Note on Effective Bandwidth of ATM Traffic [1]

Costas Courcoubetis [2] *and Jean Walrand* [3]

[2] University of Crete, Heraklion
[3] University of California at Berkeley

ABSTRACT

The objective of this note is to explore the existence of an effective bandwidth for ATM (asynchronous transfer mode) traffic. We show that such an effective bandwidth exists for a class of stationary Gaussian sources. We also show that the effective bandwidth cannot be defined for general non-stationary sources by providing an elementary counter-example.

1. Introduction

A number of recent papers have shown the existence of an *effective bandwidth* for some traffic models ([4], [5]). What this mean is that the set of acceptable numbers N_i of sources of type i, for $i = 1, 2, ..., J$, that can go through a common buffer with service rate c is such that

$$\sum_{i=1}^{J} N_i \alpha_i \leq c. \tag{1}$$

Here, acceptable means that the resulting cell loss probability because of buffer overflow is less than some specified small value. Thus, the result says that one can consider that each source of type i uses a bandwidth α_i.

This result is somewhat surprising in that one might reasonably expect that the set of acceptable numbers N_i would be characterized by a nonlinear expression. Thus, the question arises of identifying the generality of the concept of effective bandwidth.

In terms of small loss probability, the effective bandwidth has been justified for the following two models. In [5], at each time $n \geq 1$, a source of type i produces a random batch of X_n^i cells. The buffer has capacity B and it serves one cell per second. The X_n^i are independent and their distribution depends only on i. The buffer can then be viewed as a GI/GI/1 queue where the interarrival times are deterministic. It is shown in [5] that the probability that an arriving cell finds a queue length larger than some large B is smaller than a specified value if and only if (1) holds where

$$\alpha_i = \frac{B}{\gamma} \log E\{\exp[\frac{\gamma}{B} X_n^i]\}.$$

In [4] it is shown that if the sources are two-state Markov modulated fluid sources and if the buffer serves the fluid at a constant rate c, then an effective bandwidth can also be defined for each source.

In this paper, we prove the existence of an effective bandwidth for a class of stationary Gaussian sources. We also show that, in general, sources do *not* have an effective bandwidth. We exhibit an elementary counter-example in section 2. In section 3 we perform a simple calculation for i.i.d. Gaussian batches. This result, a particular case of that in [5], introduces the method that we will need for the next section. In section 4, we prove the existence of an effective bandwidth for a class of stationary Gaussian

[1] Research supported in part by a Nato Grant, by a grant from Pacific Bell, and by a matching grant from the state of California. This research was performed while the second author was visiting the University of Crete.

sources. Section 5 is an appendix that provides some details about a derivation of section 4. The results are summarized in the conclusions of section 6.

2. A counter-example

A source of type 1 produces cells at rate α between time 0 and time 1 and stops producing after time 1. A source of type 2 produces cells at rate β between time 1 and time 2 and does not produce cells at other times. There are N_1 sources of type 1 and N_2 sources of type 2. The total traffic of cells produced by all these sources goes through a buffer that serves the cells at a constant rate of c cells per second. We consider a fluid model of the evolution of the number x_t of cells in the buffer at time t. Assuming that the buffer is initially empty, we find that

$$x_1 = (N_1\alpha - c)^+$$

$$x_2 = (x_1 + N_2\beta - c)^+.$$

Let us assume that the buffer capacity is equal to B cells. We want to find conditions on N_1 and N_2 so that very few cells are lost because of buffer overflow. Specifically, we want to find the set of vectors (N_1, N_2) so that

$$x_1 \le B \tag{2}$$

and

$$x_2 \le B. \tag{3}$$

A straight-forward analysis shows that these inequalities imply that

$$N_1 \le (B + c)/\alpha \tag{4}$$

and

$$N_2 \le \frac{c + B - (N_1\alpha - c)^+}{\beta}. \tag{5}$$

The set of admissible vectors (N_1, N_2) is shown on Figure 1.

