (QoS) levels to traffic flows and groups of flows is an important challenge for future packet networks, and resource provisioning via admission control is a key mechanism for achieving this.

Consequently, a number of schemes have been devised which provide statistical services [19]. Here, a primary goal has been to admit the maximum number of flows possible (thereby efficiently utilizing system resources) subject to user requirements on throughput, loss probability, and delay. Extant algorithms achieve this goal by employing user-specified traffic parameters to estimate aggregate resource demands after accounting for the effects of statistical multiplexing. Unfortunately, this acute reliance on each flow's traffic parameters renders statistical services difficult to deploy for 1) applications that cannot accurately estimate their traffic parameters when the flow is first

established, or 2) for flows with rate variations over multiple time scales, which are not adequately characterized by standard traffic models such as the token bucket [13], [21], [22].

An alternative technique for supporting applications with ill-specified traffic characteristics is a *measurement-based* service: by basing admission control decisions on *measured* properties of traffic rather than *a priori* client-specified guesses, the effects of mistaken client traffic characterizations are largely alleviated, as is the need for a traffic model which captures the exact multiple time scale behavior of each traffic flow.

One may suppose that the problem of measurement-based admission control (MBAC) can be solved by simply measuring the unknown parameters of the traffic flows at the network routers and applying existing model-based admission control algorithms. Unfortunately, it is impractical to require routers to perform such real-time per-flow traffic measurements. Consequently, analogies with algorithms for statistical services are quite limited, as such work focuses on estimating properties of the aggregate flow from per-flow parameters, with an emphasis on, for example, computational complexity and characterization of statistical multiplexing gains.

In this paper, we design, implement, and evaluate an MBAC algorithm for multiclass networks with link sharing [12]. We develop a new theoretical framework of aggregate traffic envelopes as follows. First, to characterize the behavior of the aggregate flow, we adaptively measure its *maximal rate envelope*. Specifically, a traffic flow's *rate* is only meaningful if it is associated with a corresponding interval length. A rate envelope therefore describes the flow's behavior as a function of interval length [20]. By measuring the *maximal* rates over the corresponding interval lengths, the extreme values of the aggregate flow which are most likely to lead to packet loss are directly characterized. We will show that this envelope-based traffic characterization is not only a simple, robust, and general description of the aggregate workload, but it is also accurate enough to characterize the multiplexer's buffer dynamics, even in the case of heterogeneous flows with highly bursty (temporally correlated) traffic. Moreover, as the envelope does not employ a central-limit Gaussian traffic approximation, it can characterize scenarios with moderate statistical multiplexing gains: as detailed in [11], this is important in link sharing scenarios with a potentially moderate number of flows per traffic class.

Next, exploiting properties of this measured aggregate traffic envelope, we devise a new envelope-based MBAC algorithm as follows. First, subject to the (temporary) assumption that the past maximal rate envelope of the aggregate flow will continue to bound future packet arrivals, one could ensure that, with admission of a coarsely characterized new flow, *no* buffer overflow

Manuscript received June 8, 1998; revised May 30, 1999; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor S. Floyd. This work was supported by National Science Foundation CAREER Award ANI-9733610, National Science Foundation Grant ANI-9730104 and Grant ANI-0085842, and a grant from Texas Instruments Inc.

J. Qiu was with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA. He is now with Microsoft Corporation, Redmond, WA 98052 USA (e-mail: jingyuq@microsoft.com).

E. W. Knightly is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: knightly@ece.rice.edu).

Publisher Item Identifier S 1063-6692(01)03231-9.

would occur. Conceptually, we begin with such a framework, but quantify the uncertainty of the prediction with what we term a *schedulability confidence level*, which reflects the variation and temporal correlation of past envelope measurements, and the uncertainty of the prediction of the future workload. This concept allows us to control the QoS parameters that applications are ultimately concerned with, such as loss probability and delay-bound violation probability. For example, we derive an expression for loss probability by estimating the mean number of bits lost when the future arrivals exceed the measured envelope of the past.

To evaluate our approach and compare it to alternative schemes, we perform an extensive set of simulation experiments using traces of compressed video as well as model-generated long-range dependent traffic. With a large set of experiments, including scenarios with low capacity links in which only moderate statistical multiplexing gains are available, we illustrate the scheme's robustness and ability to accurately control the admissible region subject to the class QoS requirements.

Finally, we have implemented the scheme on a testbed of prototype routers. We describe our design's key components, namely, measurement, signaling, and admission control modules. We report measurements obtained from a set of testbed experiments and demonstrate the feasibility of aggregate envelope-based traffic monitoring and admission control.

The remainder of this paper is organized as follows. In Section II, we present the MBAC algorithm. We describe our measurement methodology, derive conditions for aggregate schedulability with an associated confidence level, derive an expression for loss probability, and introduce a scheme to incorporate temporal correlation of the envelopes themselves. The section ends with a discussion of measured rate envelopes and the algorithm's robustness to the measurement time scale. In Section III we present the results of simulation experiments. In Section IV we describe our implementation of the scheme on a prototype router and report the results of a testbed measurement study. In Section V we review related work and in Section VI we conclude.

II. ALGORITHM FOR MEASUREMENT-BASED ADMISSION CONTROL

Here, we describe a new approach for measurement-based admission control that utilizes measured values of aggregate traffic envelopes. Our scheme consists of a measurement algorithm and an admission control algorithm. The measurement algorithm continually updates the recent empirical aggregate envelope and measures the envelope's temporal variation. The admission control algorithm, invoked upon arrival of a new flow's admission request, conceptually consists of two parts. First, we check for aggregate schedulability with an associated prediction confidence level, and second, we estimate the loss probability. The new flow is admitted if the predicted performance parameters satisfy the QoS requirements of the new flow as well as all existing flows. For simplicity of explanation, we develop the admission control algorithm under the assumption that successive

measurements of traffic envelopes are uncorrelated; we remove this assumption in Section II-D.

A. Adaptive Measurement of Aggregate Rate Envelope

In characterizing a flow's rate, an associated interval length must also be specified. For example, denoting $A[s, s + I_k]$ as a flow's arrivals in the interval $[s, s + I_k]$, $A[s, s + I_k]/I_k$ is the rate in this particular interval. Moreover, the peak rate over an interval of length I_k is given by $R_k = \max_s A[s, s + I_k]/I_k$. We refer to a set of rates R_k which bound the flow's rate over intervals of length I_k as a maximal rate envelope, a traffic characterization similar in spirit to [20].

The goal of our measurement methodology is twofold. First, by measuring the maximal rate envelope of the aggregate flow, we capture the short time-scale burstiness of the traffic, which allows us to analyze the dynamics of a buffered multiplexer with admission of a new flow. Second, we measure the variation of the aggregate flow's rate envelope to characterize longer time scale fluctuations in the traffic characteristics. With the variation in the measured envelope, we can determine the confidence values of our schedulability condition and estimate the expected fraction of bits dropped should the schedulability condition fail to hold.

We consider time to be slotted with width $\tau = I_1$, the minimum interval of the measured rate envelope, which may be larger than the packet transmission time. Denoting a_t as the aggregate arrivals in time slot t such that $a_t = A[t\tau, (t + 1)\tau]$, we define the maximal rate envelope over the past T time slots from the current time t as

$$R_k^1 = \frac{1}{k\tau} \max_{t-T+k \leq s \leq t} \sum_{u=s-k+1}^s a_u \quad (1)$$

for $k = 1, \dots, T$. Thus, R_k^1 , $k = 1, \dots, T$, describes the aggregate rate over intervals of length $I_k = k\tau$ in the *most recent* $T \cdot \tau$ seconds: for notational simplicity, the dependence on t and T is implicit through the superscript. This envelope measures the short time-scale burstiness and autocorrelation structure of the aggregate flow.

Every T time slots (guidelines for setting the measurement window T , typically on the order of several seconds, are presented in Section II-E-3), the current envelope R_k^1 is measured using (1) and $R_k^n \leftarrow R_k^{(n-1)}$, $k = 1, \dots, T$ and $n = 2, \dots, N$. Thus, the variance of the measured envelopes over the past M windows of length T can be computed as

$$\sigma_k^2 = \frac{1}{M-1} \sum_{m=1}^M (R_k^m - \bar{R}_k)^2 \quad (2)$$

where \bar{R}_k is the empirical mean of the R_k^m 's, $\sum_m R_k^m/M$. Thus, we measure the *variability* of the aggregate envelope over $T \cdot M$ time slots to characterize the variation of the envelope itself over longer time scales.

