## On the Effective Bandwidth of Arbitrary on/off Sources \*

Khaled M. Elsayed Harry G. Perros Department of Computer Science and Center for Communications and Signal Processing North Carolina State University Raleigh, NC 27695-8206

January 24, 1995

#### Abstract

The effective bandwidth approximation is an attractive mechanism for performing call admission control in ATM networks. We present a methodology for evaluating the effective bandwidth of arbitrary on/off sources. An arbitrary on/off source is described by a stochastic process which alternates between on and off periods that have arbitrary probability density functions. We present two approximate methods for evaluating the effective bandwidth of an arbitrary on/off source. The first method is based on moment matching of the on and off periods. The second method is called the bandwidth matching procedure and is considerably more efficient than the moment matching procedure.

<sup>\*</sup>Supported in part by BellSouth, GTE Corporation, and NSF and DARPA under cooperative agreement NCR-8919038 with the Corporation for National Research Initiatives and in part by a gift from BNR INC.

## 1 Introduction

In this paper, we devise a methodology for calculating the effective bandwidth of an arbitrary on/off source. The effective bandwidth is a mechanism for carrying out the call admission control process in the high speed networks environment. The methodology of the effective bandwidth is in general applicable to the case of Markov modulated sources. Arbitrary on/off source are a special case of semi-Markov processes and are in general non-Markovian. Using an elaborate mapping, we find a Markov chain which can be used to characterize an arbitrary on/off source. We then apply the effective bandwidth method to the resulting Markov source.

For a Markov modulated source, the computation of the effective bandwidth involves finding the spectral radius or sometimes the largest positive eigenvalue of a matrix which is a function of the modulating Markov chain, the arrival rates at the states of the Markov chain, the multiplexer buffer size, and the required cell loss probability. This is usually a computationally intensive process. In order to be able to perform call admission decisions using the effective bandwidth scheme in real-time, we propose a heuristic matching procedure in which the original source is mapped to an *equivalent* binary Markov modulated on/off source. The equivalent binary Markov modulated on/off source, for which the effective bandwidth has a simple characterization, can then be used in place of the original source in the call admission control process. As evidenced by the numerical examples reported in the paper, the accuracy of the approximation is valid for practical purposes.

This paper is organized as follows. In section 2 we discuss the concept of effective bandwidth for Markov modulated sources. In section 3, we show how to approximately calculate the effective bandwidth of an arbitrary on/off source. We refer to this approximation as the Markovian-transformation method. We present two less computationally-intensive approximate methods for characterizing the effective bandwidth of an arbitrary on/off source in section 4. In section 5 we compare the evaluation of the effective band-

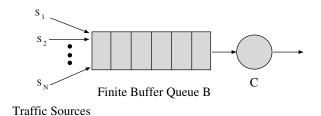


Figure 1: The multiplexer model

width using the Markovian-transformation with the two other proposed approximations. Conclusions are given in section 6.

## 2 Effective Bandwidth of Markov Modulated Sources

The primary role of a network congestion control procedure is to protect the network and the user in order to achieve network performance objectives and optimize the usage of network resources. In ATM-based B-ISDN, congestion control should support a set of ATM quality of service classes sufficient for all foreseeable B-ISDN services. Call admission control (CAC) is one of the primary mechanisms for preventive congestion control in an ATM network. CAC is one particular type of many possible resource allocation mechanisms performed by the network provider. In an ATM network, resource allocation can be identified on three different levels: call, burst, and cell levels [8].

In the context of ATM networks, the CAC process is described as follows. During the call setup phase, users declare and/or negotiate with the network their connection characteristics and their required quality of service. Some of the parameters that may be used to specify a call characteristics are sustainable (average) bit rate, peak bit rate, cell delay variation tolerance, and maximum burst length, as recommended by the ATM Forum [4].

In a loss network, where the probability of lost cells is the primary grade of service,

the effective bandwidth scheme emerged as an attractive scheme for performing call admission control. For a finite buffer multiplexer with buffer size *B* and output link capacity *C* as shown in figure 1, the effective bandwidth of a call is defined as the link capacity needed in order to keep the call's cell loss probability below a specific value  $\epsilon$ , assuming that the given call is the sole user of the multiplexer.