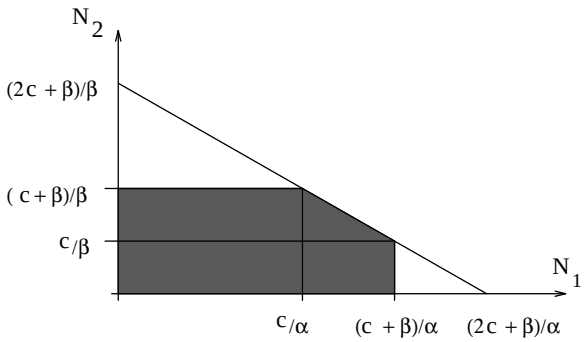

Figure 1. The set of admissible vectors.

As the figure shows, the set of admissible vectors is not characterized by a linear boundary. Consequently, no effective bandwidth can be defined for these simple sources.

The model can be made more complicated by randomizing the traffic, but the conclusion is rather clear: no effective bandwidth can be defined for general sources.

3. I.i.d. Gaussian batches

The sources of the counter-example of section 2 are highly non-stationary. We will see in section 4 that an effective bandwidth can be defined for a class of stationary Gaussian sources. Before doing this, we perform a simple calculation in the case of i.i.d. Gaussian batches. This calculation will introduce the method used in section 4 on a simple example.

A source of type i produces X_k^i cells at time k ($k \geq 1$), where the X_k^i are independent and $N(m_i, \sigma_i^2)$. (The notation $N(\mu, \sigma^2)$ designates a Gaussian random variable with mean μ and variance σ^2.) Also, the sources are all independent. (The positive part of these random variables can be taken to avoid negative numbers, but since this does not change the analysis, we will not do it.) The cells produced by all the sources go into a buffer that serves c cells at each time k. The buffer capacity is B and we want to estimate the probability that the buffer will overflow. We assume that B is large, so that overflows are very unlikely. That is, we are interested in cases where the overflow corresponds to a large deviation of the behavior of the sources. Specifically, we assume that $\sum_{i=1}^{J} N_i m_i < c$, so that it is unlikely that the queue length in the buffer will reach a large value B before becoming empty again.

To perform the analysis of the large deviations, we argue as follows. (See [1], [2], [3], [6] for related derivations.) An initially empty buffer will overflow before becoming empty again if the sources happen to produce cells at a rate larger than the average rate. Specifically, the overflow will occur if the N_i sources produce cells at a total rate $N_i a_i$ for all i in such a way that

$$\sum_{i=1}^{J} N_i a_i = \alpha + c, \tag{6}$$

for some $\alpha > 0$. In that case, the buffer will overflow at time $n := B/\alpha$ since the quantity of cells stored in the buffer grows at rate $\alpha + c - c = \alpha$.

It turns out that the most likely way for the N_i sources of type i to produce cells at a total rate $N_i a_i$ for a long time n is for each of these sources to produce cells at rate a_i. Moreover, the probability that this will occur is of the order of

$$\exp\{-nN_i I_i(a_i)\} \quad \textit{where} \quad I_i(a_i) = \frac{(a_i - m_i)^2}{2\sigma_i^2}. \tag{7}$$

(See [1] for this result.) Thus, the probability that this occurs simultaneously for all $i = 1, ..., J$ is approximately equal to

$$\exp\{-n\sum_{i=1}^{J} N_i I_i(a_i)\}.$$

Consequently, the probability that the sources of type i produce at rate a_i for $n = B/\alpha$ in such a way that $\sum_{i=1}^{J} N_i a_i = \alpha + c$ is approximately equal to

$$\exp\{-\frac{B}{\alpha}\sum_{i=1}^{J} N_i I_i(a_i)\}. \tag{8}$$

Thus, the probability that the sources will produce at a total rate $\alpha + c$ until the buffer overflows is approximately equal to the sum over all possible combinations of rates a_i such that $\sum_{i=1}^{J} N_i a_i = \alpha + c$ of the expressions (7). When B is large, this sum of exponentials is approximately equal to the exponential with the smallest factor of $-B$, i.e., to

