Many Sources Asymptotic and Effective Bandwidths: Investigation with MPEG Traffic*

C. Courcoubetis, V.A. Siris, and G.D. Stamoulis
Institute of Computer Science†
Foundation for Research & Technology - Hellas (FO.R.T.H)
Science and Technology Park of Crete
P.O. Box 1385, GR 711 10 Heraklion, Crete, Greece
Tel: +30 81 39 1726, Fax: +30 81 39 1601
Email: {courcou,vsiris,gstamoul}@ics.forth.gr

May 30, 1997

The ability to quantify and accurately measure resource usage in broadband networks is important in areas such as network dimensioning, call acceptance control, and charging for network services. Closely related is the ability to identify and measure quantities such as the relevant space and time scales (see below). This is important for accurate and efficient traffic modeling. The *many sources asymptotic* [CW96, BD95, SG95], based on the theory of large deviations, is an approach for estimating the overflow probability and identifying the relevant time scales at a buffered ATM output link which multiplexes a large number of sources. In addition, the *effective bandwidth* formula based on this asymptotic appears to be a promising approach for quantifying resource usage [Kel96, Gib96]. Specifically, the effective bandwidth of a source of type j is defined as

$$\alpha_j(s,t) = \frac{1}{st} \log \mathbf{E}[e^{sX_j[0,t]}] \,, \tag{1}$$

where $X_j[0,t]$ is the workload produced by a source of type j in an interval of length t. The parameters t (*time scale* parameter measured in, e.g. milliseconds) and s (*space scale* parameter measured in, e.g. kb^{-1}) are specific to a link's operating point. The many sources asymptotic allows us to compute the correct values of s, t, hence makes (1) more accurate than the definition based on the large buffer asymptotic [EM93, dVW95].

We investigate the accuracy of approximations based on the many sources asymptotic, and evaluate the effective bandwidth formula for real MPEG traffic. The specific questions we address include the following:

- How does the overflow probability predicted by the many sources asymptotic and its Bahadur-Rao improvement [HW94, MdV96] compare with the actual cell loss probability?
- What is the maximum utilization if we use (1) as a measure of resource usage? How is the utilization (multiplexing gain) affected by increasing the link capacity?
- How do the values of s, t computed by the theory compare with the values computed by direct estimation using simulation?
- How do the values of s, t, and subsequently the effective bandwidth, depend on the traffic mix?

The next section presents the basic results from the many sources asymptotic and its Bahadur-Rao improvement, which are used in the numerical investigations presented in Section 2. Section 3 concludes with a summary of our results.

1 Many Sources Asymptotic Approximations and Effective Bandwidths

Consider an ATM output link with buffer $B = Nb$ and capacity $C = Nc$, which multiplexes $N\rho_j$ sources of type j. N is a scaling parameter, and b, c denote respectively the amounts of buffer and capacity per source.

*This work was supported in part by the European Commission under ACTS Project CASHMAN (AC-039)
†The authors are also with the Department of Computer Science, University of Crete, Heraklion, Greece

Assume that there are J different source types. Let P[overflow] denote the buffer overflow probability, and $X_j[0,t]$ the load produced by a source of type j in interval t. If $X_j[0,t]$ has stationary increments, the following holds [CW96]:

$$\lim_{N\to\infty}\frac{1}{N}\log\ \mathrm{P[overflow]} = \sup_t \inf_s [st\sum_j \rho_j \alpha_j(s,t) - s(b+ct)], \tag{2}$$

where $\alpha_j(s,t)$ is the effective bandwidth of a source of type j and is given by (1).

The many sources (or large N) asymptotic leads to the following approximation for the overflow probability

$$\mathrm{P[overflow]} \approx e^{-NI} = e^{N\sup_t \inf_s [st\sum_j \rho_j \alpha_j(s,t) - s(b+ct)]}. \tag{3}$$

An improvement of the above can be obtained with the Bahadur-Rao theorem [HW94, MdV96]

$$\mathrm{P[overflow]} \approx e^{-NI-\frac{1}{2}\log(2\pi N\sigma^2 s^2)}, \tag{4}$$

where $\sigma^2 = \frac{\partial^2}{\partial s^2}\log \mathbf{E}[e^{sX_j[0,t]}]$ and (s,t) is an extremizing pair in (2). The term $\frac{1}{2}\log(2\pi N\sigma^2 s^2)$ can be approximated (see [MdV96]) by $\frac{1}{2}\log(4\pi NI)$, hence no addition computation (other than I) is required.

