



UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE

Repository istituzionale dell'Università degli Studi di Firenze

Admission control algorithms based on self-similar Traffic modeling for IP networks.

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Admission control algorithms based on self-similar Traffic modeling for IP networks / F. BAFFONI ; F. CHITI ; R. FANTACCI ; V. VESPRI. - STAMPA. - (2005), pp. 225-236. [10.1142/9789812701817_0021]

Availability:

This version is available at: 2158/336080 since:

Publisher:

World Scientific (Singapore, New Jersey, London, Hong Kong)

Published version:

DOI: 10.1142/9789812701817_0021

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

Publisher copyright claim:

(Article begins on next page)

Admission Control Algorithms based on Self-Similar Traffic Modeling for IP Networks

Francesco Chiti⁽¹⁾, Romano Fantacci⁽¹⁾,
Fabrizio Baffoni⁽²⁾, Vincenzo Vespri⁽²⁾

⁽¹⁾Dipartimento di Elettronica e Telecomunicazioni - Università di Firenze
via di S. Marta 3, I-50139 Firenze, Italy
{chiti, fantacci}@lenst.det.unifi.it

⁽²⁾Dipartimento di Matematica Ulisse Dini - Università di Firenze
viale M. Morgagni 67/A, I-50134 Firenze, Italy
vespri@poincare.dma.unifi.it
fabaffon@inwind.it

Abstract

Future generation networks are expected to furnish differentiated multimedia services respecting the constraints of quality of service (QoS) requirements. To accomplish this task, a widely adopted approach relies on Admission Control (AC) strategies, i.e., to keep the number of active communications under a certain threshold in order to optimize the resources allocation. As a consequence, AC protocols effectiveness strictly depends upon the accuracy of the future user needs estimation in terms of bandwidth or, equivalently, of operative load conditions. Our proposal is based on the Self Similar (SS) traffic modelling; in particular, as the backbones are likely to be IPv₄ based, we refer to the class of the second order Asymptotically SS (ASS) processes. Therefore, we derive an equivalent bandwidth evaluation criterion and apply it to a DiffServ-based scenario, highlighting a network capacity increasing together with an outage probability lowering.

1. Introduction

Accurate multimedia traffic measurements have recently shown the limits of Poisson models for describing the traffic in a telecommunication network [1] and that they are likely to going to be replaced by the self-similar (SS) models that are far apart from both conventional telephone and packet oriented applications traffic models currently considered in the literature. These models have the important property of scale-invariance with respect to time aggregation: this means that the traffic looks the same for both large and small time scales. In particular [2], it can not be defined a *burst natural length* since at every time scale, ranging from milliseconds to minutes and hours, bursts are recursively composed of bursty sub-periods.

According to [3], we have that the superposition of sufficient independent and identically distributed (*i.i.d.*) *ON/OFF* sources, each exhibiting

the so called “Noah Effect” phenomenon, that is infinite variance, results in self-similar aggregate traffic. In our investigation we adopt the *ON/OFF* source model with each source characterized by high variability.

It is worth highlighting that traffic modeling is necessary to guarantee an adequate QoS. QoS management may be decomposed into two steps: identifying a traffic QoS profile and guaranteeing the respect of these constraints by means of a dynamic resources management. To this end two traffic classes, named *Stream* and *Elastic*, could be identified. The Stream traffic class is comprised of continuous data flows having intrinsic duration and rate; this class there is error-tolerant, the packets may have different priority and there is strong dependence with the delay and jitter. Examples might be interactive services and videoconferences. The Elastic traffic class is composed by digital information delivered with a variable rate. The resulting bursty flows require low error rate, the packets are processed similarly and are delay and jitter independent. Several examples can be taken within files transfer and images transfer. Therefore for stream traffic class the time integrity must be preserved, while for an elastic traffic class the semantic integrity must be preserved.

In this scenario the QoS provisioning involves the use of two important elements. The first is represented by a support architecture. One solution is represented by the *IntServ*, which aims at integrating QoS aware services along the network devices. To guarantee delay and bandwidth needs it is foreseen a protocol of resources booking, allocating a virtual channel between source, destination and all the nodes of the route. The main disadvantage is the lack of scalability, implying a sub-exploitation of the network resources. On the other hand, *DiffServ* manage several diversified services provided by the network. The main characteristics are a greater scalability, obtained from the aggregation of individual flows, and the use of a priority field, i.e., the ToS field in IP packet header, to handle its delivering. The main drawback is the necessity of a resource pre-allocation, this implying not a true dynamic bandwidth allocation especially for highly loaded networks. The second element related to QoS provision is the use of an AC algorithm; it ensures that the admission of a new flow does not violate the QoS requirements for the already active flows. In our proposal, we focus on a capacity oriented AC procedure optimizing the throughput of each flow.

