How does an ATM switch see the traffic through the Leaky Bucket?

I. Cselényi, N. Björkman
Broadband Team, Network Research Dept., Telia Research AB
Rudsjöterrassen 2, S-13680 Haninge, Sweden,
Phone: +46-8-707 5038, Fax: +46-8-707 5596,
E-mail: istvan.i.cselenyi@telia.se

S. Molnár
HSN Laboratory, DTT, Technical University of Budapest
Stoczek utca 2, H-1111 Budapest, Hungary,
Phone: +36-1-463 3889, Fax: +36-1-463 3107,
E-mail: molnar@ttt-atm.ttt.bme.hu

Abstract

This paper describes a new way of ATM traffic characterisation based on measurement. The method, called *Leaky Bucket Analysis* can be directly used for quantitative analysis of burst structure, determination of adequate shaping rate, selection of sustainable parameter set, retrieving input parameters for source modelling, analysis of multiplexing performance and dimensioning ATM networks. The proposed method is based on the leaky bucket algorithm which is widely used in ATM context. The goal of our characterisation method is to reveal the important properties of the traffic just as the switch sees them.

The usage and applicability of the Leaky Bucket Analysis method is demonstrated both on artificial traffic patterns and on 30 real traffic traces captured from different applications such as *Internet, video conference, and LAN-LAN interconnection*. Detailed conclusions are drawn from the results of Leaky Bucket Analysis establishing the relation between the features of the source, settings in the application and the gained characteristics. The robustness of the

Leaky Bucket Analysis method is also examined; a safety margin is defined for each analysed traffic type, the impact of finite measurement trace is highlighted and the extension for the case of non-zero cell loss is illustrated. A number of application areas are illustrated at the end of the paper.

Keywords
Leaky Bucket, ATM Networks, Traffic Characterisation, Resource Allocation

1 INTRODUCTION

Source characterisation is essential for prevention of congestion in ATM networks both on call level (Call Admission Control) and cell level (Usage Parameter Control or shaping). A new call can be admitted into the ATM network only if the involved ATM nodes can provide the network resources it intends to take, without affecting the QoS of previously accepted calls (Roberts, 1996). According to the standard (ITU-T I.371) the ATM terminal has to describe its traffic by giving the so called *sustainable parameter set* (i.e. Sustainable Cell Rate, Maximum Burst Size, Peak Cell Rate and Cell Delay Variation Tolerance) to the access node. However, there are several problems arising from this concept. Firstly, this parameter set is capable only for a *very rough description* of the traffic and gives basically only the characteristics of a simple on-off source, which consumes much more network resources in terms of bandwidth and buffer space than the original traffic. The result is over-booking. Secondly, it is a *difficult task for the terminal* (or for the user in front of it) to give a realistic hint even for these few parameters.

The only way to gain more, and more reliable information about the source's behaviour in the buffer is to measure it. Our experiences verified that the result of measurement based characterisation can be used as a'priori information, since a finite threshold can be given for the deviation of repeated characterisations of the same source. Naturally, the threshold depends on the specific traffic type, e.g. Variable Bit Rate video, Internet or aggregate network traffic (see Section 5.1).

Cell level congestion control takes place after a call is admitted. The network protects itself first through policing (UPC), which enforces the contract specifying the sustainable parameter set. In addition, some of the less delay sensitive traffic classes can be shaped. The most common method both for *policing and shaping* is to apply one or two Leaky Buckets (Elwalid, 1995). However, this is often inadequate, since one or two simple Leaky Buckets (LB) can not describe and control the overall burst structure of the traffic; they are *coupled with a certain time-scale*. Therefore a more complex analysis of incoming traffic should be performed on each time-scale.

A number of different methods have been proposed (Frost, 1994, Stamoulis, 1994, Roberts, 1996) for traffic characterisation. However, based on the

aforementioned considerations a measurement based traffic analysis method is necessary which conforms with the sustainable parameter set, but easy-to fulfil and provides more accurate description of required network resources on each time scale. The access node judges and sees the incoming traffic through the LB algorithm. Therefore, it is a logical idea to characterise the traffic source by calculating its *resource consumption by the LB algorithm for all leak rates* from zero to the full link rate. Procedures for selecting shaping rate and building up traffic models using the Leaky Bucket Analysis as traffic characterisation tool were presented in previous works (Latour-Henner, 1997, Cselényi, 1997, Molnár, 1996a). It is shown in this paper that the Leaky Bucket Analysis provides besides the standardised traffic description parameters additional information which can be applied for burst structure analysis, determining the sustainable parameter set, analysis of multiplexing performance and network dimensioning.

The proposed Leaky Bucket Analysis method is described in Section 2 and applied on artificial traffic patterns in order to get a formal description in Section 3. As real validation of the method, LBA is also performed on different ATM cell streams captured by actual measurements. Several types of single and multiplexed traffic sources from different applications are characterised and the results are evaluated in Section 4. A robustness study of the Leaky Bucket Analysis method is presented giving safety margin for each analysed traffic types and showing the impact of finite measurement traces and finite buffer size, in Section 5. In Section 6, a number of examples are shown, how the proposed characterisation can be applied for different goals. Conclusions are given in Section 7.

2 LEAKY BUCKET ANALYSIS

2.1 The Leaky Bucket Curve

The points of the leaky bucket (LB) curve give the maximum queue length as a function of service rate (i.e. leak rate) in case of a G/D/1 queue fed by the source to be characterised (Figure 1). The two parameters of the leaky bucket are represented by the two axis. The points of the LB curve give the proper leak rate (r) and queue length (q) pairs. In other words, this graph shows how much bandwidth and buffer space are consumed by the source in different working points.

In case of a finite trace of bursty traffic, this is a monotonous decreasing curve, which starts from the total number of cells in the trace (n_T) and approaches zero at the peak cell rate (r_P) of the source (which is equal to the full link rate, r_L in case of unshaped traffic). Because of the bursty nature and periodicity of the traffic, the curve has several linear segments. As an approximation, the LB curve can be replaced by a set of linear sections which can be determined by their slope (s_i).

There are neighbouring sections, which slopes differ with one or more orders of magnitude, and others, which have almost equal slopes. The number of section groups (called regions) containing sections with similar slope is equal to the number of burst levels (K) in the traffic pattern. The last breaking point ($r_K - q_K$), related to the highest burst level is remarkable, since the slope following this point (s_K) is determined by the total number of cells in the trace, but this breaking point is independent of it.

In spite of the LB curve, the parameter triple of the traffic contract (r_P, r_S, n_S) gives only a line which assumes much higher resource requirement (see bold line in Figure 1).

