Trading off network utilisation and delays by performing shaping on VC ATM connections carrying LAN traffic.

P. Castelli L. Guida M. Molina
CSELT (Centro Studi E Laboratori Telecomunicazioni)
V. Reiss Romoli, 274 – 10148 Torino, Italy
Phone: +39.11.228.8784 – Fax: +39.11.228.8862
e-mail:{paolo.castelli,luisa.guida,maurizio.molina}@cselt.it

Abstract

This paper considers a scenario where the traffic of several LANs is transported on Deterministic Bit Rate (DBR) or Statistical Bit Rate (SBR) ATM Virtual Channel Connections (VCCs), that are then multiplexed into a DBR Virtual Path Connection (VPC) with fixed, dedicated bandwidth. It is investigated whether or not it is suitable to shape VCCs according to a DBR or SBR traffic contract before multiplexing. Results show that DBR shaping is rather useless, as with respect to the unshaped case no significant utilisation gain can be achieved without introducing high delays in the shapers' buffers, and that SBR shaping behaves no better, due to the impossibility of finding a typical burst duration and mapping it on SBR traffic descriptors.

Keywords

Shaping, Self Similarity, Long Range Dependence, Multiplexing, IP over ATM

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35388-3_42

H. R. van As (ed.), *High Performance Networking*

1 INTRODUCTION

When talking about ATM technology, its ability to differentiate the Quality of Service (QoS) of the carried traffic streams according to their needs is often mentioned as a graceful characteristic. Real time applications requiring stringent end to end information transfer delays and delay variations can be carried with the higher QoS class, which in the ITU-T terminology is called "QoS class 1" (ITU-T I.356, 1996). Applications more tolerant to delays, like data applications, can be carried either with QoS class 2 or QoS class U. QoS class 2 means that Cell Loss Ratio (CLR) is guaranteed to be lower that a certain bound, whereas QoS class U means that no guarantee at all is given, neither on losses nor on delays. Although data applications always have the ability to detect and recover packet losses through frame retransmission, a lot of simulation studies have shown how dramatic can be, in terms of increased end to end delays, the effect of unbounded cell losses during congestion periods (Bonaventure, 1997). In spite of being in principle "tolerant" to application delays, users of applications like Telnet, FTP, Web Browsing, etc., would greatly appreciate the performance improvement coming from limited cell losses in their data.

Right after the ability of ATM to differentiate traffic into QoS classes, it can be recalled its scalability, i.e. its suitability to be used both as a LAN and as a WAN technology. In recent years, a lot of ATM LANs have been deployed and they are nowadays being used with success. Also, trials and experiments to deploy and operate ATM in the WAN environment have been performed, and ATM backbone networks are now a reality.

In parallel, especially due to the booming growth of the Internet, IP protocol has consolidated its positions, and legacy LAN technologies like the Ethernet have been significantly improved (100 Mbit/s Ethernet switches being already widely deployed and Gigabit Ethernet being right round the corner).

In summary, ATM can be successfully used as a backbone technology also to carry data traffic relaying on the IP protocol generated on non ATM LANs. The bursty nature of this kind of traffic gives the public carriers the opportunity to achieve some statistical gain (or "multiplexing gain"), but the need to do that while meeting some QoS contract rises a lot of traffic engineering issues, and this paper addresses some of them.

The paper is organised as follows: in section 2 we describe how we performed some traffic measurements over CSELT's LANs in order to verify some characteristics of LAN traffic (Self Similarity) that had already been described in literature (Leland, 1994) and to obtain the values needed to parameterise the source traffic models we used in the study. Such models are described in section 3. In section 4 we describe the simulation scenario we implemented, while in section 5 we focus on the effectiveness of the traffic shaping as a way to meet QoS commitments. Finally, in section 6 we present some conclusion and outline the future work.

2 TRAFFIC CHARACTERISATION

A thorough characterisation of the traffic generated by IP applications over todays LANs has a great importance when performing internetworking studies. After the pioneering work at Bellcore (Leland, 1991), which put into evidence the Self Similar and Long Range Dependence (LRD) characteristics of this type of traffic, a lot of efforts all over the world was made to faithfully reproduce these characteristics by means of models more complicated than the traditional markov ones. A good review of these traffic characteristics and proposed models and can be found in (Morin, 1996).

Usually, the path followed by researchers in this area is to perform some measurement on real networks, verify Self Similar characteristics, provide an estimation of the Hurst parameter and of other parameters of interest and then use those values into some traffic models to show how good they are in reproducing the statistical characteristics of real traffic traces and/or their queuing behaviour.

This is what we did in our study too, but we were less concerned about comparing performances of "advanced models". This not because we believe we used the best possible traffic model, but because we are aware that whatever thorough the characterisation of some measured traffic is, it is probably very closely related to the network technology and to the applications and protocols used over it. This danger is very well explained in an article by Paxon and Floyd (Paxon, 1997), which also suggests the wide variation of parameter values used in the models as a method to extend the generality of internetworking studies performed. This was the approach followed in this study.

For the sake of completeness, however, we briefly describe how we collected and analysed measurements on some CSELT's LANs.

