12

A family of measurement-based admission control algorithms

Z. Turányi, A. Veres, A. Oláh
Ericsson, Traffic Analysis and Network Performance Laboratory
POB 107, 1300 Budapest, Hungary
e-mail: {Zoltan.Turanyi,Andras.Veres,Andras.Olah}@lt.eth.ericsson.se
Fax: +36-1-437 7219, phone: +36-1-437 7735

Abstract

In this paper we identify a set of the requirements for an efficient admission control algorithm and propose new algorithms. These methods need only aggregate traffic measurements, work with simple FIFO scheduling and take only minimal assumptions on the pattern of the traffic. The requirements define a family of measurement based admission control algorithms of which three key members are discussed. We give effective bandwidth formulae for buffered and bufferless systems with token bucket and peak rate limited sources. It is also shown how the utilization can be improved by measuring the variance of the traffic rate while avoiding the limitations of MBAC methods based on the Central-limit theorem.

The theoretical background of this work is the effective bandwidth definition introduced by Gibbens and Kelly which has its roots in the Chernoff bounds.

1 INTRODUCTION

Measurement Based Admission Control (MBAC) has drawn considerable attention recently as it uses only simple and probably loosely fitted descriptors and by measuring the actual traffic it increases the network utilization.

Previous approaches proposed for MBAC can be categorized into two main groups: simple heuristic methods [CKT96, JJB97, JDSZ97] and mathematical bounds [Flo96, GAN91, GibKel97]. The problem with the first group is that there is little or no clue about how to parametrize the algorithms, whether it would meet a certain QoS or not, or how it would behave if the traffic pattern changes (this happens frequently in the Internet). Members of the second group are more complex but they have the benefit of their parameters being

usually direct QoS metrics. However the underlying assumptions (such as independency, traffic models like MMPP or a certain queue length distribution [KWC93]) may limit their use. Also the bounds are usually very conservative.

It is very important that the knobs which we can tune an algorithm by be in direct contact with actual QoS parameters such as delay or loss. The resulting QoS should be independent of the traffic type and close to the required value as much as possible. This way an administrator do not have to change the settings of the AC algorithm when traffic changes. The simple reference algorithm [JJB97] for example does not fulfill these expectations because its parameters are hard to set to achieve a certain QoS (e.g loss) and depend on the burstiness.

We would like to underline the related work of Brichet and Simonian [BriSim98]. They consider a leaky bucket descriptor and make a conservative Gaussian bound based on that the long-term mean rate of the flows will be always smaller than the submitted bucket rate. This gives an upper bound on the variance and leads to a tighter bound than Hoeffding's.

If we are to implement an MBAC, we face the problem of accurate measurements. It was shown in the literature [GroTse97] that the QoS bounds are very sensitive to measurement errors. These errors can be reduced by measuring the aggregate traffic rather than each flow individually. Aggregate traffic measurements are both scalable and easy to implement.

The above requirements define a family of AC algorithms:

- the required descriptors are as simple as possible,
- as few assumptions on the traffic as possible,
- should not rely on per flow traffic measurements,
- the tuning parameters should reflect and control QoS directly.

In addition the algorithm should be easy to implement and closed form solutions are preferred even if they are suboptimal.

In this paper we develop three MBAC algorithms with different types of flow descriptors and measured parameters, each fulfilling the above requirements. The first is based on our improvement of the Hoeffding bound and uses admitted peak rates and measured average rate (Section 2). The second measures the variance as well and gives a tighter bound. (Section 3). The third algorithm utilizes a leaky bucket traffic descriptor per flow and gives bounds for guaranteeing maximum delay (Section 4). Conclusions are given in Section 5.

2   A TIGHTER BOUND FOR MBAC – F_{OPT}

One very simple member of the family was already proposed for MBAC by Sally Floyd [Flo96]. It uses the *Hoeffding bound* requiring to admit only the peak rates of the entering flows. For a supplementary parameter the mean of

the aggregate traffic rate is measured. The theoretical basis of the bound is the concept of *effective bandwidth* [Kel96]. In this section we give a tighter effective bandwidth using the same information.