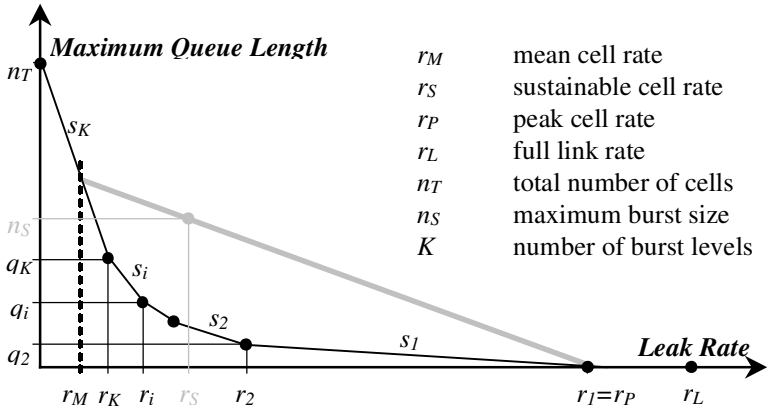

Figure 1 Leaky Bucket Curve of a finite, bursty traffic trace and operating line determined by the sustainable parameter set.

Ranging backwards from the peak rate to the mean rate, each section of the approximated curve can be described by the following equation:

$$q_i(r) = q_i + (r - r_i)s_i \qquad \text{if} \quad r_{i+1} < r \le r_i \ \text{ and } \ s_i = \frac{q_{i+1} - q_i}{r_{i+1} - r_i}. \tag{1}$$

The same equation using the leak time period ($t = 1/r$):

$$q_i(t) = q_i + \left(\frac{1}{t} - \frac{1}{t_i}\right)s_i \qquad \text{if} \quad t_i \le t < t_{i+1} \ \text{ and } \ s_i = \frac{t_i t_{i+1}\left(q_{i+1} - q_i\right)}{t_i - t_{i+1}}. \tag{2}$$

These equations can be used for finding the characteristic traffic parameters of the captured cell stream and calculate the slope of LB curve as a function of leak time period in order to analyse the burst structure. It is important to notice that

cell rates are normalised with respect to the full link rate r_L. Therefore $r_L = 1$ and $0 \le r \le 1$.

2.2 The proposed analysis method

The main steps of the Leaky Bucket Analysis are the following:

1. A sample trace should be captured from the investigated ATM source.
2. The points of the LB curve should be calculated by post-processing the captured cell stream and the result should be approximated by a set of linear sections.
3. The slope of each section should be calculated as a function of the leak time period using (2). That equation can be further simplified by a practical selection of section margins of leak time (t_i):

$$t_i = i \in \{1,2,3...K\} \quad \Rightarrow$$
$$s(t) = t(t+1)(q_{t+1} - q_t) = t(t+1)\Delta q_t \quad \Rightarrow \quad \Delta q_t \approx \frac{s(t)}{t^2} \tag{3}$$

   which equation expresses how sensitive is the buffer requirement of the source to the changes in the allocated rate.
4. The absolute value of the *s(t)* function should be plotted on a logarithmic scale, in order to highlight the characteristic sections of the LB curve. The value of t_κ can be easily read in the LB slope graph by finding the last jump of the curve. Thus r_κ can be calculated and the part of the LB curve depending on the number of captured cells can be determined.
5. Adequate *operating regions can be selected* for resource allocation and call admission control, or leak rates can be found for policing or shaping based on the LB and LB slope curves (see Section 6).
6. Other characteristic t and related Δq_t values can be determined by finding further jumps in the *s(t)* curve and applying (3). The related r_t–q_t pairs can be read from the LB graph and the burst size can be calculated using q_t (Cselényi, 1997).
7. A *quantitative description of burst structure* can be retrieved from the LB slope curve by comparison to characteristics of multilevel on-off sources. Also *source models can be established* based on the scanned burst parameters.
8. Since the LB curve gives the maximum queue length, i.e. the worst case behaviour of the analysed traffic, the maximum burst sizes can be read in the

LB slope graph. The mean burst size can be determined by plotting the mean queue length and its slope, in a similar way.

9. Apart from the exploration of burst structure of the traffic, the *buffer overflow probability* of single sources can be estimated using the three dimensional LB Analysis method (Cselényi, 1996a).

3 CHARACTERISATION OF ARTIFICIAL SOURCES

The proposed Leaky Bucket Analysis method is applied first to synthetic traffic patterns in order to get an analytical description. A simple, periodic on-off pattern is investigated first, since it was found in previous work (Cselényi, 1997) that deterministic processes producing an on-off pattern play an important role in modelling VBR video sources. Two- and multilevel on-off patterns are investigated by LBA as an extension and generalisation of the simple on-off case.

3.1 Simple on-off pattern

An on-off pattern consists of a burst at peak cell rate (r_P=1/T_P) with size n_M and duration T_B, followed by a silence period of duration T_S after which a new burst can be sent. The interarrival time of bursts is denoted by T_M, the total number of cells by n_T and the duration of traffic trace by T_T in Figure 2.

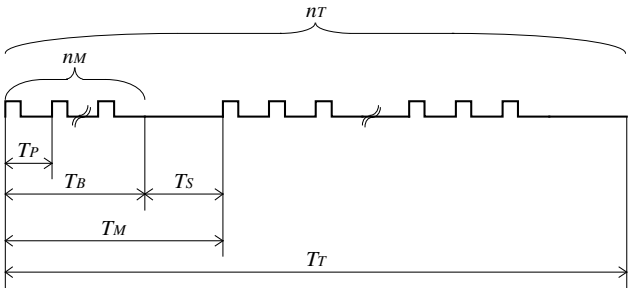

Figure 2 The traffic pattern of an on-off source.

There are three regions of the leak rate (r) where different expressions hold for the maximum queue length:

$$q_0(r) = 0 \qquad \text{if } r_P < r \le r_L . \tag{4a}$$

$$q_1(r) = r_M T_B - r T_B = n_M \left(1 - \frac{r}{r_P}\right) \qquad \text{if } r_M < r \le r_P . \tag{4b}$$

$$q_2(r) = r_M T_B\left(1-\frac{r}{r_P}\right)+\left(T_T - T_M\right)\left(r_M - r\right) = n_T - r\left(\frac{n_T - n_M}{r_M}+\frac{n_M}{r_P}\right) \text{ if } 0 < r \le r_M \text{ .(4c)}$$

These equations are visualised in Figure 3 by the LB and LB slope curves.

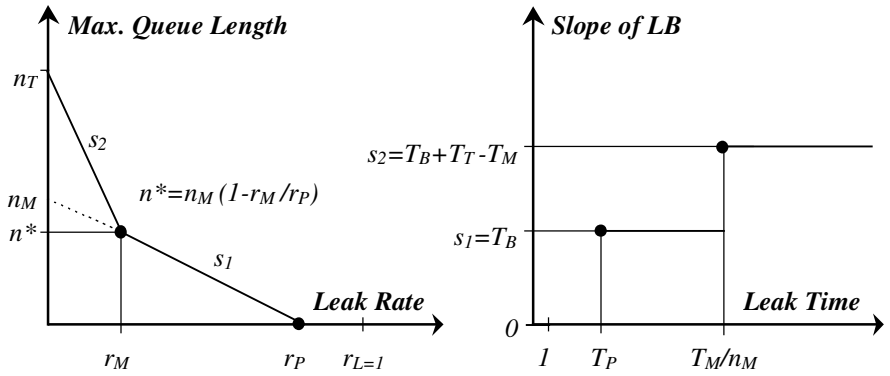

Figure 3 Leaky Bucket Curve and LB slope curve of a periodic on-off source.