We focused on a 10Mbit/s Ethernet segment collecting traffic from several hosts (mainly PCs with windows 95 and Unix workstations, mostly running ordinary network applications such as e-mail clients, FTP an Web Clients, telnet, Xwin, Sun NFS). Measurements were collected by a Sun Ultra 1 Workstation with a 200 MHz processor, with the aid of a modification of the freeware "tcpdump". The main modification consisted in the fact that no single packet information was stored on the disk but only, at fixed time intervals, the summary information about the number of Ethernet bytes seen on the segment. The time interval duration was one second, and measurements were repeatedly collected from 9.00 am to 17.00 pm, for sixteen working days. Some comparison of our data with data collected in parallel with the aid of a Wandell & Goltermann Da-30 protocol analyser, showed that packet losses by the tcpdump modification were limited, and estimated statistical parameters were not significantly affected.

Unfortunately, the 10 Mbit/s Ethernet segment under measurement only collected a limited amount (say, less than ¼) of the Intranet and Internet traffic generated/received by the 200 researchers hosted in our building. As a result, the sixteen daily collected traffic profiles often showed some evident nonstationarity. In order to increase stationarity, we superposed them four by four, thus obtaining four aggregate profiles, being more stationary and potentially more representative of the traffic generated/received by all the researchers of our building. Of course such an operation was possible only because we performed load measurements, and we didn't compute any packet interarrival time statistic.

The four "aggregated" profiles were then analysed in order to compute some statistical parameters. Among them, the more relevant to this study were

- the mean rate (in bytes/s);
- the peak factor, i.e. the ratio of the variance of the number of bytes to the mean of the number of bytes seen at each one second time interval;
- an estimation of the Hurst parameter based on the Index of Dispersion for Counts (IDC). See (Gusella, 1990).

We then chose one of the profiles as being the more representative one, and used the computed values as a starting point to parameterise our model, whose description is in the following section.

3 TRAFFIC GENERATION MODEL DESCRIPTION

We already pointed out that in order to drive conclusions not limited to the traffic characteristics of a particular LAN, we preferred not to directly perform simulations with measured traffic traces. Instead, we used a traffic model initially parameterised on the basis of measurements and then varied its parameters.

The model we implemented belongs to that category of bursty fluid models that try to achieve Self Similarity by aggregation of ON OFF sources with heavy tailed distributed on and/or off periods, as explained in (Morin, 1996) or in (Willinger, 1995). In particular, our model output consists of the rate generated by sources that can become active according to a Poisson process with parameter λ, when active generate traffic at a fixed rate R and whose active state duration h is heavy tailed distributed (see Figure 1). We will refer to that model as the Poissonian Arrival of Bursts (PAB) model. References to it can also be found in (Roberts, 1997).

Instead of investigating "a priori" whether the infinite source approximation were valid or not to correctly reproduce the traffic generated by a finite (and indeed rather limited) number of users/applications, we preferred to compare the queuing behaviour of real traces and simulated traces, usually finding a good match.

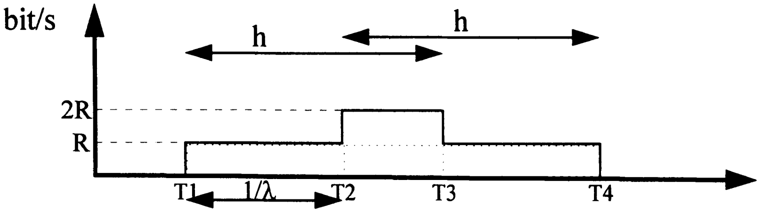

Figure 1 – Traffic generation according to the PAB model.

The exact probability density function of the burst duration h is reported in equation (1), were H is the model's resulting Hurst parameter and Tc and ε have the meaning of maximum and minimum burst duration, respectively.

$$f_h(x) = \frac{1}{1-\left(\frac{T_c}{\varepsilon}\right)^{2H-3}} \frac{1-2H}{\varepsilon^{2H-3}} x^{2H-4} \qquad \left(\varepsilon \le x \le T_c,\ 0\,otherwise\right) \tag{1}$$

Due to the presence of a nonzero ε and of a non infinite Tc, the exact expression of the Index of Dispersion for Counts (IDC) for the model generated traffic drifts from the ideal one, which would be as in reported in (2) and correspond to the IDC of asymptotically Self Similar traffic. For any nonzero ε and finite Tc, the drift becomes more and more evident as t approaches zero or tend to infinity.

$$\text{IDC}(t) = Kt^{2H-1} \qquad \text{(where K is a constant)} \tag{2}$$

The model is thus characterised by five parameters: λ, R, H, ε and Tc. We fixed the first three with the aid of three equations (not reported here) and of the three parameters extracted from measurements (mean rate, peak factor and Hurst parameter, see section 2), whereas Tc and ε can be considered as "freedom degrees" of the model. In finite time simulations, they must be set to values different from infinite and zero, respectively. We developed an algorithm that, given an interval [t1, t2] over which a "small drift" of the real IDC from the ideal expression reported in (2) is allowed, finds the best ε and Tc choices for a given maximum simulation time. In our study, we always required a "small drift" within the time range [0.1s, 50s]. The model then results as Self Similar and highly correlated (i.e. "Long Range Dependent" – LRD) at least within that range. In (Paxon, 1997), as an example, it is recalled that long term correlations of LAN traffic have frequently been observed "from hundredths of milliseconds to tens of minutes". We are not so far from this range, as even if beyond 50s the IDC curve starts to drift from the ideal one, the traffic remains correlated well above this value.

4 THE SIMULATION SCENARIO

In the following we assume that the reader is familiar with these concepts: ATM Virtual Path (VP) and Virtual Channel (VC) connections, ATM traffic contracts, Deterministic Bit Rate (DBR) transfer capability, Statistical Bit Rate (SBR) transfer capability, Peak Cell Rate (PCR), Sustainable Cell Rate (SCR), Maximum Burst Size (MBS). Definitions can be found in ITU-T recommendation I.371 (ITU-T I.371, 1996) or in ATM Forum traffic management specification 4.0 (ATMF TM 4.0, 1996).

