UPC PARAMETER ESTIMATION USING VIRTUAL BUFFER MEASUREMENT WITH APPLICATION TO AAL2 TRAFFIC

David W. Petr, Gopi K Vaddi, Yong-Qing Lu*

Information & Telecommunication Technology Center
Department of Electrical Engineering and Computer Science
The University of Kansas
Lawrence, KS 66045-2228

*Sprint Corporation, Overland Park, KS 66212

Abstract

This paper describes an efficient, measurement-based technique for estimating all minimal sets of usage parameter control (UPC) parameters for variable bit rate (VBR) ATM traffic streams. This *virtual buffer measurement* technique is applicable to simulation or live traffic measurement environments and is considerably more efficient than direct evaluation using the generic cell rate algorithm (GCRA). We derive analytically the relationships between virtual buffer measurements and minimal sets of UPC parameters as determined by GCRA, then illustrate the technique by estimating and verifying minimal UPC parameters for ATM Adaptation Layer, Type 2 (AAL2) traffic streams representing voice traffic.

1: Introduction

Traffic policing, or Usage Parameter Control (UPC) as it is known in ATM circles, is a critical component of the overall traffic management and congestion control strategy for ATM networks [1]. Associated with each Virtual Channel Connection (VCC) in an ATM network are certain traffic parameters upon which the network provider, through Connection Admission Control (CAC) procedures, can base resource allocation decisions aimed at maintaining the desired Quality of Service (QoS) for the VCC. In order to protect the QoS of each VCC from traffic misbehavior by other VCCs, each VCC's traffic stream is policed to ensure that the traffic is adhering to its stated traffic parameters. Without such policing, a VCC could consume significantly more than its allocation of network resources, thereby jeopardizing the QoS of connections that are adhering to their traffic specifications.

For Constant Bit Rate (CBR) connections, the UPC (traffic) parameters are Peak Cell Rate (PCR) and Cell Delay Variation Tolerance (CDVT). For Variable Bit Rate (VBR) connections, Sustainable Cell Rate (SCR) and Burst Tolerance (BT) are specified in addition to PCR and CDVT.

In order for the CAC and UPC mechanisms to be effective, it is fundamentally important that the specified UPC parameters accurately reflect the actual traffic stream. Parameter sets that are "too large" will tend to result in overallocation of resources by the network CAC function, resulting in unnecessarily high costs to the network user (assuming that price is based at least in part on resource allocation [2]). Parameter sets that are "too small" will tend to cause policing violations, resulting in traffic either being discarded immediately at the policer or at least being "marked" with a low discarding priority and hence targeted as expendable if the network becomes congested [1].

A fundamental question, then, is: *How can we determine or at least estimate the UPC parameters for a VCC?* The answer to this question is complicated by a number of factors. One is the uncertainty associated with the behavior of specific traffic sources. Another is the difficulty of constructing mathematical models of traffic behavior. A third is the well-known fact (see for example [3] and [4]) that there are an infinite number of UPC parameter sets that can be used to describe a given traffic stream.

In this paper, we assume that a statistically representative sample of a particular VBR traffic stream is available. Such a sample could be obtained from observation of the traffic stream at an earlier time or from observation of a different traffic stream that is known or believed to be similar in its behavior, such as a stream originating from the same type of application in the same network environment. The existence of such a representative sample nullifies the uncertainty factor, alleviates the need for construction of a traffic model, and allows for the use of measurement and/or simulation techniques. We emphasize that, although we focus here on simulation, the technique is equally valid in a live traffic measurement context. However, even with such a representative traffic sample, determining UPC parameters via direct simulation would be an arduous task, as discussed in section 2. In that section, we also describe and analyze an efficient measurement technique, *virtual buffer measurement*, that can be used to estimate all "minimal sets" (defined in section 2) of UPC parameters for a particular VBR traffic stream. Selection of a particular set of parameters can then be made based on factors such as QoS requirements and bandwidth usage, as discussed in [4] and later in the current paper. In section 3, we apply the virtual buffer measurement technique to an ATM Adaptation Layer, Type 2 (AAL2) traffic stream using simulation. Section 4 presents the virtual buffer measurement results and demonstrates the

The authors can be contacted via e-mail at dwp@ittc.ukans.edu. This research is sponsored by Sprint Corporation.

validity of the method by direct simulation with a UPC policer model. Conclusions are summarized in section 5.