For an interval of length T , we denote the distribution of the maximal rate R_k by $F_k(\cdot)$. Thus, for stationary arrivals, the (future) aggregate flow satisfies

$$P \left\{ \max_s A[s, s + k\tau]/k\tau \leq \bar{R}_k + \alpha\sigma_k \right\} = \Phi(\alpha) \quad (3)$$

where

$$\Phi(\alpha) = \int_{-\infty}^{\bar{R}_k + \alpha\sigma_k} dF_k \quad (4)$$

and (4) represents the maximum likelihood estimation of the probability that the peak rate is less than $\bar{R}_k + \alpha\sigma_k$ [27].

B. Aggregate Schedulability and Confidence Level

For a new flow requesting admission, only a coarse characterization of its traffic is required, as user-specified parameters are only used in the initial assessment of the impact of the new flow, and not in later admission decisions. We consider a general envelope for the new flow which we denote by the rates r_1, \dots, r_T so that users may specify their traffic parameters via *any* deterministic traffic model with a piecewise linear traffic constraint function [20]. In particular, a new user can specify a single “peak rate” p at setup time so that $r_k = p$ for all $1 \leq k \leq T$. Similarly, if the source characterizes its traffic with the standard dual leaky-bucket model using parameters (p, σ, ρ) , then its maximal rate over intervals of length I_k is given by

$$r_k = \frac{1}{I_k} \min(pI_k, \sigma + \rho I_k). \quad (5)$$

Then, upon arrival of a new flow request, we first perform a test for “aggregate schedulability.” This test ensures that for a certain confidence level, no packet loss will occur if the new flow is admitted. The confidence level is required here since there is no *a priori* assurance that the past envelope will continue to bound the aggregate flow, as is the case in a deterministic approach.

Schedulability Test: Consider a new flow bounded by $r_k, k = 1, \dots, T$ requesting admission to a traffic class serviced at rate C , with maximum delay requirement d , and with buffer capacity of at least $C \cdot d$. Let the aggregate flow be characterized by the peak rate envelope R_k with mean \bar{R}_k and variance $\sigma_k^2, k = 1, \dots, T$. With admission of a new flow, no loss will occur with approximate confidence level $\Phi(\alpha)$ if

$$\max_{k=1,2,\dots,T} \{k\tau(\bar{R}_k + r_k + \alpha\sigma_k - C)\} \leq Cd \quad (6)$$

and

$$\bar{R}_T + r_T + \alpha\sigma_T \leq C. \quad (7)$$

Equation (6) considers the buffer dynamics of the multiplexer and ensures that no packet is delayed beyond its required bound. Equation (7) is a *stability condition*, as it ensures that the mean rate over intervals of length T is less than link capacity with confidence level $\Phi(\alpha)$, so that the busy period is less than T also with probability $\Phi(\alpha)$. Denoting the arrivals of the new flow by \hat{A} , since

$$\begin{aligned} \max_s \{A[s, s+t] + \hat{A}[s, s+t]\} \\ \leq \max_s A[s, s+t] + \max_s \hat{A}[s, s+t] \end{aligned} \quad (8)$$

the rate envelope of the past aggregate flow multiplexed with the new flow is bounded by $(\bar{R}_k + r_k + \alpha\sigma_k), k = 1, 2, \dots, T$, also with probability $\Phi(\alpha)$.

We consider two approximations to $F_k(\cdot)$. First, motivated by extreme value theory (reviewed in Section II-E-2) [5], the Gumbel distribution describes the asymptotic distribution of the peak rate for a wide range of underlying traffic types, including log-normal, Gamma, and Gaussian distributions. Consequently, $\Phi(\alpha)$ can be approximated as

$$\Phi(\alpha) \approx \exp \left[-\exp \left(-\frac{\alpha - \lambda}{\delta} \right) \right] \quad (9)$$

with λ and δ computed from \bar{R}_k and σ_k as described in Section II-E-2.

Second, as F_k has a narrow distribution in practice, we also consider the Gaussian distribution due to its computational advantages (in Section II-D) and approximate $\Phi(\alpha)$ as

$$\Phi(\alpha) \approx \frac{1}{\sqrt{2\pi}\sigma_k} \int_{-\infty}^{\bar{R}_k + \alpha\sigma_k} \exp \left(-\frac{(r - \bar{R}_k)^2}{2\sigma_k^2} \right) dr. \quad (10)$$

Therefore, with this bound on the envelope of the aggregate flow that includes the traffic of the new flow, the probability that no loss or delay bound violations will occur is

$$P \left\{ \max_k k\tau(R_k + r_k - C) \leq Cd \right\} \quad (11)$$

i.e., this is the probability that the schedulability condition of [9] is satisfied.

Finally, (6) approximates this probability with confidence level $\Phi(\alpha)$ by employing the maximum variance approach, i.e., by considering only the dominant time scale under which loss or delay bound violation is most likely to occur. See [19] and [7] for a detailed study and justification of this approximation in a related scenario of statistical services.

We make the following observations about the schedulability test. First, for both approximations to $F_k(\cdot)$, a Gaussian heavy-traffic approximation is not required, as $F_k(\cdot)$ refers to the distribution of the peak rate, not of the underlying traffic. Moreover, as described later in Lemma 1, the peak rate distribution plays a lesser role in resource management as compared to the distribution of the underlying traffic, and we find experimentally that both choices are widely applicable.

Next, we note that even with a modest number of multiplexed flows, $\bar{R}_k + \alpha\sigma_k \gg r_k$. Consequently, if a new flow mischaracterizes its traffic parameters, the impact on the schedulability condition is relatively minor. Note further that specified traffic parameters (correct or not) for previously admitted flows are unused in the test as the impact of these flows on network performance is *measured* via the aggregate envelope.

Finally, we note that despite our use of a maximal rate envelope, our approach is able to measure and exploit the extent to which sources statistically multiplex. For example, if flows happened to synchronize in a worst-case manner, then the measured envelope R_k would be exactly the sum of the individual r_k envelopes of (5). However, when flows are statistically independent and economies of scale are present, we expect that in practice, $\bar{R}_k + \alpha\sigma_k$, the rate envelope of the ag-

gregate process, will be significantly less than the sum of the individual worst-case envelopes. We explore these observations experimentally in Section III.

C. Loss Probability

As described above, the *maximal* rate of the aggregate flow over intervals of length I_k has mean \bar{R}_k and variance σ_k^2 as given by past measurements. The schedulability test provides a no-loss schedulability condition that is satisfied with probability $\Phi(\alpha)$. However, if the future aggregate rate envelope exceeds $\bar{R}_k + \alpha\sigma_k$, then loss and delay bound violations may occur.¹ In particular, the loss probability test below enables the MBAC algorithm to ensure that the fraction of bits lost, or the loss probability, is within the class' service agreement.

Loss Probability Test: Consider an aggregate traffic flow that satisfies the schedulability test and has mean bounding rate \bar{R}_k and variance σ_k^2 over intervals of length $k\tau$. For a link capacity C , buffer size B , and schedulability confidence level $\Phi(\alpha)$, the loss probability is approximately

$$P_{\text{loss}} \approx \max_{k=1,2,\dots,T} \frac{\sigma_k \Psi(\alpha)}{\bar{R}_T} \quad (12)$$

where

$$\Psi(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-(\alpha^2/2)} - \alpha[1 - \Phi(\alpha)]. \quad (13)$$

This loss test is justified as follows. Let

$$\tilde{R}_k = \bar{R}_k + \alpha\sigma_k.$$

From the schedulability test, we have that no loss will occur unless the future bounding rate R_k exceeds \tilde{R}_k for some $k = 1, \dots, T$, which occurs with probability $1 - \Phi(\alpha)$. For a particular k and exceeding rate \tilde{R}_k , the mean number of bits lost or dropped due to buffer overflow, denoted by L_k , satisfies

$$E(L_k) \leq E(R_k - \tilde{R}_k)^+ \cdot T\tau$$

for the worst-case scenario that loss occurs over the entire window T . Further, $E(R_k - \tilde{R}_k)^+$ can be computed as

$$\begin{aligned} E(R_k - \tilde{R}_k)^+ &= \int_{\tilde{R}_k}^{\infty} (r - \tilde{R}_k) dF_k \\ &\approx \int_{\tilde{R}_k}^{\infty} (r - \tilde{R}_k) \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(r - \bar{R}_k)^2}{2\sigma_k^2}\right) dr \\ &= \sigma_k \left\{ \frac{1}{\sqrt{2\pi}} e^{-(\alpha^2/2)} - \alpha[1 - \Phi(\alpha)] \right\} \end{aligned} \quad (14)$$

under the Gaussian approximation to F_k . The fraction of bits lost is the ratio of the number of violating bits to the total number of bits sent. Considering again the dominant interval length and maximum variance, the loss probability can be approximated by (12).