To address some of the traffic engineering issues outlined in the introduction, we simulated a simple scenario, which is the multiplexing of several VC connections into a single DBR VP Connection. This may occur, for example, on an output port of an ATM switch that collects traffic from an edge ATM network or directly from ATM cards of upstream IP routers. The DBR VP connection is assigned a fixed amount of the bandwidth of the physical link, as well as a fixed amount of buffer space.

All the simulations presented in this study were performed at the fluid level, i.e. traffic sources produce as an output a sequence of couples like (Time Interval, Rate during Time Interval). The queues do not receive cells, but only the information about the intensity of an incoming workload and the duration of this intensity. Figure 2 summarises the simulation scenario we considered.

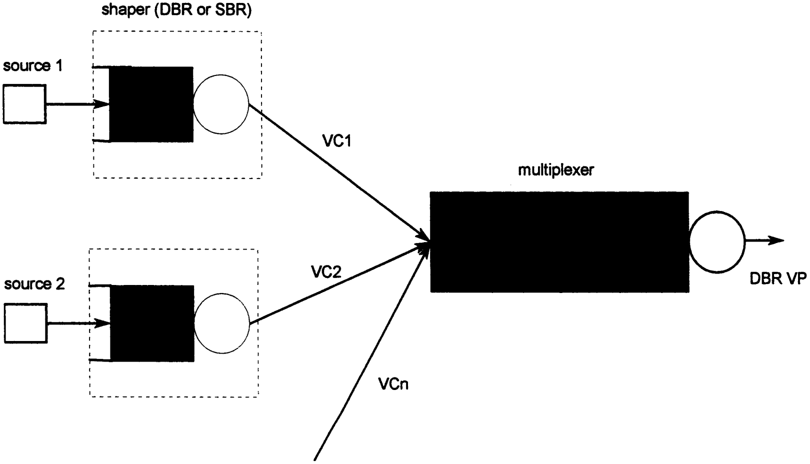

Figure 2 - The simulation scenario.

Each VC connection has a traffic contract that can be either DBR or SBR, and a PAB model as described in section 3 generates the source traffic.

In this study we considered two cases:

- each VC connection can access all the buffer space and bandwidth reserved to the DBR VP (unshaped case), i.e. its traffic contract is either not controlled or traffic contract parameters are set to values that prevent shaping devices from performing significant actions on the incoming flow (e.g. if the traffic contract is DBR, the PCR of the connection is set equal to the line rate).
- each VC connection, before accessing shared VP resources, is shaped according to a traffic contract, that can be either DBR or SBR.

In the DBR case, a "fluid" shaping device can be thought as a queue served at the Peak Cell Rate (PCR) of the connection. In the SBR case, as a queue that can be served either at PCR or at a lower rate which is the Sustainable Cell Rate (SCR) of the connection: the service rate is PCR as long as a token pool of size Maximum Burst Size (MBS) is nonempty, SCR otherwise. The token pool size is initially set at MBS, and it increases (or decrease) at a rate which is SCR minus the current rate of incoming traffic. The pool size can never exceed [0, MBS]. The size of the buffer in the shaping devices is set to infinite, i.e. no losses can ever occur in them. In both the unshaped and shaped cases the QoS class of the VC connections is QoS class 2, i.e. there is a commitment of a Cell Loss Ratio lower than a certain bound for each connection. In the following we suppose that due to the complete sharing of VP resources, achieving this commitment at the VP level (i.e. in the VP buffer) is equal to achieve it for the single connection. As VC shapers' buffers, when present, are of infinite size, the VP buffer is the only point along the connections where overflow can occur.

As pointed out in the introduction, although there's no explicit commitment for delays in QoS class 2, the network engineering should not enable delays to become intolerably high. In our simplified scenario, delays can occur both in the VP buffer and, if shaping is performed, in VC shapers' buffers. Therefore, the delay statistic we consider will be the sum of two terms: the mean delay encountered in the VP buffer and the average of the mean delays encountered in the VC shapers (if any).

Instead of taking the pure output of the PAB models, to speed up the simulations we slotted them into fixed intervals of duration 0.1s. The lowest time dynamics we will be able to observe is thus limited to this value. Time dynamics lower than 0.1s would anyway have been filtered due to the buffer size of the multiplexer, which was chosen considerably high (see later).

The sources we used are five different parameterisations of the PAB model presented in section 3. Table 1 summarises the main parameters for each one of them.

Table 1 – Parameters of the Poissonian Arrival of Bursts (PAB) model used as traffic sources

	Mean (byte/s)	*Peak Fact (bytes/0.1s)*	*H*
Parameterisation 1	682000	21655	0.8
Parameterisation 2	682000	21655	0.9
Parameterisation 3	682000	21655	0.7
Parameterisation 4	682000	43310	0.8
Parameterisation 5	682000	10827	0.8

In particular, the first parameter set was derived from the analysis of measurements (see section 2). The others are one parameter variations from the first, to study what happens with increased autocorrelation (parameterisation 2), reduced autocorrelation (parameterisation 3), increased burstiness (parameterisation 4) and reduced burstiness (parameterisation 5). The variation in the autocorrelation was obtained by varying the Hurst parameter (the higher it is, the more autocorrelated the traffic), while the variation in the burstiness was obtained by varying the peak factor (the higher it is, the more bursty the traffic).

