ATM TRAFFIC CHARACTERIZATION AND MODELING BASED ON THE LEAKY BUCKET ALGORITHM

SÁNDOR MOLNÁR[a], ISTVÁN CSELÉNYI[a], NILS BJÖRKMAN[b]

[a]High Speed Networks Laboratory, Department of Telecommunications and Telematics, Technical University of Budapest, Sztoczek u. 2, H-1111 Budapest, Hungary, e-mail: {molnar,cselenyi}@ttt-atm.ttt.bme.hu

[b]Department of System Research, Telia Research AB
Rudsjöterrassen 2, S-13680 Haninge, Sweden, e-mail: nils.p.bjorkman@telia.se

ABSTRACT In this paper, we further extend the investigation of the simple traffic characterization method based on the leaky bucket algorithm proposed in [1]. This algorithm can provide much more traffic information than has previously been examined. We focus on curves of the *maximum queue length* (bucket content) vs. the *service rate* (leak rate) to characterize traffic and discuss what useful information can be extracted from them.

We measured and analyzed real ATM traffic generated by multimedia terminals. The impact of different performance settings, protocol parameters on the characterizing curves are examined and demonstrated with examples. In addition to traffic characterization, we propose a *simple model for measured ATM traffic* based on these curves. We discuss a method of fitting the maximum queue length vs. service rate curves with curves generated from our analytical model which is a two level fluid model.

I. INTRODUCTION

The characterization of real traffic is a critical issue to the success of efficient traffic engineering in ATM networks. Research in this field has resulted in numerous models and techniques over the last decade [2, 3]. However, the developed traffic characterization methods have been, in general, rather complicated and demand intensive computation of several statistical parameters. There is a lack of simple and accurate methods that can be of practical use to network operators.

In this paper we discuss the characterization and modeling of ATM traffic generated by video conference applications. This type of video traffic is bursty VBR traffic exhibiting pronounced periodic and correlated structures in different time scales [4]. We propose a characterization framework based on the leaky bucket method and investigate the correspondence between different cell rates and related tolerance parameters through this method. Our investigated characteristics are the maximum queue length and the related service rate in the corresponding queueing model. The impacts of different performance settings, protocol stacks and background traffic on this characterization are also investigated.

Moreover, we suggest a way to set up traffic models to capture directly these characteristics. Therefore our modeling approach is different compared to the classical methods. In general traffic modeling techniques first try to characterize the cell arrival process correctly and set up a model for this process. Second, this cell arrival process model is fed to a queue and by the help of queueing theory the required queueing characteristics are computed. In this paper we avoid to use these steps and try directly to capture the important queueing behavior by applying a simple two level fluid flow model.

II. LEAKY BUCKET TRAFFIC CHARACTERIZATION

A. IDEA

The leaky bucket algorithm is widely applied in ATM context, moreover, the standardized cell conformance method (Generic Cell Rate Algorithm, GCRA [5]) is also based on this idea. However, the leaky bucket algorithm can provide much more traffic information than has previously been examined. Our purpose is in this paper to show the powerful applicability of the leaky bucket method for efficient traffic characterization and modeling.

Different names and definitions are used to describe the idea depending on the context. In this paper we use the *queueing theory terminology* to describe the leaky bucket method. Thus, we handle the leaky bucket as a G/D/1 queueing model. (Correspondence between GCRA formalism and the G/D/1 queue are discussed e.g. in [6].) We are focusing on the curves of the *maximum queue length* (x) vs. the *service rate* (c) to characterize traffic as well as having them as fundamental information for traffic modeling.

The leaky bucket algorithm can be processed for a measured cell stream. In our case the analysis means that the recorded cell streams are fed to the G/D/1 queue and the maximum queue length and the related service rate values are plotted as shown in Fig. 1.