2.1 An tighter bound on the effective bandwidth

We use the following expression for the effective bandwidth of the aggregate traffic

$$BW(s) = \sum_{k=1}^{N} \beta_k(s) + \frac{\gamma}{s} \qquad \text{where} \qquad \beta_k(s) = \frac{1}{s} \ln \left(1 + \frac{e^{sh_k} - 1}{h_k} m_k \right) \qquad (1)$$

h_k and m_k are the peak and mean rates of flow k respectively, $\gamma = -\ln(\epsilon)$, and ϵ is the saturation probability (see [Kel96, GibKel97]). In this expression we need information for the individual values of m_k, in other words we need per flow measurements. Fortunately the average of individual flows sums up so if we can find a bound where only the sum of the measurements is present then only aggregate average measurement is needed. We bound the effective bandwidth of each flow in the form $\beta_k(s) \leq \delta_k(s) = fm_k + c_k(s)$ where f must be the same for all flows and c_k depends only on s or h_k but not m_k. With this approximation the expression of the effective bandwidth of the aggregate traffic fulfils this requirement:

$$BW(s) \leq BW(s, f) = f \sum_{k=1}^{N} m_k + \sum_{k=1}^{N} c_k(s) + \frac{\gamma}{s} \qquad (2)$$

$\beta_k(s)$ is a concave function of m_k so its tangent of slope f is an upper bound. The special choice of $f = 1$ leads to the Hoeffding bound, details can be found in [GibKel97, Flo96]. The tangent of slope f is the following:

$$\delta_k(s) = fm_k - \frac{1}{s} + f \frac{h_k}{e^{sh_k} - 1} + \frac{1}{s} \ln \left(\frac{1}{fs} \cdot \frac{e^{sh_k} - 1}{h_k} \right). \qquad (3)$$

For the particular traffic mix on figure 1 the optimal choice is around $f \approx 1.4$ and at this point it is significantly less than what we get using the Hoeffding bound ($f = 1$). The question is how to find the optimal slope for a certain traffic mix with parameters h_k and m_k. $BW(s, f)$ is convex in f and so the optimal f where it is minimal can be calculated by differentiating the sum of $\delta_k(s)$ on f. Substituting this into $BW(s, f)$ we get $BW_{fopt}(s)$.

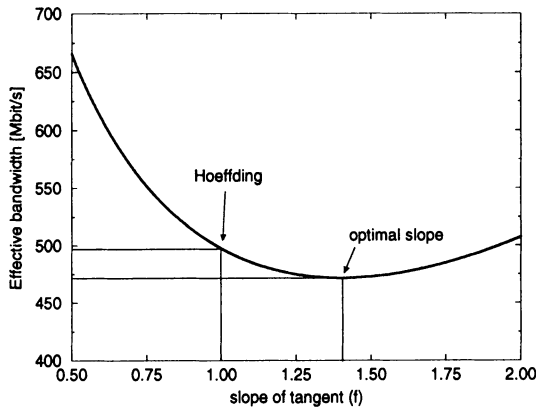$$f_{opt} = \frac{N}{s} \frac{1}{\sum_{k=1}^{N} \left(m_k + \frac{h_k}{e^{sh_k} - 1} \right)}$$

Figure 1 $BW(s, f)$ as a function of f, 100 sources $m_k = 2$, $h_k = 10$ Mbit/s, $s = 0.2$, $\gamma = 12$.

$$BW_{fopt}(s) = \frac{1}{s} \sum_{k=1}^{N} \ln \left[\frac{\sum_{j=1}^{N} m_j + \sum_{j=1}^{N} \frac{h_j}{e^{sh_j} - 1}}{N} \cdot \frac{e^{sh_k} - 1}{h_k} \right] + \frac{\gamma}{s}$$

Notice that we need the average of the aggregate traffic only. The peak-to-mean ratio influences f_{opt}. If the traffic is biased towards high or low peak-to-mean ratios then this optimal choice of f will largely differ from 1.