$$\exp\{-\frac{B}{\alpha} I^*(\alpha)\}, \tag{9}$$

where

$$I_*(\alpha) = \min \sum_{i=1}^{J} N_i I_i(a_i), \tag{10}$$

where the minimization is over the a_i subject to

$$\sum_{i=1}^{J} N_i a_i = \alpha + c. \tag{11}$$

Thus, the probability that the buffer overflows because the sources produce cells at *some* rate $\alpha + c$ is approximately equal to the sum of expressions (9) where the sum is over different values of $\alpha > 0$. When B is large, we can argue that this sum is essentially equal to $\exp\{-BK\}$ where

$$K = \min_{\alpha} \frac{1}{\alpha} I^*(\alpha). \tag{12}$$

Let us first calculate $I^*(\alpha)$. To do this, we must solve the problem (10)-(11). One finds, by Lagrange multipliers, that the optimal a_i satisfy

$$N_i I_i'(a_i) = 2\lambda N_i, \ \ i = 1, ..., J,$$

for some λ. Using (7), this implies that

$$\frac{a_i - m_i}{\sigma_i^2} = \lambda, \ \ i = 1, ..., J.$$

Substituting into (11), we can solve for λ. This specifies the optimal a_i. Substituting their values into (10), we find that

$$I^*(\alpha) = \frac{(\alpha + c - \sum_{i=1}^{J} N_i m_i)^2}{2\sum_{i=1}^{J} N_i \sigma_i^2}.$$

The minimization (12) then yields

$$K = 2\frac{c - \sum_{i=1}^{J} N_i m_i}{\sum_{i=1}^{J} N_i \sigma_i^2}.$$

The set of values of $\{N_i\}$ such that the overflow probability $\exp\{-BK\} \le \exp\{-\delta\}$ is then such that $KB \ge \delta$, i.e.,

$$2B\frac{c - \sum_{i=1}^{J} N_i m_i}{\sum_{i=1}^{J} N_i \sigma_i^2} \ge \delta,$$

i.e.,

$$\sum_{i=1}^{J} N_i \{m_i + \frac{\delta \sigma_i^2}{2B}\} \le c. \tag{13}$$

Comparing this expression with (1), we see that each source of type i can be thought of as having an effective bandwidth α_i given by

$$\alpha_i = m_i + \frac{\delta \sigma_i^2}{2B}. \tag{14}$$

It can be verified that this formula is equivalent to that given in section 1.

4. Stationary Gaussian sources

In this section, we apply the method of the previous section to analyze the effective bandwidth of stationary Gaussian sources.

The model is as follows. A source of type i produces X_k^i cells at time $k \ge 1$. The sequence $\{X_k^i, \ k \ge 1\}$ is stationary and Gaussian. Also, different sources are independent. The values $\{X_k^i, \ k \ge 1\}$

can be correlated. It will be assumed that

$$\frac{1}{n} var\{\sum_{k=1}^{n} X_k^i\} \to \gamma_i \quad \text{as} \quad n \to \infty, \quad \text{for} \quad i = 1, ..., J. \tag{15}$$

This condition holds for sources that have an auto-correlation that dies out fast enough.

To analyze this model, we first consider the evolution of the buffer over a fixed time interval $\{1, 2, ..., n\}$. We will then examine what happens for different values of n. Let $x_i = [X_1^i, ..., X_n^i]^T$ where $(.)^T$ denotes the transposition. Thus, x_i is a vector of jointly Gaussian random variables and we denote its mean by m_i and its covariance matrix by K_i. For simplicity we assume that K_i is invertible. It can be shown that the probability that x_i is close to a_i, a vector very far from the mean m_i, is of the order of $\exp\{-I_i(a_i)\}$, with

$$I_i(a_i) := \frac{1}{2}(a_i - m_i)^T K_i^{-1}(a_i - m_i).$$

(See e.g. [1].) The probability that the N_i sources of type i produce a sequence a_i of cells, for all i, is approximately equal to