2: Obtaining UPC Parameters via Virtual Buffer Measurement

2.1: Overview

Usage Parameter Control (UPC) is accomplished by using one or more traffic policers, each implemented as a Generic Cell Rate Algorithm (GCRA) [1]. For VBR traffic, a dual policer configuration is used (Figure 1), in which the first policer monitors PCR and CDVT and the second monitors SCR and BT. In the figure, T_0 = 1/PCR, T_s = 1/SCR, and the restriction of CDVT = T_0 is in accordance with standards recommendations [1]. Our goal is to find all "minimal sets"

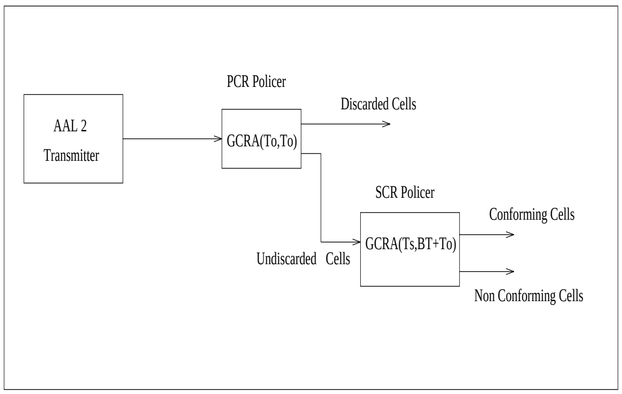

Figure 1. Dual Policer Configuration

of UPC parameters PCR, CDVT (=1/PCR), SCR, and BT for a particular VBR traffic stream, where a minimal set is defined as follows.

- 1) The traffic stream will be completely conformant (all cells conforming) when policed by a GCRA(T_0, T_0) policer (see Figure 1) followed by a GCRA($T_s, \tau + T_0$) policer, where PCR = $1/T_0$, CDVT = T_0, SCR = $1/T_s$, and BT = τ.
- 2) Any significant reduction in any one of the values PCR, SCR, or BT will result in some nonconformance when using the above dual policer configuration on the given traffic stream.

With the restriction of CDVT=1/PCR, there is a unique minimal value of PCR for any given traffic stream. However, for $every$ value of SCR between the mean rate and the peak rate of the source, there will be an associated minimal value of BT.

Even with the availability of a representative traffic stream, attempting to find even one minimal set of UPC parameters from direct simulation of this dual policer configuration would be most inefficient. This is because direct simulation would require a set of search procedures, each varying one or more parameter values until the boundary between conformance and non-conformance had been estimated to sufficient accuracy. Such a procedure is inefficient not only because the number of simulations required is relatively large, but also because the results of one simulation must be analyzed in order to set the parameter values for the next simulation. A number of such "coupled" simulations would be required to find an estimate for PCR, and then even for a given value of T_s = 1/SCR, many more such "coupled" simulations would be required to estimate the associated minimal value of BT. These difficulties are intensified if live traffic observation and measurements are attempted.

In place of this tedious direct GCRA simulation approach, we advocate here a *virtual buffer measurement* approach that is much more efficient and yet (as we will show) produces essentially equivalent UPC parameter values. The virtual buffer (VB) concept was introduced in [3], in which its equivalence with GCRA policing was assumed, but not rigorously demonstrated. The VB concept is illustrated in Figure 2.

