

Computer Communications 24 (2001) 1097-1104



www.elsevier.com/locate/comcom

Dynamic resizing of utilization target in measurement-based admission control

Y.S. Sun^a, C.-C. Chuang^b, Y.-D. Lin^{b,*}

^aDepartment of Information Management, National Taiwan University, Taipei, Taiwan, ROC ^bDepartment of Computer and Information Science, National Chiao-Tung University, 1001a Ta Hsueh Road, Hsinchu 30050, Taiwan, ROC

Received 8 December 1999; revised 24 July 2000; accepted 27 July 2000

Abstract

A major issue in design of measurement-based admission control (MBAC) algorithms relies on how to maximize link utilization while meeting applications' QoS constraints. Studies have claimed that many proposed MBAC algorithms would give the same performance if the parameters were properly tuned. In this paper, we study the effects of tuning two of the most important parameters in MBAC algorithms: *Measurement Window* and *Utilization Target*. We show that the measurement window is not a good tuning knob for enforcing rigid loss probability constraints. We propose a new parameter called Soft Utilization Target combined with a simple flow admittance policy called the *Latch* algorithm. The simulation results show that this scheme can achieve better enforcement of loss probability constraint while maintaining higher link utilization. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Admission control; Measurement-based; Integrated services; Controlled-load services

1. Introduction

The approach of measurement-based admission control (MBAC) [1] is preferred in integrated services networks [2–5] for achieving better link utilization. In the measurement-based approach, a priori source characterization is used only for incoming flows. Measurements will be used to characterize aggregated behavior of flows that have been in place. The goal is to exploit potential multiplexing gain. This is based on the observations that the flow description given at the session setup time or resource reservation phase is usually a loose bound of the traffic that the flow actually generates. The aggregated link utilization would be lower than the expectation evaluated simply by flow specifications. MBAC is proposed to overcome the problem where admission decisions and control are based on on-line measurement.

Several issues are raised in the use of MBAC. First is the feasibility of implementing real-time on-line measurement methods for high-speed networks in approximating traffic loads. The second issue is the possible violation of QoS constraints, such as the utilization target [1] and loss probability [6,7], caused by the measurement errors. The third

issue again is to maximize link utilization under the QoS constraints.

Jamin et al. [1] presented an algorithm in which three performance-related measurement parameters are defined: Measurement Interval Size S, Measurement Window T, and Utilization Target v. They showed that the way in which these parameters are determined may have significant impact on the performance. In Ref. [8], they contended that the MBAC algorithms developed so far could deliver the same performance if the parameters were properly tuned. Instead of developing new algorithms, proper tuning of the measurement parameters of the algorithm in Ref. [1] would be a good choice. However, they did not provide solutions on how to tune these parameters. Furthermore, the algorithm in Ref. [1] can satisfy the utilization target constraint, but with occasional violation of the loss probability constraint. It is appropriate for predictive services. However, it would not be suitable for applications with rigid QoS requirements.

In this paper, we adopt the evaluation method for aggregated traffic by the normal distribution approximation [6] to examine how we can achieve the loss probability constraint by resizing the measurement window and utilization target in Ref. [1]. Here we define a new parameter called *Soft Utilization Target*, denoted by v', which is dynamically adjusted to enforce packet loss constraint in contrast to the fixed utilization target used in Ref. [1]. We will show that

^{*} Corresponding author. Tel.: +886-3-573-1899; fax: +886-3-572-1490. E-mail addresses: ydlin@cis.nctu.edu.tw (Y.-D. Lin).

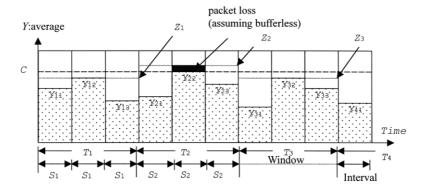


Fig. 1. Measurements in a measurement-based admission control algorithm.

soft utilization target is a finer performance knob than the measurement window. One can easily tune the value of the soft utilization target to deliver a certain loss probability constraint, while the appropriate measurement window may not exist at all. The method for tuning soft utilization target depends on the traffic whose behavior is difficult to obtain using on-line measurements. We propose a *Latch* algorithm to address the issue. In the algorithm, no new flows are admitted if the system admission controller forecasts a possible failure on loss probability constraint may occur until any flow departs.

The method proposed in the paper can also be used to support controlled-load services [4] to satisfy both utilization target and loss probability constraints. Although controlled-load services do not have explicit QoS constraints, the system manager can thus specify the QoS constraints such as utilization target and loss probability.