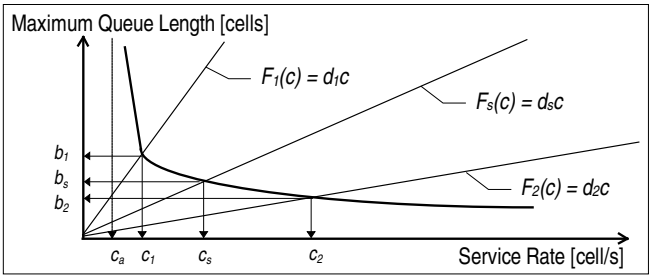

Figure 1. The Maximum Queue Length vs. Service Rate Graph

B. PROPERTIES

According to our experiments, the maximum queue length curve can be divided into three main parts in case of video traffic: a near linear part with steep slope, a knee and an other linear part with small slope. These parts are used for description of changes of the graph in the following sections. The curve can directly provide us lots of practical information. For example, from the curve we can *select Peak Cell Rate (PCR) and Cell Delay Variation (CDV) Tolerance pairs* as well as *Sustainable Cell Rate (SCR) and Burst Tolerance (BT) pairs*, we can *dimension the buffer* of a shaper or of a switch and also *determine appropriate shaping rate* as discussed below.

There are several important characteristics indicated on the maximum queue length vs. service rate graph in Fig. 1:

- c_a – the asymptote which represents the average cell rate
- b_1– indicates a feasible buffer size, which value is governed by the shaper or switch implementation
- c_2 – the service cell rate which represents the maximum shaping rate. It is determined by the required bandwidth utilization. Since the corresponding maximum queue length (b_2) is low, a low delay (d_2) belongs to this working point
- d_s – the maximum of acceptable shaping or buffering delay. This value is determined by subjective evaluation of each application

The *tolerance vs. service rate graph* can be obtained by dividing the maximum queue length values in the graph in Fig. 1 by the corresponding service rates. In this way we get a similar, new graph which represents valid tolerance–rate pair combinations for the Usage Parameter Control (UPC). From these pairs the PCR with the CDV Tolerance and the SCR with the BT can be selected. The proper setting of these parameters are necessary for appropriate resource allocation, but the way how to chose them is far from being straightforward [6, 10]. The buffer size can be dimensioned based on delay constraints (d_s) and upper buffer limit (b_1). The selected maximum queue length value can be interpreted as a minimum buffer size needed for the shaper or if there is no shaping the minimum buffer capacity needed in the network for the given source. The actual shaping rate (c_s) can be selected considering the d_s, c_2 and b_1 constraints. Therefore, the maximum queue length vs. service rate graph represents fundamental traffic characteristics of the measured source which can be used for resource dimensioning in an ATM network, i.e. bandwidth and buffer size allocation.

The advantages of this characterization can be summarized as follows:

- *simple* obtainable characteristics
- the characteristics *directly provides the required information for the UPC*, i.e. the PCR and CDV Tolerance as well as SCR and BT parameters
- the characterization *directly applicable for buffer dimensioning* as described above
- the characterization *directly applicable for determine optimum shaping rate* as described above
- traffic modeling based on fitting directly this curve has the advantage that we avoid the classical way of modeling steps (statistical analysis of arrival process - modeling of arrival process - solving the related queueing problem) but rather we are focusing on *directly to capture the queueing behavior.*

III. ATM MEASUREMENTS

A. MEASUREMENT SET-UP

The measurement set-up consisted of multimedia workstations, an ATM switch and a measurement instrument for source characterization. Each workstation was equipped with video, audio hardware and ATM interface boards. Since a repeatable input video source was required, a video tape recorder was connected to each terminal. ShowMe video conference applications were used through PVC connections among the ATM hosts [7]. The optical signal was tapped to the ATM test instrument by means of an optical splitter. In this way we captured an exact copy of the cell flow between terminals without effecting the behavior of the application in use.

The arrival time of ATM cells of interest was recorded in real-time by a module developed by Telia AB, Sweden. It resides in an ATM test instrument developed in the RACE PARASOL project. This instrument is capable of recording about 8 million cell arrivals [8]. The traffic records were analyzed with a Telia proprietary software program.

B. MEASURED DATA

Four main types of sequences were selected for the measurements – an almost still picture sequence with very low traffic intensity (Girl with Toys), a head and shoulder sequence with no camera movements and medium traffic intensity (Susie), a sequence including camera zoom and pan events (Popple) i.e. having a high and strongly varying traffic intensity and finally an "encoded noise" video sequence (Noise) producing maximum traffic intensity, since the video information is uncorrelated and thus very low amount of compression gain can be achieved. The first two correspond to a calm video conference situation while the third and fourth can be considered as an extreme case of video traffic. The average rates of the sequences are 2600, 4450, 9350, 12926 cell/s, respectively, in case of high

resolution (768x576 pixels) and 10 frames per second performance setting.