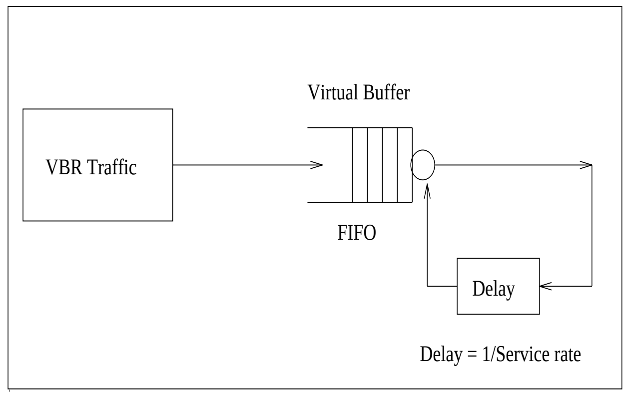

Figure 2. Virtual Buffer Model

The traffic flowing out of a source is fed into a virtual buffer (FIFO). The virtual buffer (VB) is served such that cells leave it at a fixed rate. The service rate (SR) of the VB is varied over a range of values; note that a number of parallel virtual buffers with different service rates can be implemented or simulated in parallel. For each value of service rate, the maximum buffer fill (MBF) at any instant is measured (MBF excludes the cell in service). We will show that the resulting set of (SR,MBF) pairs can be used to accurately estimate both the minimal value of PCR and a set of minimal (SCR,BT) pairs, one for each service rate simulated. Thus we need perform only one simulation for each minimal (SCR,BT) pair obtained, as opposed to several coupled simulations with the direct GCRA simulation approach. Further, there is no coupling at all between any of the simulations in the virtual buffer simulation approach.

A summary of the translations between (SR,MBF) pairs and PCR, SCR, and BT values is as follows.

1) PCR and CDVT: The minimum service rate of the virtual buffer that gives an MBF of 1 is the minimal PCR for the given traffic stream, subject to CDVT=1/PCR.

2) SCR and BT: The service rate SR corresponds to the SCR = $1/T_s$, and then BT can be found from MBF, SCR and PCR = $1/T_0$ by using the relation:

$$BT = MBF * T_s - T_0 \quad (1)$$

Once BT is found, it can be converted to maximum burst size (MBS) according to the well-known relationship [1]:

$$MBS = \lfloor BT/(T_s - T_0) \rfloor + 1 \quad (2)$$

where ⌊ ⌋ represents the integer part.

We proceed to analytically derive these relationships.

2.2: A General GCRA-VB Relationship

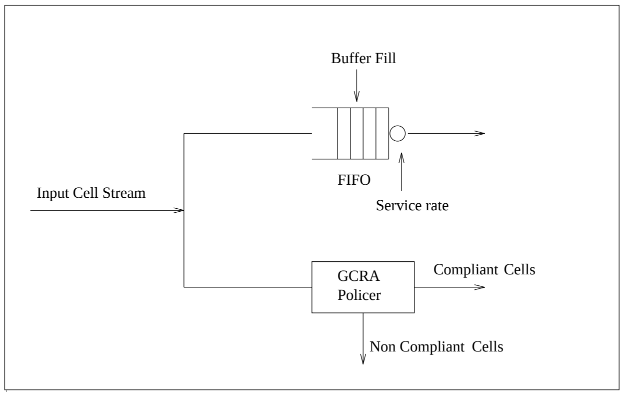

Figure 3. Comparison of the Two Systems

We first establish a general relationship between a Virtual Buffer (VB) and a GCRA policer, both acting on the same ATM traffic stream, as shown in Figure 3. Refer also to Figure 4 for details of the GCRA algorithm.

Proposition 1: A traffic stream that results in an MBF of m when processed by a Virtual Buffer served at a rate of 1/I will be GCRA(I, L) conforming if $L = mI$ and will be nonconforming if $L < (m-1)I$.

We begin the proof by noting that the MBF will result from one or more specific busy periods of the VB. We analyze one such busy period, letting cell n of that busy period be the one that results in the given MBF, which we designate m. The situation is shown in Figure 5, where the T_i values are interarrival times of cells in the traffic stream. Without loss of generality, we let t=0 correspond to the arrival time of the first cell in this busy period.

We first note that for cell n to result in an MBF of m, there must have been exactly $n-m-1$ service completions between time t=0 and the arrival of cell n at time $T_1 + T_2 + T_3 + ... + T_{(n-1)}$. This yields:

$$(n-m-1)I \leq T_1 + T_2 + ... + T_{(n-1)} \leq (n-m)I \quad (3)$$

Furthermore, since the buffer fill does not exceed m, the following must hold for every arrival $k < n$:

$$(k-m-1)I \leq T_1 + T_2 + ... + T_{(k-1)} \quad (4)$$

We now follow the GCRA(I, L) algorithm for each arrival and let X'_k be the GCRA variable X' after the arrival of cell k. Assuming that all cells through $k-1$ are conforming and that X' never reaches zero (which cannot happen during a virtual buffer busy period), it is easy to show that $X'_k = (k-1)I - (T_1 + T_2 + ... + T_{k-1})$ and so cell k will be conforming if $X'_k = (k-1)I - (T_1 + T_2 + ... + T_{k-1}) \leq L$ or $(k-1)I - L \leq T_1 + T_2 + ... + T_{k-1} \geq (k-1)I - L$. Comparing this last with inequality (4) above, we see by induction that conformance of every cell in this (and every) busy period will be assured if $L = mI$.