Guérin, Ahmadi and Naghshineh [6], and also Gibbens and Hunt [5], provide an efficient method to evaluate the effective bandwidth of a single call and the aggregate bandwidth of multiplexed calls. The arrival process is modeled as a binary Markov modulated on/off fluid source. The results of Anick et al. [1] were used to calculate the effective bandwidth.

Elwalid and Mitra [2] generalize these results to general Markovian traffic sources for both fluid and point processes models. The effective bandwidth is shown to be the maximal real eigenvalue of a matrix directly obtained from the source characteristics and admission criterion. The approximation is mostly valid in the asymptotic regime as buffer size approaches infinity and the cell loss probability tends to zero.

Kesidis, Walrand, and hang [10] use large deviations theory to find an approximation of the effective bandwidth for a variety of traffic sources. We use their results to handle the case of discrete-time Markov modulated sources. The main result for discrete-time sources can be summarized as follows. Consider a traffic source modeled as an L-state source  $(\mathbf{P}, \vec{\lambda})$  where  $\mathbf{P}$  is the probability transition matrix of the modulating Markov chain and  $\vec{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_{L-1})$  is the vector of peak arrival rates at the various states. Let  $\bar{\lambda}$  be the mean rate and  $\hat{\lambda}$  be the maximum rate. The buffer is served by a channel of capacity C (C = 1 after normalization). Define  $G(B) = Pr[x \ge B]$ , where x is the stationary buffer content. Then the grade of service is  $G(B) \le \epsilon$ . The effective bandwidth e of a source  $(\mathbf{P}, \vec{\lambda})$  was shown to be [10]:

$$e = \log(\Omega\{exp(\delta \boldsymbol{\Lambda})\boldsymbol{P}\})/\delta$$
(1)

where  $\mathbf{\Lambda} = diag(\mathbf{\vec{\lambda}})$  and  $\delta = -\frac{log(\epsilon)}{B}$ , and  $\Omega\{M\}$  is the spectral radius of matrix M. The value of e satisfies the relation:  $\mathbf{\vec{\lambda}} \leq e \leq \mathbf{\hat{\lambda}}$ . When N sources are multiplexed the total effective bandwidth is approximated by  $\sum_{i=1}^{N} e_i$ .

For a 2-state Markovian source described by the triplet (R, r, b), where R is the peak rate in cells/sec, r is the average rate in cells/sec, and b is the mean burst length in cells, the effective bandwidth, e, of the source can be shown to be equal to:

$$e = \log\left(\frac{1}{2} \left[ P_{11} + exp(\delta R)P_{22} + \sqrt{(P_{11} + exp(\delta R)P_{22})^2 + 4exp(\delta R)(1 - P_{11} - P_{22})} \right] \right) / \delta$$
 (2)

where  $P_{11} = ((R - r)b - r)/(R - r)b$  and  $P_{22} = 1 - 1/b$ .

Other related work is Kelly [9] and Whitt [13].

## 3 Effective Bandwidth of an Arbitrary on/off Source

Consider a traffic source which alternates between on and off periods where the lengths of the on and off periods have an arbitrary probability distribution. The distributions of the off and the on periods are given by:

$$f^{off}(i) = Pr[t^{off} = i \text{ slots}] = \alpha_i \quad 1 \le i \le K$$
  
$$f^{on}(i) = Pr[t^{on} = i \text{ slots}] = \beta_i \quad 1 \le i \le L$$
(3)

where  $t^{off}$  ( $t^{on}$ ) is a random variable indicating the length of an off (on) period. Arrivals occur periodically every *T* slots during the on period.