In all the performed simulations, all the multiplexed sources belonged to the same parameterisation, i.e. we didn't consider the case of mixed types of traffic sources.

Note that the value of the Peak Factor at 0.1s did not come directly from the analysis of measurements (that for technical reasons were taken with a period of 1s), but was extrapolated from the value computed at 1s, as described in the following.

If the traffic has Self Similar characteristics over a certain timescale range [t1, t2], then its index of dispersion for counts, over this range, has the expression reported in (2). The value of the Peak Factor at time t corresponds to IDC(t). In our 1Hz frequency measurements we could verify Self similar Characteristics over a range [1s, t2], were t2 depended on the measurement's day but was always of the order of hundredths of seconds. We also estimated a Hurst parameter H close to 0.8, computed the peak factor at 1s (i.e. IDC(1s)) and finally computed the value of k in equation (2). If the hypothesis that the traffic has Self Similar characteristics also on lower timescales is made, then the computation of IDC(0.1s) i.e. the Peak Factor at 0.1s is straightforward. This hypothesis is supported from a lot of empirical data analysed in several studies, see e.g. (Paxon, 1997). It wouldn't make sense to extend it to timescales lower than 0.1s.

In each simulation run we considered at least a simulated time span of 10^{+5} seconds (more than a day), in order to ensure the correct statistical behaviour of our sources, which have Long Range Dependent characteristics.

In all the simulations, the bandwidth of the DBR VP was fixed at 155 Mbit/s (thus representing the case of a single OC3 link dedicated to this kind of traffic) and its buffer space was fixed at 477000 bytes, i.e. 9000cells. This buffer space value is quite large (even if not unrealistic for today switches), and we choose it in order to better observe the effects of LRD traffic (the longer the buffer, the more relevant the impact of correlations properties in the traffic).

In Figures 3 and 4 we present the losses vs. utilisation and the mean delay vs. utilisation plots of the unshaped case for the five parameterisations of Table 1.

From Figure 3 it can be noted that the variation of the Hurst parameter (from 0.7 to 0.9) does not significantly impact the loss ratio in the VP buffer. This is partly due to the finiteness of the buffer. Performing other simulations with larger buffer space (may be unrealistic for an ATM switch) differences between the curves with different H value started to be more observable. However, it should not be deduced that Self Similar properties in the traffic do not affect performances, (indeed, in

Figure 3 there's no comparison with non Self Similar models). Only, whenever using heavy tailed ON OFF models in finite time simulations of finite buffer systems, the value of the Hurst parameter may not significantly affect the results. Mean delays are even less sensible to Hurts parameter variation (see Figure 4).
On the contrary, the variation of the peak factor parameter, which is directly related to the burstiness of the sources on low timescales, significantly affects the performances, and should therefore receive much consideration when setting model parameters for engineering purposes.

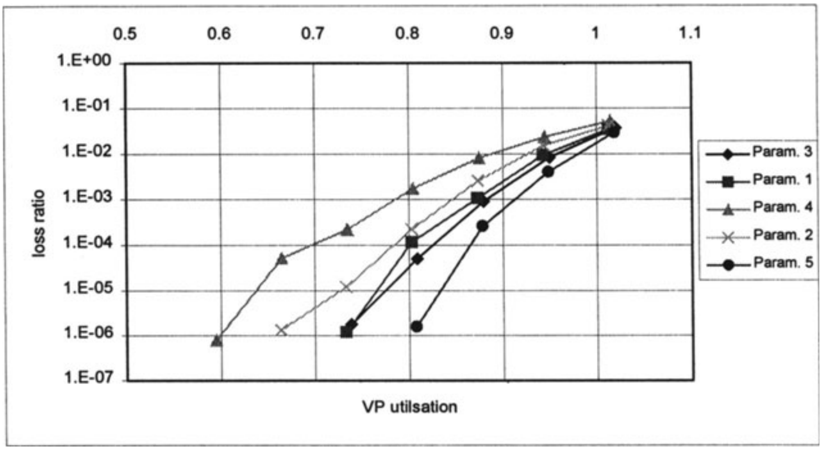

Figure 3 – Losses vs. utilisation plots for the five source parameterisations considered – all sources are unshaped.

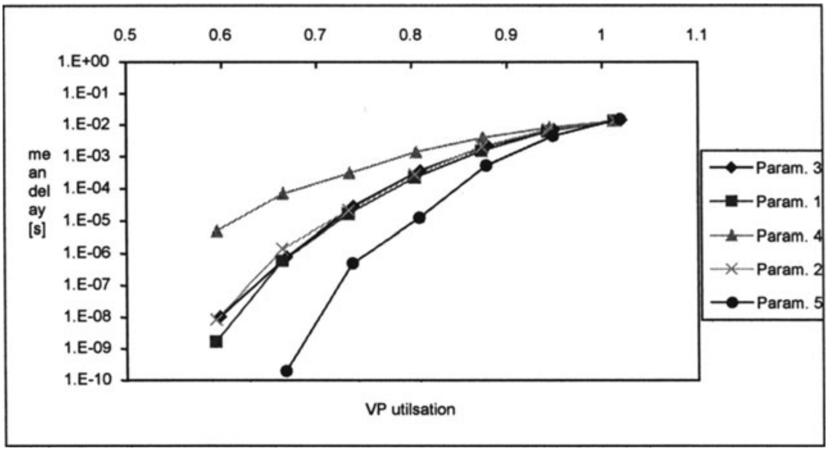

Figure 4 –Mean delay in multiplexer vs. utilisation plots for the five source parameterisations considered – all sources are unshaped.