Noticeable are the three plateau in the LB slope curve which correspond to the three regions of the LB curve ($r_P - r_L$, $r_M - r_P$and $0 - r_M$). The breaking points of the LB curve represent the peak and mean rate, while the jumps in the LB slope graph highlight the burst interarrival times. The region with a slope equal to zero represents the lowest "burst" level (i.e. cell level), where the leak time period is less than T_P. This segment of the LB curve does not exist in case of unshaped sources. One can read the burst size (n_M) on the next burst level by extending the first section (s_1) of the LB curve beyond r_M toward the q axis. The duration of on period (i.e. the first level burst, T_B) is indicated by the first plateau (s_1) in the LB slope graph. The junction point of first and second sections represent the mean cell rate, maximum burst size parameter pair. The slope of last section (s_2) is determined mainly by the duration of the traffic trace.

3.2 Two-level on-off pattern

As an extension of the previous case, one can consider a state machine with three states. The source generates n_2 cells at peak cell rate of duration T_{B2} in the first state, than waits T_{S2} cell time until generating the next burst. It repeats this procedure n_3 times and waits T_{S3} cell time before starting again. The resulting traffic pattern has consequently four burst levels i.e. the LB curve has four sections (see Figure 4). The buffer requirement has the following expression in the four leak rate regions:

$$q_0(r) = 0 \qquad \text{if} \quad r_1 < r \le r_L . \tag{5a}$$

$$q_1(r) = n_2\left(1 - \frac{r}{r_P}\right) \qquad \text{if} \quad r_2 < r \le r_1 . \tag{5b}$$

$$q_2(r) = n_3 n_2 - r\left(\frac{n_2\left(n_3 r_P - r_P + r_2\right)}{r_2 r_P}\right) \qquad \text{if} \quad r_3 < r \le r_2 . \tag{5c}$$

$$q_3(r) = n_T - r\left(\frac{n_T - n_3 n_2}{r_3} + \frac{\left(n_3 - 1\right)n_2}{r_2} + \frac{n_2}{r_P}\right) \qquad \text{if} \quad 0 < r \le r_3 . \tag{5d}$$

Similarly to the previous case, the LB slope curve expresses the duration of bursts on each burst level. The burst interarrival times can be also easily read in that graph.

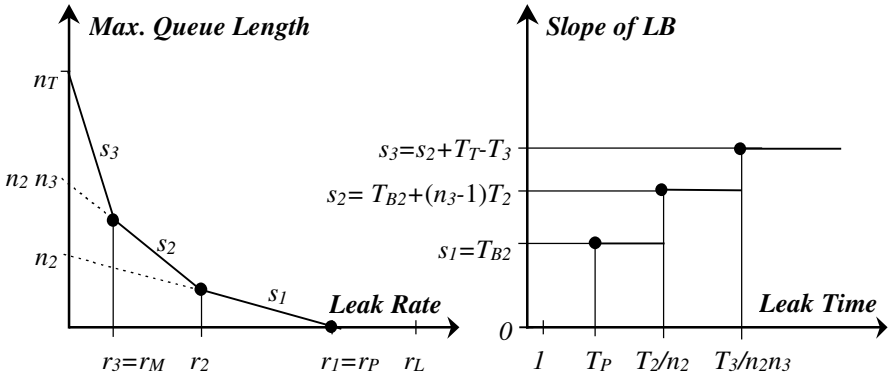

Figure 4 Leaky Bucket characteristics of a two-level on off pattern.

3.3 Multilevel on-off pattern

Further extending the number of burst levels, yields in a multilevel on-off pattern (Figure 5), of which LB curve consists of linear sections just as the approximated LB curve of bursty traffic drawn in Figure 1. The number of plateau in the LB slope curve corresponds to the number of burst levels (K). The first burst level is the cell-scale where one cell makes the "burst" (i.e. n_1=1) followed by a silent period of T_{s1} (which is zero in case of unshaped sources). An interesting way of modelling bursty sources is to consider the linear sections of its LB curve as an equivalent multilevel on-off source, calculate the burst parameters according to equation (5) and extend this model with information about the time domain behaviour of the source (e.g. sequence order of bursts) (Cselényi, 1997).

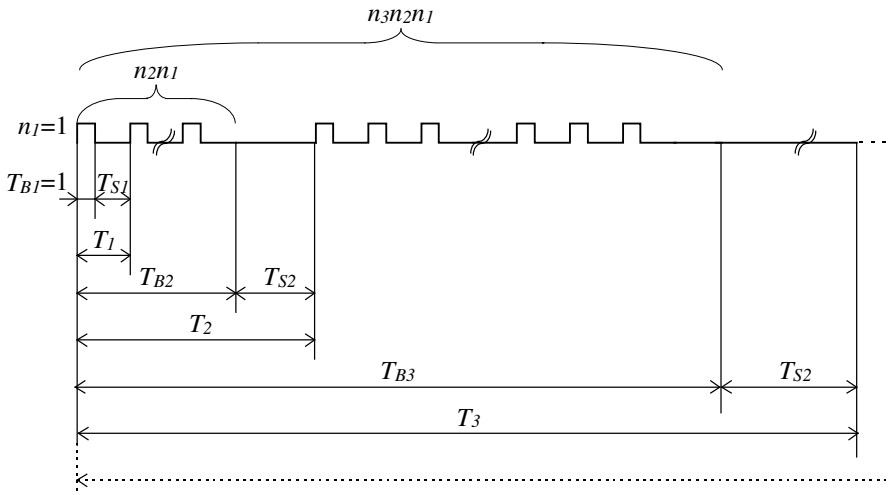

Figure 5 Multilevel on-off traffic pattern.

The traffic pattern and the main parameters of a multilevel on-off source are presented in Figure 5, while the mapping between the parameters of the presented on-off patterns and this general case are given in Table 1.

Table 1 Mapping the parameters of on-off traffic patterns

K	T_M	T_P	T_{B1}	n_1
2	T_2	T_1	1	1
3	T_3	T_1	1	1
K	T_K	T_1	1	1

4 CHARACTERISATION OF REAL ATM TRAFFIC

Several hundreds of traces from many different ATM traffic types were captured by measurements in Telia Research, Sweden during the last couple of years (Björkman, 1995, Molnár, 1996b, Cselényi, 1997). These cell streams were analysed by the proposed method. In order to verify the applicability, exemplary results are presented and evaluated giving an explanation of the results.