IV. Analysis Of Measured Traffic

A. Performance Setting Dependence

The effects of altering the main parameters of video performance on the multimedia terminal (e.g. frame rate and picture resolution) were tested. The change of maximum queue length vs. service rate graph in case of medium resolution (384x288 pixels) and three frame rate values (10, 20 and 25 fps) for the three sequences is presented in Fig. 2. The following phenomena can be observed in this Figure:

- The steep part of the graphs (and the average rate – c_a) is shifted towards the higher cell rates by increasing the frame rate.
- The flat, linear part is not significantly changed, and it represents the same maximum queue length values for each sequence.

B. Protocol Dependence

In order to check the effect of changing the protocol stack, the default setting (AAL 5, LLC) was changed and cases of AAL 3/4 and null encapsulation were investigated. The three maximum queue length vs. service rate graphs are depicted in case of the Popple sequence with high resolution and 10 fps in Fig. 3. Observations:

- AAL 3/4 requires slightly higher (8-10 %) cell rate to be allocated than AAL 5.
- Encapsulation requires also slightly higher (5-7%) cell rate to be allocated than null encapsulation.
- The maximum queue length graphs are mostly just shifted along the service rate axes, similarly to the previous trials where the video performance was changed. The explanation is that both the increasing of protocol overhead and increasing of frame rate induces more cells generated by the source.

C. Dependence On Background Traffic

Snap shot and white board are frequently used together with video conference application. Since the white board tool generates a traffic of very low intensity, snap shot was used as a background traffic source for the Popple sequence. The performance setting was high resolution and 10 frames per second. A video snap shot produces in the order of 10 000 cells. The cells are transmitted at a rate of approximately 15 000 cells/s. The results in Fig. 4. shows that the flat and steep part of the maximum queue length vs. service rate graph remained almost unchanged, while the knee was different in each case. Also the effect of snap shot on the curve is practically negligible.

V. Traffic Modeling

A. Two Level Fluid Model

Various video models have been investigated in the last decade ranging from the popular ARMA (AutoRegressive Moving Average) models to the different Markov type models, e.g. the family of MMPPs (Markov Modulated Poisson Process) [2, 3]. Our intention was to find the simplest model which can capture the basic characteristics of video traffic. Video traffic is generated by encoding subsequent video frames composed by several packets. The data rate is in the order of several megabits per second, while the packet length is less than 100 kilobits (in our situation a packet consisted of 180 cells). Thus, we can move from cell scale to larger time scales and the discrete nature of data can be ignored. So we chose the fluid flow model for our modeling purposes where a source generates continuous bit streams at different data rates. Moreover, as our first model, the simplest *two level fluid flow model* has been selected.

In this model the fluid arrival rate at time t is denoted by $\Lambda(t)$ has the form $\Lambda(t)=r(Z(t))$ where $Z(t)$ is a Markov process and $r(z)$ is a deterministic function. Now assume that the Markov process $Z(t)$ has two states with generation rate r_0 and r_1, and $r_0 < c < r_1$ where c is the service rate. The transition intensities of $Z(t)$ are denoted by λ (from state 0 to state 1) and μ (from state 1 to state 0). By solving the Chapman-Kolmogorov equations for the partial distribution functions finally the complementary queue length distribution can be obtained [9]:

$$Q(x) = \frac{\lambda}{\lambda+\mu}\frac{r_1 - r_0}{c - r_0}e^{-\zeta x} \quad \text{with} \quad \zeta = \frac{(c - r_0)\mu - (r_1 - c)\lambda}{(r_1 - c)(c - r_0)}$$

B. Parameter Fitting

There are four parameters to set in the model: the two *transition rates* (λ and μ) and the two *generation rates* (r_0 and r_1). Our purpose to set these parameters such that the model can capture the curve on Fig. 1. A problem arises from the nature of stochastic models, namely there is no point to speak about maximum queue length because there is a non zero probability of having any queue length in the model. Thus, here we relate the curve corresponding to a specified quantile of the queue length (the queue length which is exceeded by a specified probability) in the model to the maximum queue length curve of the leaky bucket analysis of the recorded data.