The rest of our paper is organized as follows. In Section 2, we first review the MBAC parameters in Ref. [1] and introduce the basic model we use. Then we show the weakness of measurement window in enforcing loss probability constraint and present our solution idea which resizes the soft utilization target. In Section 3, we derive the method to resize the soft utilization target for loss probability constraint and introduce the latch algorithm. We also arranged the pseudo codes of the admission control in this section. The simulation results are given in Section 4, we will discuss the result of loss probability constraint and performance of the admission control. Section 5 concludes our paper.

2. Measurement window and loss probability constraint

2.1. Measurements in an MBAC

In Ref. [1], the admission control algorithm works as follows. Measurement window T specifies the time span of each measurement; measurement interval of size S specifies the time interval between two measurements. First, the MBAC computes the average load Y of a measurement

interval. For a measurement window containing k measurement intervals, the *highest* average load of the k intervals is taken as the estimated load of the next measurement window. For example, in Fig. 1, the size of the measurement window is three. There are three measurement intervals — S_{11} , S_{12} and S_{13} — in measurement window T_1 . At the end of the period, the highest average load Y_{12} is obtained and used as the estimated load of the link in the admission control decisions during the next measurement period. Note that in this method, the load estimation is not dynamically updated during the period. When there are sudden load changes due to the arrival of traffic bursts, without immediate update on the network state, the admission decisions may cause violation of the OoS constraints.

A new flow requesting the rate r is admitted if

$$Z < vC - r \tag{1}$$

where Z is the load estimate obtained from the previous measurement window, v is the link utilization target and C is the link capacity. Here we notice two important relationships between Z, v and the performance of the MBAC. First, because the load estimate used is based on the measurement of the last period, the longer the duration of the measurement window, the lesser the responsiveness of the admission control, and thus the higher the possibility of violating the QoS constraints. Second, with a smaller value of the utilization target, the admission control decisions would be more conservative in admitting new flows. Here, we propose to dynamically resize the measurement window or the utilization target to adjust the resulting loss probability.

In this paper, we consider a high-speed link with a large number of flows passing through it. It is therefore assumed that the averaged loads of the measurement intervals are in steady state and independently and identically distributed with a normal distribution. The probability density function (p.d.f.) of the standard normal random variable, *x*, is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right),\,$$

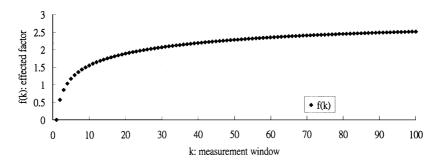


Fig. 2. A coefficient function of load variance.

and the standard normal cumulative distribution function is

$$N(y) = \int_{-\infty}^{y} \phi(x) dx.$$

For the sake of simplicity, we assume that links have no buffers. An arriving packet is dropped if it finds the outgoing link busy. The dark shadow area shown in Fig. 2 corresponds to the amount of packets dropped during T_2 . Let the loss probability constraint be denoted by $P_{\rm qos}$. The goal of the system is to ensure

$$P(\tilde{Y} > C) \le P_{\text{gos}}$$

Let the mean and standard deviation of \tilde{Y} be denoted by $\mu_{\tilde{Y}}$ and $\sigma_{\tilde{Y}}$, respectively. We have

$$1 - N\left(\frac{C - \mu_{\tilde{Y}}}{\sigma_{\tilde{Y}}}\right) \le P_{\text{qos}}.$$
 (2)

We will use this equation to evaluate the performance of the tuning knobs: measurement window and utilization target.

2.2. Measurement window

To evaluate how the size of the measurement window affects the accuracy of the load estimation, one would need to find the relationship between the expected value of \tilde{Z} and the measurement window size k, i.e. the function F that renders $E[\tilde{Z}] = F(k)$. Since the load estimate of a measurement window is set to the maximum average load of the k measurement intervals, we have $\tilde{Z} = \max\{\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3, ..., \tilde{Y}_k\}$, where $\tilde{Y}_1, \tilde{Y}_2, \tilde{Y}_3, ..., \tilde{Y}_k$ are independently identically distributed as \tilde{Y} . Given $\tilde{Y} - \mu_{\tilde{Y}}/\sigma_{\tilde{Y}} \sim \text{Normal}(0, 1)$, we can obtain the p.d.f. of \tilde{Z} as

