Mark Bromirski & Krzysztof Łysek
Military University of Technology
Warszawa, Poland

New effective bandwidths concept for Call Admission Control in ATM networks

ABSTRACT

This paper presents a new method of approximation of effective bandwidths for CAC function in ATM networks. The accuracy of the Extended Effective Bandwidths method proposed in the paper was examined through simulation experiments.

INTRODUCTION

One of the important problems in designing ATM systems is implementation of *Connection Admission Control* (CAC). The stringent performance requirements for ATM networks (e.g. 10^{-9} cell blocking probabilities) have led several authors to approach the important problem of connection admission control via asymptotics for steady state distributions in queueing models [1]. The idea is to consider the steady-state blocking probability *p(x)* as the buffer size *x* gets large. For Markovian traffic processes (MP), it is possible to show that

$$p(x) \sim \alpha e^{-\eta x} \quad \text{as} \quad x \to \infty \qquad (1)$$

or the weaker results

$$\log p(x) \sim -\eta x \quad \text{as} \quad x \to \infty \qquad (2)$$

where *x* runs through the integers (number of cells) η is a positive constant called the *asymptotic decay rate* and α is a positive constant called *asymptotic constant*.

An appealing simple approximation based on the asymptotics in (1) or (2) is

$$p(x) \approx e^{-\eta x} \qquad (3)$$

Approximation (3) can be a reasonable substitute for (1) if we are primarily interesed in the buffer size producing a target blocking probability, rather than the blocking probabilities themselves. For example, if we seek x_p such that $p(x_p) = p$, and apply (1) for this purpose, then we get

$$x_p = \frac{\ln\alpha - \ln p}{\eta} \qquad (4)$$

When *p* is very small and α is not to different from 1, then *ln α - ln p* will be well approximated by *-ln p*. Then (3) will be a good approximation for determining x_p.

Approximation (3) is appealing because the asymptotic decay rate η is relatively easy to determine, exactly or approximately, while the asymptotic constant α in (1) is not. Approximation (1) is also appealing because under (3) the bandwidth requirement of sources is additive, so that (3) leads to a relatively simple algorithm for admission control using a concept of *"effective bandwidths"* [1]. Hence, we call (3) the *effective bandwidth approximation* (EB).

On the other hand recent traffic measurement studies from a wide range of working packet networks have convincingly established the presence of significant statistical features that are characteristic of *fractal traffic processes* (FP), in the sense that these features span many time scales. Of particular interest in packet traffic modelling is a property called *long-range dependence* (LRD) which is marked by the presence of correlations that can extend over many time scales. Leland et al. [2] observed the Ethernet traffic seems to look the same in the large scales (min, h) as in the small (s, ms). A number of quantities have been evaluated to demonstrate the invalidity of the Markovian models:

Index of dispersion for counts (IDC) is given by the variance of the number of arrivals in an interval of length *t* divided by the mean number of arrivals:

$$IDC(t) = \frac{Var[N_t]}{E[N_t]} \qquad (5)$$

where N_t is the number of arrivals in an interval of length *t*. The IDC has been defined in order that a Poisson process the value of IDC(t) = 1 for all t. We see in [2] that for FP IDC(t) increases monotonicaly throughout a time span of 6 orders of magnitude. In contrast all finite MP have indices of dispersion to fixed values over time scales.

Hurst Parameter - Let $X = (X_t\!: t = 0,1,2,...)$ be a covariance stationary stochastic process and define

$$X^{(m)} = \frac{1}{m}(X_{km-m+1} + ... + X_{km}), k \geq 1 \qquad (6)$$

The process *X* is called second order self-similar with self-similarity parameter $H = 1-\beta/2$ if for all $m = 1,2,...$, $var(X^{(-m)}) = \sigma^2 m^{-\beta}$ with 0<β<1 (σ - variance for m=1). Estimating β, it is possible to deduce

Hurst parameter H. β is given by the slope of the diagram $\log_{10} \frac{\mathrm{var}(X^{(m)})}{\sigma^2}$ to $\log_{10}$(m).

Table 1 compares main differences between the Markovian Process and the Fractal Process.