4.1 Multiplexed VBR traffic sources

Traffic of multimedia workstations connected via the Stockholm Gigabit Network (an ATM MAN) was multiplexed with CBR background traffic (Molnár, 1996b). Long traces of both traffic types were captured before and after multiplexing. LB curves of a shaped single multimedia source (a) and the aggregate traffic of four workstations (b) are depicted in Figure 7. The parameters and traditional traffic characteristics – such as mean cell rate and burstiness (i.e. squared coefficient of cell interarrival time variation) – of the traces are presented in Table 2.

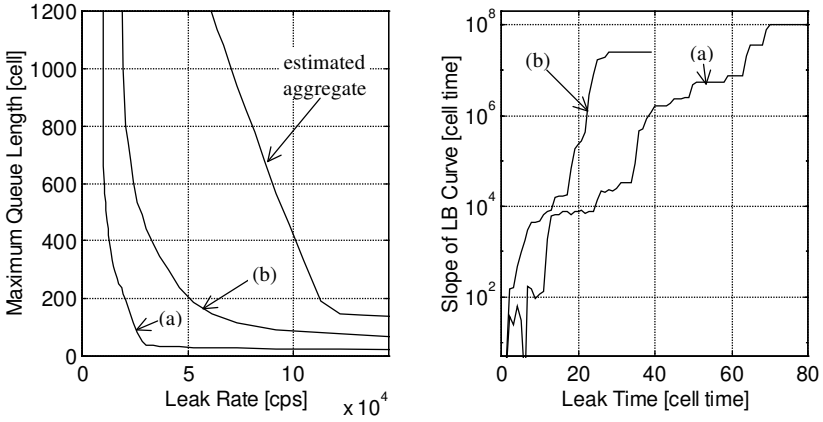

Figure 6 Multiplexing VBR video sources: (a) single VBR source before multiplexing (b) aggregate of four sources after multiplexing.

The four source were shaped to 34 Mbps each, in order to avoid congestion in the first switch. Naturally, this peak rate limitation does not hold for the aggregate traffic (b) as it is pronounced in the LB slope graph by missing 0^{th} level plateau. The applied shaping rate can be identified by the place of first jump in the LB slope curve of trace (a). There were several ATM switches and SDH add/drop multiplexers on the way of the traffic from the source to the measuring point. These active ATM and SDH devices destroyed the burst levels of the original VBR traffic (e.g. sequences c,d,e) by splitting and merging the bursts on different levels.

4.2 Multiplexed CBR and VBR traffic

VBR traffic produced by multimedia workstations and CBR traffic as a background load were fed into a FIFO multiplexer with no prioritising. Several traces were captured from the VBR traffic of the same input video sequence after

multiplexing. In spite of the increasing background CBR load the characteristics of traces (b-d) coincide very well in both graphs. Four plateau i.e. burst levels can be identified in the LB slope curve of each sequence and only one of them differs significantly, actually the third region in the leak time region of 18 – 27 cell time. An other noticeable difference is that as more the background CBR load increases as lower the slope curve starts. A possible explanation of this observation is that the heavy CBR stream spaced out the small, bursty VBR traffic. Since no cell loss occurred, the mean cell rate (r_M) remained unchanged, but the other burst levels (0^{th} and 3^{rd}) were changed.

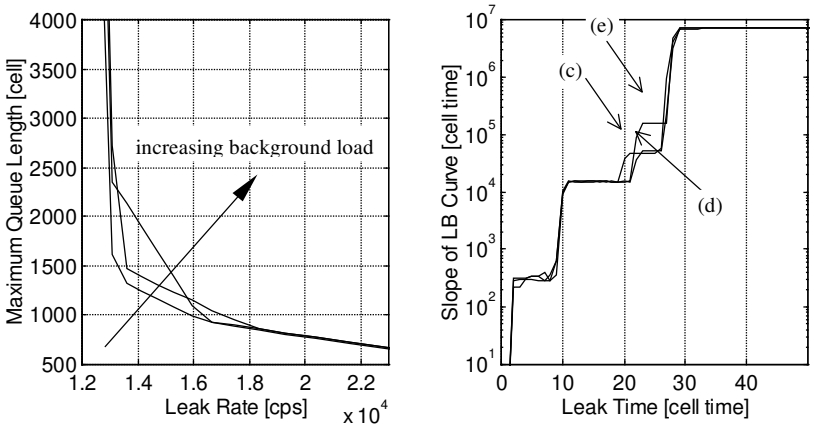

Figure 7 VBR video source multiplexed with CBR traffic in case of (c) @100 Mbps, (d) @120 Mbps and (e) @140 Mbps background load.

4.3 VBR video traffic

The effect of changing input video sequence were analysed by measuring the traffic of a TCP/IP over ATM based video conference application transferring several standardised CCIR video sequences which had equal duration but different burst structure corresponding to the picture content (Figure 8). The mean cell rate and the r_K - q_K point of sequence (f), which is an almost "still picture video", are far from the other sequences. However, the first two plateau of the LB slope curves coincide in case of each video sequence. A possible explanation is that the VBR video source has a very deterministic nature. The video application produces video frames on a given frame rate. These frames are sliced first into Maximum Transfer Units (MTU), than packed into IP packets and segmented into ATM cells by the protocol stack. This results in a multilevel burst structure, where only the size and timing of top-level burst (i.e. video frame) is

different. The number of captured cells (see Table 2) determines the s_κ values, as expected.

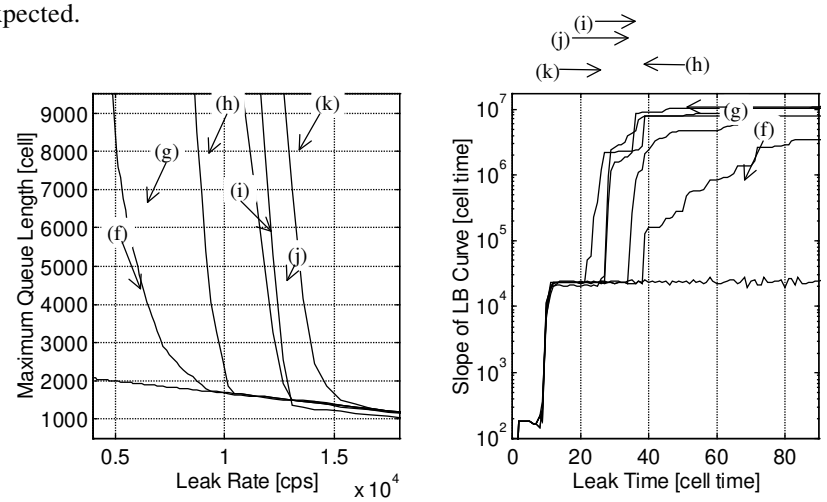

Figure 8 Leaky Bucket Curves of several video sequences: (f) Girl with Toys, (g) Sussie, (h) Table Tennis, (i) Tempest, (j) Flower Garden, (k) Popples.