2.2   Finding an optimal value for s

In this section we give a closed form solution for s. For obtaining an esti-mation of the optimal s we approximate $BW_{fopt}(s)$ using Taylor series and differentiate in s.

$$BW_{approx}(s) = \sum_{j=1}^{N} m_j + \left(\frac{1}{8} \sum_{j=1}^{N} h_j^2 - \frac{1}{2} \cdot \frac{1}{N} \left(\sum_{j=1}^{N} m_j - \frac{1}{2} \cdot \sum_{j=1}^{N} h_j \right)^2 \right) \cdot s + \frac{\gamma}{s}$$

$$s_{opt} = \sqrt{\frac{\gamma}{\frac{1}{8} \sum_{j=1}^{N} h_j^2 - \frac{1}{2} \cdot \frac{1}{N} \left(\sum_{j=1}^{N} m_j - \frac{1}{2} \cdot \sum_{j=1}^{N} h_j \right)^2}}$$

In an implementation first we calculate s_{opt} to get an estimate for optimal s and substitue it to $BW_{fopt}(s)$. Then we admit the new flow if there is enough capacity between BW and the link rate C. In figure 2 we can see a typical BW_{fopt} and its approximation as an example.
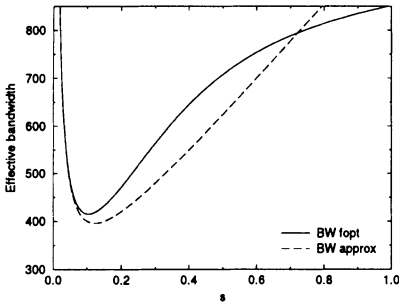
Figure 2 BW_{fopt} and BW_{approx} as a function of s, 100 sources $m_k = 2$, $h_k = 10$ Mbit/s.
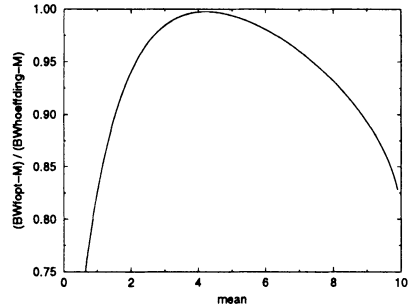
Figure 3 The ratio of the margins *Hoeffding* and f_{opt} methods add to the mean as a function of the mean. 100 sources, $h_k = 10$ Mbit/s.

If we want to compare our bound with the Hoeffding bound it is clear that the greatest difference arises when f is far from 1. This is only the question of the traffic mix. On figure 3 it is visible that for small or large peak-to-mean ratios the gain can be in the range of 5-15%. Large peak-to-mean ratios mean bursty traffic or loosely fitted descriptors; these are the situations where this MBAC has its strength over traditional AC algorithms and algorithms based on the Hoeffding bound.

3 MBAC USING MEASUREMENT OF THE VARIANCE

So far we used only the measured aggregate mean and the admitted peak rate of the flows. Theoretically the more infomation we have about the distribution of X_ks the tighter bound can be given. In this section we give a closed form effective bandwidth formula with an extra parameter which is the measured variance of the aggregate traffic rate. The validity of this bound does not depend on the number of flows unlike bounds based on the Central-limit theorem [GAN91]. The variance - similarly to the mean - sums up, so an expression with the sum of the variances of the flows can be replaced with the variance of the rate of the aggregate traffic.

The bound (1) containing only the mean and the peak rates $(\beta_k(s))$ is the same as if we assumed on/off sources [GibKel97]. On/off sources have the worst possible variance for a given peak and mean rate and apparently a large part of the real-life traffic will not be on/off, so a measurement on the variance may lead to a tighter bound.

First rewrite $\alpha_k(s)$ as (for definition of $\alpha_k(s)$ see [Kel96]):

$$\alpha_k(s) = \frac{1}{s}\ln\left(\mathbb{E}\left[e^{s(X_k - m_k)}\right]e^{sm_k}\right) = \frac{1}{s}\ln\left(\mathbb{E}\left[e^{s(X_k - m_k)}\right]\right) + m_k.$$

Then $e^{s(X_k - m_k)}$ is bounded by

$$e^{s(X_k - m_k)} \leq 1 + s(X_k - m_k) + \frac{e^{sh_k} - sh_k - 1}{h_k^2}(X_k - m_k)^2 \qquad (4)$$

for $h_k \geq m_k, X_k \geq 0$. Using the above formula the effective bandwidth is smaller than

$$\alpha_k(s) \leq \beta_k^{var}(s) = \frac{1}{s} \ln \left(1 + \frac{e^{sh_k} - sh_k - 1}{h_k^2} \sigma_k^2 \right) + m_k.$$