It can be recalled that in order to meet a QoS commitment on CLR for multiplexed DBR or SBR ATM connections, four methods may be used.

The first one is simply to reduce the shared VP utilisation, i.e. to move the working point of curves like the ones of Figure 3 towards the left bottom corner. This is always possible and the multiplexing delay is even reduced, but it leads to lower incomes for the network operator.

The second one is to increase the level of multiplexing, i.e. to increase both the VP bandwidth and the number of admitted sources. Anyway, this is only possible if there are enough flows to multiplex, and in any case the VP bandwidth cannot exceed the one of the physical link. For some analytical considerations on the benefits of multiplexing with Fractional Brownian Motion traffic, see (Erramilli, 1996).

The third one is to increase the buffer space assigned to the VP. This always results in an increased multiplexing delay too, and there are cases were due to traffic source characteristics the buffer growth can be unacceptable. This is often referred as the “buffer ineffectiveness” for Long Range Dependent traffic.

The fourth one is to perform shaping on the flows before multiplexing according to some traffic contract parameters, that should be carefully chosen. Supposing to leave the VP utilisation the same, this always results in the creation of a second delay component (due to the queuing into the shaper’s buffers), while the delay component due the multiplexer is expected to be reduced. The effectiveness of such a method depending on traffic source characteristics has already been questioned in (Erramilli, 1996) and is further investigated in the following of this paper.

5 EFFECTIVENESS OF SHAPING IN REDUCING LOSS RATIO AND SIDE EFFECTS ON DELAYS

5. 1 DBR shaping

We start considering the case of VC connections being shaped before multiplexing according to a DBR traffic contract: whenever the source has a peak whose intensity exceeds a given Peak Cell Rate, it is limited to that PCR and the excess work is buffered.

In the first simulated case, we multiplexed as many sources belonging to the first parameterisation listed in Table 1 as necessary to push, in the unshaped case, the loss ratio above 10^{-5} (QoS class 2 target). This required 23 sources and led to a VP utilisation of 0.81.

While keeping the generated traffic the same, we then varied the PCR of the shapers in order to evaluate their effectiveness in reducing the loss ratio below the QoS class 2 target.

Results are reported in Figure 5 (for the moment, refer to the “PAB” plot only). In Figure 5(a) there is the value of the loss ratio in the VP buffer vs. the ratio of mean source rate to PCR, while in Figure 5(b) the sum of the mean delay in the VP buffer with the average of the mean delays in the shapers' buffers vs. the same mentioned ratio. As the delay monotonically increases with a tighter shaping, it's clear that the increasing of the delay component introduced by the shaper always dominates the decreasing of the delay component in the VP buffer.

Also, it should be noted that losses start to differ from the unshaped case only when the shaping becomes "tighter enough" (e.g., referring to the PAB plot of Figure 5(a), only when mean source rate / PCR > 0.45) This means that the sharpest peaks, which are the only ones eliminated by a loose shaping, are not the main cause of losses in the multiplexer.

This result certainly depends on the rather big (9000 cells) size of the VP buffer, but also on the LRD characteristics of the PAB traffic sources. For comparison, always in Figure 5, we plotted the effect of shaping on "traditional" ON OFF sources with on and off periods exponentially distributed, with mean rate equal to the mean rate of PAB sources and peak rate during on periods equal to four times the mean source rate[1]. It's evident that in such a case even a looser shaping is more effective, and that a significant reduction in the loss ratio can be achieved without a dramatic increase in the delays.

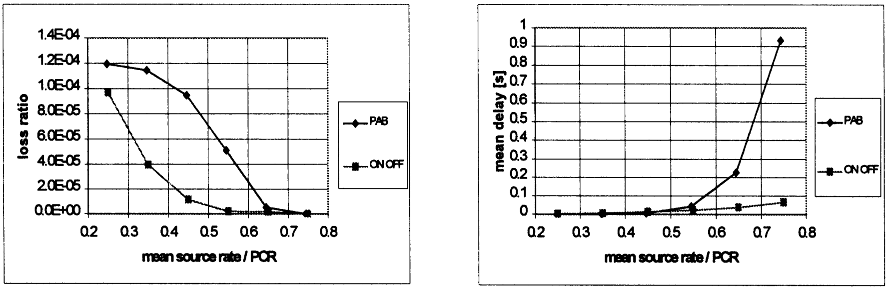

Figure 5(a) and 5(b) – Effects on loss ratio and mean delay of DBR shaping for PAB parameterisation 1 sources (Table 1) and for ON OFF exponential sources. Abscissa values represent the ratio of the mean source rate to shapers' PCR.

In Figures 6 through 9 we present plots analogous to Figure 5 for the other parameterisations of Table 1. The attempt was always to evaluate the effectiveness of the shaping to reduce the loss ratio from a starting value, in the unshaped case, above 10^{-5}. In order to do so, for each parameterisation we varied the VP utilisation by adding or removing an integer number of sources, and this is the reason why the leftmost values in the loss plots may not be exactly the same.

Results show that the main drawback of performing DBR shaping on LRD traffic is the same outlined for parameterisation 1: the region where the shaping begins to

---

[1] Mean on and off period duration was chosen in order to obtain a loss ratio close to the one of the PAB sources in the unshaped case.

be effective on losses corresponds to the "sensible" region where mean delays in the shapers' buffers start to (rather sharply) increase.

