A Robust Adaptive Effective Bandwidth Allocation Scheme

Houjin Li, ChangCheng Huang
Department of Systems and Computer Engineering
Carleton University
Ottawa, ON, Canada

Michael Devetsikiotis
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC, USA

Yanick Champoux
Alcatel
Ottawa, On, Canada

Abstract—**Efficient bandwidth allocation has been a popular research topic during the last few years. It is a challenging issue to provide guaranteed quality of service (QoS) for network applications while still obtaining high network utilization. As a promising approach to achieve tradeoff between network utilization and the provisioning of QoS, the concept of effective bandwidth has been widely accepted. However, it requires a full characterization of the underlying process to calculate its effective bandwidth, which is not trivial. And it is also well known that the pure effective bandwidth allocation is conservative. To bypass modeling the underlying traffic and overcome the conservative nature of effective bandwidth, we propose a robust adaptive effective bandwidth allocation (AEBA) algorithm. We study the performance of the robust AEBA method under the dynamic weighted round-robin (DWRR) scheduling with a set of simulations using both self-similar traffic and traditional Poisson traffic as input. The simulation results show that our approach allows different QoS requirements to be satisfied while effectively exploiting the statistical multiplexing gain occurring among multiple traffic classes at the same time. The simulation results also show that our approach is robust in that it does not need any assumptions about the underlying traffic.**

Keywords—QoS; Effective Bandwidth; Performance

I. INTRODUCTION

Effective Bandwidth, or its synonyms, equivalent bandwidth or equivalent capacity, has been the subject of intense research activities since the idea was proposed by Hui in the seminal paper [1] in1988. Afterward, many researchers have contributed to develop the appropriate notion of effective bandwidths by investigating different traffic models.

The general notion of effective bandwidth is that: given the system parameters such as the buffer space B and channel capability C, we wish to know how to estimate the required bandwidth, say α_j, for a set of traffic sources with the QoS requirement in terms of buffer overflow probability ε, such that P{*queue length* $\geq B$} $\leq$ ε if

$$\sum_j \alpha_j N_j \leq C \qquad (1)$$

where N_j is the number of sources of type j and α_j is the effective bandwidth of source type j.

From a practical standpoint, two essential characteristics of effective bandwidths are defined in [2]:

- Independence property: the effective bandwidth of a traffic stream is independent of the other streams with which it is mixed; it is only a function of the traffic stream's parameters.

- Additive property: Effective bandwidths are additive, i.e., the sum of the effective bandwidths of a set of independent traffic streams is equal to the effective bandwidth of their superposition.

In the literature, there are many approaches to estimate effective bandwidth for a bursty source. In general, we can categorize them into two classes: the first one includes the analytic approaches based on traditional queueing theory. By hypothesizing the traffic models, an explicit expression for the effective bandwidths for some traffic sources (such as Markov and fBm processes) can be obtained. There are rich researches about such approaches, see, for example, [3] [4] [5] etc.

The second categorization is to use Kelly's mathematical definition to calculate the effective bandwidths for different kinds of traffic. Based on large deviation theory, Frank Kelly defined a mathematical framework [6] for effective bandwidth of a stationary arrival process as follows:

$$\alpha(s,t) = \frac{1}{st} \log E[e^{sX[0,t]}] \quad 0<s,t<\infty, \qquad (2)$$

where s is the space-scale parameter and t is the time-scale parameter, $X[0,t]$ denotes the amount of data that arrives from a source during the interval of length t.

Formula (2) has been the most widely accepted effective bandwidth form since Kelly defined it. The problem with respect to its practical usage is to find the appropriate values for s and t (network operating point). In general, there are two methods to calculate this operating point. One is many sources

asymptotic [7] method, which assumes that as the number of independent input increases, the buffer size and service rate per input stay fixed; the other one is the large buffer asymptotic, which is concerned with how buffer overflow probability decays as buffer size increases.

In many sources asymptotic method, the following formula is used to estimate the values of *s* and *t*.

$$\lim_{N \to \infty} \frac{1}{N} \log Q(Nc, Nb, Nn) = \sup_{t} \inf_{s} \left[st \sum_{j=1}^{J} n_j \alpha_j(s,t) - s(b+ct) \right] = -I \quad (3)$$

where *N* is the number of combined sources entering the link, *b*, *c* are the buffer and capacity per source, *I* is called the asymptotic rate function. The following two facts about many sources asymptotic approach make it only suitable for off-line effective bandwidth estimation: One is that, to estimate the effective bandwidth, the whole trace of the traffic source is needed beforehand. The other one is that *sup-inf* calculation can be very computationally intensive, particularly with long traffic traces.