To describe an arbitrary on/off source as a Markov modulated source, we define the

following quantities:

$$a_{i} \stackrel{\triangle}{=} \Pr[t^{off} > i | t^{off} > i - 1] = 1 - \alpha_{i} / (1 - \sum_{1}^{i-1} \alpha_{j}), \quad 1 \le i \le K$$
  
$$b_{i} \stackrel{\triangle}{=} \Pr[t^{on} > i | t^{on} > i - 1] = 1 - \beta_{i} / (1 - \sum_{1}^{i-1} \beta_{j}), \qquad 1 \le i \le L$$

We can then represent the off and on periods as a Markov chain with L and K states respectively. The states are represented by the pair (i, s),  $0 \le s \le 1$ . States with s = 0represent the off period, where  $1 \le i \le K$ , and states with s = 1 represent the on period, where  $1 \le i \le L$ . The variable i in the state descriptor (i, s) indicates that the Markov chain has reached slot i in the current period. When the source is in state (i, s), the only two possible transitions are:

- 1. The source moves to state (i + 1, s) with probability  $a_i$  respectively  $b_i$  if s = 0 respectively s = 1. This means that the source stays for at least one more slot in its current state.
- 2. The source moves to state  $(1, \bar{s})$ , where  $\bar{s} \neq s$  is the other type of period, with probability  $1 a_i$  respectively  $1 b_i$  if s = 0 respectively s = 1. This represents the end of the current period after *i* slots and the start of the other type of period in the following slot.

The discrete-time Markov chain representing the source is then given by:

	0 a <sub>1</sub>		$1-a_1  0  \cdots  0$
	$0 a_{2}$		$1-a_2  0  \cdots  0$
	··. ··.		i i ·. i
		0 <i>a</i> <sub><i>K</i>-1</sub>	$1-a_{K-1}$ $0$ $\cdots$ $0$
P _		0	$1  0  \cdots  0$
1 —	$1-b_1$	0 0	$0 \ b_1$
	$1 - b_{2}$	0 0	$0  b_2$
	:	÷ ·. :	··. ··.
	$1 - b_{L-1}$	0 0	$b_{L-1}$
	1	0 0	0

Now let us define  $O_{i \times j}$  is matrix of *i* rows and *j* columns with entries equal to zero,  $A = \begin{bmatrix} O_{K \times K} & O_{K \times L} \\ O_{L \times K} & RI_{L \times L} \end{bmatrix}, R = 1/T, \text{ and } \delta = -log(\epsilon)/B.$  The effective bandwidth of the source is *approximated* by

$$log(\Omega\{exp(\delta A)P\})/\delta$$

where  $\Omega{M}$  is the spectral radius of matrix *M*.

#### 3.1 The Eigenvalue Problem

As noted above, we need to solve for the spectral radius of the matrix  $exp(\delta A)P$  This matrix is very sparse, since only 2(K + L - 1) of the whole  $(K + L)^2$  matrix elements are non-zero. This suggests that using sparse matrix techniques can be useful here, specially when *K* and/or *L* are large. However, for our purposes, we preferred to use the inverse-

iteration method which proved very successful in solving for the effective bandwidth. In order to use the inverse-iteration method, we need to invert the matrix  $exp(\delta A)P$ . The inversion process requires the storage of the whole matrix.

Other methods of interest, which make use of the sparsity of the matrix, include the Krylov subspace method [12] which does not require the storage of the whole matrix. This method evaluates the characteristic equation of a given matrix. (The roots of the characteristic equation are the eigenvalues of the matrix.) A possible solution of the problem would then be to solve for the root of the characteristic equation in the period [r, R], where r is the average rate of the source. It has been reported in Faddeeva [3] that the accuracy of the method is not satisfactory and poor-conditioning can occur easily. This is due to the fact that the coefficients of the characteristic equation, which are quantities of different order of magnitude, are solved by the same system of linear equations. A good candidate for solving this eigenvalue problem is the simultaneous iteration method for obtaining the set of eigenvalues of largest absolute magnitude as described by Stewart and Jennings [11]. This method is particularly suitable for large sparse matrices.

The power method, a method for finding the eigenvalue with the largest magnitude which is known to be very suitable for sparse-matrix calculations, can not be used here efficiently. This is because of the highly periodic nature of the matrix P.

# 4 Approximations for the Effective Bandwidth of an Arbitrary on/off Source

The disadvantage of the methodology presented in the previous section is clearly the computational and storage complexity. In this section, we propose two approximation methods for characterizing the effective bandwidth of an arbitrary on/off source. The first

method is based on *moment matching* of the statistics of the off and on periods. The second method finds a binary Markov modulated on/off source which, for some carefully chosen values of the buffer size and required cell loss probability, has the same peak rate, average rate, and effective bandwidth as the arbitrary on/off source. The resulting binary Markov modulated on/off source is then used in place of the original source for performing call admission control. We call this method the *bandwidth matching* method.