The paper is organized as follows. In Section 2, we characterize the network traffic presenting two simple approximations for the equivalent capacity for a single traffic class scenario and a DiffServ scenario. In Sec-

tion 3 the numerical results of the proposed approach are presented. Finally, in Section 4 we summarize the contribution of the paper outlining possible extensions.

2. Proposed Approaches

Guaranteeing QoS requirements of multimedia applications is the main challenge in future broadband networks. When several flows are statistically multiplexed into a single channel, it is not easy to argue if and how the telecommunication system can guarantee their QoS constraints. In other words the problem is how many calls of a given type can be admitted at a given time.

As already said, QoS management requires a support architecture and an AC algorithm. We have seen that there are two major QoS frameworks defined within the Internet community, i.e., Intserv and DiffServ, both providing QoS with different approaches. In this section we focus instead on algorithms. In particular, we derive and compare two different AC algorithms based on equivalent bandwidth evaluation for a class of aggregated traffic. Afterward, the more efficient approach is implemented within DiffServ scenario comprised of two priority traffic classes.

2.1. Traffic Analysis

In this section we randomly create n *ON/OFF* connections, each with different peak rate and transition, or state, probabilities. The peak rate is chosen from the uniform distribution on $[0, 1]$, while *ON* and *OFF* probabilities are Pareto distributed with parameters able to obtain a Self-Similar traffic with Hurst parameter [2] $H = 0.8$.

The analysis of the obtained network traffic has shown that as the number of connections n increases the distribution of the aggregated traffic tends to have a Gaussian behavior. This result has allowed to consider an aggregated traffic composed by 50 connections, to derive the cumulative distribution function, as in Fig. 1, and to calculate the outage probability concerning those connections as follows:

$$P_{outage}^{Init}(i) = 1 - CDF(B(i)) \quad (1)$$

where i and $B(i)$ are, respectively, the number of activated connections and the bandwidth assigned to them. We may define the outage probability as the probability that a flow is admitted but that it is not always able to take advantage of the available resources. The value of that bandwidth is easily

obtainable from Eq. (1) once the outage probability is proactively chosen to limit the portion of interested aggregated traffic.

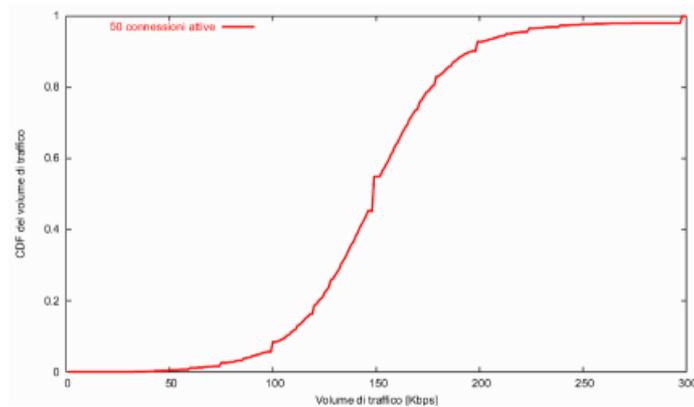


Figure 1. CDF del traffico.

2.2. AC Algorithms for a Single Traffic Class Scenario

In this section, we resort to the equivalent capacity approach for a class of an aggregate of self-similar sources. This approach is supported by the Internet architecture who encourage the use of variable bandwidth applications whenever possible; furthermore this approach is able both to guarantee a minimum bandwidth and to bound the maximum packet delay.

The algorithms estimating the equivalent capacity on which our investigation is focused on are based on the approach introduced in [4]. According to the the first model (AC-Normal Equivalent Capacity, AC-NEC), the instantaneous aggregated arrival rate has a normal distribution [5], whilst the second (AC-Hoeffding Equivalent Capacity, AC-HEC) gives an upper bound for the equivalent capacity based on a peak rate policy, taking advantage of a result derived from Hoeffding [6].