We set the following fitting rules:

1. the *average arrival rate* (m) should be fitted (this is the asymptote of the maximum queue length vs. service rate curve)
2. the *sustainable cell rate* (s) should be fitted with the *burst tolerance* or the related queue length value (b) at a given *cell loss probability* (q)
3. the r_0 and r_1 values can be tuned such that we obtain the best fitting

This parameters can be directly read or determined from the curve (Fig. 1) of the recorded data with a specified cell loss requirements. Thus, we have the following fitting equations where the cell loss is approximated by the overflow probability of the queue:

$$m = r_0 + (r_1 - r_0)\frac{\lambda}{\lambda+\mu} \quad \text{and} \quad q = \frac{\lambda}{\lambda+\mu}\frac{r_1 - r_0}{s - r_0}e^{-\zeta b}$$

Through the calculation of parameter ζ we obtain the transition rates:

$$\zeta = -\frac{1}{b}\ln\left(q\frac{s - r_0}{m - r_0}\right)$$

$$\lambda = \zeta\frac{(r_1 - s)(s - r_0)}{(s - r_0)\left(\frac{r_1 - m}{m - r_0}\right) - (r_1 - s)} \quad ; \qquad \mu = \frac{r_1 - m}{m - r_0}\lambda$$

There are two parameters, r_0 and r_1 to be tuned in order to get the best curve fitting, but we might also set them as r_0 to be zero and r_1 to be the peak cell rate.

The above model has the advantage that it is a rather simple model with very simple fitting procedure. It also has the nice feature that it directly captures the important queueing behavior instead of only the statistics of the arrival process coupled with a nontrivial queueing problem.

VI. Verification Of The Model

The Susie and Noise video traces were investigated for modeling purposes in case of high resolution, 10 fps performance setting (Fig. 5, 6). After performing the leaky bucket analysis of the measured video traces to get the maximum queue length curves and choosing the service rate related to the queue length b=500 as sustainable cell rate with q=10^{-5} cell loss requirement, the following model parameter settings were obtained:

	m [cps]	s [cps]	λ	μ	r_0 [cps]	r_1 [cps]
Susie	4 450	50 252	9.74	269.44	2 224	66 022
Noise	12 926	26 863	3.40	218.80	12 500	40 347

Table 1. Model Parameters

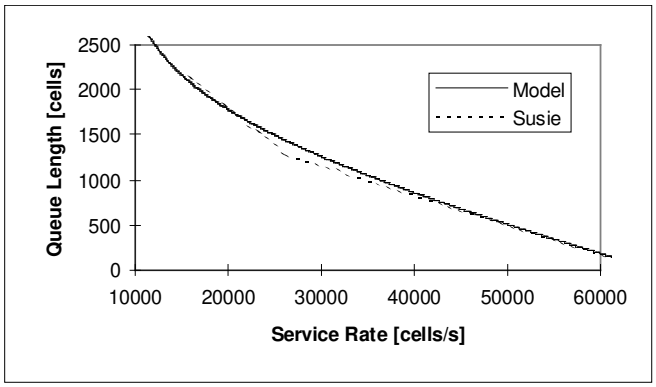

Figure 5. Curve Fitting Results For The Susie Sequence

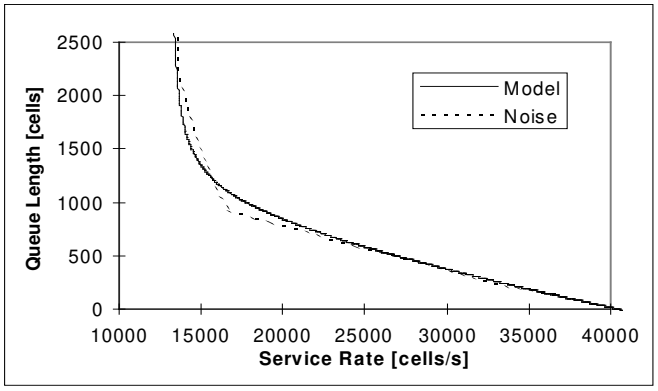

Figure 6. Curve Fitting Results For The Noise Sequence

In Figure 5 and 6 the maximum queue length vs. service rate curve of the measured data (Susie and Noise) and the queue length (related to the 10^{-5} quantile) vs. service rate of the model are plotted. We can conclude that the model can accurately capture the curves and can be used as source traffic model of the investigated video traffic.