#### 4.1 The Moment Matching Procedure

From the distributions of the off and on periods  $f^{off}(i)$  and  $f^{on}(i)$ , we evaluate the mean and squared coefficient of variation of each period. We can then use standard techniques for fitting phase-type distributions to given moments. The arrival rate during the on period is set equal to R = 1/T. In this case, an on/off source with phase-type distribution for the on and off periods is used to substitute the original source. The effective bandwidth of the Markov modulated phase-type on/off source is calculated as we have discussed in section 3.

Alternatively, we can use the means of the on and off periods and the peak rate of the source to define a binary Markov modulated on/off source. The on and off periods have an exponential distribution with mean equal to the mean on and off periods of the original arbitrary on/off source. This source is referred to as the *plain* binary Markov modulated on/off source when it is used to substitute for the original source.

#### 4.2 The Bandwidth Matching Procedure

Consider a binary Markov modulated on/off source specified by the triplet (R, r, b). The effective bandwidth of such a source can be obtained using equation 2 for a given value

of the required cell loss probability  $\epsilon$  and the buffer size *B*. The proposed bandwidth matching procedure is based on finding a binary Markov modulated on/off source which has the same effective bandwidth as that of the arbitrary on/off source, for a given  $\epsilon$  and *B*. The peak rate and average rate of the binary source are set equal to those of the original source. We argue that the two sources have approximately the same effective bandwidth over a range of values of  $\epsilon$  and *B*. The binary Markov modulated on/off source is used instead of the complex original source for call admission control decisions at different points of the ATM network. Our results strongly support the validity of this approximation.

The matching procedure is outlined as follows:

- Find the effective bandwidth *e* of the arbitrary on/off source for a chosen value of *B* and *e* using the method developed in section 3. Our experience indicates that choosing *B* = [*b* = mean length of the on period] provides a good approximation (which can be seen from the results in section 5.4).
- Find the mean length of the on period *b*' of the binary Markov modulated on/off source as follows:

$$b' = \frac{1}{1-Q_{22}},$$

$$Q_{22} = \left[\eta^2 - (1-\omega)\eta - exp(\delta R)\omega\right] / \left[(\omega + exp(\delta R))\eta - (\omega + 1)exp(\delta R)\right],$$

$$\omega = \frac{r}{R-r},$$

$$\eta = exp(e\delta), \text{ and}$$

$$\delta = -\log(\epsilon)/B$$

$$(4)$$

The above relation is obtained from equation 2 relating the effective bandwidth of a binary Markov modulated on/off source to the parameters (R, r, b'), the required cell loss probability  $\epsilon$ , and the buffer size *B*.

 Set the peak rate and average rate of the binary Markov modulated on/off source equal to those of the arbitrary on/off source. The accuracy of the bandwidth matching procedure is better than the moment matching procedure (using one or two moments fitting). This will be demonstrated in section 5.

A related methodology was suggested by Gün [7] for approximating a complex source by a simple binary on/off source. In this method, however, it is necessary to obtain the complete queue length distribution through the numerical solution of the multiplexer with the complex source at the input. Explicit expressions for the queue length distribution of the multiplexer with the binary Markov modulated source are known from the work of Anick, Mitra, and Sondhi [1]. Gün used these expression to find the parameters of a binary Markov modulated on/off source that would make the queue length distribution of the original source and that of the binary Markov modulated source equal at a given value of the queue length. This method can only be applied to Markov modulated sources. Our method is more general in this regards as it can handle non-Markovian sources and avoids solving for the entire queue length distribution.

## 5 Results

In this section, we provide some numerical examples with a view to validating the methodologies of sections 3 and 4.