In the first algorithm, the equivalent capacity based on the normal distribution, \hat{C}_N , is given by:

$$\hat{C}_N(\mu_S, \sigma^2, \epsilon) = \mu_S + \alpha\sigma \quad (2)$$

provided that:

$$\alpha = \sqrt{2 \ln \frac{1}{\epsilon} + \ln \frac{1}{2\pi}} \quad (3)$$

where μ_S , σ^2 and ϵ are, respectively, the mean value, the variance and the outage probability of the aggregated traffic for a given class. Eq. (2) means that if the arrival rate S is characterized by a normal distribution, then the arrival rate after a time interval T , i.e., S_T , is likely to exceed the estimated equivalent capacity \hat{C}_N with a probability at most ϵ .

It is easy to notice from Eq. (3) that there are some limitations concerning the outage probability values which is upper bounded to about 40%:

$$0 \leq \epsilon \leq \frac{1}{\sqrt{2\pi}}, \quad (4)$$

this is not however a problem because ϵ is a parameter that we intend to minimize.

According to the second algorithm, the equivalent capacity \hat{C}_H , resorting to the Hoeffding bounds, is given by:

$$\hat{C}_H(\mu_S, \{p_i\}_{1 \leq i \leq n}, \epsilon) = \mu_S + \sqrt{\frac{\ln(1/\epsilon) \sum_{i=1}^n p_i}{2}} \quad (5)$$

where μ_S is the mean arrival rate of the S^{th} traffic class, $\{p_i\}_{1 \leq i \leq n}$ are the peak rates of the n admitted flows and ϵ is the outage probability.

In every approach, for each incoming flow α , the AC algorithm verifies that:

$$\hat{C}_E + p^\alpha \leq B_{max} \quad (6)$$

where p^α is the peak rate requested by flow α , B_{max} is the link bandwidth and $E \in \{N, H\}$. Eq. (6) states that a new flow is admitted if the equivalent capacity of the admitted flows plus the peak rate of the new flow is less than the allocated bandwidth for that class. Depending on the admission of the incoming flow, the load estimation is to be updated.

We give now some details of the proposed protocol. In particular, we first consider a *soft preemptive* approach, based on an upper bound for the peak rate allocation for each connection, with minimum bandwidth and maximum delay always guaranteed. We refer to soft preemption instead of real preemption, since the real-time traffics we take into account present adaptive playback times, being so able to match the time-varying delays along the network.

Besides, we consider an approach without preemption in which the last admitted flow has an unfair allocation due to the presence of already activated connections (50 in our assumption) with a great amount of assigned resources, implying more strict upper bounds on the peak rates for the incoming flows, but with minimum bandwidth and maximum delay still guaranteed. This approach is useful in the presence of short connections with a high bit-rate, whilst the former policy is necessary to assure an average rate.

2.3. Generalization to Differentiated Traffic Classes Scenario

Whenever several traffic classes are to be addressed, a generalization of the above protocols is needed.

As shown in Figs 2, 3 and 4 a soft preemptive mechanism guarantees the admission to a limited number of flows with a greater peak rate with respect to the non-preemptive mechanism. Furthermore, an estimate of the equivalent capacity based on the Hoeffding bounds is more effective than the estimate based on the normal distribution. As a consequence, the more convenient approach is represented by AC-HEC with soft preemption.

In the present scenario two traffic classes with different priority are introduced. The presence of two traffic classes and the unfairness of this algorithm reduce the available bandwidth that is given by:

$$C_{H,Classe_1} + p_1^\alpha \leq B_{max}, \quad (7)$$

$$C_H + p_2^\alpha \leq B_{max} - C_{H,Classe_1} \quad (8)$$

where p_j^α are the estimated peak rates of the requesting flows belonging to the j^{th} ($j \in \{1, 2\}$) traffic class and $C_{H,Classe_1}$ the equivalent capacity allocated to the first traffic class that has a greater priority.

This circumstance also limits the peak rates of each traffic classes:

$$p_1^\alpha \leq B_{max} - C_{H,Classe_1}, \quad (9)$$

$$p_2^\alpha \leq B_{max} - C_{H,Classe_1} - C_H. \quad (10)$$

However the traffic class with a greater priority behaves as the second is not present.