Here we have both the mean and the variance in the expression. Following the logic of the calculations in the previous section we can obtain the aggregate effective bandwidth and the corresponding optimal s as

$$BW^{var}(s) = \frac{1}{s} \sum_{k=1}^{N} \ln \left[\frac{\sum_{j=1}^{N} \sigma_j^2 + \sum_{j=1}^{N} \frac{h_j^2}{e^{sh_j} - sh_j - 1}}{N} \cdot \frac{e^{sh_k} - sh_k - 1}{h_k^2} \right] + \sum_{k=1}^{N} m_k + \frac{\gamma}{s}$$

$$s_{opt}^{var} = \sqrt{\frac{\gamma}{\frac{1}{2} \sum_{j=1}^{N} \sigma_j^2 + \frac{1}{18} \sum_{j=1}^{N} h_j^2 - \frac{1}{18} \cdot \frac{1}{N} \left(\sum_{j=1}^{N} h_j \right)^2}}$$

See figure 4 for an example how σ influences the effective bandwidth. As we can see the closed form approximation we use to obtain s_{opt}^{var} is not very good for small σ values. For small variances numerical optimization should be used.

Another effective bandwidth estimation is used in the literature based on the normal distribution [GAN91, GroTse97]. They assume that the aggregate traffic rate can be approximated with the normal distribution and so the effective bandwidth is given as $\widetilde{BW} \approx m + \sigma\sqrt{2\gamma - \ln 2\pi}$, where m is the average and σ^2 is the variance of the aggregate traffic rate.

This expression uses the same information as BW^{var}, but $\widetilde{BW}$ is only an approximation while BW^{var} is an upper bound. Also unlike BW^{var}, $\widetilde{BW}$ assumes normal distribution (number of flows is large) which reduces robustness.

4   MBAC WITH LEAKY BUCKET

In the previous sections we investigated the bufferless case. Now we analyze a buffered system. If the sources are able to provide more information than the peak rate such as a leaky bucket, we can do more efficient admission control.

Denote $X_k[t]$, $k = 1 \ldots N$ the number of bits sent by flow k into the network
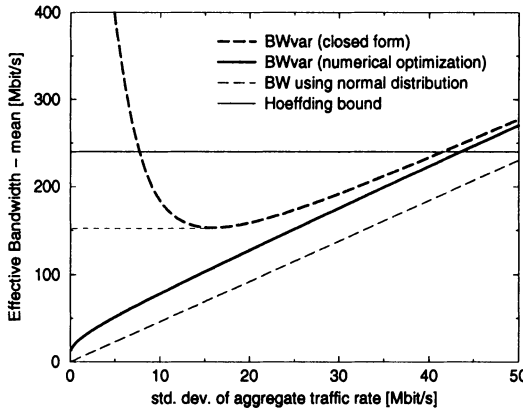
Figure 4 The effective bandwidth - mean calculated using Hoeffding bound, normal distribution and BW^{var} as the function of the aggregate σ. 100 sources, $h_k = 10$ Mbit/s, $\epsilon = 10^{-5}$.

during a time interval of length t. If flow k is controlled by a leaky bucket policer, we can give an upper bound on the maximum number of bits entering the network during any time interval of length t: $0 \leq X_k[t] \leq \sigma_k + \rho_k t$, or if a peak rate is given as well: $0 \leq X_k[t] \leq \min(h_k t, \sigma_k + \rho_k t)$.

4.1 Effective bandwidth for leaky bucket policed sources

If we assume that $X_k[t]$ are independent and stationary random variables then we can give a very simple bound for the probability that the aggregate traffic exceeds a certain number of bits B entering during a time interval of length t, i.e. $\Pr\left(\sum X_k[t] \geq B\right) \leq \epsilon$.

Using the Chernoff bound the effective bandwidth for ϵ can be written as:

$$BW(s,t) = \sum \frac{1}{t}\theta_k(s,t) + \frac{\gamma}{ts} \qquad \text{where} \qquad \theta_k(s,t) = \frac{1}{s}\ln\left(\mathbb{E}\left[e^{sX_k[t]}\right]\right).$$

$\theta_k(s,t)$ relates to the number of bits sent by flow k to the network [Kel96].