The effect of changing the frame rate as performance parameter of the video application was also investigated. It can be seen from the result in Figure 9 that the LB curves are shifted along the x-axis but the slope values are equal. The slope curves show that the burst structure differs only on the highest level, i.e. video frame level, while the MTU and IP levels are equivalent.

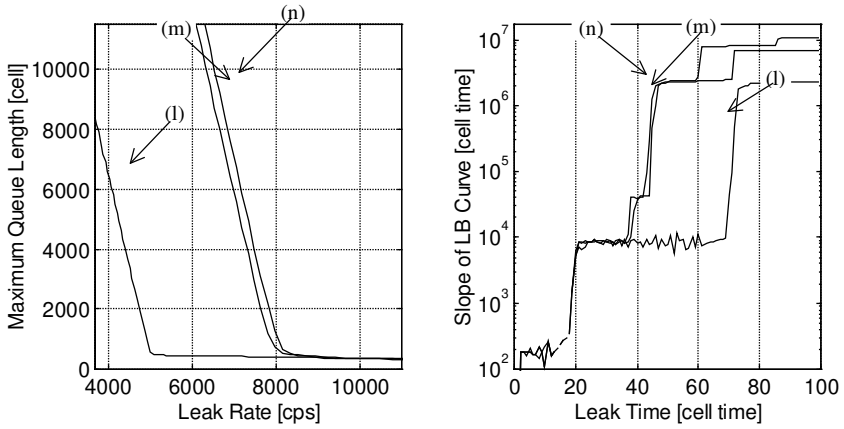

Figure 9 Effect of changing frame rate (l) @10 fps, (m) @20 fps, (n) @25 fps.

One group of media parameters has no impact on the network resources (e.g. brightness) while the other parameter group changes the amount of data to be transferred. The frame rate parameter belongs to the second group, as it can be seen in Figure 9. Although the amount of data is linearly proportional to the frame rate, the (m) and (n) curves are much closer to each other than to curve (l). A possible explanation to this phenomenon is that there is a finite upper limit of the performance of multimedia workstation, and it can not really support real-time pictures (25 fps). The LB slope curve of trace (n) support the hypothesis regarding the saturation of workstation performance.

Table 2 Traditional characteristics of VBR traffic sources

Trace	*Trace Length (s)*	*Total Number of Cells*	*Mean Cell Rate (cps)*	*Burstiness*	*Comments*
a	294.42	1 627 720	5500	140.3063	shaped
b	68.92	1 000 000	14 500	133.3298	aggregate
c	19.54	253 627	12 900	700.6972	bgr. 100 Mbps
d	20.38	264 258	12 900	693.0733	bgr. 120 Mbps
e	20.14	260 080	12 900	680.9803	bgr. 140 Mbps
f	51.00	100 309	1960	179.8622	10 fps
g	45.60	138 453	3030	258.6941	10 fps
h	36.86	151 544	4110	345.9633	10 fps
i	45.68	307 471	6720	483.3476	10 fps
j	38.27	250 142	6540	486.7423	10 fps
k	49.67	293 180	5890	452.2486	10 fps
l	35.73	87 353	2440	197.9074	10 fps
m	32.41	117 899	3640	245.1639	20 fps
n	45.29	186 966	4120	226.8030	25 fps

4.4 Internet traffic

Internet traffic is in the spotlight of many recent research activities. The traffic of fundamental TCP/IP based applications such as WWW Browsing, FTP and Ping and were analysed. The traditional traffic characteristics are given in Table 3.

During capturing traces (o,p) a large size image was downloaded several times with a WWW browser. The $r_\kappa - q_\kappa$ and $r_2 - q_2$ points of curve (o) are very close to

each other, which means that it has a very simple burst structure determined by only a few burst levels (Figure 10).

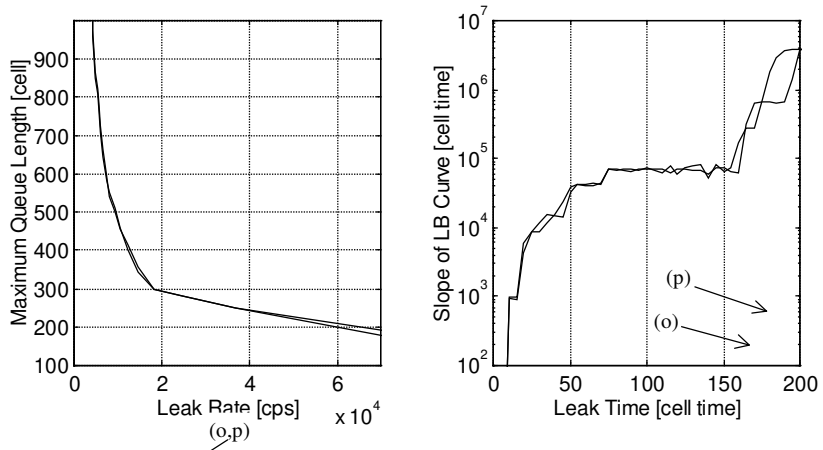

Figure 10 LB Curves of WWW browsing: (o) real and (p) emulated user.

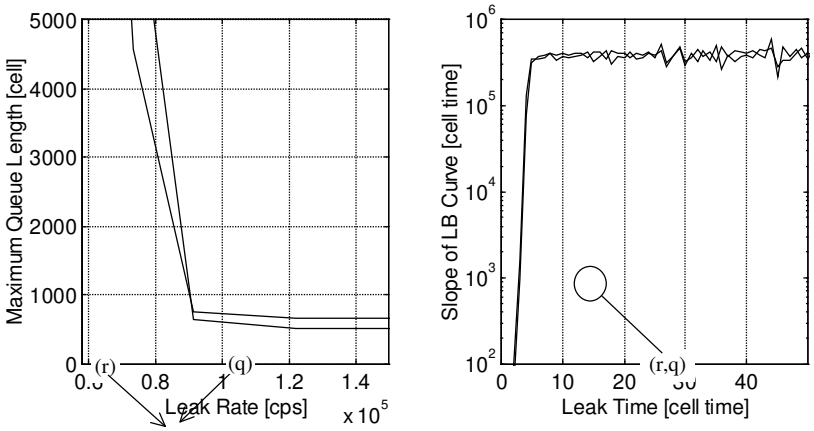

Figure 11 Leaky Bucket Curves of file transfer sessions.

A large size file was downloaded during the FTP session (Figure 11). The LB slope curves of these sources (q,r) indicates the very simple bust structure. The similar characteristic of FTP traffic for repeated measurements – what we experienced for much more trials (Molnár, 1996b) – is also remarkable.