In large buffer asymptotic, originally, the effective bandwidth is calculated as the time parameter *t* tends to be infinite [8]. In practice, we may choose a suitable time interval for *t* for on-line measurement-based effective bandwidth estimation. The space parameter is calculated as follows:

$$s = -\ln(Ploss/B) \quad (4)$$

where *B* is the buffer size.

As a promising approach to achieve the tradeoff between network utilization and the provisioning of QoS, the concept of effective bandwidth has been widely accepted. However, from (2), we can see that it requires a full characterization of the underlying process to calculate its effective bandwidth, which is not trivial. And it is also well known that the pure effective bandwidth allocation is conservative [9]. To bypass modeling the underlying traffic and overcome the conservative nature of effective bandwidth, we propose a robust adaptive effective bandwidth allocation (AEBA) algorithm. To provide the service differentiation for different traffic classes, we adopt the dynamic weighted round robin (DWRR) scheduling using the robust AEBA to study the statistical QoS assurance for multiple traffic classes.

The rest of this paper is organized as follows: In Section II, we describe the dedicated and robust AEBA methods. Then in Section III, we evaluate the performance of the robust AEBA using a set of simulations with Poisson, ON-OFF, and self-similar traffic as input. Finally, in Section IV, we draw the conclusions.

II. DEDICATED AND ROBUST EFFECTIVE BANDWIDTH ALLOCATION

A. Dynamic Weighted Round Robin Scheduling

To provide service differentiation for multiple traffic classes, we study the performance of effective bandwidth under DWRR scheduling instead of FIFO queueing discipline in a self-sizing network [10]. We consider a class-based DWRR
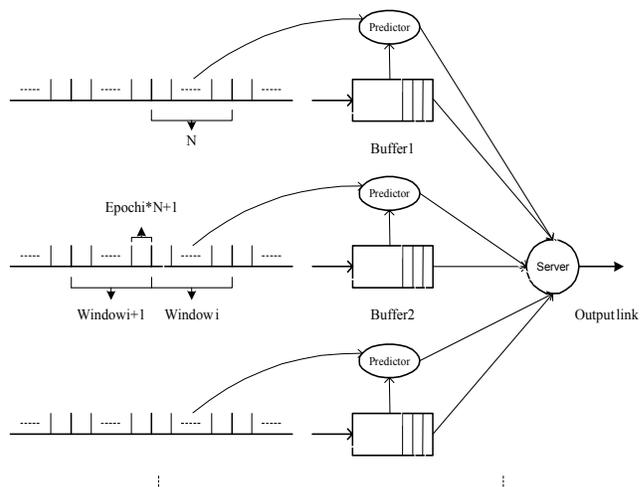

Figure 1. Dynamic weighted round robin scheduling

scheduling where packets of different classes are scheduled according to their measured arrival rates, as shown in Fig. 1.

At the incoming edge of the network, the incoming traffic streams are classified into different classes and isolated into separate buffers. Each class of traffic has similar traffic characteristics and same QoS requirement. Here, we consider the packet loss probability as the QoS requirement of interest. It is difficult to achieve *a priori* link dimensioning in self-sizing networks, a traffic prediction method is required. Noting that the most effective way to predict the traffic is to use the latest second to predict the next second, the latest minute to predict the next minute, there is not too much benefit to increase the number or lengths of the accounted time intervals in the past [5]. We use the effective bandwidth estimated in the current time window to predict the bandwidth to be allocated in the next time window. According to the additive and independence properties of effective bandwidths, we may wish to allocate the overall bandwidth according to the sum of the effective bandwidth of individual classes and adjust the weights assigned to each class according to its effective bandwidth estimated in the current time window. The scheduler visits each queue in a weighted round-robin fashion. According to the buffer occupancy and the allocated bandwidth of each traffic class, the weights are adjusted dynamically.