$$f_{\tilde{Z}}(z) = kN \left(\frac{z - \mu_{\tilde{Y}}}{\sigma_{\tilde{Y}}}\right)^{k-1} \phi \left(\frac{z - \mu_{\tilde{Y}}}{\sigma_{\tilde{Y}}}\right) \frac{1}{\sigma_{\tilde{Y}}},$$

by the order statistics [9]. Now, we can evaluate the expected value of \tilde{Z} as

$$E[\tilde{Z}] = \int_{-\infty}^{\infty} z f_{\tilde{Z}}(z) dz$$

$$= \int_{-\infty}^{\infty} z k N \left(\frac{z - \mu_{\tilde{Y}}}{\sigma_{\tilde{V}}} \right)^{k-1} \phi \left(\frac{z - \mu_{\tilde{Y}}}{\sigma_{\tilde{V}}} \right) \frac{1}{\sigma_{\tilde{V}}} dz$$

Let *x* be equal to $z - \mu_{\tilde{Y}}/\sigma_{\tilde{Y}}$ and rewrite the equation, we have

$$E[\tilde{Z}] = \int_{-\infty}^{\infty} (\sigma_{\tilde{Y}}x + \mu_{\tilde{Y}})kN(x)^{k-1}\phi(x)dx = \sigma_{\tilde{Y}}f(k) + \mu_{\tilde{Y}}$$
(3)

where

$$f(k) = \int_{-\infty}^{\infty} xkN(x)^{k-1}\phi(x)dx.$$
 (4)

From Eq. (3), we can see that the expectation of the largest average load during a measurement window period is the mean traffic load plus the standard deviation times a factor which is a function of the measurement window size. Given k, Eq. (4) can be evaluated using numerical approximation. Fig. 2 shows a sketch of f(k). One can see that f(k) grows fast when k is small. After around 25, it starts to saturate. It indicates that indefinitely increasing the duration of the measurement window (observing a longer period of time) only has limited effect on the accuracy of the load estimation.

Next, we will show how to apply this result on run-time measurement. Let μ and σ denote the mean and standard deviation of the measured load \tilde{Y} , respectively. We are interested in the heavy load behavior, namely when the link utilization approaches the utilization target ν , i.e.

$$E[\tilde{Z}] \approx vC \tag{5}$$

From Eqs. (4) and (5), we have

$$\sigma f(k) + \mu \approx \nu C. \tag{6}$$

Let the utilization target v be set to its maximum value 1 and assume the equality of Eq. (6) holds. Then we have $f(k) = C - \mu/\sigma$. With Eq. (2), we have $1 - N(f(k)) \le P_{\text{qos}}$. Solving for f(k), as $N(\cdot)$ increases monotonically, we have $f(k) \ge N^{-1}(1 - P_{\text{gos}})$. We take

$$f(k) = N^{-1}(1 - P_{qos}), (7)$$

or

$$k = f^{-1}[N^{-1}(1 - P_{\text{gos}})], \tag{8}$$

if f^{-1} exists.

Table 1 presented some numerical results to illustrate the

Table 1
The relationship between measurement window size and loss probability constraint

Loss probability (P_{qos})	10^{-1}	10^{-2}	10^{-3}	10^4
$N^{-1} (1 - P_{\text{qos}})$	1.28	2.32	3.15	3.96
Window size (k)	7	61	≫ 500	≫ 500

relationship between the size of the measurement window and the loss probability constraint. Note that very large window sizes will be needed to achieve finer degrees of loss probability constraint that may require long periods of measurement. Unfortunately, we could not find the inverse function of f. This means, given the loss probability constraint, we cannot obtain the appropriate size of the measurement window.

The measurement window-based scheme is conceptually simple. But it may not be feasible to support links with loss probability constraint greater than 10^{-3} due to the large sampling overhead incurred in the run-time measurement. If the sampling interval is fixed, larger measurement window size means longer observation period. In this approach, the scheme may not be able to respond quickly to load changes, thus compromising the link utilization and its serviceability. (A new flow request might be rejected due to the inaccurate load estimation.) On the other hand, if one fixes the time duration of the measurement window, one would need to increase the sampling frequency. In this case, the measurement overhead can become a concern to the system performance.