#### 5.1 Example 1

For our first example, we choose an arbitrary on/off source with the distribution of the on and off periods shown in tables 2(a) and 2(b). We obtain the effective bandwidth using four different approximations: the Markovian-transformation procedure of section 3, the bandwidth matching binary on/off source, the phase-type on/off source obtained by matching the first two-moments of the off and on periods, and the plain binary

Length	Probability		
1	0.1		
50	0.05		
100	0.05	Length	Probability
150	0.2	1	0.6
200	0.6	100	0.4
(a) Off Period		 (b) (	On Period

Table 1: The distribution of the off and on periods

on/off source obtained by matching the means of the off and on periods. The bandwidth matching is carried out at a cell loss probability equal to  $10^{-6}$  and a buffer size equal to [b] = [40.6] = 41. As it can be seen from figure 2, the bandwidth matching provides the most accurate approximation over the range of the required cell loss probability and buffer sizes. It is noted here that as the buffer size increases, all the methods converge to an effective bandwidth equal to the average rate.

#### 5.2 Example 2

We now consider an arbitrary on/off source where the distribution of the off period is a uniform distribution in the period [1,130]. The on period has a probability of 0.71 of being of length 1 and a probability of 0.29 of being of length 40.

We obtain the effective bandwidth using four different approximations: the Markoviantransformation procedure of section 3, the bandwidth matching binary on/off source, the phase-type on/off source obtained by matching the first two-moments of the off and on periods, and the plain binary on/off source obtained by matching the means of the off and on periods. The bandwidth matching is carried out at a buffer size equal to [b] = [12.31] = 13 and a cell loss probability equal to  $10^{-6}$ . The results are plotted in

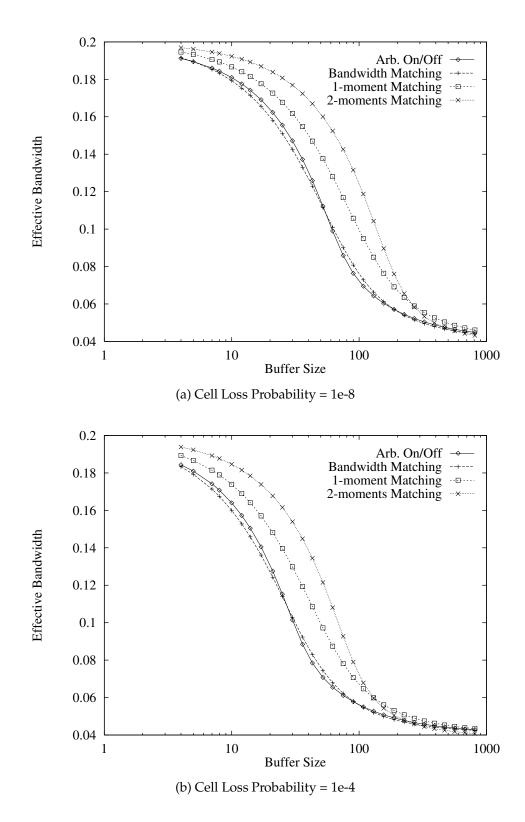


Figure 2: Comparison of various effective bandwidth calculation schemes

figure 3. Again, we note here that the bandwidth matching procedure provides the best overall accuracy. As observed in the other examples, as the buffer size grows, the effective bandwidth estimate of all the methods tends to the average rate of the source.

#### 5.3 Example 3

We now use an on/off source with a phase-type distribution for the off and on periods. The mean and squared coefficient of variation of the off and on periods are given by the pairs (150.0,5.0) and (100.0,10.0) respectively. The peak rate during the on period is equal to 0.025. The mean and squared coefficient of variation are used to fit a hyper-exponential distribution of the off and on periods. In figure 4, we plot the effective bandwidth as obtained by the effective bandwidth method for the phase-type on/off source, the bandwidth matching binary on/off source, and the plain binary on/off source in figure 4. The bandwidth matching was done at a buffer size equal to  $\lceil b \rceil = 100$  and a cell loss probability  $\epsilon = 10^{-6}$ .

In this example too, it can be seen that the bandwidth matching procedure provides an accurate estimate of the effective bandwidth. This is not the case with the plain binary on/off source whose estimate of the effective bandwidth is far lower than its actual value for a wide range of buffer sizes. If we use the plain binary source in the call admission control decision process, the required cell loss probability may not be achievable since we are admitting a larger number of sources than what can be handled.