We do not know the probability distribution of $X_k[t]$ but we know that it is limited by the leaky bucket and we can measure the average number of bits entering the network during time t ($\sum m_k t$). Considering that

$$e^{sx} \leq 1 + \frac{e^{s\min(h_k t, \sigma_k + \rho_k t)} - 1}{\min(h_k t, \sigma_k + \rho_k t)}x \qquad \text{if} \qquad 0 \leq x \leq \min(h_k t, \sigma_k + \rho_k t)$$

$\theta(s,t)$ can be bounded as

$$\theta_k(s,t) \le \zeta_k(s,t) = \frac{1}{s} \ln \left(1 + \frac{e^{s \min(h_k t, \sigma_k + \rho_k t)} - 1}{\min(h_k t, \sigma_k + \rho_k t)} m_k t \right).$$

Note that a similar line of reasoning led to (1).

The requirement for aggregate mean rate measurement suggests to bound the effective bandwidth with a slope f linear function of m_k. For simplicity we write the final effective bandwidth of the aggregate traffic for $f = 1$ which leads to a form of the Hoeffding bound:

$$BW_1(t) = \sum_k m_k + \frac{1}{t} \sqrt{\frac{\gamma}{2} \sum_k (\sigma_k + \rho_k t)^2} \qquad (5)$$

or if the peak rate is available as well:

$$BW_2(t) = \sum_k m_k + \frac{1}{t} \sqrt{\frac{\gamma}{2} \sum_k \min(h_k t, \sigma_k + \rho_k t)^2} \qquad (6)$$

The value of t is the length of the interval over which the amount of incoming load is bounded. Furthermore t also gives the interval over which we limit the peak rate of the flows and measure the average rate.

If t is small, the effective bandwidth depends only on the admitted peak rates and we get back the original Hoeffding bound using only the peak rates. On the other hand a larger t leads to smaller effective bandwidth as the averaging interval is longer.

4.2   Guaranteeing maximum delay

As we have a buffered system we may allow larger bursts into the network and temporarily the link rate can be exceeded. A certain choice of t limits the maximum amount of bits to be accumulated in the buffer and so the maximum buffering delay. In the next sections we give three admission control methods that build on the concept of $BW(t)$ and statisfy a certain maximum delay requirement d_{max} with probability ϵ.

(a)   Method 1
This very simple method is applicable to both cases when we have or do not have information on the peak rate ($BW_1(t)$ and $BW_2(t)$). We try to control the delay by limiting the length of the busy periods. Let C be the link rate. If the length of a busy period (during which the server is continuously busy) is longer than t then during the first t seconds of the busy period more than Ct

bits arrived. Thus if the arriving number of bits during any time interval of length t is smaller than Ct then all busy periods are shorter than t seconds. In other words the length of the busy periods is bounded by t with probability ϵ if $BW(t)t \leq Ct$. If the length of the busy periods are bounded by d_{max} then the maximum delay is also bounded by d_{max}, thus the guarantee is met if:

$$BW(d_{max}) \leq C. \tag{7}$$

On figure 6 we can observe that under a certain delay limit the peak rate is in effect and the number of admitted flows is much higher than if only the leaky bucket descriptor were available. If we know the peak the result is identical with the bufferless case for low delays. However if the allowed maximum delay is larger then the number of admitted flows increases significantly and the statistical multiplexing gain is larger than in the bufferless case.

(b) Method 2

In methods 2 and 3 we control the delay without limiting the length of the busy period. Method 2 applies if we do not have information on the peak rate (we use $BW_1(t)$). The link can carry at most Ct bytes during time t. During a busy period t seconds after the start of the busy period the buffer occupancy is $\sum_{k=1}^{N} X_k[t] - Ct$. The buffer occupancy cannot exceed Cd_{max} to guarantee the delay limit so if within all busy periods and for all t

$$b_1(t) = (BW_1(t) - C)t \leq Cd_{max} \tag{8}$$

then the buffer overflow probability will not exceed ϵ. If we find the maximum of $b_1(t)$ in t, we can decide whether the flows fit into the link or not. It can be shown that $b_1(t)$ is a convex function of t. If $\lim_{t \to \infty} b_1'(t) < 0$ (eq. 9) then its supremum is at $t \to 0$. If $\lim_{t \to 0} b_1(t) < Cd_{max}$ as well (eq. 10) then the buffer occupancy will always be under Cd_{max}. The above conditions can be written as:

$$\sum_{k=1}^{N} m_k + \sqrt{\frac{\gamma}{2} \sum_{k=1}^{N} \rho_k^2} < C \tag{9}$$