Table 3 Traditional characteristics of Internet traffic sources

Trace	*Trace Length (s)*	*Total Number of Cells*	*Mean Cell Rate (cps)*	*Burstiness*	*Comments*
o	11.45	22 889	1900	142.7399	real user
p	12.20	22 888	1880	150.4459	SUE
q	1.51	81 117	6843	680.8446	
r	1.48	81 116	6718	759.9904	
s	30.37	230 732	7573	448.6227	128 cells
t	29.94	286 945	9583	444.6371	256 cells
u	29.98	233 723	7772	471.4246	512 cells
v	29.83	290 948	9726	458.3453	1024 cells

The ping UNIX command was used with four different sizes of transferred message (128, 256, 512 and 1024 cells). The value of s_1 and q_2 are proportional to the message size in case of these four sources (s-v). It is noticeable in Figure 12 that the LB slope curve of each Ping trace begins with a value which corresponds to the packet size setting of the trace, i.e. 164, 240, 475 and 1180 cell times for the (s-v) traces, respectively. Although these values does not exactly follow the 1:2:4:8 ratio, the error is due to the fact that LBA captures the worst case behaviour of the traffic (exactly as the switch).

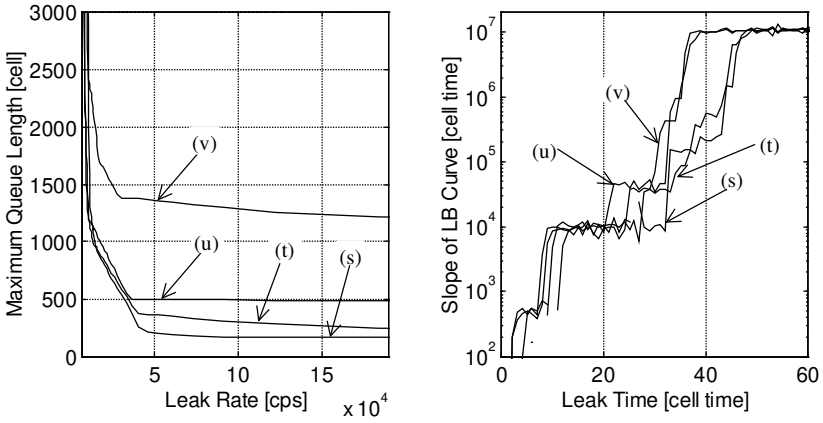

Figure 12 Leaky Bucket Curves of traffic from ping sessions. Message size: (s) @128, (t) @256, (u) @512, (v) @1024 cells.

4.5 Aggregate traffic

Internet traffic of interconnected LANs was also measured on the Swedish University Network (SUNET). Since these traces contain more than 83 million cells, these curves have large q_κ and large slope values (Figure 13). Although the traces were taken at different time, it is interesting to recognise the similarity among the curves from independent measurements in both graphs. The mean cell rate varied between 3000 and 9000 cps. The burst structure is very disperse, similarly to the other aggregate trace of four video sources (b) in Figure 7.

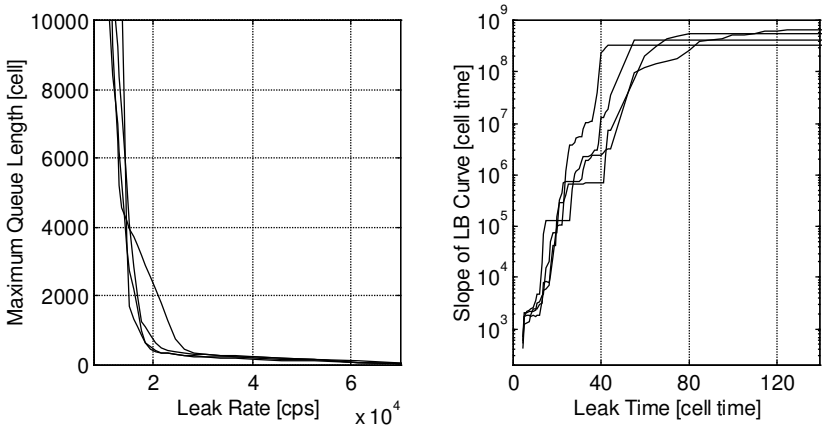

Figure 13 Leaky Bucket Curves of aggregate traffic measured on SUNET.

5 ROBUSTNESS OF THE LEAKY BUCKET ANALYSIS

An important feature of a practicable characterisation tool that it highlights the essential characteristics of the traffic and hides the unnecessary variations. Therefore an ideal tool gives exactly the same descriptors to each traces of a specific traffic, independently from the length of the trace and the time when it was captured. The impacts of these two factors on the result of the proposed Leaky Bucket Analysis method are presented in this section.

5.1 Repeatability

Many traces were captured from the same traffic type, using the same quality settings (e.g. video frame rate) in order to examine the repeatability of the method. The traffic types are different from the perspective of determinism of cell generation. The VBR video traffic is easily reproducible, since the input video sequence can be repeated and most of the processes transforming and transferring the data inside of the multimedia workstation are of deterministic nature (Cselényi, 1997). The only non-deterministic factor is the software application that gets the coded input video from the video card and sends the processed data unit to the protocol stack according to the scheduler of the operating system. That is why it can be assumed that traces taken from a VBR video traffic source have the same significant characteristics. This fact is noticeable in Figure 14 by showing the relative error ratio of LB curves calculated for traces of the same source from different trials:

$$E_{j,k}(t) = \frac{2\left|q_j(t) - q_k(t)\right|}{\mathrm{E}\left\{q_j(t) + q_k(t)\right\}} \qquad \text{for} \quad \forall\ j,k \in \{1,2,\ldots N\};\quad j \neq k. \tag{6}$$

Safety margins can be defined by giving accuracy thresholds and the leak rate (or leak time) regions where they are valid. The safety margins are indicated by bold lines and the maximum, minimum and mean error curves are drawn in Figures 14-17. The relative error has a peak between 20 and 30. This means about 40 ms i.e. the video frame rate set in the video application. As described above, the real frame rate is determined by the scheduling system and the upper limit of terminal's performance. This is a reasonable explanation for the degradation of repeatability on this track.

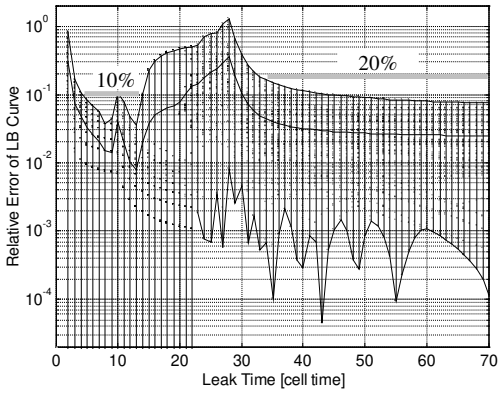

Figure 14 Repeatability of VBR video traffic.