B. Adaptive Effective Bandwidth Allocation

When using the sum of per-class effective bandwidth as the overall bandwidth requirement, we impose a linear relationship on the effective bandwidth allocation under DWRR scheduling where such linearity does not hold. Due to the effect of statistically multiplexing multiple classes together, the bandwidth required to carry a set of classes with a certain QoS demand for each class might be less than the sum of the bandwidths that would be needed to carry each class separately with the same QoS requirement. However, due to the coupled interactions among different classes, it is very difficult to calculate the multiplexing gain. To exploit the statistical multiplexing gain among multiple traffic classes, we develop the adaptive effective bandwidth allocation approach as shown in Fig. 2.
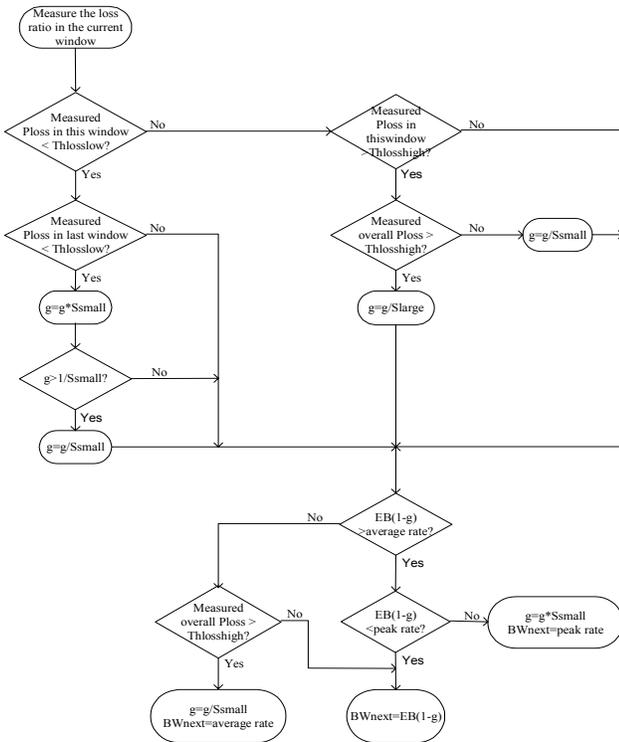
116

Figure 2. Adaptive effective bandwidth allocation

It is a two-step bandwidth allocation procedure: (1) we use the effective bandwidth as a rough approximation of the bandwidth to be allocated; (2) we adjust the bandwidth to be allocated according to the measured QoS.

According to the measured loss ratios in each window and the measured overall loss ratio, we adjust the value of multiplexing gain factor g, which adjusts the bandwidth to be allocated for the next window in the following way:

$$BWnext=Min\{Max\{EB(1\text{-}g),average\ rate\},peak\ rate\}, \quad (5)$$

where *BWnext* is the bandwidth to be allocated for the next window and *EB* is the effective bandwidth estimated in the current window.

There are two important thresholds related to the AEBA operation, which characterize each class: *Thlosshigh*, the upper threshold of the packet loss ratio for each class; and *Thlosslow*, the lower threshold of the packet loss ratio for each class.

$$0<Thlosslow<Thlosshigh< Target\ loss\ ratio \quad (6)$$

We also define two step control parameters, *Ssmall* and *Slarge*, for adjusting the value of *g*:

$$Slarge>Ssmall>1 \quad (7)$$

If the measured loss ratios are lower than *Thlosslow* in two successive windows, the over-allocation may have occurred. We increase the value of *g* by multiplying *g* with *Ssmall*, which will reduce the bandwidth allocation according to (5). If the measured loss ratio in the current window is higher than *Thlosshigh*, with high probability, *g* is too large. We reduce the value of *g* to the former value by dividing it with *Ssmall*. If the measured overall ratio is higher than *Thlosshigh* at the same

time, we need to reduce the loss ratio in the next several windows to reduce the overall loss ratio to the target loss ratio. To achieve this goal, we reduce the value of *g* by dividing it with *Slarge*. If the measured loss ratio in a window falls between [*Thlosslow*, *Thlosshigh*], we don't change the value of *g*. By this way, we can measure the multiplexing gain:

$$Gain=\frac{MeasuredEffectiveBandwidth-ActuallyAllocatedBandwidth}{MeasuredEffectiveBandwidh} \quad (8)$$

C. Dedicated Adaptive Effective Bandwidth Allocation Method

The effective bandwidth of one source is determined by its traffic characteristic and its associated QoS requirement. Since the notion of effective bandwidth was originally introduced to simplify the call admission procedure in the packet-switched networks, a lot of researchers have shown the existence of effective bandwidths for a large class of traffic models. For example, in [8], the following effective bandwidth formula for a Poisson traffic source with average arrival rate of λ events per time unit is derived:

$$EB_{Poisson}=\lambda\ \frac{e^{s}-1}{s}\ , \quad (9)$$

where $s=-\ln Ploss/B$.