Similarly, we tested the case when the on and off periods have a coefficient of variation which is smaller than one. The bandwidth matching procedure provided good accuracy whereas the plain binary on/off source provided an upper bound of the effective bandwidth.

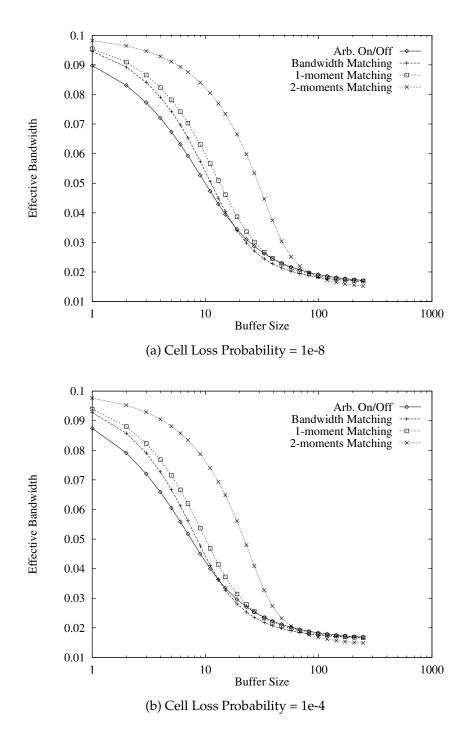


Figure 3: Comparison of various effective bandwidth calculation schemes – off period: uniform(1,130) – on period: P[1] = 0.71, P[40] = 0.29

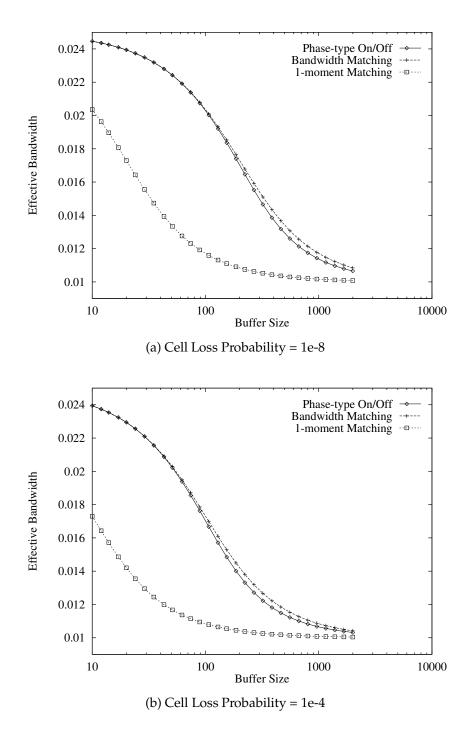


Figure 4: Comparison of various effective bandwidth calculation schemes – the source has a hyper-exponential on and off periods

#### 5.4 Selection of Buffer Size for the Bandwidth Matching Procedure

The selection of the buffer size, at which the effective bandwidth of the arbitrary on/off source is evaluated, affects the accuracy of the bandwidth matching procedure. We use the on/off sources of examples 1, 2, and 3 to demonstrate this relationship. For the sources of examples 1 and 2, we obtain the effective bandwidth by the Markovian transformation method. For the source of example 3, we obtain its effective bandwidth by directly applying the effective bandwidth method since the source is already Markovian. The calculations of the effective bandwidth is compared with three different binary Markovian on/off sources obtained by bandwidth-matching. These three sources are obtained by applying the bandwidth-matching procedure at three different values of the buffer size. The three different values of the buffer size are designated as low, medium, and large. The low buffer value is equal to maximum(1, [ mean length of the on period/10 ]), the medium value is equal to [20 × mean length of the on period].