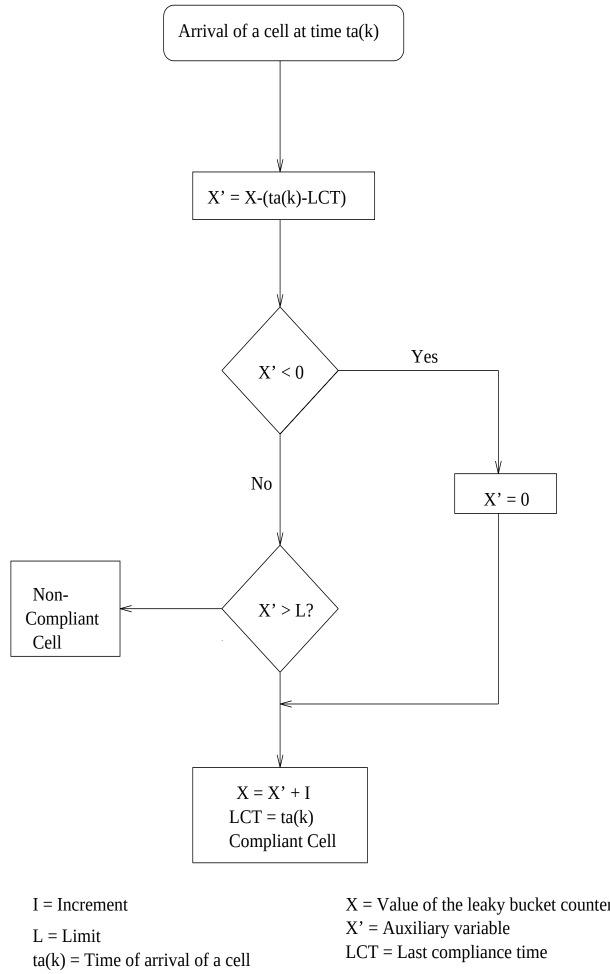

Figure 4. The GCRA Leaky Bucket Policer

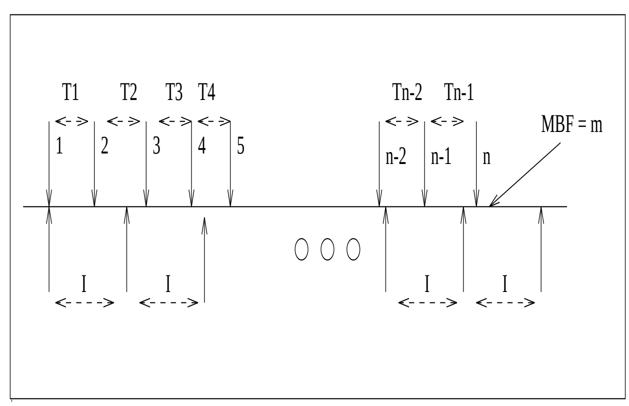

Figure 5. Proposition 1

Now consider cell n, which causes the maximum buffer fill of m. The arrival of this cell will result in $X'_n = (n-1)I - (T_1 + T_2 + ... + T_{n-1})$. Multiplying inequality (3) above by -1 and then adding $(n-1)I$, we conclude that $(m-1)I \leq X'_n \leq mI$. Since GCRA conformance requires $X' \leq L$, we conclude that a value of $L < (m-1)I$ would result in non-conformance for cell n and hence for the entire traffic stream. This completes the proof of Proposition 1.

2.3: Application to UPC Parameter Estimation

We first apply Proposition 1 to the problem of determining the minimum PCR for a given traffic stream from virtual buffer observations. As noted previously, PCR will be policed with a GCRA(T_0, T_0) policer, where T_0 = 1/PCR. Using Proposition 1 with $I = L = T_0$, we see that the only possible MBF values for GCRA(T_0, T_0) conformance are 0 and 1. Since we seek the minimum value of PCR, we conclude immediately that this will be the minimum VB service rate that will result in an MBF of 1 in the virtual buffer. This can be determined (at least approximately) from the MBF vs. service rate data derived from VB simulation.