Moreover, it should be noted that for parameterisations 1, 2 and 3 (variation of the Hurst parameter, see Figures 5(b), 6(b) and 7(b)) the absolute values of delays for the same PCR are very different (the higher H, the higher the delays). We verified that these delays are dominated by the delay components introduced by the shapers. This means that when shaping at the source level, the value of the Hurst parameter plays a dominant role in the system performances. On the contrary, Figure 3 and even more Figure 4 showed that in an unshaped scenario the performance differences due to the Hurst parameter were negligible. So, another drawback of performing shaping is that more information about correlation properties of the sources (e.g. a reliable estimation of the Hurst parameter) would be needed.

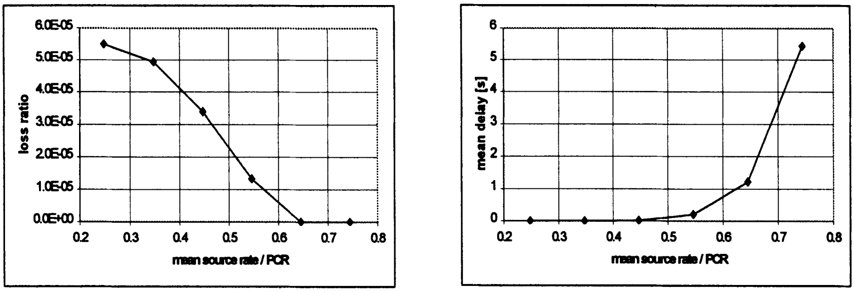

Figure 6(a) and 6(b) – Effects on loss ratio and mean delay of DBR shaping for PAB parameterisation 2 sources (see Table 1). Abscissa values represent the ratio of the mean source rate to shapers' PCR.

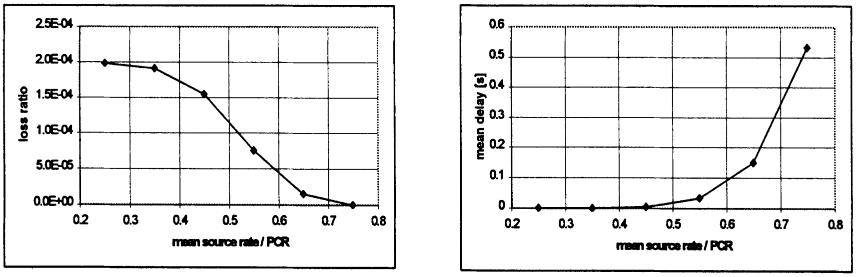

Figure 7(a) and 7(b) – Effects on loss ratio and mean delay of DBR shaping for PAB parameterisation 3 sources (see Table 1). Abscissa values represent the ratio of the mean source rate to shapers' PCR.

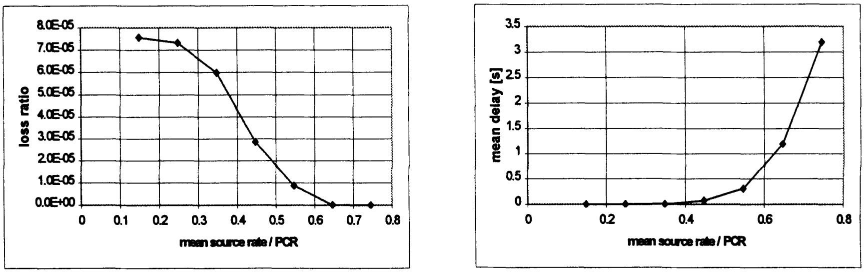

Figure 8(a) and 8(b) – Effects on loss ratio and mean delay of DBR shaping for PAB parameterisation 4 sources (see Table 1). Abscissa values represent the ratio of the mean source rate to shapers' PCR.

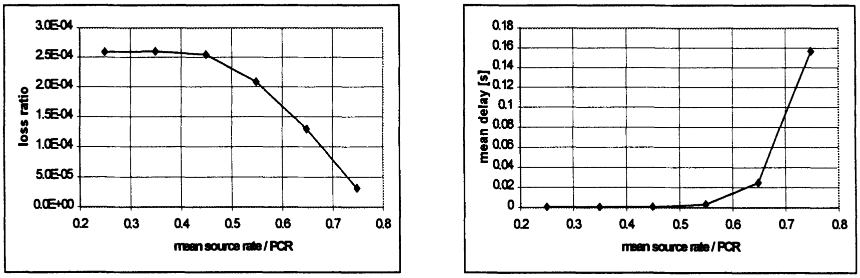

Figure 9(a) and 9(b) – Effects on loss ratio and mean delay of DBR shaping for PAB parameterisation 5 sources (see Table 1). Abscissa values represent the ratio of the mean source rate to shapers' PCR.

All these consideration lead to the conclusion that it's better not to alter the characteristics of the sources by performing a DBR shaping, (or at most perform only a loose, cautelative shaping), especially because the multiplexing gain can be considerable even with LDR sources (Erramilli, 1996).

5.2 SBR shaping

As the main drawback of DBR shaping seems to introduce significant delays in the shaper's buffer, we then investigated whether this could be overcome by using a more sophisticated shaping. The SBR traffic contract allows the source to transiently exceed a certain Sustained Cell Rate (SCR) rate, but on the long term it will be limited to it. The Maximum Burst Size (MBS) parameter limits the amount of this transiently generated excess traffic. Anyway, the source rate can never exceed a Peak Cell Rate (PCR). In order to limit the complexity of the analysis, we fixed the PCR at a rate higher than the maximum rate ever generated by a source. With such a choice, our implementation of an SBR shaper with a certain SCR and MBS = 1 cell is equivalent (for practical purposes) to a DBR shaper with PCR equal to this SCR.