$$\sqrt{\frac{\gamma}{2} \sum_{k=1}^{N} \sigma_k^2} < Cd_{max}. \tag{10}$$

Thus we separate the meaning of the leaky bucket parameters. ρ_k is used to determine if the link can carry our traffic on the long run as a kind of an asympthotic test, and σ_k controls the delay. The AC algorithm is very simple, it just checks (9) and (10), and if both pass then the new connection can be
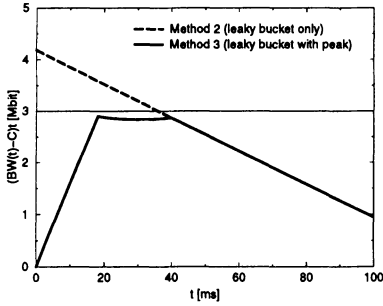
Figure 5 $b_{1,2}(t)$ for a particular traffic mix. The traffic mix contains two kinds of sources 2300 and 200 from each type respectively: $h = 1, 7$; $\sigma = 0.02, 0.1$; $\rho = 0.5, 1.5$; $m = 0.3, 1$; $\gamma = 12$; $C = 1000$ in Mbit.
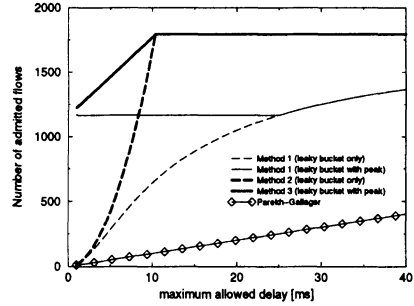
Figure 6 The number of admitted flows as a function of the delay requirement. The sources are identical $h_k = 5$ Mbit/s, $\sigma_k = 0.1$ Mbit, $\rho_k = 1$ Mbit/s, $m_k = 0.5$ Mbit/s. $\gamma = 12$, $C = 1$ Gbit/s.

admitted. We can observe some similarities between (9) and the Hoeffding bound. See figure 5 for an example of $b_{1,2}(t)$.

(c)   Method 3

If we have information about the peak rate of the flows as well, we can give a tighter bound than (10). The problem is that $b_2(t) = (BW_2(t) - C)t$ is not a convex function. We must look for its maximum and see if it is under Cd_{max}. We partition the domain of the function into intervals by points $t_k = \sigma_k/(h_k - \rho_k)$. We call these *breakpoints*. This way within each interval for any flow either $h_k t$ or $\sigma_k + \rho_k t$ is greater. Within one interval we can write an expression for $b_2(t)$ that does not contain the *min* operator. We define two sets, those flows belong to set A where $h_k t < \sigma_k + \rho_k t$ within that interval, and those belong to set B where it is the other way. Then in that interval

$$b_2(t) = \left(\sum_{k=1}^{N} m_k - C\right) t + \sqrt{\frac{\gamma}{2} \left(\sum_{k \in A} h_k^2 t^2 + \sum_{k \in B} (\sigma_k + \rho_k t)^2\right)}.$$

It can be shown that this is a convex expression. Thus within an interval $b_2(t)$ is always below the value at one of the two breakpoints. So $b_2(t)$ takes its global supremum value either

1. at $t = 0$ (then the supremum is 0) or
2. at one of the breakpoints or
3. it has no supremum, it diverges to positive infinity after the last breakpoint.

After the last breakpoints $b_2(t) = b_1(t)$, so item 3 can be checked as before in (9). By checking $b_2(t_k) < Cd_{max}$ for all $k = 1 \ldots N$ we can decide if $b_2(t) < Cd_{max}$ for $t > 0$.

On figure 5 some properties of $b_2(t)$ can be observed. $b_2(t) \leq b_1(t)$ especially for small t values. In this case the supremum is not necessarily in $t = 0$ (as in the case of $b_1(t)$) but in a breakpoint instead.

Checking the value of $b_2(t)$ in several breakpoints may be computationally expensive. The number of breakpoints is not greater than the number of different descriptors, so if several flows share common descriptors then calculation takes less time. On the other hand one class of flows can have several leaky bucket descriptors for several timescales giving a more detailed traffic description [ZhaKni94]. For example the ATM VBR descriptors (PCR,CDVT) and (SCR,MBS) are one example of such a multilevel descriptor.