We next consider the problem of determining (SCR,BT) pairs from VB observations. We note that the SCR policer in Figure 1 has GCRA parameters $I = T_s$ and $L = BT + T_0$, where T_s = 1/SCR. Direct application of Proposition 1 yields the following result. Every given VB service rate is a valid SCR (within the reasonable bounds of mean_rate_of_traffic_stream < SCR < PCR). For a given service rate (SCR=1/T_s) with given MBF, Proposition 1 states that the following value of BT will result in the traffic stream being declared conforming: BT = MBF$*T_s - T_0$. (Note that we have already been able to determine T_0 from VB measurements.) Proposition 1 also states that a value of BT less than (MBF-1)$*T_s - T_0$ will result in non-conformance. Thus, BT = MBF$*T_s - T_0$ may not be the absolute minimum value of BT for the given SCR, but it is within T_s of being the minimum value, which is the best that can be done from VB measurements.

Having obtained a value for BT, the corresponding value of MBS can be obtained from equation (2).

3: Application to AAL2 Traffic

In this section and the next, we use the procedure and results of Section 2 to obtain traffic descriptors (UPC parameters) for AAL2 multiplexed voice traffic streams, then verify their accuracy. All modeling and simulation was done with the BONeS Designer simulation package [9].

The virtual buffer simulation setup is as shown in Figure 6. A previously designed AAL2 transmitter [5] is used as a source for generating AAL2 traffic. The AAL2 output is fed to a number of parallel virtual buffers, each served at a different rate. The output rate of the transmitter is given a practically infinite value for this traffic stream (10 Mbps) so that there is no traffic shaping done at the transmitter.

3.1: Simulation Models

3.1.1: Sources

Each individual voice source is modeled as an On-Off source. The On and Off times are assumed to be exponentially distributed, and the sources have a constant rate when they are ON. This model for the On-Off sources is derived from [6]; it has been verified by analyzing additional recordings of telephone conversations [7]. The values of mean On and Off times estimated from these recent recordings differ somewhat from the values in [6].

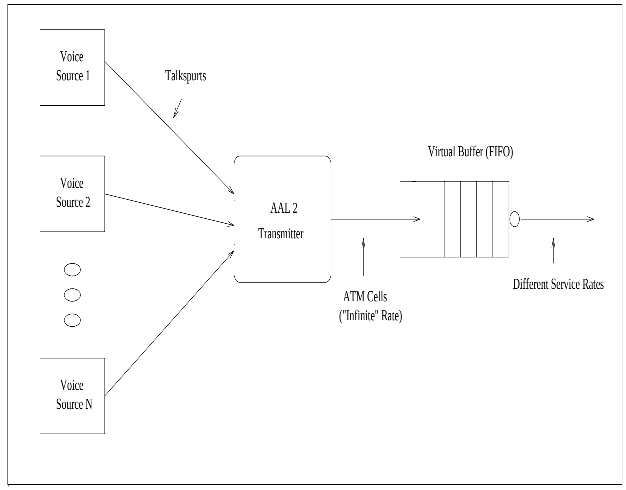

Figure 6. Simulation Model

3.1.2: AAL2 Transmitter

A previously designed AAL2 transmitter is used for multiplexing the voice traffic into a single AAL2 ATM traffic stream. The transmitter has been designed on the basis of the ITU-T draft specification I.362.3. It has been used in previous studies of AAL2 performance characterization [5] and finding the maximum number of users subject to a 95^{th} percentile delay constraint [8].

3.2: Parameters Used in Simulation

3.2.1: Fixed Simulation Parameters

- Mean ON time (1.230s)
- Mean OFF time (1.373s)
- CPS packet size (20 bytes).
- Voice coding rate (32 kbps).
- CU Timer (5.1 ms).

The combination of CPS packet size and voice coding rate results in a packetization time of 5 ms. The CU timer value (time the transmitter waits before sending a cell that is not full) is selected as 5.1 ms to make sure that the cells are densely packed (two packets are packed in a cell even for a single active user). The maximum time delay for a packet is 10.1 ms (5 ms of packetization delay and 5.1 ms CU Timer). The mean On and mean Off times are taken from the recent analysis of speech files [7]. The number of CPS packets simulated for each parameter set is 60,000.