VII. Conclusion

A traffic characterization framework based on the leaky bucket method using the maximum queue length vs. service rate curves is presented. We discussed that numerous useful information can be obtained from this simple characterization which directly supports PCR and CDV Tolerance or SCR and BT selections, buffer dimensioning and determination of appropriate shaping rates. Our intention is to use these curves for resource dimensioning to achieve efficient statistical multiplexing.

The method is demonstrated using measured ATM traffic and the impacts of different performance settings, protocol stacks and background traffic on these curves are presented. The impact of video traffic intensity and the performance settings of the terminal found to be significant.

We also introduced a new method for fitting these curves to an analytical model which is a simple two level fluid flow model with its verification results. Our basic point is in the modeling to directly characterize the important queueing behavior instead of describing the cell arrival process with a complicated queueing problem.

Acknowledgment

We would like to thank Alexander Latour-Henner, Urban Hansson, Aziz Miah, and Ola Pers from Telia Research for the help during our measurements and analysis. Thanks are also due to the Board of Hungarian State Eötvös Found.

References

[1] N. Björkman, A. Latour-Henner, U. Hansson, O. Pers, A. Miah, "Practical ATM Resource Dimensioning Based on Real-Time Traffic Measurements and Analysis", *GLOBECOM*, Vol 1, 1995, pp. 399-403.

[2] G. D. Stamoulis, M. E. Anagnostou, A. D. Georgantas, "Traffic Source Models For ATM Networks: a Survey",

Computer Communications, Vol. 17, No 6, June 1994, pp. 428-438.

[3] V. S. Frost, B. Melamed, "Traffic Modeling For Telecommunications Networks", *IEEE Communications Magazine*, March 1994, pp. 70-81.

[4] S. Molnár, I. Cselényi, N. Björkman, C-G. Perntz, M. Boda, "ATM Traffic Measurements and Analysis on a Real Testbed", *10th ITC Specialist Seminar*, Lund, Sweden, September 17-19, 1996.

[5] ITU-T Recommendation I.371: Traffic control and congestion control in B-ISDN

[6] COST 242, Cell Delay Variation in ATM Networks, (eds. A Gravey and S. Blaabjerg), Interim Report, October, 1994.

[7] The Complete Guide to ShowMe 2.0.1, SunSolutions, Sun Microsystems, Inc., June 1994

[8] O. Melteig, "Introduction to the PARASOL Project" *In The 9'th Nordic Teletraffic Seminar*, Norwegian Telecom, Research Department, August 21-23, 1990

[9] COST 224, Performance Evaluation and Design of Multiservice Networks, (ed. J. Roberts), Final Report 1991.

[10] S. Blaabjerg, S. Molnár, "Methods for UPC Dimensioning of a CDV Perturbated Cell Stream", *RACE BRAVE Workshop*, Milano, Italy, June 14-15, 1995.

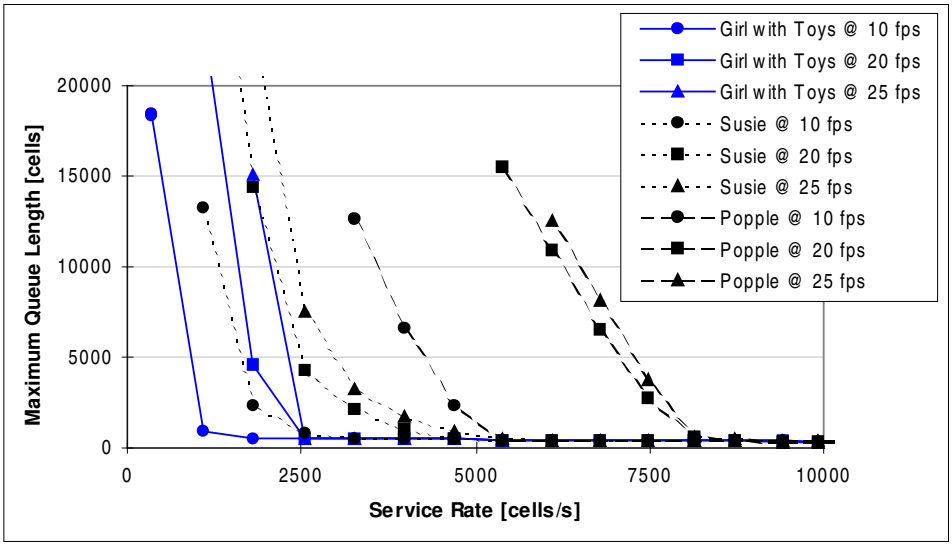

Figure 2. Impact Of Change In Performance Setting (Medium Resolution)