In spite of VBR video traffic, one can only assume that the sources have similar behaviour in a statistical manner in case of Internet traffic. Therefore to have a robust statistical average, hundreds of traces should be taken from the same traffic. Repeating the measurement is quite easy in case of simple "Internet sessions", like Ping and FTP. According to Figure 15, the behaviour of these sources is very deterministic. The only jump of relative error can be seen around the r_K breaking point. Repeating more complex sessions, like WWW browsing is more problematic. Thus the examined WWW session was repeated using a software tool called Service User Emulator (Cselényi, 1996b), which can record and play back the interactions of user and this way it generates repeatable traffic. The emulated user could repeat the WWW session quite well, as it can be seen in Figure 10.

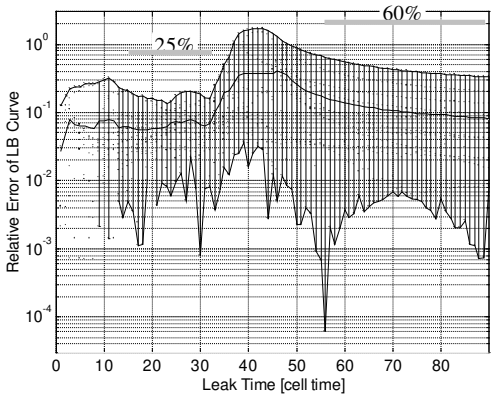

Figure 15 Repeatability of ping sessions.

The safety region of acceptable error threshold (50%) is much shorter for aggregated Internet traffic, than for simple sources. The main reason is the much higher number of independent processes which are generating or influencing the traffic. Considering the other safety margin, it is interesting that any bandwidth can be allocated if the buffer requirement is doubled.

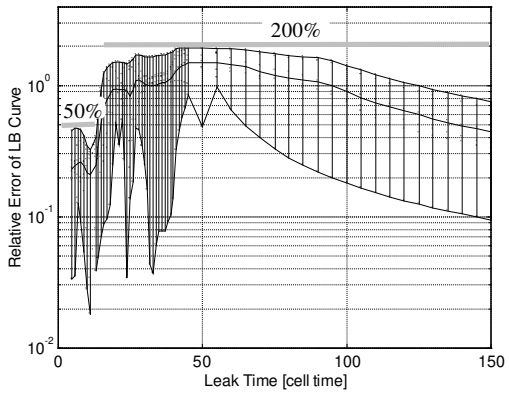

Figure 16 Repeatability of aggregate Internet traffic.

5.2 Effect of trace length

One trace of measured video traffic, captured after multiplexing with CBR background traffic, were sliced into two and four parts in order to investigate the effect of total cell number on the shape of the LB curve. The LB slope curve of these traces are depicted on a linear scale in Figure 17.

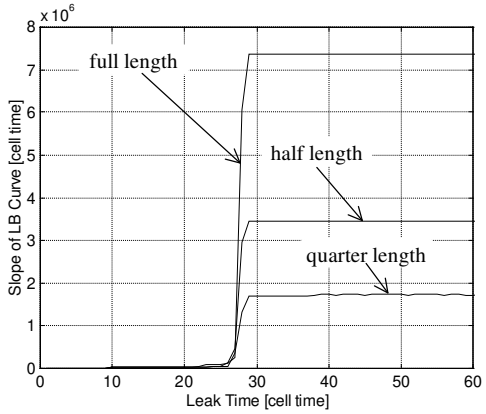

Figure 17 Effect of the number of cells in the measured trace.

The results confirm the theoretical assumptions presented in Section 2. The value of last slope (s_K) of each traces are proportional to the number of cells in the trace (Table 2). However the LB curves are almost identical for the $r_K < r$ (or $t_K > t$) part, supporting the proposed steps of our method.

5.3 Cell loss probability estimation

The Cell Loss Probability can be estimated by the overflow probability calculated by post-processing of measurement traces. In this way a set of LB curves can be retrieved corresponding to different loss probabilities (see Figure 18). Instead of the original zero loss curve, the adequate LB curve can be applied for the analysis and resource allocation task. More details can be found in (Cselényi, 1996a).

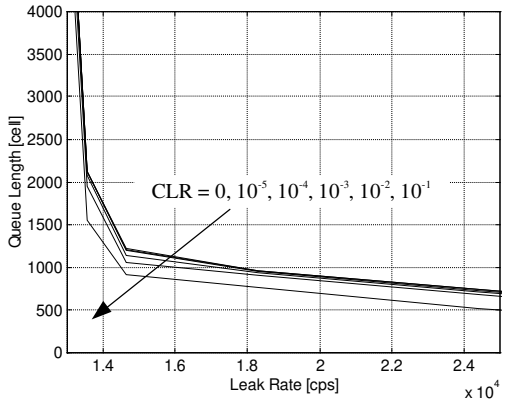

Figure 18 LB type curves of trace (b) parameterised by estimated CLR.

6 APPLICATION OF THE LEAKY BUCKET ANALYSIS

The aim of this section is to give examples how the proposed leaky bucket algorithm can be applied in practice. Based on the results presented in Section 4, different characterisation goals are achieved. Other applications, such as determining shaping rate, retrieving parameters for VBR source model and network dimensioning, were described in previous works (Björkman, 1995, Molnár, 1996a, Latour-Henner, 1997, Cselényi, 1997).

6.1 Burst structure analysis

One possible application of the LBA is to detect the burst structure of the analysed traffic source. The Leaky Bucket curve and the LB slope graph of real traffic sources are very similar to that of the two-level on-off traffic drawn in Figure 4, hence it seems to be logical to apply the equations (5a-d) for determining the parameters of burst structure (i.e. duration and interarrival time of bursts at each level). The parameters read in the graph are written in the left column while the calculated parameters in the right:

$$T_{B2} \approx 300 \quad \Rightarrow \quad n_2 \leq 300 \tag{6a}$$

The value of n_2 can be more precisely read in the LB graph and calculated using (5b). With that value:

$$\frac{T_2}{n_2} = 10 \text{ , } n_2 \approx 180 \quad \Rightarrow \quad T_2 = 1800 \tag{6b}$$

$$T_{B2} + (n_3 - 1)T_2 \approx 15800 \quad \Rightarrow \quad n_3 \approx 9.7 \tag{6c}$$

$$20 \leq \frac{T_3}{n_2 n_3} \leq 21 \text{ , } n_2, n_3, \quad \Rightarrow \quad 35000 \leq T_3 \leq 37000 \tag{6d}$$

Therefore there are active periods of duration 300 cell time followed by silent period of 2700 cell time on the first burst level, according to observations (6a) and (6b). The size of this burst (n_2) is determined by the Maximum Transfer Unit of the video application (app. 8200 bytes). On the next level, about nine of first level bursts make a longer burst of duration 15800 (6c). This is the largest video frame in the trace. The range of interarrival time of these frames given by (6d) corresponds to the frame rate setting in the application, i.e. 10 fps. These observations support results of previous work (Molnár, 1996b).