3. Resizing soft utilization target and latch algorithm for loss probability constraint

3.1. Soft utilization target

In the previous section, we have analyzed the relationship between different loss probability constraints and the sizes of the measurement window. In this section, we look at another important performance knob in the MBAC schemes, i.e. the utilization target. We will study how to the utilization target can be set to meet the loss probability constraint. First, a new parameter called soft utilization target, denoted by ν' , is used to enforce better the loss probability constraint. It is different from the utilization target whose value is typically fixed and set by the system administrator according to some bandwidth sharing policy between the QoS-constrained traffic and best-effort traffic.

Since the goal is to ensure that the packet loss probability, at all times, is upper bounded by the loss probability constraint as given in Eq. (2), by rewriting it, we have $\mu_Y \le [1 - (\sigma_Y/C)N^{-1}(1 - P_{qos})]C$. Let the soft utilization target be set as follows:

$$v' = 1 - \frac{\sigma}{C} N^{-1} (1 - P_{\text{qos}}),$$
 (9)

Note that σ is the standard deviation of the measurements $\{\tilde{Y}\}$. If we replace ν as ν' in Eq. (1), the admission control decision inequality equation becomes as follows:

$$E[Y] < v'C - r \tag{10}$$

The σ in Eq. (9) only describes traffic variation of the existing flows. To compute the variance of the flow aggregated with the newly requested flow, we assume that the existing flows and the new one are independent and that the variance of the aggregated traffic is the sum of σ^2 and σ^2_r if the flow is admitted, where σ^2_r is the variance of the newly requested flow. For links with hundreds and thousands of flows, one can assume the value of σ^2 will be much larger than that of σ^2_r . Thus, we will ignore the variation effect introduced by the new call.

3.2. The latch algorithm

One of the major considerations in admission control is not to let the admittance of a new flow affect the promised QoS to existing flows. In the measurement-based approach, it relies on accurate estimation of the mean and standard deviation of the existing traffic loads. However, due to measurement uncertainty, especially in heavy load conditions, correcting or handling estimation errors becomes one of the major issues in the MBAC. Assuming the estimated parameters are the real parameters could be dangerous. For example, an underestimation of the flow parameters may deteriorate the QoS guarantee to admitted flows. Overestimation of the number of permissible flows may also result in inefficient use of network resources and unnecessary blocking of call requests.

To address the issue, we propose a method called the latch algorithm to dynamically adjust the admission control policy. First, if there is a violation of the packet loss constraint during a measurement interval, the system will stop admitting new flows until an existing flow departs. This is to prevent further violation of the packet loss constraint. Moreover, we consider that an occurrence of loss constraint violation may imply further occurrence of the event. It certainly shows that the total traffic load was greater than expected. Therefore, even though the current load estimate is low, one would like to temporarily close flow admittance to reduce the uncertainties introduced by admitting more flows into the system. Generally, such violation is owing to the fluctuations or burstiness of the aggregated traffic. In practice, it is still very difficult to quantize such effect at the flow admission time based on the flow priori traffic descriptions. By doing so, it will certainly help to compensate the estimation error.

Second, if the system rejects a flow, the flow admission will be temporarily closed until a flow departs. The goal is to compensate possible underestimation of the aggregated traffic load.

The two policies proposed are simple and effective in enforcing loss probability constraints. The simulation

```
(a)
extern float Accumulate;
                                 // Accumulated packet size
void Packet_Measurements(float Packet_size)
   Accumulate+=Packet_size; // Accumulate the total packet
size
(b)
                            // Accumulated packet size
extern float Accumulate;
extern float Traffic Data[];
                                  // Traffic data storage
extern int Traffic Data Index;
                                 // Latch flag
extern int Latch;
void On line Measurements()
   float Bufferless_Loss=
          (Accumulate-LINK BANDWIDTH)/Accumulate;
   if (Bufferless Loss>P QoS)
                             // Latch policy 1
   Traffic Data Index=(Traffic Data Index+1)%ROTATE STORAGE;
   Traffic_Data[Traffic_Data_Index] = Accumulate;
   Accumulate=0;
                             // Reset the accumulation
extern float Traffic_Data[];
                                       // Traffic data storage
extern int Traffic Data Index;
extern int Latch;
                                    // Latch flag
int New Flow Arrival(float Rate)
   float Mean, Std Dev, Soft UT;
   if (Latch) return(REJECT);
                                    // Reject if being latched
   Std Dev=Moving_Std_Dev();
                                    // Get the standard
deviation statistic
   // Calculate the Soft Utilization Target
   Soft_UT=1-(Std_Dev/LINK_CAPACITY)*Un_Normal(1-P_QoS);
   // Make the admission decision
   if
(Traffic_Data[Traffic_Data_Index]<Soft_UT*LINK_CAPACITY-Rate)
      return(ACCEPT);
                               // Accept a new flow
   else {
      Latch=1; return(REJECT);
                             // Latch policy 2
(d)
extern int Latch;
void Flow_Departure()
   Latch=0;
```

Fig. 3. (a) Packet measurement procedure; (b) on-line measurements procedure; (c) new flow arrival procedure; and (d) flow departure procedure.