3.2.2: Variable Simulation Parameters

The simulations are done for 1 through 12 users in steps of 1 and from 12 through 48 users in steps of 6. The service rate is varied from a value slightly larger than effective mean rate given by,

$$eff.\ \ mean\ \ rate = n(\frac{1.230}{1.230+1.373})(32)(\frac{23}{20})(\frac{53}{47})kb/s \quad (5)$$

to a value more than twice effective peak rate, where

$$eff.\ \ peak\ \ rate = n(32)(\frac{23}{20})(\frac{53}{47})kb/s \quad (6)$$

Twenty rates in this range are taken to obtain the curves.

4: Results and Discussion

4.1: UPC Parameter Results

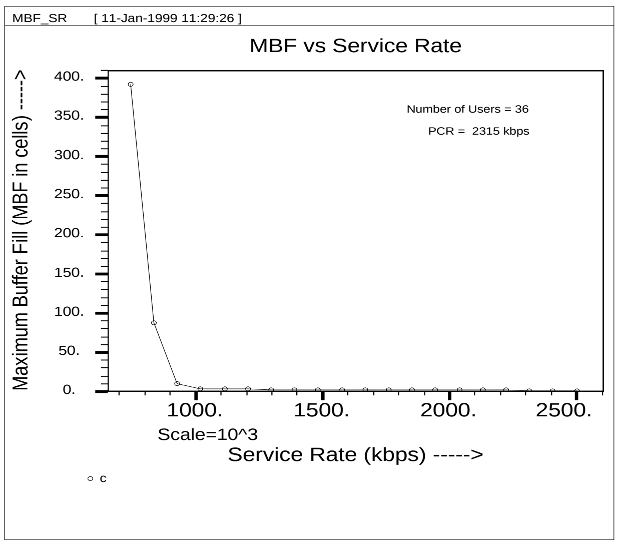

Figure 7. MBF vs. Service Rate for 36 users

The Virtual Buffer simulations yield a curve between MBF and the service rate (see Figure 7). From this curve we obtain the following.

- PCR and CDVT : As established in Section 2, the minimum service rate that corresponds to MBF=1 is taken as the minimum PCR when CDVT equals $1/PCR = T_0$. This choice ensures that the traffic is GCRA(T_0, T_0) conforming. Figure 8 shows two estimates of PCR vs. the number of users. The value "calculated PCR" corresponds to the minimum virtual buffer service rate for MBF of 1 cell. Twice the effective_peak_cell_rate (equation 6, expressed in cells/sec) is shown as "estimated PCR." We see that it may be possible to use twice the effective_peak_cell_rate to estimate the required PCR without resorting to simulation.
- Curves BT vs. SCR and MBS vs. SCR (see Figures 9 and 10): Each value of virtual buffer service rate is taken as SCR and the corresponding maximum buffer fill (MBF) is converted to the near-minimum burst tolerance (BT) using equation (1), resulting in Figure 9. By interpolation in Figure 9, we can find an infinite number of near-minimal (SCR,BT) pairs that describe the given traffic stream. We can also obtain Figure 10 which shows the curve of near-minimal MBS vs. SCR using equation (2).