$$\exp\{-\sum_{i=1}^{J} N_i I_i(a_i)\}.$$

Thus, the probability that the sources produce a total number of cells equal to

$$B + nc \tag{16}$$

is approximately equal to

$$\exp\{-I^*(n)\} \quad \textit{where} \quad I^*(n) := \min \sum_{i=1}^{J} N_i I_i(a_i) \tag{17}$$

where the minimization is over the vectors a_i such that

$$\sum_{i=1}^{J} N_i a_i^T 1 = B + cn \tag{18}$$

where 1 is the column vector with n ones. To solve this minimization problem, we use Lagrange multipliers and we find that

$$\nabla I_i(a_i) = \lambda 1.$$

Using the expression for $I_i(.)$, this gives

$$K_i^{-1}(a_i - m_i) = \lambda 1,$$

i.e.,

$$a_i = m_i + \lambda K_i 1. \tag{19}$$

Substituting in (18) we get

$$\sum_{i=1}^{J} N_i m_i^T 1 + \lambda \sum_{i=1}^{J} N_i 1^T K_i 1 = B + cn, \tag{20}$$

which gives

$$\lambda = \frac{B + cn - \sum_{i=1}^{J} N_i m_i^T 1}{\sum_{i=1}^{J} N_i 1^T K_i 1}. \tag{21}$$

We now substitute the expressions (19) and (21) in the expression to be minimized in (17) to conclude that

$$I^*(n) = \frac{(B + cn - \sum_i N_i m_i^T 1)^2}{\sum_{i=1}^{J} N_i 1^T K_i 1}. \tag{22}$$

Now, by stationarity,

$$m_i^T 1 = \beta_i n. \tag{23}$$

Also, from (15),

$$1^T K_i 1 \simeq n\gamma_i. \tag{24}$$

We show in the appendix that (15) enables us to replace the approximate sign in (24) by an equality. Using (23)-(24) into (22), we find

$$I^*(n) = \frac{(B + cn - n\sum_i N_i \beta_i)^2}{n\sum_{i=1}^J N_i \gamma_i}.$$

We now minimize this expression over n and find

$$K := \min_n I^*(n) = 4B\frac{c - \sum_{i=1}^J N_i \beta_i}{\sum_{i=1}^J N_i \gamma_i}. \tag{25}$$

Consequently, the set $\{N_i,\ i = 1, ..., J\}$ such that the probability of overflow is less than $\exp\{-\delta\}$ is such that

$$\exp\{-K\} \le \exp\{-\delta\}, \textit{ i.e. }, k \ge \delta,$$

which gives

$$\sum_{i=1}^J N_i \alpha_i \le c \quad \textit{where} \quad \alpha_i := \beta_i + \frac{\delta\gamma_i}{4B}. \tag{26}$$

This expression establishes the existence of an effective bandwidth α_i for a source of type i.

5. Appendix

It remains to justify replacing the approximate sign in (24) by an equality. More precisely, we need to justify (25).

Let

$$f(n, B) := \frac{(B + nD)^2}{\gamma_n} \tag{27}$$

where

$$D := c - \sum_{i=1}^J N_i \beta_i \tag{28}$$

and

$$\gamma_n := \sum_{i=1}^J N_i E\{[\sum_{k=1}^n (X_k^i - \beta_i)]^2\}. \tag{29}$$

From (15),

$$\frac{\gamma_n}{n} \to \gamma \text{ as } n \to \infty. \tag{30}$$

We will show that

$$\lim_{B\to\infty} \frac{1}{B} \min_n f(n, B) = \frac{4D}{\gamma}, \tag{31}$$

which will justify (25).