Table 1 A comparison of some statistic parameters of MP and FP

Parameter	MP	FP
$\mathrm{var}(X^{(m)})$	$\sim m^{-1}$, $m \to \infty$	$\sim m^{1-\alpha}$, $m \to \infty$
IDC	const, $T \to \infty$	$\sim T^{2-\alpha}$, $T \to \infty$
H	0.5	> 0.5

where α is an parameter of Pareto distribution which is discussed in next Section.

Willinger et al. [3] revisit the Bellcore Ethernet LAN traffic and extract from the aggregate traffic the traces generated by individual source-destination pairs. Statistical analysis of these traces reveals that:

- the traffic generated by each pair is consistent with an ON/OFF model;
- the distribution of the sojourn times in the ON/OFF states can be accurately described using Pareto-type distributions which exhibit infinite variance.

Thus, the examined traffic data are not only consistent with self-similarity of aggregate packet traffic, but they are also in full agreement with given below explanation. It is reasonable to assume, that LAN traffic measured on Ethernet can be examined at three major levels of behaviour corresponding to certain resolution of time:

- The connection level describes the human behaviour. The connection duration is determined by the file sending time and file length. In tactical LAN networks both parameters are additionally determined by specific requirements and limitations. The duration between calls on an Ethernet is typically in time range of 10 - 1000 s.
- The TCP/IP level describes the transport level. The traffic sent on the network depends of an uncontrollable number of parameters but the major influences on it is the network behaviour. The transmission duration of a TCP/IP packet varies typically from 0.01 - 10 s.
- The Ethernet network level where the sent traffic depends essentially on the local traffic flowing on the network. The time between sending and not sending a frame is typically in the range 1 - 50 ms.

Above considerations come to conclusion that *Classic Effective Bandwidth Approximation (CEB)* done by equation (3) may be not adequate in case of FPs. This paper describes a concept of *Extended Effective Bandwidth Approximation* (EEB) which is formulated especially for implementation CAC function in ATM networks with self-similar traffic. The rest of the paper is organized as follows. In Section 2 the proposed EEB method is discussed. Section 3 gives a comparison of both CEB and EEB approximations for different traffic mixes. Section 4 concludes with a summary of the paper.

EEB METHOD

When the number of traffic sources in the system is increased, then parameter α in (1) can be very small or very large (for bursty sources, α is typically less than 1), and CEB approximation in (3) is much too conservative; e.g., α can be 10^{-5} or less. Then CEB method can underestimate the number of sources the system can handle by a factor two or more.

Since CEB approximation performs well in some parameter regions, above analysis does not completely rule out this method. Moreover, it may still be possible to exploit asymptotic results in a different way to obtain an effective simple CAC algorithm [4].

For detailed examination of the problem discussed in the paper, as a surrogate for the steady-state blocking probability *p(x)* in a system with capacity *x*, we actually consider the tail probability *P(W > x),* where *W* is the steady-state waiting time until beginning service in an infinite-capacity system. When the service times are deterministic with mean 1, as is the case with ATM cells, the steady-state waiting time coincides with the steady-state queue length or buffer content. We are thus approximating the steady-state blocking probability in the finite-capacity system by the steady-state probability that the buffer content at an arrival epoch exceeds the capacity level in the finite-capacity system. In particular, analysis is based on the $\sum_{i=1}^{n} G_i / \mathrm{G}/1$ queue, which has a single server, unlimited buffer, FIFO discipline and i.i.d. service times that are independent of a superposition arrival process. The complicated feature of this model is the arrival process; it is the superposition of *n* independent general arrival process. Our analysis methods permit these component arrival process to be both *Batch Markovian Arrival Processes* (BMAPs), as in [5] or *Batch Self-Similar Arrival Processes* (BSSAPs), as in [6]. Since superposition of independent BMAPs are again BMAPs, and superposition BSSAPs are again BSSAPs, it suffices to consider both the BMAP/G/1 and the BSSAP/G/1 queues. The proposed method permit comparison of effectiveness of both CEB and EEB approximations examined in experiments.