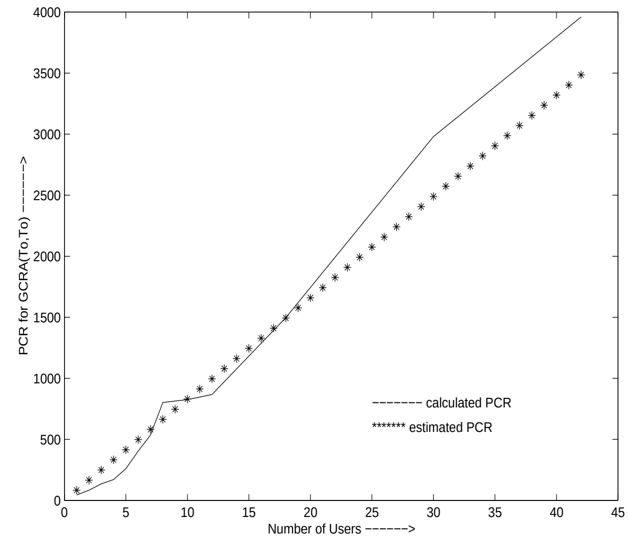

Figure 8. PCR Estimates vs. Number of Users

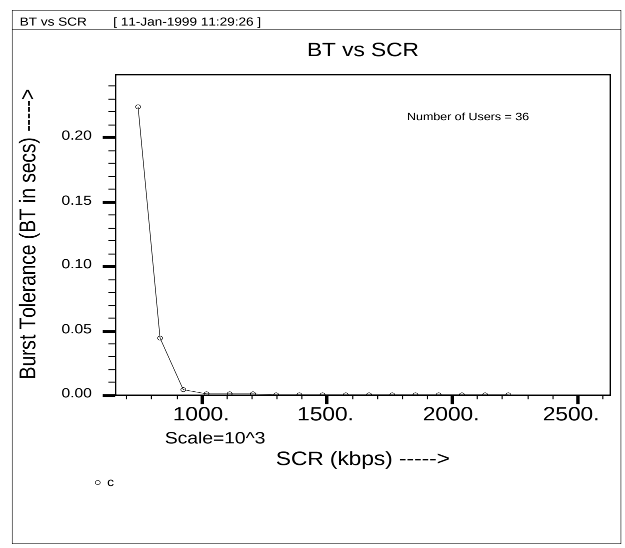

Figure 9. BT vs. SCR for 36 users

Any point on the curve of Figures 9 or 10 will result in GCRA conformance. The selection of a particular point on the SCR, MBS (or BT) curve can be done by different methods.

Method 1: Choose the (SCR,MBS) or (SCR,BT) pair so as to minimize the effective bandwidth requirement. The effective bandwidth might be minimized at larger values of MBS and correspondingly smaller values of SCR. This may not always be an ideal choice. The trade-offs involved in selectiing a bandwidth efficient (SCR,BT) pair are discussed in [4].

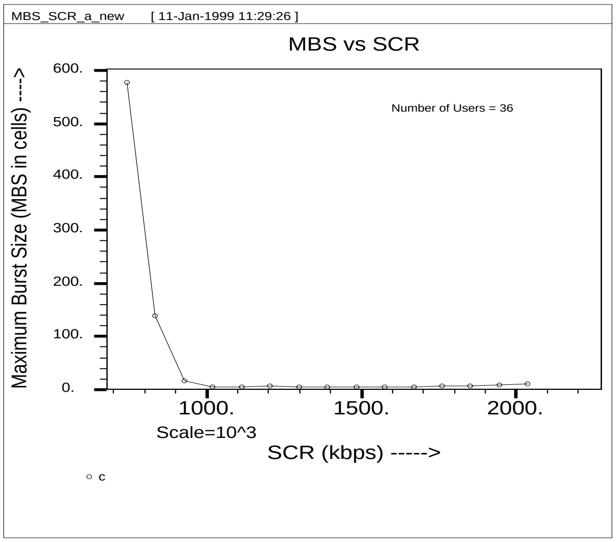

Figure 10. MBS vs. SCR Estimates for 36 users

Number of Users	SCR (kb/s) at MBS = 50 cells	SCR (kb/s) per User
1	40.5	40.50
2	78.5	39.25
3	113	37.67
4	150.5	37.63
5	191	38.20
6	215	35.83
12	358	29.80
18	535	29.70
24	699	29.13
30	832	27.70
36	900	25.00
42	1060	25.20
48	1158	24.13

Table 1. SCR Values for MBS of 50 Cells

Method 2: Limit MBS to a reasonable value and select the corresponding SCR value. The SCR values for MBS of 50 cells are given in Table 1. The incremental SCR for each additional user is also given. Table 1 shows that the incremental SCR required to support each additional user tends to a constant value as the number of users increases. It is clear that this incremental SCR must be lower-bounded by the effective mean rate per source of 19.6 kbps (from equation (5) with n = 1).