Another effective bandwidth formula which is often mentioned in literature is Gaussian approximation [3]. Assuming that the distribution of the aggregate arrival rate of a set of traffic streams can be accurately approximated by a Gaussian distribution, the effective bandwidth for such aggregate traffic is derived as:

$$C=m+\sigma\sqrt{-2\ln(Ploss)-\ln(2\pi)}\ , \quad (10)$$

where m is the mean aggregate rate, σ is the standard deviation of the aggregate rate. We should note that Gaussian approximation is a bufferless model, (10) is obtained from the approximation for the inverse of Gaussian distribution of the aggregate rate, of which the cumulative tail probability beyond C does not exceed *Ploss*.

And in [5], Norros modeled the aggregated connectionless traffic as the fractional Brownian motion traffic and derived its associated effective bandwidth.

Since the traffic of different applications will exhibit different statistical characteristics, for example, some exhibit short-range dependence (SRD) characteristic and others exhibit long-range dependence (LRD) characteristic, we may wish to use different effective bandwidth approaches for different classes of traffic to get better performance. We name this method as dedicated AEBA approach in the following discussion.

D. Robust Adaptive Effective Bandwidth Allocation

However, it is often not feasible to get the reliable traffic models that can capture the statistical characteristics of the actual traffic, the assumptions made about the traffic in effective bandwidth approach may not be in agreement with the actual traffic. Recent research on statistical QoS has attempted

to estimate the required bandwidth without assuming a specific traffic model [11]. Therefore, we prefer to use a robust AEBA approach, which would bypass modeling the underlying traffic.

We choose (2) to calculate the effective bandwidth in AEBA algorithm. For practical purpose, the block estimator [12] is applied to (2) to make it suitable for real-time on-line measurement:

$$\alpha(s,t) = \frac{1}{st} \log \frac{1}{N} \sum_{i=1}^{N} e^{sX[(i-1)t,it]}, \qquad (11)$$

where N is the window size.

III. SIMULATION RESULTS

To evaluate the performance of our proposal, we do simulations with the heterogeneous traffic as input. We use both SRD traffic and LRD traffic as input.

We set the initial value of the multiplexing gain factor g=0.05, upper loss ratio threshold *Thlosshigh*=0.98**target loss ratio*, lower loss ratio threshold *Thlosslow*=0.8**target loss ratio*, small step control parameter *Ssmall*=1.1 and large step control parameter *Slarge*=1.3.

A. Traffic Profiles

Table I gives the traffic profiles for each class. Without loss of generality, we use the Poisson traffic source, the superposition of the Fractal Renewal Point Process (Sup_FRP) [13] and the superposition of ON-OFF sources to generate packet-based traffic. To obtain general results, we use a variable packet size to make the traffic more realistic.

TABLE I.   TRAFFIC PROFILES

Traffic Sources		Class 1	Class 2	Class 3
Target loss ratio		0.001	0.01	0.05
Traffic models		Poisson	Sup_FRP (H=0.7)	ON-OFF (50) ON: 0.362s OFF: 0.65s
Arrival rate (packet/s)		1000	1000	50 when ON
Buffer size (Packet)		50	100	20
Packet size (Byte)		Poisson (100)		
Effective Bandwidth Method	*Dedicated*	Poisson	Block Estimator	Gaussian Approximation
	Robust	Block Estimator		

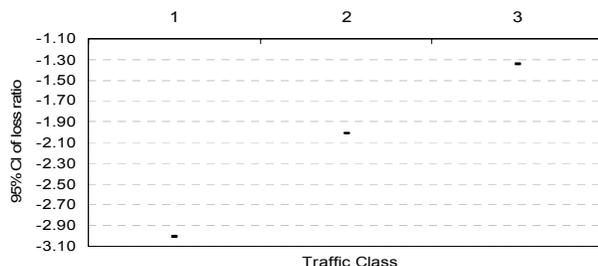#### B. Simulation Results with Dedicated AEBA Approach

Figure 3. Measured loss ratio with dedicated effective bandwidth approach
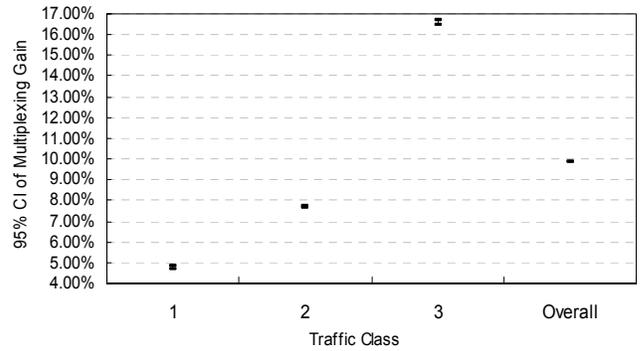

Figure 4. Measured multiplexing gain with dedicated AEBA approach

The simulation results in Fig.3 and Fig.4 show that the different QoS requirements of all classes can be satisfied at the same time.