¹We consider that packets that violate their delay bound are dropped and consider only loss.

Note that while the above loss probability test also uses a Gaussian approximation for the distribution of R_k , the result can easily be extended to other distributions F_k . (Recall that F_k refers to the distribution of the *maximum* rate over intervals of length k , as opposed to the distribution of the arrival sequence itself.) For example, the loss probability estimate can be refined for the Gumbel distribution as

$$\begin{aligned} E(R_k - \tilde{R}_k)^+ &= \int_{\tilde{R}_k}^{\infty} (r - \tilde{R}_k) dF_k \\ &= \sigma_k \int_{\alpha}^{\infty} (x - \alpha) \frac{1}{\delta_0} \\ &\quad \cdot \exp\left(-\frac{x - \lambda_0}{\delta_0} - \exp\left(-\frac{x - \lambda_0}{\delta_0}\right)\right) dx \\ &\approx \sigma_k \int_{\alpha}^{\infty} (x - \alpha) \frac{1}{\delta_0} \exp\left(-\frac{x - \lambda_0}{\delta_0}\right) dx \\ &= \sigma_k \delta_0 e^{-(\alpha - \lambda_0/\delta_0)} \end{aligned}$$

with λ_0 and δ_0 obtained from the mean and variance, as described in Section II-E-2.

Finally, we observe that by considering buffered multiplexers and hence the temporal correlation structure of the traffic, this admission control test has approximately T times the computational complexity of a bufferless test, where T is the number of slots in the measurement window, i.e., the computational complexity is increased by the order of the correlation function.

D. Conditional Prediction of Traffic Envelopes

Here, we describe a simple technique for measuring and exploiting the correlation structure of the envelopes themselves. In other words, a traffic envelope $\{R_k\}_{k=1}^T$ describes the flow's correlation structure up to intervals of length T ; with flows continually arriving and departing as well as possible long range dependence of the aggregate flow, we incorporate correlation at time scales greater than T , i.e., temporal correlation of the envelope itself.

As an illustrative example, consider a number of consecutive measurement windows of length T in which the network is admitting new flows without any leaving. In contrast, consider the same traffic load, with time reversed, such that flows are continually leaving the network. In both cases \bar{R}_k and σ_k^2 are the same so that if an MBAC algorithm uses only this information (the envelope and its variation), it would make the same decision in both cases, potentially overadmitting flows in the former case and unnecessarily rejecting flows in the latter case.

To solve this problem without assuming a specific call process and traffic model, we measure the correlation structure of the aggregate traffic's empirical mean rate over time windows of length T . In particular, denoting R_T^j as the mean rate over the j th window of duration T , R_T 's correlation function

$$\gamma(i) = E \left[R_T^j R_T^{j+i} \right] \quad (15)$$

characterizes the long-time-scale traffic dynamics including the impact of flow arrivals and departures. Below, we show how application of standard conditional prediction techniques [17] to

R_T (the rate over intervals of length T) can improve predictions of the entire traffic envelope.

First, we wish to predict \hat{R}_T^i using the measured current mean rate R_T^{i-1} as well as previous measured rates $R_T^{i-2}, \dots, R_T^{i-M}$. Define $\Sigma_{i,j}$ as

$$\Sigma_{i,j} = \gamma(i-j) - \overline{R_T}^2 \quad (16)$$

and define Σ as an $M-1$ by $M-1$ matrix with elements $\Sigma_{i,j}$, $i, j = 1, \dots, M-1$. Restricting this conditional prediction to linear functions of $R_T^{i-1}, R_T^{i-2}, \dots, R_T^{i-M}$, we estimate \hat{R}_T^i as

$$\hat{R}_T^i = \overline{R_T} + \vec{\Sigma}_2 \Sigma^{-1} \vec{\psi} \quad (17)$$

where $\vec{\Sigma}_2$ is a row vector $[\Sigma_{1,M}, \Sigma_{2,M}, \dots, \Sigma_{M-1,M}]$ and $\vec{\psi}$ is a column vector $[R_T^{i-1} - \overline{R_T}, R_T^{i-2} - \overline{R_T}, \dots, R_T^{i-M} - \overline{R_T}]^T$. When the peak rate R_T is a Gaussian process with correlation function $\gamma(i)$, the mean square error of this prediction is

$$\Gamma^2 = \sigma_1^2 - \vec{\Sigma}_2 \Sigma^{-1} \vec{\Sigma}_1 \quad (18)$$

where $\vec{\Sigma}_1$ is a column vector $[\Sigma_{M,1}, \Sigma_{M,2}, \dots, \Sigma_{M,M-1}]^T$ and σ_1^2 is given by (2).

Thus, \hat{R}_T^i is the linear least squares estimation of R_T^i given $R_T^{i-1}, R_T^{i-2}, \dots, R_T^{i-M}$. We estimate \hat{R}_T^i in this way because 1) it is computationally simple, and 2) if R_T is a correlated Gaussian process, then computing \hat{R}_T^i as above is an optimal prediction, minimizing

$$E \left[\left(\hat{R}_T^i - R_T^i \right)^2 \middle| R_T^{i-1}, R_T^{i-2}, \dots, R_T^{i-M} \right] \quad (19)$$

the prediction's mean square error.

Now, we turn to incorporating this improved prediction of R_T^i into an improved estimate of the entire envelope $R_1^i, R_2^i, \dots, R_{T-1}^i$ to refine the prediction of loss probability given above.

To achieve this, we define a *normalized envelope* as the peak-to-mean ratio $r_k = R_k/R_T$ for intervals of length I_k . With an estimate of the mean and variance of r_k , computed in an analogous manner, as in Section II-A, as $\bar{r}_k = (1/M) \sum_{m=1}^M r_k^m$ and $\xi_k^2 = (1/(M-1)) \sum_{m=1}^M (r_k^m - \bar{r}_k)^2$, our refined estimate of the maximal rate envelope and its variance are given by

$$\bar{R}_k = \bar{r}_k * \hat{R}_T \quad (20)$$

$$\sigma_k^2 = (\bar{r}_k^2 + \xi_k^2) \left(\hat{R}_T^2 + \Gamma^2 \right) - \bar{R}_k^2. \quad (21)$$

We note that by writing $R_k \equiv r_k * R_T$, the dynamics of R_k are decomposed into two parts, one the normalized envelope r_k which reflects the burstiness over intervals of length I_k in each window of length T , and the other the mean rate $\overline{R_T}$ in intervals of length T which, with its autocorrelation structure, describes the flow dynamics at time scales longer than T . Thus, the goal of employing conditional prediction to envelope estimation is to efficiently incorporate slow time scale behavior into estimation of the future traffic envelope. Finally, we note from the experiments of Section III-E that with this relatively minor computational overhead of the above scheme (the computational complexity of the conditional prediction algorithm is small as R_T^j is

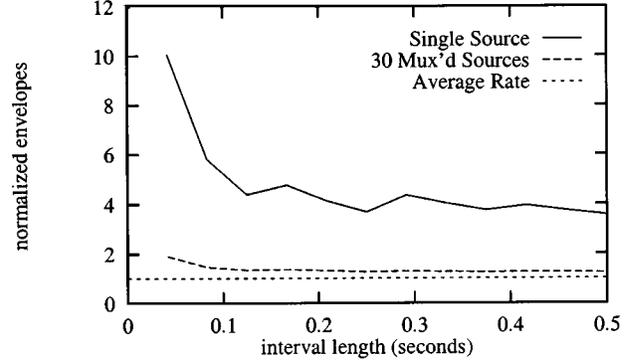


Fig. 1. Peak rate envelope for multiplexed video sources.

updated on the time scale of several seconds), the MBAC algorithm can prevent overadmission of flows when the system load is increasing over longer time scales due to successive arrivals.