4.2: Verification Using the Dual Policer Configuration

This section verifies that the values of PCR, CDVT, SCR and BT found above form a near-minimal UPC parameter set, using a simulated dual policer configuration shown in Figure 1. The conformance test has been done in 3 stages after selecting a point from the (SCR,BT) curve for a fixed PCR (see Figure 9). The following point has been selected for verification:

PCR (kbps)	Number of violating cells with PCR policer	Number of violating cells with SCR policer
2315.0	0	0
2222.5	1	0
2130.0	1	0
2037.5	1	0
1945.0	21	0
1852.5	75	0

Table 2. PCR violation with SCR = 835 kb/s and BT = 0.044502 s

SCR (kbps)	Number of violating cells with PCR policer	Number of violating cells with SCR policer
835	0	0
785	0	74
735	0	332
685	0	1056
635	0	2105

Table 3. SCR violation with PCR = 2315 kb/s and BT = 0.044502 s

PCR: 2315 kb/s, SCR: 835 kb/s, BT: 0.044502 s

1) Conformance test with variation in PCR: The values of SCR and BT are kept constant and the value of PCR is varied. PCR is decreased from its minimal value of 2315 kb/s in steps of the servicing rates used in the virtual buffer simulation. Table 2 shows that even a slight decrease in PCR gives violations in the PCR policer. As expected, SCR policer conformance is not strongly affected by PCR variations.

2) Conformance test with variation in SCR: The values of PCR and BT are kept constant and the value of SCR is varied. SCR is decreased from its minimal value of 835 kb/s in steps of 50 kbps (smaller than the virtual buffer service rate increment). Table 3 shows that even a slight decrease in SCR gives violations in the SCR policer.

3) Conformance test with variation in BT: The values of SCR and PCR are kept constant and the value of BT is varied. BT is decreased from its minimal value of 0.044502 s in steps of T_s first and by larger values later. Table 4 shows that even slight decrease in BT gives violations with the SCR policer.

These tests verify that the traffic descriptors (PCR = 2315 kb/s, SCR = 835 kb/s, BT = 0.044502 s) found are a near-minimal set for the given traffic stream; that is, reduction in any one value results in GCRA violation (either PCR or SCR).

5: Conclusions

In this paper, we have developed an efficient virtual buffer measurement method for finding the near-minimal sets of UPC parameters of a given ATM traffic stream. This method has been analytically derived and verified with simulation. Near-minimal sets of UPC parameters were found using this method for AAL2 multiplexed voice traffic. The vir-

BT seconds	Number of violating cells with PCR policer	Number of violating cells with SCR policer
0.044502	0	0
0.044085	0	1
0.043485	0	2
0.042885	0	3
0.030	0	20
0.025	0	31

Table 4. BT violation with SCR = 835 kb/s and PCR = 2315 kb/s

tual buffer method can be used for determining UPC parameters for any ATM traffic stream for which a representative sample can be obtained. Virtual buffer measurement is amenable both to simulation and live traffic measurement environments.

References

[1] "Traffic Management Specification," The ATM Forum Technical Committee, Version 4.0, af-tm-0056.000, April 1996.

[2] L. A. DaSilva, D. W. Petr, N. Akar, "Equilibrium Pricing in Multiservice Priority-Based Networks," Proceedings of the IEEE Globecom Conference, pp. S38_6.1 - S38_6.5, November 1997.

[3] Kunyan Liu, David W. Petr, Cameron Braun, "A Measurement-Based CAC Strategy for ATM Networks," ICC 1997, pp 1714-1718.

[4] Catherine Rosenberg, Josee Mignault, Fabrice Guillemin, "On Characterizing an ATM Source via the Sustainable Cell Rate Traffic Descriptor," Infocom 1995, pp 1129-1136.

[5] Prema Sampath, Raghushankar R. Vatte, D. W. Petr, "AAL2 Transmitter Simulation Study: Revised," Technical Report, ITTC-FY98-TR-13110-01, March 1998.

[6] Paul T. Brady, "A model for generating ON-OFF speech patterns in two-way conversations," Bell System Technical Journal, Vol. 48, pp. 2445-2472, September 1969.

[7] Gopi K. Vaddi, D. W. Petr, "Voice Activity Statistics," ITTC Technical Report, September 1999.

[8] D. W. Petr, Prema Sampath, Raghu R.Vatte, Yong-Qing Lu, "Performance Comparison Between AAL1, AAL2 and AAL5," ICC'99, June 1999.

[9] "BONeS Designer: Modeling Reference Guide," Comdisco Systems Inc., April 1992.