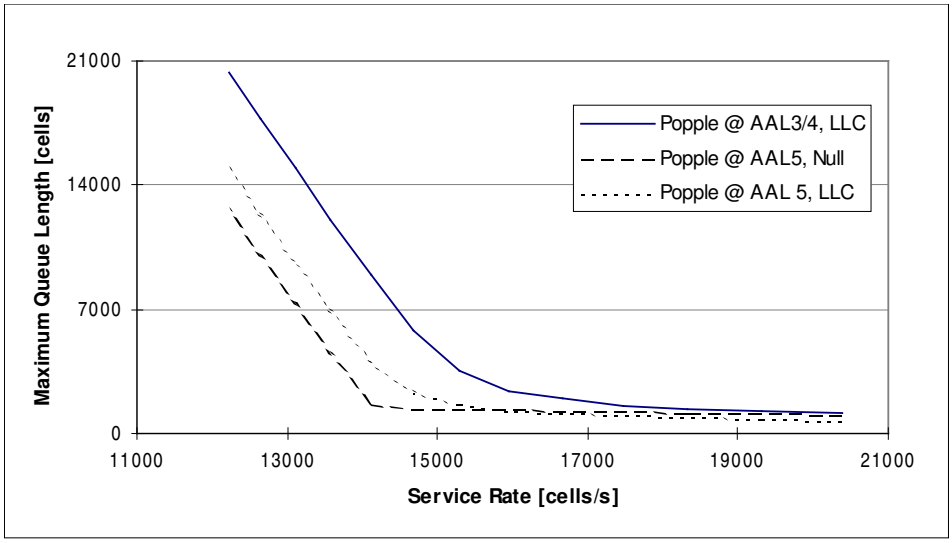

Figure 3. Impact Of Change In The Protocol Stack (High Resolution)

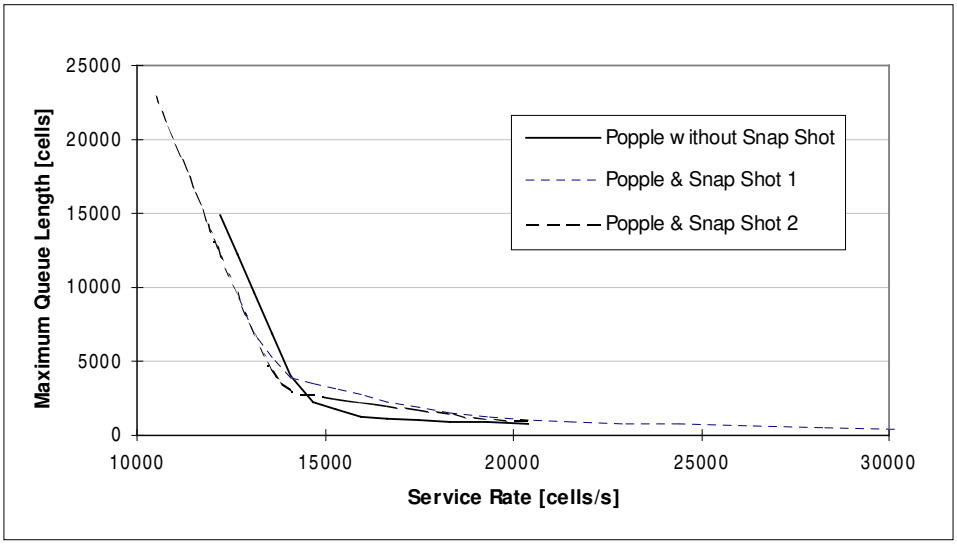

Figure 4. Impact Of Snap Shot In The Background (High Resolution)