Meanwhile, the overall multiplexing gain of around 9.85% is achieved. For class 1, the multiplexing gain is around 4.75%, for class 2, it is around 7.68%, and for class 3, it is around 16.57%.

C. Simulation Results with Robust AEBA Approach

The simulation results in Fig.5 and Fig. 6 show that the different QoS requirements of all classes can also be satisfied with the robust AEBA approach.

Meanwhile, the overall statistical multiplexing gain of around 5.56% is achieved. For class 1, the multiplexing gain is around 4.25%, for class 2, it is around 7.77%, and for class 3, it is around 5.56%.
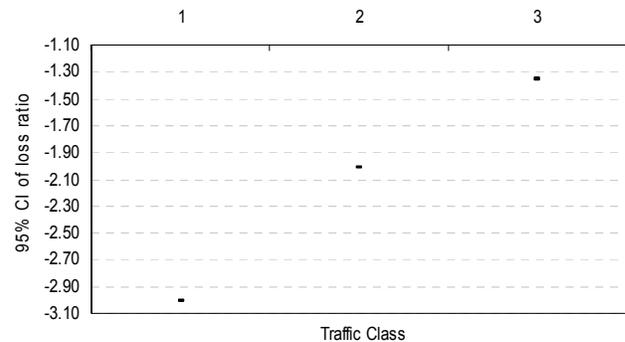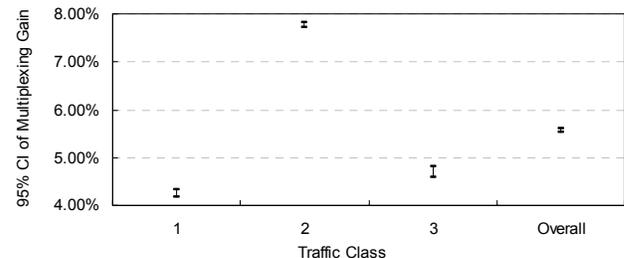

Figure 5. Measured loss ratio with robust AEBA approach

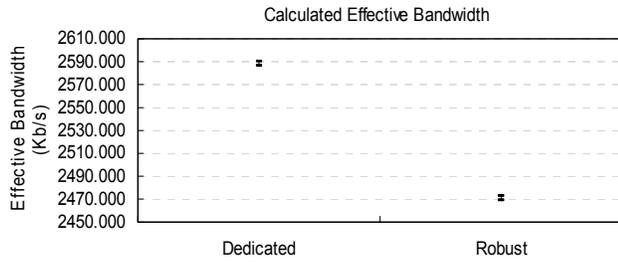Figure 6. Measured multiplexing gain with robust AEBA approach

118

Figure 7. Comparison of the calculated effective bandwidth of two approaches
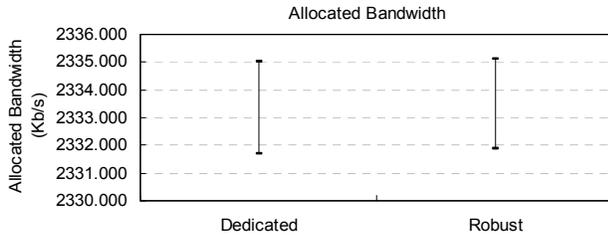

Figure 8. Comparison of the actually allocated bandwidth of two approaches

D. Comparison of Dedicated and Robust AEBA Approaches

The simulation results show that there is a significant difference of the overall multiplexing gains of these two approaches. However, we should note that, as shown in Fig.7, the calculated effective bandwidths are different in these two cases since we use different effective bandwidth methods to measure them. And as shown in Fig. 8, the actually allocated bandwidths are almost the same with these two approaches. Since the QoS requirements of all classes can be satisfied with both methods and the bandwidth consummation are almost the same with these two methods, we may conclude that the robust AEBA approach has the same performance as the dedicated AEBA approach.