We calculate the root mean square relative error and average relative error between the three approximating sources and the arbitrary on/off source. The root mean square relative error and average relative error are defined as follows. Let N be the total number of readings, x be the actual value of a reading, and  $\hat{x}$  be an estimated value for x. The root mean square relative error is equal to

$$\sqrt{\sum_{i=1}^{N} \left(rac{\hat{x}_i - x_i}{x_i} imes 100
ight)^2}/N,$$

and the average relative error is equal to

$$\sum_{i=1}^{N} \left( \frac{\hat{x}_i - x_i}{x_i} \times 100 \right) / N.$$

The calculations are performed at logarithmically equidistant values of buffer sizes in the range [maximum(1, [mean length of the on period/10]),  $[20 \times \text{mean length of the on period}]]$ 

Table 2: Effect of performing the bandwidth matching procedure at different values of the buffer size

BWM Buffer Size	RMS Error	Average Error
4	0.30%	0.64%
41	0.24%	-0.60%
812	0.31%	-2.57%

BWM Buffer Size	RMS Error	Average Error
1	1.36%	-15.61%
13	0.46%	-0.21%
248	0.52%	1.76%

(a) Example 1

BWM Buffer Size	RMS Error	Average Error
10	0.19%	2.14%
100	0.14%	1.44%
2000	0.11%	0.99%

(b) Example 2

(c)	Example	3
$(\mathbf{U})$	LAIMPIC	$\mathcal{O}$

and 7 equi-distant points for the required cell loss probability in the range  $[10^{-9}, 10^{-3}]$ . The results are shown in tables 3(a), 3(b), and 3(c) for the sources of examples 1, 2, and 3 respectively. In the tables the column BWM buffer size indicates the buffer size at which the bandwidth matching procedure is performed. Matching at the buffer size value equal to [mean length of the on period] provides the best overall performance of the three alternatives.

## 6 Conclusions

We have introduced the Markovian-transformation methodology for characterizing the effective bandwidth of an arbitrary on/off source. We proposed the bandwidth matching procedure as an effective mechanism for performing the effective bandwidth calculation in real-time. The bandwidth matching procedure provides good accuracy for the set of sources we tested. An interesting extension would be to investigate if it is feasible to provide upper and/or lower bounds for the cell loss probability in a statistical multiplexer with arbitrary on/off sources input using the bandwidth-matching procedure.

## References

- D. Anick, D. Mitra, and M. M. Sondhi. Stochastic Theory of a Data-Handling System with Multiple Sources. *Bell Sys. Tech. J.*, 61:1871–1894, 1982.
- [2] A. I. Elwalid and D. Mitra. Effective bandwidth of general markovian traffic sources and admission control of high speed networks. *IEEE Transactions on Networking*, 1:329–343, 1993.
- [3] D. K. Faddeev and V. N. Faddeeva. *Computational Methods of Linear Algebra*. W. H. Freeman and Company, 1963.
- [4] ATM Forum. ATM User-Network Interface Specification, Version 2.2, June 1991.
- [5] R. J. Gibbens and P. J. Hunt. Effective Bandwidths for the Multi-type UAS Channel. *Queueing Systems*, 9:17–28, 1991.
- [6] R. Guérin, H. Ahmadi, and M. Naghshineh. Equivalent Capacity and Its Application to Bandwidth Allocation in High-Speed Networks. *IEEE Journal on Selected Areas in Communications*, 9:968–981, 1991.
- [7] L. Gün. An Approximation Method for Capturing Complex Traffic Behavior in High Speed Network. *Performance Evaluation*, 19:5–23, 1994.
- [8] J. Hui. Resource Allocation for Broadband Networks. *IEEE Journal on Selected Areas in Communications*, 6:1598–1608, 1988.
- [9] F. P. Kelly. Effective Bandwidth at Multi-Class Queues. *Queueing Systems*, 9:5–16, 1991.

- [10] G. Kesidis, J. Walrand, and C.-S. Chang. Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE Transactions on Networking*, 1(4):424–428, August 1993.
- [11] W. J. Stewart and A. Jennings. A Simulatenous Iteration Algorithm for Real Matrices. *ACM Transactions on Mathematical Software*, 7:184–198, 1981.
- [12] F. Stummel and K. Haines. *Introduction to Numerical Analysis*. Scottish Academic Press, Edinburgh, 1980.
- [13] W. Whitt. Tail Probabilities with Statistical Multiplexing and Effective Bandwidth for Multi-Class Queues. *Telecommunications Systems*, 2:71–107, 1993.