E. Discussion

1) *Maximal Envelopes and Aggregate Traffic*: In this section, we explore the use of a maximal rate envelope to capture the important multiplexing properties of the traffic flows. For two scenarios, Fig. 1 depicts a flow's maximal rate envelope normalized to the flow's mean rate. In the figure, the upper curve depicts the normalized envelope of a single 30-min trace of an MPEG-compressed action movie. Plotting the bounding rate to mean ratio r_k versus the interval length I_k , the figure shows how the traffic characterization captures the maximum rates and durations of the flow's bursts. For example, for small interval lengths, R_k approaches the source's peak rate, which is about ten times the mean rate. For longer interval lengths R_k decreases toward the long-term average rate, which is the total number of bits in the MPEG sequence divided by the length of the sequence.

The second curve of Fig. 1 shows the normalized envelope of 30 multiplexed copies of this same trace. If all 30 flows are exactly synchronized, then this second curve would overlap exactly with the first. However, when the flows have statistically independent phases, the peaks of each flow do not line up exactly, and when normalized to the aggregate traffic's mean rate, the measured normalized envelope of the aggregate flow is significantly less than that of the individual flow. Hence, even with a traffic characterization which describes the flow's maximal rates, the extent to which flows statistically multiplex is evident.

We also note from Fig. 1 that the autocorrelation structure of the flow is revealed from the traffic envelope. For example, with a single flow, the source's alternation between large intracoded frames and smaller intercoded frames is evident from the shape of the normalized rate envelopes, with its sharp drop from one to two frame times and its oscillation thereafter. However, when these video sources multiplexed, this quasiperiodicity is largely removed as evidenced by the near monotonicity of the envelope of the aggregate flow. Thus, with larger N , the normalized envelope of the aggregate flow approaches the long-term average rate.

Finally, we note that in addition to characterizing the extreme values of the traffic flow which can be exploited for resource allocation, the maximal rate envelope has the desirable property

that the variation of the *maximum* rate tends to be less than the variance of the flow itself. The following lemma shows that this is asymptotically true for uncorrelated flows with bounded rate.

Lemma 1: Denote ϕ_t as the arrival rate in window t and let the maximum rate in T windows be $\phi_{*T} = \max\{\phi_1, \phi_2, \dots, \phi_T\}$. Denoting the distribution of ϕ_t as $G(x)$, if ϕ_t is bounded by capacity C so that $G(x) = 1$ for $x \geq C$ and ϕ_s is independent of $\phi_t, s \neq t$, then $\lim_{T \rightarrow \infty} \{E[\phi_{*T}^2] - E[\phi_{*T}]^2\} = 0$.

Proof: The distribution of ϕ_{*T} is given by $G^T(x)$ under the independence assumption so that the limiting distribution is

$$\lim_{T \rightarrow \infty} G^T(x) = \begin{cases} 0, & x < C \\ 1, & x \geq C = U(x - C) \end{cases}$$

where $U(x)$ is a step function. Thus, the limiting variance of ϕ_{*T} is given by

$$\lim_{T \rightarrow \infty} \text{var}(\phi_{*T}) = \int_{-\infty}^{+\infty} x^2 dU(x) - \left(\int_{-\infty}^{+\infty} x dU(x) \right)^2$$

which is $C^2 - C^2 = 0$, while the variance of the rate itself asymptotically remains $\text{var}(\phi_t)$.

2) *Distribution of the Peak-Rate Envelope:* To motivate the application of extreme value theory, consider a sequence X_1, X_2, \dots of independent and identically distributed random variables with distribution $F(x)$. The maximum of n such random variables has distribution

$$P\left(\max_{1 \leq i \leq n} X_i \leq x\right) = F^n(x). \quad (22)$$

Extreme value theory addresses the asymptotic distribution of $\max_{1 \leq i \leq n} X_i$: analogous to how the central limit theorem describes the distribution of sums of random variables without requiring knowledge of their exact underlying distributions, extreme value theory describes the distribution of the *extremes* of sequences of random variables for a general class of underlying distributions. In particular, for a large class of distributions $F(x)$, including Gaussian, exponential, log-normal, Gamma, Gumbel, and Raleigh distributions

$$\lim_{n \rightarrow \infty} P\left(\max_{1 \leq i \leq n} X_i \leq x\right) = \exp\left[-\exp\left(-\frac{x - \lambda}{\delta}\right)\right] \quad (23)$$

where $\exp[-\exp(-(x - \lambda)/\delta)]$ is a Gumbel distribution with mean $\mu = \lambda + 0.577728\delta$, and variance $\sigma^2 = \pi^2\delta^2/6$. Moreover, even if X_1, X_2, \dots are *dependent*, for most correlation structures and the same class of distributions above, the asymptotic distribution of $P(\max_{1 \leq i \leq n} X_i \leq x)$ is still Gumbel [5]. Observe that for the measurement algorithm, the mean and variance of the Gumbel distribution are measured, yielding λ and δ for (9). We find experimentally that the asymptotic approximation works well, even for moderate values of T .

3) *Setting the Measurement Window T :* Here, we discuss how to set the measurement window T . With (6) and (7), the applications' specified QoS will always be satisfied for any choice of the measurement window T . As we will illustrate below, a

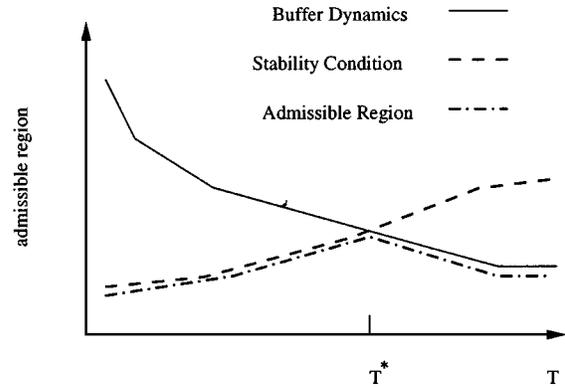


Fig. 2. Setting the measurement window T .

poorly set T can only underutilize network resources. For a particular T , either the delay test or the stability condition will restrict the admission of a new flow, so that the admissible region is the smaller of the two admissible regions calculated from (6) and (7). Fig. 2 illustrates the relationship between the admissible region (number of admitted flows) and the measurement window T . If T is set too small, σ_T , the rate variation over intervals of length T , will be large, so that the stability condition cannot be satisfied. Alternatively, as T is set larger, the measured maximum rate \bar{R}_k will also increase as the “maximum” is taken over a larger data set. Consequently, under a workload characterized by a high rate with low variance, the buffer size in the buffer test can be effectively assumed to be zero, and the available capacity for the aggregate traffic becomes the bottleneck.

In between these two extremes, there exists an optimal T^* , where the maximal admissible region is achieved. From our experiments, we find that our proposed MBAC algorithm can achieve within 5% of the best admissible region over an order of magnitude of choices of T . Regardless, we note that setting of the measurement time scale is a fundamental one to MBAC, and further discussions of its proper setting can be found in [4], [16], and [18], for example.

III. SIMULATION EXPERIMENTS

In this section, we evaluate the performance of the new algorithm for measurement-based admission control and compare with [11], [18]. The workload consists of a set of twenty 30-min traces of MPEG and JPEG compressed video from [26]. In addition, we perform simulations using heavy-tailed on-off sources, which also form a long-range dependent traffic flow in aggregate. With this collection of traces and an implementation of our MBAC algorithm, we perform a set of trace-driven simulation and admission control experiments for a wide variety of traffic mixes and network capacities.

A. Workload and Experimental Scenario

The MPEG traces from [26] exhibit both fast-time-scale rate variations due primarily to the coder's alternation between large intracoded frames and smaller intercoded frames, and slow-time-scale rate variations due to scene changes. The 20 traces exhibit considerable heterogeneity and burstiness

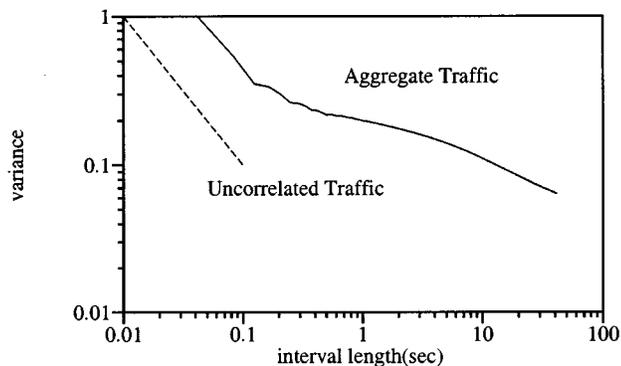


Fig. 3. Variance–time plot.

with mean rates ranging from 175 to 960 kb/s and standard deviations from 269 to 892 kb/s.