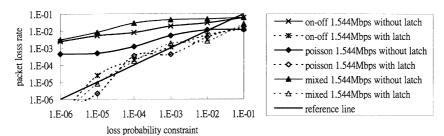


Fig. 4. Loss probability constraint and simulated loss rate in the 1.544 Mbps link bandwidth.

results shown in the next section will show the performance of maximizing link utilization without loss probability violation by employing this algorithm.

The pseudo codes of the proposed admission control policies are shown in Fig. 3a.

In Fig. 3a, the code for packet measurement triggered by the event of packet arrival is shown. In Fig. 3b, is shown the code for the on-line traffic load estimation. This function is triggered by clock alarm, once per measurement interval. It also detects the violation on the loss probability constraint. Fig. 3c is the code for handling new flow arrivals. Fig. 3d shows the function for handling flow departures.

In this function, when a new flow request is received, the admission control unit will first check if the latch flag is off. If yes, it computes the current estimate of the standard deviation of the traffic load to obtain a new value of soft utilization target. Then the admission control process is executed based on the new utilization target. If the flow is rejected, the latch flag is turned on and the flow admittance is closed. The function of "Flow Departure" is invoked by a flow departure event. It will turn off the latch flag and resume the admission control process.

4. Simulation results

In this section, we study the performance of the proposed method to dynamically resize the parameter of the target utilization in MBAC in enforcing loss probability constraints. In the simulation, links with two different bandwidths are considered: 1.544 Mbps (T_1) and T_2 and T_3 (Ethernet). Three types of traffic source are studied. The

first one is the ON-OFF traffic source. In the ON-OFF model, a flow alternates between the ON and OFF states; the duration of both states is assumed to be exponentially distributed with the mean of 2 s. The flow sends 8 kbps constant bit rate data in the ON state and is idle when in the OFF state. The second model is the Poisson process with the mean of 8 kbps. Third, we study the effect under heterogeneous traffic. Here we have traffic mixed of the above two types of flow with sending rate varying from 8 to 16 kbps. The measurement interval is set to 1 s. We use Moving-8 method for the on-line calculation of the mean and standard deviation of the traffic load. Namely, the memory of the online measurement is eight measurement intervals. The mean μ and standard deviation σ of \tilde{Y} are computed as follows: $\mu = \sum_{n=1}^{8} Y_n/8$ and $\sigma = (\sum_{n=1}^{8} (Y_n - \mu)^2/8 - 1)^{1/2}$. The loss rate, $E[(\tilde{Y} - C)^+]/E[\tilde{Y}]$ is evaluated every 8000 measurement intervals, where $(\cdot)^+$ evaluates the positive part of the result. It is assumed that the transmission link is buffer-less.

4.1. Enforcing loss constraints

Figs. 4 and 5 show the resulting packet loss probability vs. the target loss constraint for the MBAC, with or without the use of the latch algorithm, for links with 1.544 and 10 Mbps capacity, respectively. A reference line is shown to indicate whether the loss probability constraint is satisfied or not. The data points that lie below the reference line correspond to those experiments where loss constraints are not satisfied.

From the figures, one can see that for the transmissions under the same type of traffic sources, those using admission

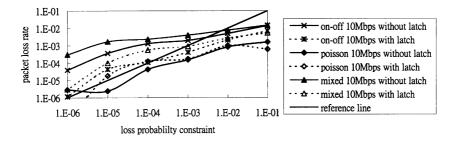


Fig. 5. Loss probability constraint and simulated loss rate in the 10 Mbps link bandwidth.

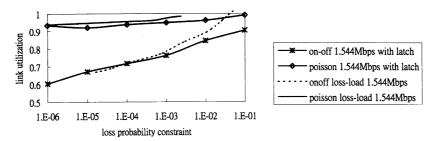


Fig. 6. Performance of the admission control in the 1.544 Mbps link bandwidth.

control with the latch algorithm achieve better performance in enforcing packet loss constraint than the ones not employing the latch algorithm. This is mainly due to the blocking of the new flows when the link load approaches the target utilization. Next, we compare the performance of the transmissions using the proposed method under different types of traffic source. From the simulation results, one can see that because the Poisson traffic sources generate the least degree of traffic burstiness, the chances of having loss constraint violation are lowered, thus having the best performance, especially for cases with more rigid loss constraint. Comparison of the performances for the same transmissions, but under different link capacities, it therefore hard.