Comparing the simulation results shown in figures 4 and 6, we can find that the difference between the multiplexing gains achieved by these two approaches is mainly caused by class 3, which is shown more explicitly in Fig.9. The reason is that, the Gaussian approximation does not fully exploit the amount of statistical gain that can be achieved since it does not take into account the effect of buffer. Since the buffer will absorb some of the traffic fluctuations, the actual allocated bandwidth could be much less than the upper bound calculated with (10).

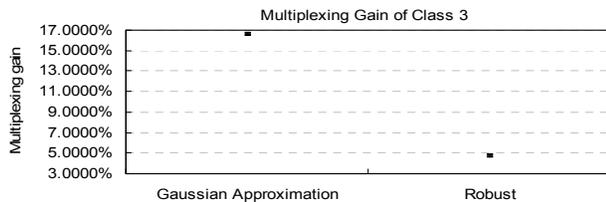The reason that the actually allocated bandwidths are

Figure 9. Comparison of the measured multiplexing gain of class 3 with two approaches

almost the same with these two methods is due to the AEBA algorithm which will adjust the bandwidth to be allocated according to the measured loss ratio adaptively.

IV. CONCLUSION

We have evaluated the performance of the dedicated AEBA approach and the robust AEBA approach with heterogeneous traffic combined with both LRD and SRD traffic as input. The simulation results show that the robust AEBA approach can also allow the different QoS requirements of all traffic classes to be satisfied at the same time with almost the same network resources consumed by the dedicated AEBA method. And both methods can overcome the conservative nature of effective bandwidth and exploit the multiplexing gain among multiple traffic classes effectively.

The robust AEBA approach is robust in that it does not require the *a priori* knowledge about the underlying traffic, which makes it more practically useful than the dedicated AEBA approach.

REFERENCES

[1] J. Y. Hui, "Resource Allocation for Broadband Networks," IEEE Journal on Selected Areas in Communications, vol. 6, No.9, pp.1598-1608, December 1988.

[2] J. Roberts, U. Mocci, J. Virtamo(Eds.), Broadband Networks Teletraffic: Final REPORT of Action COST 242, Springer, 1996.

[3] R. Guérin, H. Ahmadi, M. Naghshineh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High-speed Networks," IEEE Journal on Selected Areas in Communications, vol. 9, no. 7, pp.968-981, September 1991.

[4] A. Elwalid, D. Mitra. "Effective Bandwidth of General Markovian Traffic Sources and Admission Control of High Speed Networks," IEEE/ACM Transactions on Networking, vol. 1, no.3, pp.329-343, June 1993.

[5] I. Norros, "On the Use of Fractional Brownian Motion in the Theory of Connectionless Networks," IEEE Journal on Selected Areas in Communications, vol. 13, No.6, pp.953-962, August 1995.

[6] F. Kelly, "Notes on Effective Bandwidths," Stochastic Networks: Theory and Applications, pp.141-168, Clarendon Press, Oxford, 1996.

[7] C. Courcoubetis, V.A. Siris, G.D. Stamoulis, "Application of the Many Sources Asymptotic and Effective Bandwidths to Traffic Engineering," Telecommunication Systems 12, pp.167-191, 1999.

[8] G. Kesidis, J. Walrand, C.S. Chang, "Effective Bandwidths for Multiclass Markov Fluids and Other ATM Sources," IEEE/ACM Transactions on Networking. vol. 1, No.4, pp.424-428, Aug. 1993.

[9] M. Schwartz, Broadband Integrated Networks, Prentice-Hall, 1996.

[10] H. Li, C. Huang, M. Devetsikiotis and G. Damm, "Extending the Concept of Effective Bandwidth to Diffserv Networks," Processings of 17th annual IEEE Canada Conference on Electrical and Computer Engineering, Niagara Falls, Ontario, Canada pp.1669-1672 May 2004.

[11] R. Boorstyn, A. Burchard, J. Liebeherr, C. Oottamakorn. "Effective envelopes: statistical bounds on multiplexed traffic in packet networks," Proceedings of INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol.3, pp.1223 – 1232,March 2000.

[12] N.G. Duffield, J.T. Lewis, N.O'Connell, R. Russell and F. Tomey, "Entropy of ATM Streams: A Tool for Estimating QoS parameters,"IEEE Journal on Selected Areas in Communications, vol. 13, no. 6, pp.981-990, August 1995.

[13] B.K. Ryu and S.B.Lowen, "Fractal traffic models for Internet simulation," Computers and Communications, pp. 200-206, 2000.