2.4. Application to DiffServ QoS Scenario

Finally, the aforementioned approaches is applied to a DiffServ QoS management scenario, deriving an equivalent bandwidth criterion resorting to the Hoeffding bounds in the case of two traffic classes. As a consequence, a new connection is accepted if:

$$C_{H,j} + p_j^\alpha \leq B_{max} - \sum_{i=1}^{j-1} C_{H,i}, \quad j = 1, 2, \dots, n \quad (11)$$

where p_j^α and $C_{H,j}$ are, respectively, the estimated peak rates of the requesting flows belonging to the j^{th} traffic class and the related equivalent allocated capacity, whilst n represents the number of traffic classes. Eq. (5) highlights that the higher priority class ($j = 1$) is not affected by the presence of the other class.

The obtained results show that the equivalent bandwidth approach is well suited for low-to-medium initial link load and, moreover, the AC-HEC is always more convenient.

3. Numerical Results

In this section, we provide several numerical results obtained via computer simulations to point out the equivalent capacity concept and evaluate its accurateness.

In particular, we first compare the estimated equivalent capacity \widehat{C}_N based on a normal distribution and the estimated equivalent capacity \widehat{C}_H based on Hoeffding bounds with an average arrival rate policy, for the case of one traffic class. We show also that a soft preemptive mechanism application allows to develop a more efficient approach. After that, the better policy is implemented within a scenario composed by two traffic classes.

3.1. Single Traffic Class Scenario

We refer here to a single traffic class with n admitted *ON/OFF* sources, comparing AC-HEC and AC-NEC with an approach based on an average arrival rate policy. We have created two different scenarios: the first in which the link is initially empty, and the second where there 50 already active connections. Moreover, we take into account a worst case scenario neglecting the connections termination. A proper outage probability ϵ range values $[10^{-6}, 2 \cdot 10^{-2}]$ depending on the QoS constraints and the link bandwidth (100MB) is introduced. Each source consists of n *ON/OFF* sources

randomly created, each with a different peak rate p_i and *ON* probability o_i . Fig. 2 highlights an approach without preemption and a link initially

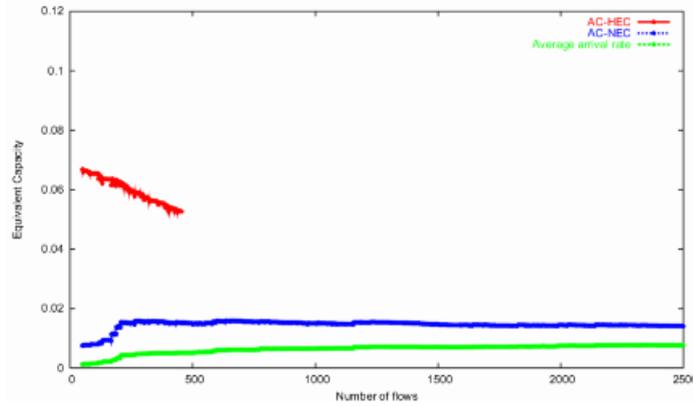


Figure 2. Assigned capacity in the case without preemption and initially occupied link.

occupied. In the subsection 2.1 we calculate the initial aggregated traffic by superposing $n = 50$ sources with an *i.i.d.* heavy-tailed distribution, hence obtaining an Asymptotically Self-Similar (A-SS) traffic with a Hurst parameter $H = 0.8$; here this traffic is used to initially bias the link. Besides, each for the generic i^{th} *ON/OFF* connection the peak rate p_i is chosen from the uniform distribution on $[0, 1]$, and the probability o_i is chosen from the Pareto distribution such that the aggregated traffic is A-SS with $H = 0.8$. It is important to notice that the peak rates are upper bounded to the following value:

$$p \leq B_{max} - (C_X - C^{Init}), \quad (12)$$

where C^{Init} and C_X are, respectively, the equivalent capacity calculated for the initially activated connections and the equivalent capacity estimated for the already admitted flows; thus an incoming flow is admitted if the requested bandwidth does not exceed the available bandwidth.