Define ε_n by

$$\gamma_n = n(\gamma + \varepsilon_n),\ n \ge 1.$$

Let us fix $\varepsilon > 0$ arbitrarily small. We will show that for B large enough,

$$\left| \frac{1}{B} \min_n f(n, B) - \frac{4D}{\gamma} \right| \le \frac{4D}{\gamma - \varepsilon} - \frac{4D}{\gamma + \varepsilon}. \tag{32}$$

This will prove (31). To show (32), let us fix n_0 large enough so that

$$|\varepsilon_n| \le \varepsilon \text{ for } n \ge n_0. \tag{33}$$

Let us then choose B_0 large enough so that

$$n_0 \le \frac{(B + D)^2}{4BD(\gamma + \alpha)^2} \text{ for } B \ge B_0 \tag{34}$$

where

$$\gamma + \alpha = \max_n \frac{1}{n} \gamma_n. \tag{35}$$

It follows from (27) and (33) that

$$\frac{(B + nD)^2}{n(\gamma + \varepsilon)} \le f(n, B) \le \frac{(B + nD)^2}{n(\gamma - \varepsilon)}, \text{ for } n \ge n_0. \tag{36}$$

Therefore, by taking the minimum over n,

$$\frac{4BD}{\gamma + \varepsilon} \le \min_{n \ge n_0} f(n, B) \le \min_{n \ge n_0} \frac{(B + nD)^2}{n(\gamma - \varepsilon)} \tag{37}$$

for all B. Now, trivially,

$$f(n, B) \ge \frac{(B + D)^2}{n_0(\gamma + \alpha)}, \text{ for } n \le n_0. \tag{38}$$

But

$$\frac{(B + D)^2}{n_0(\gamma + \alpha)} \ge \frac{4BD}{\gamma + \alpha}, \text{ for } B \ge B_0, \tag{39}$$

by (34). Therefore, (38) and (39) imply

$$\min_{n \le n_0} f(n, B) \ge \frac{4BD}{\gamma + \alpha}, \text{ for } B \ge B_0, \tag{40}$$

Similarly, for $n \le n_0$ and $B \ge B_0$,

$$\frac{(B + nD)^2}{n(\gamma - \varepsilon)} \ge \frac{(B + D)^2}{n_0(\gamma - \varepsilon)} \ge \frac{4BD}{\gamma - \varepsilon}, \tag{41}$$

which implies that

$$\min_{n \ge n_0} \frac{(B + nD)^2}{n(\gamma - \varepsilon)} = \min_n \frac{(B + nD)^2}{n(\gamma - \varepsilon)} = \frac{4BD}{\gamma - \varepsilon}, \text{ for } B \ge B_0. \tag{42}$$

The inequalities (37), (40), and (42) show that, for $B \ge B_0$, the minimum of $f(n, B)$ is achieved for $n \ge n_0$ so that we can conclude that

$$\frac{4BD}{\gamma + \varepsilon} \le \min_n f(n, B) \le \frac{4BD}{\gamma - \varepsilon}, \tag{43}$$

which proves (32) and terminates our derivation.

6. Conclusions

We have shown that an effective bandwidth cannot be defined for general non-stationary sources. We also showed that such an effective bandwidth can be defined for stationary Gaussian sources (modulo a weak technical assumption). It remains to determine whether an effective bandwidth can be defined for general stationary sources.

7. References

[1] J.A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation.* John Wiley, New York, 1990.

[2] M. Cottrell, J-C. Fort, and G. Malgouyres, "Large deviations and rare events in the study of stochastic algorithms," *IEEE Trans. on Aut. Control,* Vol. **AC-28**, no. 9, pp. 907-920, 1983.

[3] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand, and R. Weber, "Admission control and routing of ATM networks based on measured buffer occupancy," EECS, Univ. California, preprint.

[4] R.G. Gibbens and P.J. Hunt, "Effective bandwidths for the multi-type UAS channel,"" Statistical Laboratory, University of Cambridge, preprint.

[5] F.P. Kelly, "Effective bandwidth at multi-class queues," Statistical Laboratory, University of Cambridge, preprint.

[6] S. Parekh and J. Walrand, "A quick simulation method for excessive backlogs in networks of queues," *IEEE Trans. on Aut. Control,* Vol. **AC-34**, no. 1, pp. 54-66, 1989.