In the next step we have developed algorithms for calculating the tail probabilities *P(W > x)* exactly and refined three-term approximation proposed for CEB method in [7] of the form

$$P(W > x) \approx \alpha_1 e^{-\eta_1 x} + \alpha_2 e^{-\eta_2 x} + \alpha_3 e^{-\eta_3 x} \qquad (7)$$

where α_1 and η_1 are the asymptotic constant and asymptotic decay rate in (1).

For EEB approximation, we use an exactly self-similar model, based on *Fractional Brownian Motion* (FBM) which has been proposed by Norros [8]. In this model the total amount of traffic arriving to a system until time t is given by

$$A(t) = mt + \sqrt{cm}Z(t), \quad t \in (-\infty, \infty) \qquad (8)$$

where Z(t) is normalized FBM characterized by the self-similarity parameter H ∈ (0.5, 1). Norros uses a scaling analysis to derive analytic expression with regards to the *Quality of Service* (QoS) criteria. In particular Norros shows that the complementary queue distribution is asymptotically bounded by a stretched exponential or Weibull form

$$P(L > x) \approx ae^{-\gamma x^{\beta}}, \quad 0 \le \beta \le 1 \qquad (9)$$

where $\gamma = f(c, m, H)$ and $\beta = 2 - 2H$. This form of the queue length distribution for H > 0.5, is much heavier than the exponential decay predicted by traditional model.

According to (7) and (9) we propose refined three-term approximation for EEB method of the form

$$P(L > x) \approx a_1 e^{-\zeta_1} + a_2 e^{-\zeta_2} + a_3 e^{-\zeta_3} \qquad (10)$$

where a_1 are the asymptotic constant in (9), and

$$\zeta_n = -\gamma_n x^{\beta} \qquad (11)$$

Using our exact numerical algorithms for both the BMAP/G/1 queue and the BSSAP/G/1 queue, we have investigated the approximations in (1), (3), (7) and (10).

SIMULATION EXPERIMENTS

In this section we consider properties of both CEB approximation and EEB approximation in case when two kinds of ON/OFF sources produces the traffic offered to the examined system. These traffic sources are described as follows:

Markovian source (MS) - This ON/OFF source has exponentially distributed *on* and *off* periods. During the *on* periods, arrivals occur according to a Poisson process; during the *off* period there are no arrivals. Each MS is characterized by three parameters, the mean *on* period ω_{MS}, the mean *off* period ξ_{MS} and constant rate r_{MS} during the *on* period.

Fractal source (FS) - Each FS can be in one of two states, active or idle. In active state, a source generates cells at a constant rate r_{FS}. In the idle state, a source does not generate cells. The time spent by FS source in active state is the random variable τ which has distribution such that

$$P\{\tau > x\} \cong x^{-\alpha}, \ \ \mathrm{x} \to \infty, \ \ 1 < \alpha < 2 \qquad (12)$$

Equation (12) means that τ has a Pareto-type distribution with a finite mean ω_{FS} and infinite variance. The off period of the FS is exponentially distributed with the mean period ξ_{FS}.

Figure 1 shows a set of four variance-time plots which were obtained by plotting log{var[$X^{(m)}$]} against log m, where for each m = 1, 2, ..., the aggregated process $X^{(m)} = \{X^{(m)}{}_k\}$ is obtained by averaging the original traffic process X over nonoverlapping intervals of size 10m miliseconds. The plot compares measured traffic (data from the delta modulation voice sources and Bellcore LAN traffic) with simulated traffic generated by number of the MSs as well as the FSs.

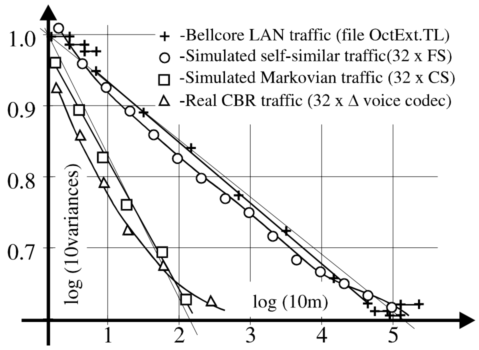

Fig.1. Variance-time plot for different traffic traces

As depicted in Figure 1, short-range dependent processes (generated by voice codecs and CSs) are characterized by an asymptotic slope of -1. For both self-similar processes (Bellcore LAN traffic and FSs traffic), the asymptotic slope parameter is readily estimated to be about - 0.45, resulting Hurst parameter *H* estimate of H ≈ 0.78. The result of this experiment confirms good properties of the simulation tools which are used in the next phase of experiments. This phase first repeats examination of refined three-term approximation proposed for CEB method in [7]. Obtained results are presented in Figure 2.