In addition to the variability and autocorrelation of the traffic for an individual trace, these traffic flows retain a significant temporal correlation structure when they are multiplexed. Fig. 3 illustrates the nature of the autocorrelation structure of the aggregate flow by depicting an example variance-time plot of 40 multiplexed flows. The figure depicts the normalized variance of the aggregate arrivals as a function of interval length. As described in [23], an uncorrelated arrival sequence would have a variance-time characterization given by a line with slope -1 on the figure’s log–log scale. In contrast, the figure indicates that the aggregate flow is highly bursty with a significant autocorrelation structure over a wide range of interval lengths.

In the following experiments, flows are randomly chosen from the 20 traces with randomly shifted initial phase. Each flow has exponential holding time with mean 600 seconds, and flows arrive with exponential interarrival times with mean three seconds. A new flow provides the MBAC algorithm with its leaky-bucket parameters (σ, ρ) and QoS requirements (delay, P_{loss}). Our MBAC algorithm is then invoked to measure the aggregate traffic and perform admission control as described in Section II with a measurement window of two seconds and slot duration of $\tau = 10$ ms. We first consider a first-come-first-serve multiplexer with buffer size B , the video traces as the workload, and a link capacity of $C = 45$ Mb/s. We set the buffer size to be C times the required delay bound so that the delay-bound violation probability is the same as the loss probability. We measure the fraction of bits that are dropped from the aggregate flow as the empirical loss probability,² and measure the fraction of time that the multiplexer is busy as the average utilization of the multiplexer. We perform repeated simulations for each scenario and report average results along with 95% confidence intervals where applicable.

B. MBAC Performance with Trace-Driven Simulations

A key goal of an admission control algorithm is to admit the maximum number of traffic flows possible without violating their QoS requirements, i.e., to maximize resource utilization subject to some QoS constraints. Consequently, the effective-

²See [19] for a comparison of this measure with the more commonly used “tail probability,” which is the fraction of time an infinite buffer queue exceeds a threshold B .

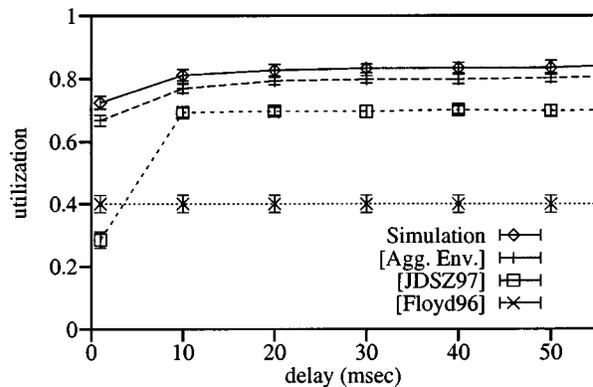


Fig. 4. Admissible regions from simulations and MBAC algorithms.

ness of an admission control algorithm can be evaluated in two ways. One is by comparing the admissible region (number of admitted flows) achieved by the MBAC algorithm with the true admissible region (obtained experimentally) under the same set of QoS constraints; the other is by comparing the QoS predictions of the MBAC algorithm with those obtained experimentally under a certain traffic load. In the experiments below, we evaluate the effectiveness of our MBAC algorithm in these two ways.

In addition to evaluating the aggregate envelope MBAC algorithm, we also compare with two schemes from the literature. First, we consider the approach of Jamin *et al.* [18], which uses an equivalent token bucket model to characterize the aggregate traffic flow of each class. Admission control is performed by measuring and controlling the average bandwidth utilization $\hat{\nu}$ and the experienced maximal queueing delay \hat{D} to target values. In our experiments, we use the recommended performance tuning parameters of [18] with v set to 0.9, $\lambda = 2$, $S = 1/24$ s, and $T = 3$ s.

We also compare with Floyd’s approach [11], which, like the aggregate envelope MBAC, addresses scenarios with moderate numbers of multiplexed traffic flows. As pointed out in [11], such a scenario is important in a link-sharing environment in which the available capacity is partitioned to support a number of services and traffic classes. The algorithm in [11] is based on the Hoeffding bound which utilizes the measured mean rate of the aggregate flow together with the individual peak rates as specified and policed for each flow. In the experiments, we smooth the video source over twelve frames to reduce the peak rate and maximize [11]’s admissible region.

Fig. 4 depicts the results of our experiments for target $P_{\text{loss}} = 10^{-4}$ and link capacity $C = 45$ Mb/s. In the figure, a point on the “Aggregate Envelope” curve represents the average utilization achieved by our MBAC algorithm, subject to the QoS constraints as given by the delay bound depicted on the horizontal axis and a loss probability of 10^{-4} . The “Simulation” curve depicts the experimental admissible region, or the maximum utilization achievable subject to these QoS constraints. We observe that the aggregate envelope algorithm accurately controls the admissible region subject to the QoS constraints, capturing most of the available statistical multiplexing gain with the difference between the utilizations of the “Simulation” and “Aggregate Envelope” curves less than 5%. Moreover, the target QoS objectives

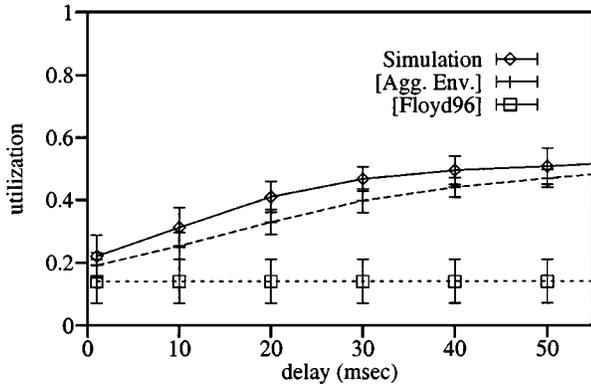


Fig. 5. Admissible regions with a moderate number of flows, $C = 5$ Mb/s.

are met and measured loss probabilities are within one or two orders of magnitude below the target value. For example, for a 20 ms delay, the mean measured loss probability is 0.8×10^{-5} . We also observe that the algorithm distinguishes among different QoS requirements, properly increasing the admissible region as the delay requirement becomes less restrictive.

Further observe from Fig. 4 that the aggregate envelope MBAC algorithm attains higher utilizations than both “JDSZ97” [18] and “Floyd96” [11], while still satisfying the target QoS. In particular, for small delay requirements below 10 ms, [18] is conservative, rejecting admission of new flows because the delay test fails (in [18, eq. (7)]). For larger delay-bound requirements, its utilizations are approximately 70% compared to simulated utilizations above 80%. A further advantage of the proposed approach is our control of QoS parameters directly via the delay and loss probability, rather than indirectly through the “target utilization” parameter.

Finally, the admissible region of [11] is approximately 40% for all delays. The difference is due to two factors. First, we perform admission control using maximal rates of the *aggregate* flow rather than user-specified per-flow peak rates: these experiments indicate that such aggregate control is important for exploiting the full statistical multiplexing gain. Second, we have incorporated buffering into our MBAC algorithm which, as evidenced by Fig. 4 and quite significantly in Fig. 5, further increases system utilization.