Next we derive the the effective bandwidth constraint when the target QoS is P[overflow] $\le e^{-\gamma}$. If (s,t) is an extremizing pair in (2), then locally at (s,t) the following holds [Kel96]

$$\sum_{j\in J} n_j\alpha_j(s,t) \le C + \frac{1}{t}\left(B - \frac{\gamma}{s}\right) = C^*, \tag{5}$$

where C^* is the amount of effective capacity. Next we derive the effective bandwidth constraint using the Bahadur-Rao improvement (4). Because $\frac{1}{2}\log(2\pi N\sigma^2 s^2)$ can be approximated by $\frac{1}{2}\log(4\pi NI)$, (4) gives $NI \ge \gamma - 1/2\log(4\pi NI)$, which can be approximated[1] by $NI \ge \gamma - \frac{1/2\log(4\pi\gamma)}{1+1/2\gamma} = \gamma'$, from which we get the following effective bandwidth constraint

$$\sum_{j\in J} n_j\alpha_j(s,t) \le C + \frac{1}{t}\left(B - \frac{\gamma'}{s}\right) = C^*_{\text{B-R}}. \tag{6}$$

An important observation is that in both (5) and (6) the formula for the effective bandwidth is the same.

We conclude this section with an interpretation of the time scale t and space scale s parameters. Parameter t represents the most probable duration of the busy period prior to buffer overflow [CW96], and s indicates the degree of multiplexing: a smaller value of s corresponds to a larger degree of multiplexing. This can be seen from (2) where the extremizing value of s becomes smaller as the capacity C increases. From (5), we have that $s = \frac{\partial\gamma}{\partial B}$, where $\gamma = -\log$ P[overflow], based on which we can directly estimate s using the ratio of differences

$$s = \frac{\Delta\gamma}{\Delta B}. \tag{7}$$

2 Numerical Investigations

In our experiments we use the popular MPEG-1 trace of the *Star Wars* film [Ros95, Gib96] and modeled voice traffic. The MPEG stream has the frame pattern IBBPBBPBBPBB (12 frames) and rate 25 fps (frames per second). The total duration is approximately 30 minutes and the mean rate approximately 0.26 Mbps[2]. In

[1] By setting $NI = \gamma + \epsilon$ and taking the expansion of $\log[4\pi(\gamma+\epsilon)]$

[2] This is the mean rate after the stream has been segmented into ATM cells.

the simulation model, we produce independent streams of the single MPEG trace by randomly selecting the starting frame for each stream. Also, we assume that frames from individual streams are aligned on frame boundaries, and that cells belonging to the same frame are evenly distributed within that frame.

For modeling a voice source (used only in the last experiment), we use an on/off Markov modulated fluid with peak rate 64 kbps, mean "on" phase 352 milliseconds, and mean "off" phase 650 milliseconds (this gives mean rate 24.4 kbps). Finally, in all the results presented in this note, the link capacity is 155 Mbps.

Cell Loss Probability (CLP) estimation

Consider homogeneous traffic where all sources produce *Star Wars* MPEG-1 traffic. Figure 1(a) shows the overflow probability estimated by the two approximations (large N and large N with the Bahadur-Rao improvement) and the overflow probability estimated using simulation. Observe that the large N approximation overestimates the cell loss probability by approximately three orders of magnitude, while the B-R asymptotic by two orders of magnitude. Also observe that the asymptotic approximations are closer to estimating the probability of overflow in a frame[3]: the large N asymptotic overestimates the frame overflow probability by approximately one order of magnitude. Finally, the cell loss probability is smaller than frame overflow probability by approximately two orders of magnitude. This is because when a frame overflow occurs, typically only a few cells belonging to the frame are lost.

Achievable utilization using the effective bandwidth formula

Figure 1(b) shows the average achievable utilization[4], for different buffer sizes and target Cell Loss Probability (CLP) equal to 10^{-4}. Observe that for link capacity 155 Mbps, statistical multiplexing is efficient, hence a high utilization (over 88%) can be achieved. Furthermore, the achievable utilization with the large N asymptotic, even for small buffers (e.g., for a buffer size equal to 351 cells, corresponding to a 1 msec delay), is 78%, which is 89% of the maximum utilization (obtained from simulation). For larger buffer sizes (over 8 msec or 2800 cells), the achievable utilization using the effective bandwidth formula is 95% of the maximum utilization. The B-R improvement consistently achieves a higher utilization, exceeding that of the large N asymptotic by an additive 2.5%. We note here that, for small buffer values, the utilization is related to the (optimistic) assumption that cells belonging to the same frame are evenly spaced within that frame. Any other distribution would decrease the achievable utilization.