On figure 6 we plotted how many identical flows can be admitted using the various methods. One can observe that supplying peak rate besides a leaky bucket improves the utilization considerably. It can be seen that a deterministic guarantee [PaGa93] is much worse as it does not exploit statistical multiplexing gain.

5   CONCLUSION

A set of requirements for efficient admission control algorithms was given. These requirements define a family of measurement-based admission control algorithms of which we worked out a few representative members. These MBAC algorithms differ in the set of information required from the sources upon flow setup and in the measurements performed on the multiplexed traffic. The algorithms are based on the Chernoff-bound with direct QoS metrics such as loss and delay as tuning parameters. In sections 2 and 3 we used a model which does not give explicit guarantees on delay. These algorithms can offer a *Controlled Load Service*. In section 4 we used a buffered model and gave probabilistic delay bounds. This method can be applied for the IETF *Guaranteed Service*.

In all methods we assume only FIFO buffering discipline so there is no need for expensive hardware with sophisticated scheduling algorithms. While the information of the admitted flows are stored for each flow, the per-packet work –the measurement– is done on the aggregate traffic making these methods scalable to tens of thousands of flows. To decrease the computational costs and delay of admission control decisions all bounds are expressed in closed form.

REFERENCES

[BriSim98] F. Brichet, A. Simonian, "Conservative Gaussian models applied

to Measurement-based Admission Control", *submitted to IEEE INFO-COM '98*, San Francisco, March. 1998

[CKT96] C. Casetti, J. Kurose, D. Towsley, "A New Algorithm for Measurement-based Admission Control in Integrated Services Packet Networks", emphProtocols for High Speed Networks '96, INRIA, Sophia Antipolis, Oct. 1996

[Flo96] S. Floyd, "Comments on Measurement-based Admissions Control for Controlled-Load Service", *unpublished*, available at ftp://ftp.ee.lbl.gov/papers/admit.ps.Z

[GAN91] R. Guerin, H. Ahmadi, M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-spped networks", *IEEE Jurnal on Selected Areas in Communications*, 9(7), pp. 968-981, Sep. 1991

[Kel96] F. P. Kelly, "Notes on Effective Bandwidths", *In F. P. Kelly, S. Zachary and I. B. Ziedins, Stochastic Networks: Theory and Applications, Royal Statistical Society, Lecture Note Series 4, p 141-168, Oxford Univ. Press*

[GibKel97] R. J. Gibbens, F. P. Kelly, "Measurement-Based Connection Admission Control", *International Teletraffic Congress 15*, Jun. 1997

[GroTse97] M. Grossglauser, D. Tse, "Towards a Framework for Robust Measurement-based Admission Control", *SIGCOM '97*, Cannes, September 1997.

[JamShe97] S. Jamin, S. Shenker, "Measurement-based Admission Control Algorithms for Controlled-load Service: A Structural Examination", *Internal report*, Apr. 1997

[JDSZ97] S. Jamin, P. Danzig, J. Shenker, L. Zhang, "A Measurement-Based Admission Control Algorithm for Integrated Service Packet Networks", *IEEE/ACM Transactions on Networking*, vol. 5. no. 1. Feb. 1997

[JJB97] S. Jamin, C. Jin, L. Breslau, "A Measurement Based Admission Control Algorithm for Controlled-Load Service with a Reference Implementation Framework", *Internet draft*, Nov. 1997

[KniZha95] E. Knightly, H. Zhang, "Traffic Characterization and Switch Utilization using a Deterministic Bounding Interval Dependent Traffic Model", in *Proc. IEEE INFOCOM '95*, Boston, April 1995

[ZhaKni94] H. Zhang, E. Knightly, "Providing end-to-end statistical performance guarantee with bounding interval dependent stochastic models", *Proc. ACM SIGMETRICS '94*, pp. 211-220.

[KWC93] G. Kesidis, J. Walrand, C. Chang, "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources", *IEEE Trans. Networking*, Vol. 1, No. 4, pp. 424-428, Aug. 1993.

[PaGa93] A. Parekh, R. Gallager, "A generalized processor sharing approach to flow control in integrates services networks: The multiple node case", *Proc. INFOCOM'93*, San Francisco, CA, Mar. 1993.