C. From Moderate to Large Numbers of Traffic Flows

As discussed above, the capacity allocated to a particular class of traffic can be quite small in a link sharing environment, and a Gaussian approximation of the aggregate traffic distribution may be inaccurate. Both the aggregate envelope algorithm and [11] address this scenario. To compare these two MBAC algorithms with a moderate number of flows, we perform experiments with a link capacity of 5 Mb/s, which allows only 5 to 10 active video sources. Fig. 5 depicts the resulting admissible regions for target $P_{\text{loss}} = 10^{-4}$ and new users arriving with a mean interval of 30 s. Observe that at this small capacity, the simulated utilization is much lower than for a 45 Mb/s link, as significant statistical multiplexing gains are not possible with so few sources. Regardless, the “Aggregate Envelope” curve indicates that the proposed algorithm is still able to control the admissible region, closely following the simulated

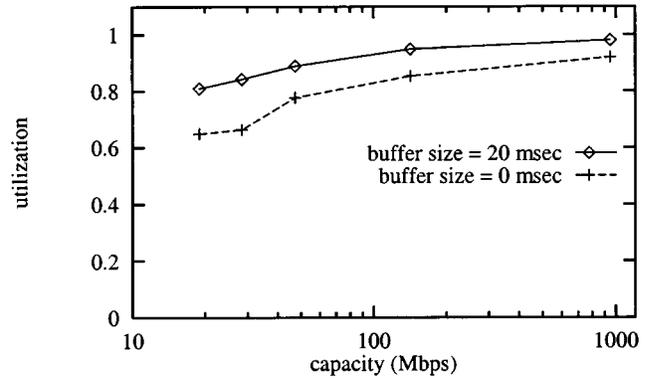


Fig. 6. Economies of scale with a large number of sources.

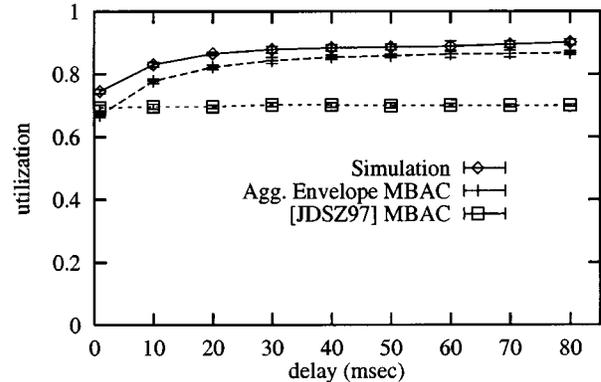


Fig. 7. Experiments with Pareto on-off sources.

curve and incorporating the region’s increasing trend with larger delay bounds; moreover, measured loss probabilities are within an order of magnitude below the target value.

In Fig. 6, we study the economies of scale with a large number of sources. The experiments depicted in the figure explore the ability of the algorithm to exploit statistical multiplexing gains available from multiplexing a large number of sources. For a fixed loss probability of 10^{-4} , the figure depicts the achievable utilization versus the link capacity C at delay bounds 0 and 20 ms. As shown, with larger link capacities, further economies of scale can be exploited (i.e., fewer network resources per source are required) and higher utilization can be achieved, and the buffering gain is significant even at large capacities.

Together, Figs. 4–6 indicate that the aggregate envelope MBAC algorithm performs well over a wide range of link capacities and buffer sizes, from less than ten to thousands of traffic flows.

D. Experiments with Multiplexed Pareto On-Off Sources

In [30], Willinger *et al.* showed that on-off sources with heavy-tailed distributions exhibit self-similarity when aggregated. Motivated by this observation and following [18], we explore the ability of the algorithm to support the Pareto on-off sources described in [18] (sources LRD1 and LRD2 in [18, Table 1]). In particular, we use a link capacity of 10 Mb/s and have flows characterize their traffic by only their peak rate, i.e., a token bucket depth of one token and a token rate of 256. In

TABLE I
IMPACT OF CONDITIONAL PREDICTION ALGORITHM ON MBAC

delay	1 msec	10 msec	20 msec	40 msec	60 msec
MBAC <i>without</i> conditional prediction, target $P_{loss} = 10^{-4}$					
$P_{mbac} (*10^{-4})$	0.02	0.6	4.0	5.0	6.0
MBAC Util. (%)	69.0	84.0	88.6	90.0	91.0
Simul. Util. (%)	75.2	84.8	86.6	87.8	88.8
MBAC <i>with</i> conditional prediction, target $P_{loss} = 10^{-4}$					
$P_{mbac} (*10^{-4})$	0.002	0.02	0.26	0.61	0.54
MBAC Util. (%)	63.8	79.2	84.9	87.7	88.5
Simul. Util. (%)	74.1	84.4	86.9	88.3	89.4

the experiments, a heterogeneous traffic mix is considered with LRD1 and LRD2 sources chosen with equal probability.

Fig. 7 depicts representative results of these experiments again with a target loss probability of 10^{-4} and the delay is depicted on the horizontal axis. Observe that for the small delay bounds, both the aggregate envelope approach and [18] are quite close to the true admissible region. However, for larger delays, the rate test of [18] remains the bottleneck so that potential utilization gains due to buffering are not exploited. In contrast, the aggregate envelope approach exploits this significant utilization gain due to buffering. This is in part due to an envelope’s ability to simultaneously account for the interaction among a flow’s bandwidth and buffer (delay) requirements without the need for a conservative separation of these two resources. In addition, our notion of a schedulability confidence level provides a finer level of control over QoS parameters such as the delay-bound-violation probability.

E. Impact of Flow Arrivals and Departures

In Section II-D, we described a technique by which we exploit the temporal correlation of successive traffic envelopes to incorporate the effects of flow arrivals and departures as well as to capture traffic dynamics at time scales larger than that of the envelope and measurement window (typically on the order of several seconds). In the experiments below, we illustrate the effectiveness and importance of this aspect of the MBAC algorithm. Moreover, while the above experiments employ the Gumbel approximation to the distribution of the peak rate envelope as in (9), here we employ the Gaussian approximation of (10) for its computational advantages in conditional prediction.

Table I shows the utilizations and loss probabilities obtained by the MBAC algorithm both with and without the conditional prediction scheme of Section II-D. In the table, the target loss probability is 10^{-4} and P_{mbac} is the loss probability measured from simulations; “MBAC Util.” is the utilization achieved with the MBAC algorithm, and “Simul. Util.” is obtained from simulation.

We observe from the table that both with and without conditional prediction, the utilization achieved by the MBAC algorithm is quite close to that of the simulation. However, notice that without conditional prediction, the MBAC algorithm can slightly underestimate the loss probability, or equivalently overadmit traffic flows. In particular, the table indicates that at a delay bound of 20 ms, the MBAC algorithm overadmits

flows by approximately 2% utilization if conditional prediction is not used. (We note that under the refined Gumbel approximation, the trend is similar, although such overadmission was not observed.) As described in Section II-D, the reason for this is that successive traffic envelopes are correlated and this must be captured in admission control. By accounting for this property using conditional prediction techniques, the table indicates that the MBAC algorithm can control the admissible region with a higher degree of accuracy.

IV. IMPLEMENTATION AND MEASUREMENTS

In this section, we describe our implementation and testbed experiments on a network of prototype FreeBSD v3.2 routers and hosts inter-connected via 10 and 100 Mb/s links.

A. Design Issues

A key principle for our implementation is a modular design that divides the architecture into separate components. The objective is to develop the scheme’s building blocks independently, with simple interfaces among elements. In this way, the solution is applicable to systems with a broad range of router architectures and capabilities, including high-speed hardware implementations. These components are the traffic measurement module which assesses the current per-class workload, the signaling protocol by which new flows are established, and the admission control module which accepts or rejects requests to establish new real-time flows.

1) *Measurement Module:* The measurement module consists of mechanisms for data collection and envelope computation.

We have utilized two techniques for collection of arrival data. First, we employ tcpdump to record each packet’s arrival time, size, and class over the previous one-second interval, information which is later utilized in computing the maximal rate envelope. The advantage of this approach is that tcpdump is a standard traffic measurement tool with well-tuned implementations; we utilize tcpdump in our testbed experiments reported below. For higher speed routers beyond the 10 Mb/s link capacities in the testbed, we have devised an implementation of both the data collection and envelope computation mechanisms using so-called network processors. In particular, we used the TI TMS320C6000 family of network processors, which consist of a C6000 DSP platform combined with high-speed network I/O.

Such high-speed measurement and traffic analysis provides a mechanism for the scheme to scale to high-speed routers.

To compute a class' maximum rate envelope, the method described in Section II determines the flow's peak rate for a given interval length. An analogous method which we use in our implementation is to instead measure the minimum interval length over which a certain number of bytes are transmitted. We use the latter approach because it reduces computational complexity by removing divisions by the interval lengths from the algorithm. In other words, instead of determining the maximum number of bytes to arrive in an interval of given size (via a sliding window) or discretizing the time scale, we calculate the minimum time required for a certain number of bytes to arrive, or discretize the data scale. In our implementation, we discretize the arrival's byte scale logarithmically, with eight levels beginning at 1 kB.