6.2 Giving hints for the sustainable parameter set

Assuming that the sustainable parameter set should be determined for a VBR video source like (c-e), first a region is to be selected which is indifferent for changes in the environment (i.e. background load in this case). In the LB slope graph of Figure 6, the region related to leak times between 11 and 18 cell time seems to be adequate. After checking the safety range in Figure 14, this interval can be further reduced to $t_s = 12$, i.e. the corresponding sustainable cell rate is $t_s = 30566$ cps. The maximum burst size q_s and the peak cell rate r_p can be read in the LB graph. However, the former should be increased according to the safety

threshold (25% for this t_s) for a robust decision. Therefore r_p= 366792 cps and q_s= 439 cell.

6.3 Analysing multiplexing performance

LBA is adequate for analysing the effect and performance of active network elements, such as mutliplexer and shaper. Figure 6 presents the LB characteristics of a video stream before multiplexing (a) and of the aggregate traffic of four multiplexed video streams (b). The third curve illustrates the theoretical maximum resource requirement of the four sources, when they would consume the most bandwidth and buffer space. It is visible, that statistical multiplexing performed quite well, since the LB curve of real aggregate traffic (b) is much lower than for the estimated worst case. Naturally these observations can be quantified based on the LB graphs as well.

7 CONCLUSIONS

In this paper, we introduced a simple and practical traffic characterisation method called Leaky Bucket Analysis. The procedure of LBA starts with capturing a trace from the examined traffic and calculating the maximum buffer requirement as a function of LB leak rate. Approximation of LB curve by a set of linear sections and plotting the slope of these sections as a function of leak time are proposed for highlighting the burst structure of the source and allow quantitative analysis. The main advantage of this method is that it gives the *resource requirement* and describes the *burst structure* of the source *on each time-scale* in one graph, while the standardised source descriptors and usual burstiness measures are time scale dependent. Moreover, the Leaky Bucket Analysis method provides a quick visual impression about the traffic structure.

The method has been described and *demonstrated on 30 traces taken from actual traffic measurements* of VBR video, Internet, ATM LAN and MAN traffic. Both single sources and traces from aggregate traffic (e.g. after multiplexing) has been analysed. The results show that besides the basic traffic characterisation parameters additional information can be gained about the burst structure and periodicity of ATM traffic. The *robustness of LBA is analysed* and safety margin, error threshold pairs are defined for presented traffic types. *Several application area* of LBA, such as burst analysis, selection of sustainable parameter set, analysis of multiplexing performance and dimensioning ATM networks are also presented in the paper.

Because of its simplicity, it would be possible to implement the LBA method as an integral part of ATM network elements. The practical applicability of the method for network dimensioning, and CAC design are in the focus of our recent and future research activities.

8 ACKNOWLEDGEMENTS

The authors wish to thank Alexander Latour-Henner, Urban Hansson, Olle Pers, Aziz Miah, and Joakim Berkvist from Telia Research AB for their contributions to this paper and help during the measurements.

REFERENCES

Björkman, N., Latour-Henner, A., Hansson, U., Pers, O., Miah, A. (1995) Practical ATM resource dimensioning based on real-time traffic measurements and analysis. *Proc. Globecom '95*, Singapore, 399-403.

Cselényi, I. (1996) On measurements and analysis of VBR video traffic. *Proc. High Speed Networks Laboratory Workshop*, Budapest

Cselényi, I., Björkman, N., Baumann, F., Haraszti, P., Henk, T., Latour-Henner, A. and Miah, A. (1996) Service User Emulator for generating a repeatable network load. *Proc. IASTED Conference on Networks*, Orlando, Florida, 229-32.

Cselényi, I., Molnár, S., Björkman, N. and Latour-Henner, A. (1997) Hierarchical VBR video source model. *Proc. 5th International Conference on Telecommunication Systems Modeling and Analysis*, Nashville, 401-10

Elwalid, A., Mitra, D. and Wentworth, R.H. (1995) A new approach for allocating buffers and bandwidth to heterogeneous, regulated traffic in an ATM node. *IEEE Journal on Selected Areas of Communications*, **13** (6), 1115-27.

Frost, V.S. and Melamed, B. (1994) Traffic modeling for telecommunications networks. *IEEE Communications Magazine*, **3**, 70-81.

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

Latour-Henner, A., Björkman, N., Cselényi, I. (1997) ATM traffic shaping at the multimedia terminal. *Proc. International Conference On Telecommunications*, Melbourne, **3**, 1193-8

Molnár, S., Cselényi, I. and Björkman, N. (1996) ATM traffic characterisation and modeling based on the Leaky Bucket Algorithm. *Proc. IEEE Singapore International Conference on Communication Systems*, Singapore, **1**, 185-9

Molnár, S., Cselényi, I., Björkman, N., Perntz, C.G. and Boda, M. (1996) ATM traffic measurements and analysis on a real testbed, *Proc. 10th International Teletraffic Congress Specialist Seminar*, Lund, Sweden, 237-50.

Roberts, J., Mocci, U. and Virtamo, J. (Eds.) (1996) COST 242 -Performance Evaluation and Design of Broadband Multiservice Networks. Final Report. Springer-Verlag, Berlin.

Stamoulis, G.D., Anagnostou, M.E. and Georgantas, A.D. (1994) Traffic source models for ATM networks: A survey. *Computer Communications*, **17** (6), 428-38.

BIOGRAPHIES

István Cselényi

István Cselényi received his M.Sc. degree in electrical engineering from the Technical University of Budapest, Hungary in 1994. He started Ph.D. studies in the Department of Telecommunications and Telematics, Technical University of Budapest in 1994. He was board member and application co-ordinator in the High Speed Networks Laboratory and participated in research activities in the area of broadband multimedia services and ATM traffic measurements, analysis and modelling. Since 1997 he has been working for Telia Research. His current research interests include resource management for distributed multimedia services based on IP and ATM.

Nils Björkman

Nils Björkman received his M.Sc. degree in engineering physics from the Royal Institute of Technology, Stockholm in 1985. The same year he joined Swedish Telecom to work with the introduction of emerging technologies for fibre optical transmission systems, SDH and ATM. Since 1992, he has been with Telia Research where his task has been to establish the ATM laboratory and to lead the various activities in the laboratory - currently spanning from ATM traffic control to secure and chargeable high performance IP services. He has published original research on various experimental investigations on ATM networks, including both objective measurements and subjective Quality of Service assessments.

Sándor Molnár

Sándor Molnár received his M.Sc. and Ph.D. in electrical engineering from the Technical University of Budapest, Budapest, Hungary, in 1991 and 1996, respectively. He is an Assistant Professor at the Department of Telecommunications and Telematics, Technical University of Budapest and the research leader of the teletraffic research program of the High Speed Networks Laboratory. Dr. Molnár has been participated in the European project COST 242 and now he is engaged in project COST 257 on “Impacts of New Services on the Architecture and Performance of Broadband Networks”. His main interests include teletraffic analysis and performance evaluation of modern communication networks with special interest in B-ISDN.