Parameters s, t: theory versus direct estimation

Next we compare the values of s, t estimated theoretically (by numerically solving the supinf of equation (3)) with the values estimated directly using (7): $s = \frac{\Delta\gamma}{\Delta B}$, where $\gamma = -\log(\mathrm{CLP}_{\mathrm{sim}})$, and $\mathrm{CLP}_{\mathrm{sim}}$ is the cell loss probability estimated using simulation. The value of t is estimated by taking the average duration of the busy period prior to buffer overflow. The corresponding results are shown in Figures 2(a),(b), from which we see that the values of s, t estimated theoretically are in good accordance with the values estimated using simulation.

Dependence of the effective bandwidth on the traffic mix

Our final investigation deals with the dependence of the effective bandwidth with the traffic mix containing MPEG compressed video sources and voice traffic. In Figures 3(a),(b), the horizontal axis depicts the percentage of voice connections, where each connection is a bundle of 30 individual voice channels, each modeled by the on/off Markov model defined in the beginning of this section. From the figure we observe that the effective bandwidth does not exhibit frequent and large changes when the traffic mix varies. The small "jumps" of the effective bandwidth are due to the MPEG frame pattern (a further discussion of this falls outside the scope of this note). These observations suggest that we can use a different pair (s, t) to characterize periods of the day during which the traffic mix remains relatively constant.

[3] A frame is said to have an overflow if at least one cell belonging to that frame is lost.

[4] Defined as $N \times m/C$, where m is the mean rate (0.26 Mbps for the *Star Wars* stream), N is the number of sources, and C is the available capacity.

3 Conclusions

We have presented numerical results investigating, for MPEG traffic and modeled voice traffic, approximations based on the many sources asymptotic and the resulting effective bandwidth formula. Our results are summarized as follows:

- For MPEG traffic, the many sources approximations can underestimate the cell loss probability by two orders of magnitude.
- Despite the previous, using the effective bandwidth formula (1), we can achieve a high utilization of link capacity. This is enabled by the efficiency of statistical multiplexing for the link capacity considered (155 Mbps).
- The values of the space scale s and time scale t parameters estimated theoretically follow the values estimated directly very well.
- The dependence of the parameters s, t on the traffic mix are such that the effective bandwidth does not exhibit frequent and large changes. Hence, a set of (s, t) pairs, one per period of day during which the traffic mix is relatively constant, can be used to characterize a network's operating point. Such a set can be used to build effective tariff structures [CKW97].

References

[BD95] D. D. Botvich and N. Duffield. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems*, 20:293–320, 1995.

[CKW97] C. Courcoubetis, F. P. Kelly, and R. Weber. Measurement-based charging in communications networks. Technical Report 1997-19, Statistical Laboratory, University of Cambridge, 1997.

[CW96] C. Courcoubetis and R. Weber. Buffer overflow asymptotics for a switch handling many traffic sources. *Journal of Applied Probability*, 33, 1996.

[dVW95] G. de Veciana and J. Walrand. Effective bandwidths: Call admission, traffic policing and filtering in ATM networks. *Queueing Systems*, 20:37–59, 1995.

[EM93] A. I. Elwalid and D. Mitra. Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *IEEE/ACM Trans. on Networking*, 1(3):329–343, October 1993.

[Gib96] R. J. Gibbens. Traffic characterisation and effective bandwidths for broadband network traces. In F. P. Kelly, S. Zachary, and I. Zeidins, editors, *Stochastic Networks: Theory and Applications*. Oxford University Press, 1996.

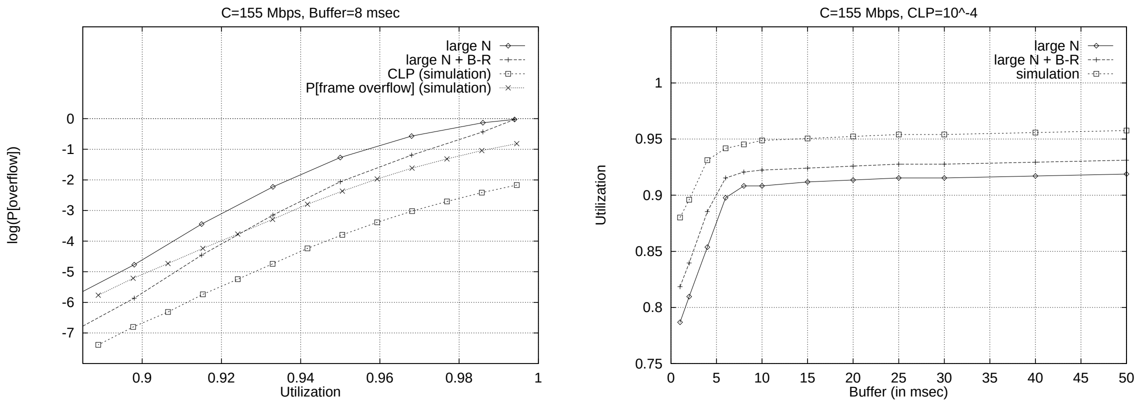
[HW94] I. Hsu and J. Walrand. Admission control for ATM networks. In *IMA Workshop on Stochastic Networks*, 1994.