Figure 2 shows results for a P_{outage}^{Init} for the initially activated connections and a P_{outage} for the incoming flows, set to $2 * 10^{-2}$. The x -axis shows n , the number of active connections, while the y -axis is normalized and it shows the assigned bandwidth. The red line shows the AC-HEC approach, the blue line shows the AC-NEC approach and the green line shows the average arrival rate policy; in every approach the equivalent capacity is expressed

as a fraction of the sum of the allocated peak rates. The Figure suggest that a minimum bandwidth is always guaranteed and a gain of about 50% is provided for the AC-NEC approach, while AC-HEC seems to be 10 times better than the average arrival rate policy. Decreasing the P_{outage} for the incoming flows we guarantee better QoS, but with less admitted flows.

Fig.s 3, 4 show, instead, an approach with soft preemption, where the link is initially empty. This approach is more effective as it is shown by comparing the assigned bandwidth in Fig.s 2, 3 and 4, even if it depends on the initial conditions. However, to obtain better QoS, we need to reduce the number of admitted users.

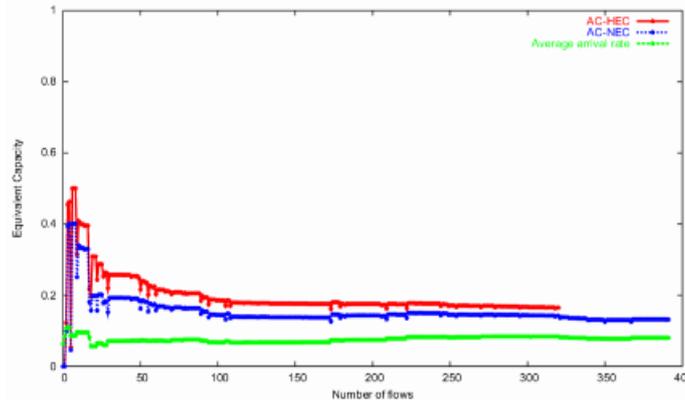


Figure 3. Assigned capacity in the case with soft preemption and initially empty link.

The peak rate for the i^{th} connection p_i and the probability o_i are uniformly distributed in the range $[0, 1]$ and $[0, \frac{1}{2}]$. Then the peak rates are upper bounded to the following value:

$$p \leq B_{max} - C_X, \quad (13)$$

where C_X is the equivalent capacity estimated for the already admitted flows.

In Fig. 3 a scenario where the P_{outage}^{Init} is fixed to $2 * 10^{-2}$ while in Fig. 4 it is decreased to 10^{-6} . and it refers to the incoming flows since there are no initially active connections.

Fig. 3 suggest that a minimum bandwidth is still guaranteed, furthermore, for high loaded link both the methods, AC-HEC and AC-NEC, get

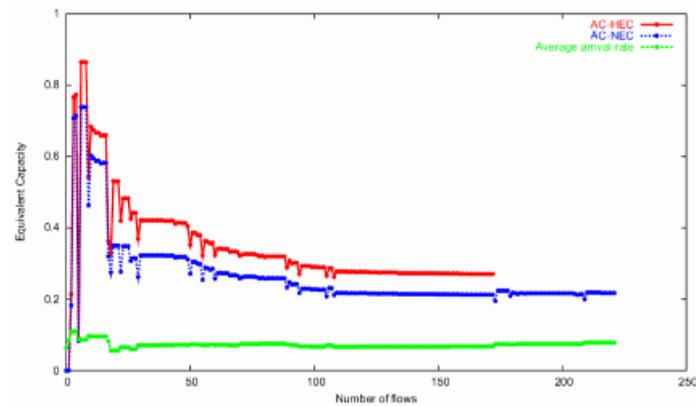


Figure 4. Assigned capacity in the case with soft preemption and initially empty link.

closer to the average. We get better QoS if we reduce the outage probability: in Fig 4, we notice an important property of the AC-HEC approach: for low traffic load, that is few connections, it would be equivalent to a peak rate AC. Moreover the proposed algorithms guarantee a maximum and a minimum capacity, this upper bounding the delivering delay as well.

The value chosen for the outage probability ϵ represents a key parameter for a generic AC-xEC procedure. In the case of a link initially empty, it is likely to chose a lower value for ϵ , resulting in a more conservative AC procedure, whilst it can be increased if the admission of a new flow points out that a less conservative AC policy would be adequate. In the opposite case, it would be useful to start with an initial greater value for ϵ for the active flows to avoid reaching the link saturation, without decreasing it in dependence with the new admissions.