We then let MBS assume the values 1, 500, 5000 and 50000 (in cells) and for each of them we performed the same analysis of the shaping effectiveness as before by varying the SCR value. In Figures 10(a) and 10(b) the results for parameterisation 1 are reported.

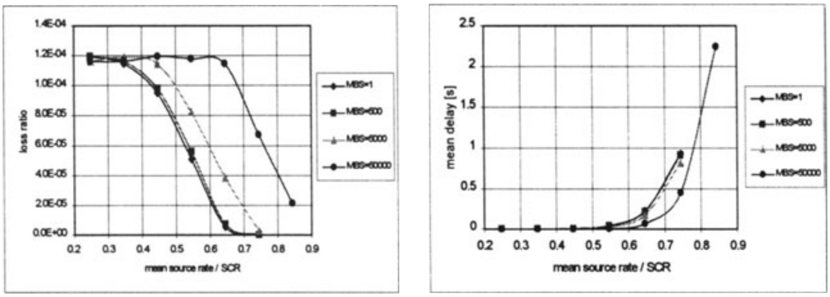

Figure 10(a) and 10(b) – Effects on loss ratio and mean delay of SBR shaping for PAB parameterisation 1 sources (see Table 1). Abscissa values represent the ratio of the mean source rate to shapers' SCR.

As expected, for the same SCR larger values of MBS lead to reduced delays (Figure 10(b)), as the main delay component (i.e. the one in the shapers) benefits for the increased "tolerance" of the shaping algorithm. Conversely, that tolerance throws into the VP buffer a traffic that is less filtered, and losses increase (Figure 10(a)). So, the right way to state whether SBR shaping is more or less effective than DBR shaping is to choose a loss ratio target value on Figure 10(a), read for each MBS curve what is the SCR needed to achieve it, and then find on Figure 10(b) the corresponding delay value. In Figure 11 we reported the results of such an operation for all the parameterisations considered.

As all the curves are monotonically growing, the conclusion is that SBR shaping is always less effective than DBR shaping (i.e. the MBS=1 point), probably because with these type of Self Similar sources no "typical burst length" can be identified. Note that absolute values of these curves should not be compared, as they depend also on the VP utilisation, which may differ from one parameterisation to another as explained before.

On the contrary, for the exponentially distributed ON OFF model considered before (see Figure 5), SBR shaping shows to be more effective than DBR shaping when MSB has a value around 500 cells (see Figure 12). The model had indeed an ON period average duration of 11ms, and an ON rate of 51471 cells/s: 500 cells is thus near to the "typical burst duration" of this type of source.

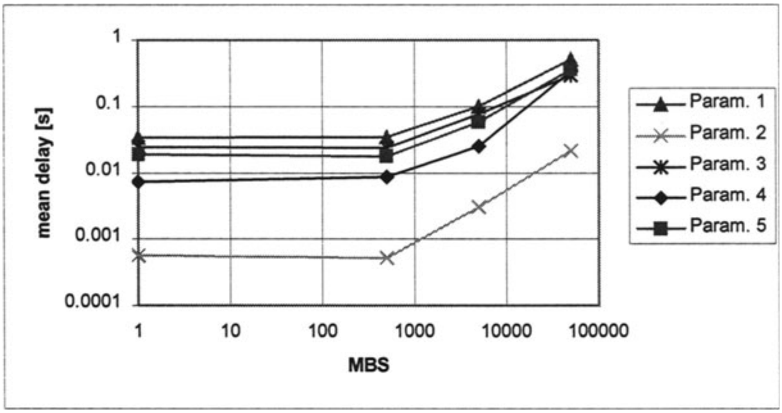

Figure 11 – Effect on mean delay of SBR shaping for PAB parameterisation 1 through 5; varying MBS, for a fixed loss ratio target. Absolute values across the different parameterisations should not be compared.

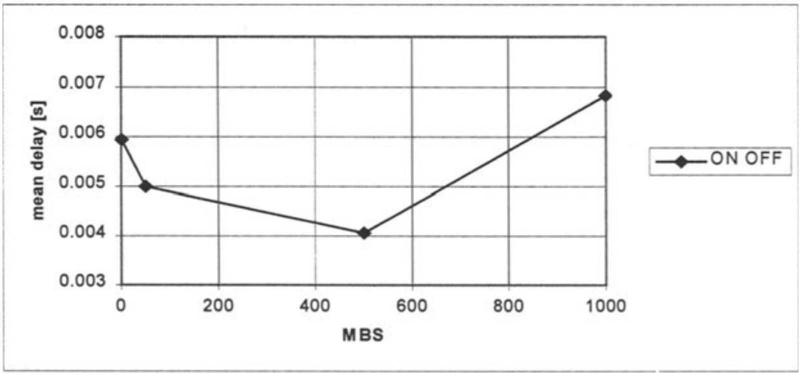

Figure 12 – Effect on mean delay of SBR shaping for an ON OFF exponential sources, varying MBS, for a fixed loss ratio target. – Note that the range of MBS values of interest is different from the one of Figure 11

6 CONCLUSION AND FUTURE WORK

In this paper we studied, by means of simulation, a simplified scenario where several VC ATM connections carrying QoS class 2 LAN traffic are multiplexed on a bandwidth resourced DBR VP ATM connection. Traffic was generated on the VC connections with a model having Long Range Dependent behaviour, parameterised on the basis of real aggregate LAN traffic measurements. The goal was to study the trade off between reduced loss ratio in the VP buffer (or, equivalently, increased VP utilisation) and increased transfer delays when the VC connections were shaped according to some traffic contract. Results showed that when shaping tighter and tighter according to a DBR traffic contract, the reduction of losses in the VP buffer becomes significant only in the region where the mean delays in the shapers' buffers have started to increase, i.e. where the risk of having huge peaks in the delays is probably high (see Figures 5 through 10).