4.2. Link utilization

Next, we would examine the link utilization performance using the proposed method. In Figs. 6 and 7, we compare the link utilization for transmissions under different link capacities. In addition, the *loss-load curves* [10] are shown in each figure for comparison. The loss-load curve gives the maximum utilization obtained to satisfy the loss probability constraint. One can see that the performance of admission control with the latch algorithm is very close to the loss-load curves.

In Fig. 7, the distance between the Poisson performance curve and the corresponding loss—load curve is about 100 kbps. Given that each Poisson flow is of 64 kbits/s average rate, the difference is about a flow. But one can achieve better loss constraint when using the latch algorithm. In the case of the ON–OFF sources, the proposed method appears to be conservative in admitting new flows. It performs better link utilization in the range of tighter loss constraints.

5. Conclusions

It has been shown that tuning certain performance knobs, such as the measurement window size, is important for the performance of the MBAC. In this paper, we first analyze how the size of the measurement window affects the accuracy of the load estimation, assuming the load samples follow normal distribution. We show that increasing the number measurement points of a measurement window period, namely a longer period of observation, has only limited effect on the accuracy of the load estimation. Next, we studied the tuning performance of another performance — utilization target — in the guarantee of the loss probability. We propose a new parameter called soft utilization target whose value is computed based on the mean of the load measures. Instead of using a fixed target utilization, the soft utilization target is used when the link load is heavy. The latch algorithm is proposed to throttle the flow admission when the link load is heavy and the aggregated traffic is bursty. Generally, the occurrence of the violation of loss probability is because of the fluctuations or burstiness of the aggregated traffic. It makes the load estimation and admission control even more difficult when the load is heavy and resources scarce. The two policies proposed are simple and effective in enforcing loss probability constraints. The simulation results show that transmissions using admission control with the latch algorithm achieve better performance in enforcing packet loss constraint. They also achieve the link utilization very close to the maximum utilization obtained to satisfy the loss probability constraint.

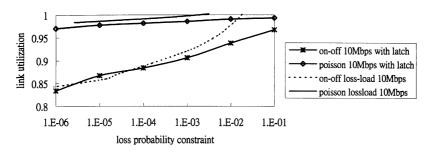


Fig. 7. Performance of the admission control in the 10 Mbps link bandwidth.

References

- S. Jamin, P.B. Danzig, S.J. Shenker, L. Zhang, A measurement-based admission control algorithm for integrated service packet networks, IEEE/ACM Transactions on Networking 5 (1) (1997) 56–70.
- [2] R. Braden, L. Zhang, S. Berson, S. Herzog, S. Jamin, Resource ReSer-Vation Protocol (RSVP) — Version 1 Functional Specification, RFC2205, September 1997.
- [3] S. Shenker, C. Partridge, R. Guerin, Specification of Guaranteed Quality of Service, RFC2212, September 1997.
- [4] J. Wroclawski, Specification of the Controlled-Load Network Element Service, RFC2211, September 1997.
- [5] J. Wrocławski, The Use of RSVP with IETF Integrated Services, RFC2210, September 1997.
- [6] M. Grossglauser, D. Tse, A framework for robust measurement-based

- admission control. Proceedings of ACM, SIGCOMM '97, 1997, pp. 237-248
- [7] D. Tse, M. Grossglauser, Measurement-based call admission control: analysis and simulation. Proceedings of IEEE, INFOCOM '97, Kobe, Japan, 1997, vol. 3, pp. 981–989.
- [8] S. Jamin, S. Shenker, Measurement-based Admission Control Algorithms for Controlled-load Service: A Structural Examination, University of Michigan, CSE-TR-333-97, April 1997. Available at http://irl.eecs.umich.edu/jamin/papers/mbac/clmbac.ps.gz.
- [9] S. Ross, A First Course in Probability, 4th ed., Prentice Hall, Englewood Cliffs, NJ, 1994, pp. 279–282.
- [10] C.L. Williamson, D.R. Cheriton, Loss-load curves: support for rate-based congestion control in high-speed datagram networks, SIGCOMM '91, 1991, pp. 17–28.