Observe that in both cases, each packet arrival is measured (we do not consider packet sampling), and traffic analysis is performed on a per-class basis. The resulting envelope is then communicated via a shared memory interface to the admission control module.

2) *Signaling and Control Plane*: We have implemented the MBAC scheme within two architectures. The first is the IntServ architecture [8] that uses RSVP signaling as specified in [8], [32]. Within this framework, the algorithm is providing a multiclass controlled-load-like service [31], yet with particular latency and loss targets for each class, and hence, a stronger service model than specified by [31]. The second architecture, described in [6], achieves scalability by processing signaling messages only at edge routers, and managing services on a per-path basis. In this paper, we describe our measurements in the former architecture, with measurements in the edge-based architecture described in [28].

For this implementation, hosts request a QoS session by first generating an RSVP message with their reservation request. Such messages are forwarded toward the destination and processed by the RSVP daemon at each node along the path. Each RSVP router has several local procedures for reservation setup and enforcement. Policy control determines whether the user has administrative permission to make the reservation. Admission control keeps track of the system resources and determines whether the node has sufficient resources to supply the requested QoS. The RSVP daemon invokes both procedures before accepting the new flow. If either test fails, the RSVP daemon returns an error notification to the application that originated the request. If both checks succeed, the RSVP daemon sets parameters in the packet classifier and packet scheduler to obtain the requested QoS. The packet classifier determines the QoS class for each packet and the packet scheduler orders packet transmission to achieve the promised QoS for each stream [32]. Our implementation consists of a modification of the RSVP daemon to call the admission control procedure described in Section II and as described below. (In contrast, the architecture of [28] simplifies the signaling protocol itself as only edge routers are signaled.)

While a detailed measurement study of the performance of RSVP signaling can be found in [24], below we present measurements of the latency incurred due to both the admission control procedure and signaling. Here, we observe that signaling

TABLE II
TESTBED MEASUREMENTS

Delay Reqst (msec)	Number Flows	Mean Delay (msec)	Maximum Delay (msec)	% Outside Bound
5	16.0	1.52	17.9	1.25
10	16.3	1.83	22.5	1.20
20	18.0	2.35	36.8	0.56
60	21.1	12.85	124.7	6.16

is performed for each individual flow, whereas measurement, storage, and forwarding is aggregate, i.e., neither traffic measurements, control plane state, nor scheduling is performed on a per-flow basis.

3) *Admission Control Module*: The admission control procedure is invoked by the RSVP daemon when a router receives an RSVP packet. The procedure's parameters are a target delay bound and loss probability (which in turn constrains α), and the available capacity C . The measurement window T must also be specified. With a new admission request, the RSVP daemon accesses the quantized peak rate envelope via shared memory communication with the measurement module as described above. The test ensures that for the eight discrete "byte" levels of the envelope, the schedulability and loss probability tests are passed, i.e., for the target α , the empirical minimum arrival time (mean minus α times the standard deviation) is sufficiently large so that the delay and loss requirements are satisfied.

In our implementation, the admission control procedure required 0.19 to 0.58 ms to perform, measured as the time between when the RSVP daemon calls the admission control procedure, to when the decision is returned. This corresponds to approximately 2.4% of the 8 to 22 ms total flow setup time, measured as the time between when the originating host generates an RSVP message, until it receives its accept or reject message.

B. Testbed Measurements

Our testbed consists of three routers, each connected to the two other, with multiple hosts connected to each router. Maximum link rates are 100 Mb/s for hosts and 10 Mb/s for routers, as manually configured by the multiport Ethernet cards. The buffer size of the routers is 250 packets, which for a link capacity of 10 Mb/s and MTU of 1500 bytes corresponds to a maximum queuing delay of 300 ms. We have performed an extensive measurement study on the testbed and below report a set of measurements from a configuration in which multiple flows are established across a single router.

For traffic sources, we designed a Pareto on-off traffic generator that transmits packets only with successful admission signaled by RSVP. Packets are generated according to the Pareto on-off model with the following parameters: packet size 1000 bytes, mean burst time 250 ms, mean idle time 250 ms, peak rate 800 kb/s, and average rate 400 kb/s. The Pareto shape parameter is 1.9 and the flow lifetime is five minutes. The traffic generator also handles communication with RSVP. In order to communicate the user's request to the network, a module in the traffic generation program on the source host side generates a path message, while a module on the destination side sends a

reserve message with the QoS request. By calling the RSVP application interface function, the sender side receives the admission control result.

Table II shows the results of the experiments. Each row corresponds to experiments with a particular delay target of 5 to 60 ms as specified in the first column, and the remaining columns show the measured performance averaged over three to five experiments. In all cases, the measured link capacity using netperf is 9 Mb/s indicating that the range of admissible flows is 11 to 22, corresponding to peak-rate and average-rate allocation, respectively.

We make the following observations about the experiments. First, the algorithm has exploited statistical multiplexing gains, even in this scenario of a moderate number of traffic flows. The average link utilizations are in the range of 67% to 94%. As a consequence of overbooking, violations of the target delay occur, and the fifth column indicates that the violations range from 0.56% to 6.16% of packets. Second, observe that assigning different delay targets has the desired impact on measured performance, allowing mean delays in the range of 1.52 to 12.85 ms, and maximum delays in the range of 17.9 to 124.7 ms. Hence, the algorithm provides the basic mechanisms for performance differentiation in multiclass networks. Finally, we observe that the targeted violations due to statistical multiplexing cannot be precisely met, as the percentage of violations differs in the four cases despite having the same α of 1.0 for all experiments. The differences arise from a number of factors: the quantization of the measured arrival process; the discrete nature of flows themselves (a discrete number of flows is admitted, whereas to achieve the precise QoS target may require between N and $N + 1$ flows); the strong impact on QoS for each new flow in the regime of a moderate number of flows; and the extreme burstiness of the traffic itself. Thus, these measurements illustrate the difficulties of achieving precise control of class QoS measures, but do indicate that the algorithm can control admissions so that empirical QoS has a strong correspondence to the targeted values.

V. RELATED WORK

Aggregate statistical traffic envelopes for QoS have been studied in [1], [25]. The focus of [1] is statistical multiplexing of adversarial flows while the focus of [25] is interclass resource sharing. Both studies address statistical multiplexing in a number of multiclass schedulers such as strict priority and earliest deadline first. In contrast, our MBAC algorithm utilizes *measurements* to characterize and control the aggregate traffic envelope.

Several approaches to MBAC perform admission control using *a priori* per-flow traffic models such as on-off in addition to aggregate measurements, e.g., [11], [14], [15]. While all schemes require at least an *initial* coarse description of a new flow, we have found that later discarding all specified parameters (even if they are accurate) and performing aggregate control is important for efficient service provisioning. Namely, we find that using prespecified parameters to estimate properties of the aggregate flow introduces inaccuracies that can be avoided by directly observing and controlling the aggregate flow.

Other studies have used a simplified system model, e.g., bufferless multiplexers [10], [11], [14]–[16], and/or simplified traffic models, e.g., Gaussian arrivals [10], [16], to study particular aspects of MBAC. For example, a study of the role of the measurement window and the impact of measurement errors is found in [16]. In contrast, our goal was to design and implement an algorithm that can exploit buffering, is applicable to both moderate and large numbers of flows, and applies to a general class of traffic types. Regardless, such studies complement our work, and can be used for example to determine the ideal setting of the measurement window.

The MBAC algorithm in [18] measures an equivalent token bucket model to characterize the aggregate traffic flow of each class, and performs admission control by measuring and controlling the aggregate bandwidth utilization and the experienced maximal queueing delay. Our work generalizes this approach in two distinct ways. First, an envelope is a more general and accurate traffic characterization than the token bucket, as the token bucket captures flow behavior at only a single time scale. Second, by monitoring variations in the observed envelope and introducing a schedulability confidence level, we develop a way for network service providers to directly control QoS measures such as loss probability and maximum delay. This contrasts with [18] which requires proper setting of a “target utilization” parameter, which may require dynamic retuning under different traffic types.