In addition, results showed that when shaping according to a SBR traffic contract, it's impossible to find a couple of parameters (SCR, MBS) leading to delays lower than the ones obtained with a DBR shaping, with the same loss ratio target level. This shows the difficulty to identify a "typical burst length" for these type of Long Range Dependent sources (Figure 11).

The results remain similar even if varying the parameters of the Long Range Dependent traffic generations models. On the contrary, they are rather different if other types of traffic generation models are used, such as a traditional ON OFF model with exponentially distributed ON and OFF periods. In this case DBR shaping is more effective in reducing loss ratio and its side effects (delay increase) are less disturbing (Figure 5). Always in this case, we showed that with an appropriate choice of MBS, SBR shaping can be more effective than DBR shaping, and therefore it makes sense looking for a "typical burst duration" (Figure 12). Unfortunately, LAN traffic characteristics are certainly closer to the ones showed by LRD models.

The relation between losses and delay curves, shaping parameters, traffic model generation parameters and buffer space was not addressed in this paper, but the results suggest some caution with setting tight shaping parameters until more light on the subject is shed.

7 ACKNOWLEDGEMENTS

The authors wish to acknowledge Mr. Damiano Inguaggiato and Luca Viale for cooperating in the measurement collection phase, and Mr. Marco Canova for writing part of the measurement analysis code.

8 REFERENCES

ATM Forum TM4.0 (1996): Traffic Management specification - Version 4.0.

Bonaventure, O. (1997) – A simulation study of TCP with the proposed GFR service category – Presented at Dagstuhl Seminar 9725 on High Performance Networks for Multimedia Applications, June 15-20, 1997, Schloss Dagstuhl, Germany.

Erramilli, A. - Narayan, O. and Willinger, W. (1996) – Experimental Queueing analysis with Long Range Dependent packet traffic – IEEE/ACM Transactions on networking, Vol. 4 No.2 – pp. 209-223 – April 1996.

Gusella, R. (1990) – A measurement study of diskless workstation traffic on an Ethernet. - IEEE transactions on communications – Sep. 1990

ITU-T Recommendation I.356 (1996) – B-ISDN ATM Layer Cell Transfer Parameters.

ITU-T Recommendation I.371 (1996) - Traffic control and congestion control in B-ISDN.

Leland, W.E. and Wilson, D.V. (1991) – High time resolution measurements and analysis of LAN traffic: Implications for LAN interconnections – IEEE Infocomm'91, paper 11D.3.1

Leland, W.E. - Taqqu M.S. - Willinger, W. and Wilson D.V. (1994) - On the self similar nature of Ethernet traffic - IEEE/ACM Transaction on networking, Vol 2, pp. 1-15, 1994.

Morin, P.R. (1996) - The Impact of Self-Similarity on Network Performance Analysis – Publicly available on the WEB at http://www.scs.carleton.ca/~morin/performance/selfsim/paper/paper.html

Paxon V. and Floyd, S. (1997) – Why we don't know how to simulate the Internet – Proceedings of Winter Simulation Conference – Atlanta 1997.

Roberts, J. – Mocci, U. and Vitramo J. (1997) editors - Broadband Network Traffic - Final report of action COST 242 - Springer Verlag – Chapter 13.

Willinger, W. - Taqqu, M.S. - Sherman, R and Wilson, D.V. (1995) - Self Similarity through high variability: statistical analysis of Ethernet LAN traffic at the source level. Proc. of SIGCOMM '95, pp.100 – 113.

9 BIOGRAPHY

Paolo Castelli received the degree in electronic engineering from Polytechnic of Turin, Italy, in 1985. After some years spent in the field of industrial automation, in 1989 he joined the Network Planning and Management department in CSELT, where he is now the head of the Traffic Engineering research unit. His area of activity includes the analysis of traffic management aspects over different networking technologies, with special attention to ATM networks. He participated to several EURESCOM and ACTS projects, and to the standardisation of the B-ISDN both in ITU-T SG 13 and in ETSI NA5.

Luisa Guida received the degree in telecommunication engineering from the University of Pisa, Italy, in 1997. During her thesis activity in CSELT, for which she received an award from the "Fondazione M. Bellisario", she investigated on performance of ATM shapers in presence of Self Similar traffic. Currently she is a researcher of the Traffic Engineering research unit in CSELT and her interests include LAN traffic measurements, characterisation and modelling.

Maurizio Molina received the degree in electronic engineering from Polytechnic of Turin, Italy, in 1993. The thesis activity, for which he received an award from Telecom Italia, concerned a simulation study of connectionless services supporting through ATM. He has been working as a researcher in the Traffic Engineering research unit in CSELT since 1995, and his current activities are in the area of traffic modelling for management purposes. He participates to the standardisation of the B-ISDN in ITU-T SG 13 and in an ACTS project.