Finally, in these approaches, we have considered the three way relationship who binds outage probability, bandwidth and number of users; unfortunately it is not possible to optimize the three parameters simultaneously, so we have tried to jointly optimize two of them once the outage probability is chosen: for the approach without preemption we have optimized outage probability and number of users, while in the soft preemptive approach outage probability and bandwidth are optimized .

3.2. Two Traffic Classes Scenario

This section is dedicated to generalize the equivalent capacity approximation developed in the previous section to a scenario compound by two traffic classes with different priority.

We resort to an AC-HEC algorithm with a soft preemption mechanism. We have considered a scenario with a link initially empty to apply our results to applications like fast streaming or video-conferences.

Fig. 5 shows results for an outage probability of (10^{-6}) , resulting in a not conservative algorithm. The x -axis shows n , the number of active connections, and the y -axis is still normalized and it shows the assigned bandwidth. The red and the blue lines show, respectively, the performance of the higher priority class and the lower priority class. The results has

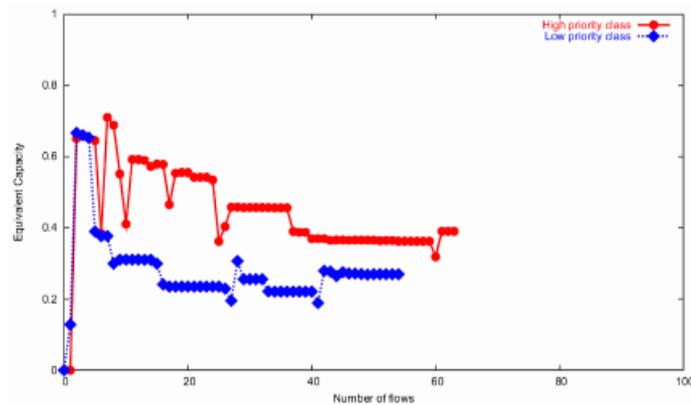


Figure 5. Assigned capacity in the case with soft preemption and initially empty link with two traffic classes.

pointed out that the algorithm behaves well also for the lower priority class even if the other is favored. Qualitatively the approach improves if the outage probability is reduced.

4. Conclusions

In this paper we apply the Self-Similar models, particularly the aggregated *ON/OFF* model to describe the real network traffic. After an admission control protocol optimized for a packet oriented telecommunication network has been proposed, after an analysis of the most significant proactive

or reactive algorithms. However, the effectiveness of an admission control algorithm strictly depends on the exactness of the evaluation of the conditions of operating load in terms of necessary bandwidth, so we focused our attention to an Equivalent Capacity approach. Such approach has been considered using two different estimate methods (AC-HEC and AC-NEC) acting with and without a preemptive mechanism. Moreover, we have compared the above methods in different scenarios in order to highlight impairments and possible improvements and we have extended the more qualitative method to a DiffServ scenario. It has also been ascertained, by means of the simulations results, that the performance of both the approaches can be improved using a *shortest remaining processing time* (SRPT) scheduling discipline. The implementation of SRPT in the case a single link even if it is complex might provide remarkable benefits to both both users and network provider by employing a flow control protocol which discriminates in favour of short documents [7].

REFERENCES

- (1) V. Paxson and S. Floyd, Wide-Area Traffic: the Failure of Poisson Modeling, *IEEE/ACM Trans. Networking*, vol. 3, 1995, 226-244.
- (2) W. Leland, M. S. Taqqu, W. Willinger and D. Wilson, On the Self-Similar Nature of Ethernet Traffic (extended version), *IEEE/ACM Trans. Networking*, vol. 2, 1994, 1-15.
- (3) W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, Self-Similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level, *IEEE Trans. Networking*, vol. 5, 1997, 71-86.
- (4) S. Floyd, Comments on Measurement-Based Admission Control Algorithms for Controlled Load Service, *Lawrence Berkeley National Laboratory Technical Report*, Jul. 1996.
- (5) R. Guerin, H. Ahmadi, and M. Naghshineh, Equivalent Capacity and its Application to Bandwidth allocation in High-Speed Networks, *IEEE Journal*, vol. 9, 1991, 968-981.
- (6) W. Hoeffding, Probability Inequalities for Sums of Bounded Random Variables, *JASA*, vol. 58, 1963, 13-30.
- (7) J. W. Roberts and L. Massouli, Bandwidth Sharing and Admission Control for Elastic Traffic, *Proc ITC*, 1998.