The performance study of [2], [3] found a large class of MBAC algorithms (including this one) to obtain identical performance. Our conclusions are quite different as we consider “performance” to be the system’s ability to properly control the admissible region under a set of prespecified QoS parameters. In contrast, [2] views QoS parameters as tunable parameters rather than prespecified service objectives, and finds that all algorithms *measure* the same loss probability under a given utilization, although their *input* QoS parameters can be vastly different. Alternatively, we compare algorithms’ measured loss probability under *identical* input QoS parameters; further justification of our experimental methodology was provided in Section III.

Finally, in recent work, we have used the approach presented in this paper as a foundation for studying scalable multiclass admission control. In particular, in [6], [28], we design and implement a scheme in which edge nodes passively monitor network paths to assess their available service. Admission control functionality is then fully distributed by signaling only a flow’s egress router.

VI. CONCLUSION

In this paper, we introduced a novel algorithm for measurement-based admission control (MBAC) that exploits measured peak rate envelopes of the aggregate traffic flow to allocate network resources. Our approach uses the ability of the aggregate flow’s envelope to reveal the critical characteristics of the traffic for admission control, such as the extent to which flows are statistically sharing network resources and the autocorrelation structure of the aggregate flow. Since there is no assurance that the aggregate flow will continue to be bounded by its past

behavior, we developed new theory to quantify the confidence level of a schedulability condition that incorporates the randomness of the aggregate envelope. Moreover, we devised new techniques to estimate the loss probability by deriving an expression for the expected number of bits lost when the schedulability condition is violated. Our goal was to design an accurate, robust, and general MBAC algorithm that encompasses a wide range of traffic types, buffer sizes, and link capacities, including situations with a relatively modest number of multiplexed flows per traffic class.

ACKNOWLEDGMENT

The authors would like to thank L. Breslau, S. Floyd, P. Key, C. Li, R. Riedi, S. Shenker, and the anonymous reviewers for insightful comments and discussions.

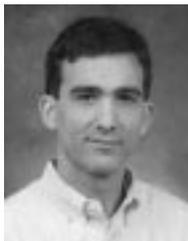
REFERENCES

- [1] R. Boorstyn, A. Burchard, J. Liebeherr, and C. Ottamakorn, "Effective envelopes: Statistical bounds on multiplexed traffic in packet networks," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.
- [2] L. Breslau, S. Jamin, and S. Shenker, "Measurement-based admission control: What is the research agenda?," in *IEEE/IFIP IWQoS*, London, U.K., May 1999.
- [3] —, "Comments on the performance of measurement-based admission control algorithms," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.
- [4] G. Casetti, J. Kurose, and D. Towsley, "A new algorithm for measurement-based admission control in integrated services packet networks," in *Proc. 5th Int. Workshop Protocols for High-Speed Networks*, Antipolis, France, Oct. 1996, pp. 13–28.
- [5] E. Castillo, *Extreme Value Theory in Engineering*. New York: Academic, 1988.
- [6] C. Cetinkaya and E. W. Knightly, "Scalable services via egress admission control," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.
- [7] J. Choe and N. Shroff, "A central limit theorem based approach to analyze queue behavior in ATM networks," *IEEE/ACM Trans. Networking*, vol. 6, pp. 659–671, Oct. 1998.
- [8] R. Braden, D. Clark, and S. Shenker *et al.*, "Integrated services in the internet architecture: An overview," Internet RFC 1633, 1994.
- [9] R. Cruz, "Quality of service guarantees in virtual circuit switched networks," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1048–1056, Aug. 1995.
- [10] Z. Dziong, M. Juda, and L. Mason, "A framework for bandwidth management in ATM networks—Aggregate equivalent bandwidth estimation approach," *IEEE/ACM Trans. Networking*, vol. 5, pp. 134–147, Feb. 1997.
- [11] S. Floyd, "Comments on measurement-based admissions control for controlled-load services," Lawrence Berkeley Laboratory, Tech. Rep., July 1996.
- [12] S. Floyd and V. Jacobson, "Link-sharing and resource management models for packet network," *IEEE/ACM Trans. Networking*, vol. 3, pp. 365–386, Aug. 1995.
- [13] M. Garret and W. Willinger, "Analysis, modeling and generation of self-similar VBR video traffic," in *Proc. ACM SIGCOMM*, London, U.K., Aug. 1994, pp. 269–280.
- [14] R. Gibbens and F. Kelly, "Measurement-based connection admission control," in *Proc. 15th Int. Teletraffic Cong.*, Amsterdam, The Netherlands, 1997, pp. 781–790.
- [15] R. Gibbens, F. Kelly, and P. Key, "A decision-theoretic approach to call admission control in ATM networks," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1101–1114, Aug. 1995.
- [16] M. Grossglauser and D. Tse, "A framework for robust measurement-based admission control," *IEEE/ACM Trans. Networking*, vol. 7, pp. 293–309, June 1999.
- [17] S. Haykin, *Modern Filters*. New York: Macmillan, 1989.
- [18] S. Jamin, P. Danzig, S. Shenker, and L. Zhang, "A measurement-based admission control algorithm for integrated services packet networks," *IEEE/ACM Trans. Networking*, vol. 5, pp. 56–70, Feb. 1997.
- [19] E. W. Knightly and N. Shroff, "Admission control for statistical QoS: Theory and practice," *IEEE Network*, vol. 13, pp. 20–29, Mar. 1999.
- [20] E. W. Knightly and H. Zhang, "D-BIND: An accurate traffic model for providing QoS guarantees to VBR traffic," *IEEE/ACM Trans. Networking*, vol. 5, pp. 219–231, Apr. 1997.
- [21] M. Krunz and S. Tripathi, "On the characterization of VBR MPEG streams," in *Proc. ACM SIGMETRICS*, Seattle, WA, June 1997, pp. 192–202.
- [22] A. Lazar, G. Pacifici, and D. Pendarakis, "Modeling video sources for real time scheduling," *ACM Multimedia Syst. J.*, vol. 1, pp. 253–266, Apr. 1994.
- [23] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.
- [24] A. Neogi, T. Chiueh, and P. Stürpe, "Performance analysis of an RSVP-capable router," *IEEE Network*, vol. 13, pp. 56–63, Sept. 1999.
- [25] J. Qiu and E. W. Knightly, "Interclass resource sharing using statistical service envelopes," in *Proc. IEEE INFOCOM*, New York, Mar. 1999, pp. 1404–1411.
- [26] O. Rose, "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems," in *Proc. IEEE Conf. Local Computer Networks*, Minneapolis, MN, Oct. 1995, pp. 397–406.
- [27] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
- [28] J. Schlembach, A. Skoe, P. Yuan, and E. Knightly, "Design and implementation of scalable admission control," in *Proc. Int. Workshop QoS in Multiservice IP Networks*, Rome, Italy, Jan. 2001.
- [29] M. Siler and J. Walrand, "On-line measurement of QoS for call admission control," in *IEEE/IFIP Int. Workshop QoS*, Napa, CA, May 1998.
- [30] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self-similarity through high-variability: Statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Trans. Networking*, vol. 5, pp. 71–86, Feb. 1997.
- [31] J. Wroclawski, "Specification of the controlled-load network element service," Internet RFC 2211, 1997.
- [32] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A new resource ReSerVation Protocol," *IEEE Network*, vol. 7, pp. 8–18, Sept. 1993.



Jingyu Qiu received the B.S. degree in applied physics from Fudan University, Shanghai, R.O.C., and the M.S. degree in electrical and computer engineering from Rice University, Houston, TX.

He was a member of the Rice Networks Group studying traffic measurement and control for multi-class networks. He is currently a Software Engineer with Microsoft Corporation, Redmond, WA.



Edward W. Knightly (S'91–M'96) received the B.S. degree from Auburn University, Auburn, AL, in 1991, and the M.S. and Ph.D. degrees from the University of California, Berkeley, in 1992 and 1996, respectively, all in electrical engineering.

Since 1996, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, Rice University, Houston, TX. His research interests are in the area of quality-of-service theory, algorithms, and architectures.

Dr. Knightly received the National Science Foundation CAREER Award in 1997 and the Sloan Fellowship in 2001. He is an Editor of the *Computer Networks Journal* and *IEEE/ACM TRANSACTIONS ON NETWORKING*, and formerly of *IEEE Network Magazine*. He served as Co-Guest Editor of *IEEE Network Magazine*'s special issue on integrated and differentiated services for the Internet. He served as Co-Chair for the 1998 IEEE/IFIP International Workshop on Quality of Service and currently serves on its steering committee.