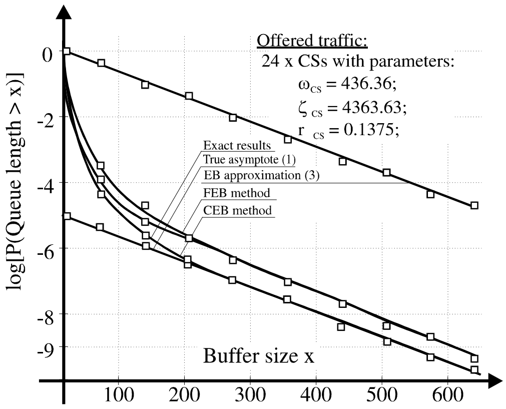

Figure 2. A comparison of approximations and exact method of the buffer overflow for pure Markovian traffic case

Figure 2 displays the exact tail probabilities $P(W > x)$ for $x \leq 600$ and the approximations (1), (3), (7) and (10). Note that indeed α in (1) is very different from 1 ($\alpha \approx 1.5 \text{ x } 10^{-5}$). The plot from Figure 2 indicates that (1) and (7) are still in error by a factor 4 at x = 600. This error is relatively small, though, compared to the error in (3) which is by factor of 10^5. The proposed EEB approximation obviously is even significantly better than all other compared methods.

In next experiment all approximations studied in the paper were compared for the case of pure fractal traffic which was generated by 12 FSs. The parameters of these sources are chossen (as in the case above), so that

$$P(W > 600) \approx 10^{-9} \qquad (13)$$

For pure self-similar traffic the EEB approximation preserve small constant error for all values of the buffer length. In this case, the buffer overflow is greatly underestmated by both (1) and (7) approximations. The EB (3) formula predicts exactly the cell losses for buffer length $x \approx 550$, but for other *x* values significantly overestimates.

CONCLUSIONS

The Extended Effective Bandwidth approximation is proposed in the paper. The simulation experiments shows that EEB method is fully applicable for ATM networks with Markovian traffic as well as self-similar traffic. The practical implementation of the proposed method is easy and relies on having only three statistical parameters of each traffic source model. The evaluation on real life traffic streams shows, that the RMS error was usually less than half an order of magnitude. The present results can thus be used as a starting point for further studies of effective bandwidth methods in ATM systems. However, as with any approximation further tests are always required.

REFERENCES

[1] W.Whitt, ”Tail Probabilities with Statistical Multiplexing and Effective Bandwidth for Multi-Class Queue”, Telecommunication Systems, vol. 3, 1995.

[2] W.Leland, ”High Time Resolution Measurements and Analysis of LAN Traffic”, IEEE Infocomm’91, 1991.

[3] W.Wilinger, ”Statistical Analysis of Ethernet LAN Traffic at the Source Level”, Proc. Signalcomm’95, Boston, 1995.

[4] R.Guérin, ”A Unified Approach for Bandwidth Allocation and Access Control”, Proc.Infocom’92, 1992.

[5] D.Lucantoni, ”New Results on the Single Server Queue with a BMAP”, Stochastic Models 7, 1991.

[6] M.Bromirski et al. ”Batch Self-Similar Arrival Process in Effective Bandwidth Approximation”, to appear.

[7] G. Choudhury et al. ”Heavy-Traffic Approximation for the Asymptotic Decay Rate”, Stochastic Models 5,1994.

[8] I.Norros, ”A Storage Model with Self-Similar Input”, Queueing Systems Theory and Applications, vol. 16, 1994.