[Kel96] F. P. Kelly. Notes on effective bandwidths. In F. P. Kelly, S. Zachary, and I. Zeidins, editors, *Stochastic Networks: Theory and Applications*, pages 141–168. Oxford University Press, 1996.

[MdV96] M. Montgomery and G. de Veciana. On the relevance of time scales in performance oriented traffic characterizations. In *Proceedings of IEEE INFOCOM'96*, pages 513–520, April 1996.

[Ros95] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. Technical Report 101, University of Wuerzburg, February 1995.

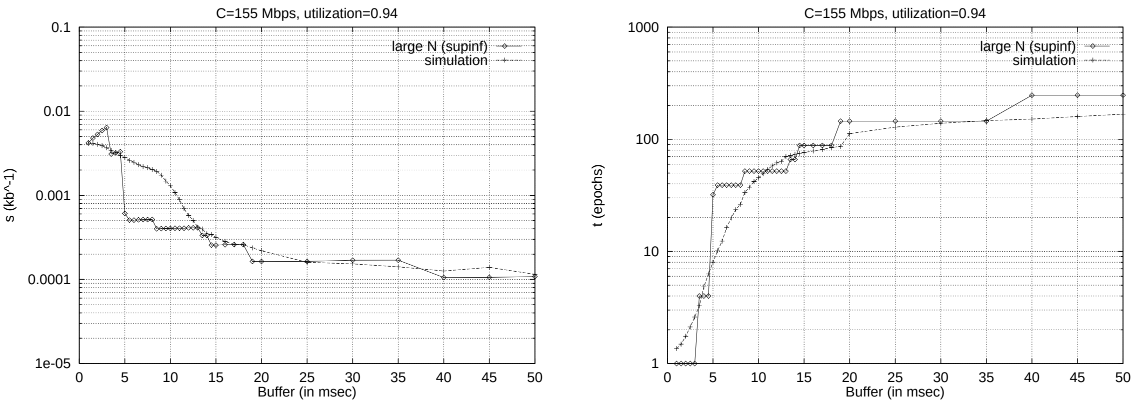
[SG95] A. Simonian and J. Guibert. Large deviations approximations for fluid queues fed by a large number of on-off sources. *IEEE J. Select. Areas Commun.*, 13(7):1017–1027, September 1995.



(a) Overflow probability

(b) Utilization

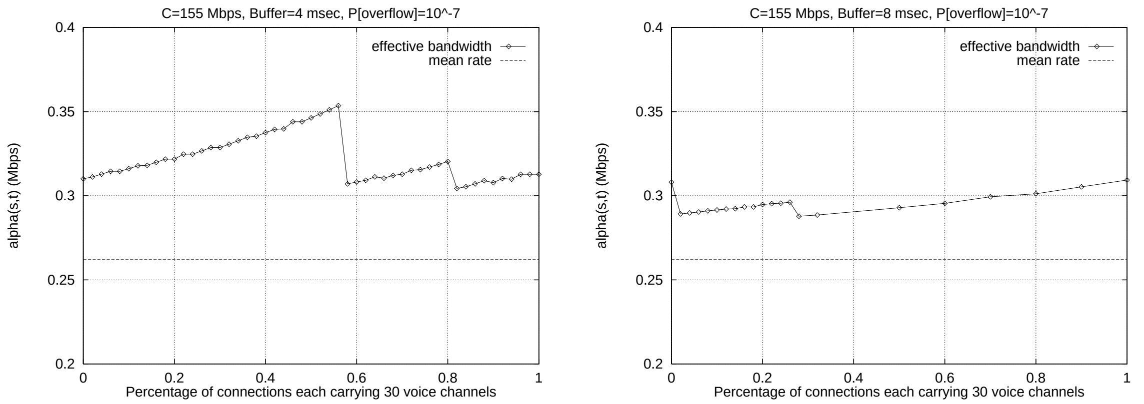
Figure 1: Many sources asymptotic approximations.



(a) Parameter s

(b) Parameter t (1 epoch = 40 msec)

Figure 2: Parameters s, t: theory vs. direct estimation.



(a) Buffer=4 msec

(b) Buffer=8 msec

Figure 3: Dependence of the effective bandwidth of the *Star Wars* stream on the traffic mix.