HIERARCHICAL VBR VIDEO SOURCE MODEL

István Cselényi[a], Sándor Molnár[a], Nils Björkman[b], Alexander Latour-Henner[c]

[a]High Speed Networks Laboratory, Dept. of Telecommunications and Telematics, TU Budapest
Sztoczek u. 2. H-1111 Budapest, Hungary, e-mail: {cselenyi, molnar}@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

[c]Telia Instrument, Telia AB
Vitsandsgatan 9. S-12386 Farsta, Sweden, e-mail: alexander.a.latour@telia.se

Abstract

There are many ways to build up traffic models for different sources in ATM networks. A frequently applied methodology is to use mathematical analysis based on realistic assumptions to set up a source model that generates cell streams according to a stochastic process. In this case, the critical issue is the validation of the generated traffic by comparing statistics to results obtained from measurements on the real source. In this paper, we choose a different and more practical approach to investigate the detailed behavior of the real traffic source. A number of video sequences were measured and analyzed in case of a large number of performance settings to understand the nature of Variable Bit Rate (VBR) video traffic. Using the knowledge of generation process and results of the analysis, a *hierarchical source model* is set up for modeling the multimedia terminal. Thereby our model imitates the generation of video frames and the inner working of each level of protocol hierarchy and tries to reproduce the complex behavior of the real source. Moreover, we use the *leaky bucket analysis* for verification in order to capture directly the behavior of the traffic in a queue. The analysis results of measured VBR video sources and the related model building technique are presented with the verification of the model.

Keywords: ATM Traffic Modeling and Analysis, VBR Video Traffic

1. Introduction

Variable Bit Rate video is expected as one of the major source of B-ISDN traffic. This is because a substantial portion of forecast B-ISDN traffic is produced by multimedia sources (e.g. teleconferencing terminals, video-on-demand servers, etc.). The characteristics of VBR video traffic based on measurement studies and the related networking problems have been of increasing interest in the last decade of teletraffic research [14,20].

Source models of VBR video are needed to dimension networks and control methods to achieve acceptable quality and optimal usage of network resources. A number of different models have been proposed for VBR video modeling (for an overview see [8,20]). The large variety of modeling approaches can be divided into the following three categories: *Markov models* [4,12], *autoregressive processes* [10,15] and *fractal models* [1]. However, all of these approaches are common in the modeling technique that a specific stochastic process is chosen and the parameters are set by a specific method to fit some statistics of the real source. This methodology can be regarded as a “black box” approach, which is based merely on the characteristics of the measured traffic. An alternative method is the so called *“white box” approach,* which attempts to reproduce the detailed behavior of the source by imitating the inner working of the physical source and end system [9]. This methodology has not received much attention so far from the video modeling community. However, we believe that this non-traditional modeling concept can be very successful in practice.

In this paper, we introduce a simple and practical VBR video model, which is built up using the "white box" approach. We describe a simple technique to model the video frames and introduce a hierarchical generating process which emulates the impact of layers in the protocol stack on the ATM traffic. The measurements are taken from multimedia communication between two workstations. Moreover, we suggest to capture directly the queuing behavior related parameters of the traffic (i.e. maximum queue length versus service rate) by the leaky bucket analysis method [13]. This information is built into our model.

The paper is organized as follows. Section 2 introduces the measurement tools, scenario and measured traffic sources. Considerations about the factors determining the structure of an ATM cell stream are given in Section 3. Section 4 provides a detailed analysis of the VBR video traffic stream and hints for setting the model parameters. The proposed hierarchical modeling technique and its verification are described in Section 5 and the conclusions are summarized in Section 6.

2. Measurement

2.1 Scenario and Tools

Video and audio traffic were sent from a multimedia workstation through an ATM switch to an other workstation. A video cassette recorder provided the video and audio signals, in order to have a repeatable measurement procedure. This scenario is adequate to traffic characterization of multimedia sources in case of distributed multimedia applications such as video conference, movie teleswitching or video on demand. SunSPARC 10 workstations equipped with video, audio hardware and ATM interface boards were used as end stations. Permanent virtual connections were established among them with classical IP over ATM protocol stack. Internet Protocol (IP) datagrams were encapsulated using IEEE 802.2 LLC/SNAP and segmented into ATM cells using ATM Adaptation Layer 5 (AAL5). The video and audio information were multiplexed on a single Virtual Channel. SDH physical layer were applied in the network interface card. The ATM traffic was not shaped. The optical signal from the multimedia server was copied by means of an optical splitter to an ATM test equipment without affecting 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 [2]. The traffic records were post-processed with software developed by the authors.

2.2 Multimedia Source

Several standard CCIR video sequences recorded in VHS quality were used as video source. Three of them were selected for this paper:

- Girl With Toys: A sequence of almost still pictures – no motion dynamics
- Sussie: Head and shoulder scene, no camera movements – low motion dynamics
- Popple: Zoom of moving object – high motion dynamics

The video sequences were coded using CellB compression and packetized by the desktop multimedia application so that a constant rate of frames of varying sizes was generated. The CellB compression algorithm is a low-cost video compression scheme used mostly on Sun platform. In comparison with standard video compression algorithm such as MPEG or H.261, it has the advantage of cheap hardware support and less variable output video rate. The audio information was coded and transferred in a stream of 64 Kbps constant bit rate. Various settings of performance parameters, such as image resolution, frame rate, were tested using these video sequences. The cases examined in this paper are given in Table 1.

Notation	Name of Video Sequence	Frame Rate [frame/sec]	Resolution [pixels]	Length of Trace [sec]	Average Rate [Kbps]
GT10	Girl with Toys	10	384x288	21.41	452.93
SU10	Sussie	10	384x288	37.17	582.75
PL10	Popple	10	384x288	24.04	1500.01
GT25	Girl with Toys	25	384x288	35.89	922.26
SU25	Sussie	25	384x288	34.55	1026.76
PL25	Popple	25	384x288	35.19	2251.11

Table 1 Parameters of the Measured Video Sequences

Our investigations have been carried out on a general and common terminal architecture in order to have results of general validity. The workstation platform, video card and coding scheme, desktop multimedia application and ATM network interface card are widely used in public as a multimedia terminal of reasonable price. The TCP-UDP/IP over ATM protocol stack is also common and in the spotlight of several studies [18,21].

3. Traffic Generation Process

The generating and coding mechanisms that have an impact on the produced ATM cell stream are described in this section. This a priori knowledge about the traffic generation process will be tuned according to the results of our traffic analysis in Section 4. Our hierarchical modeling concept introduced in Section 5 is based on this process.

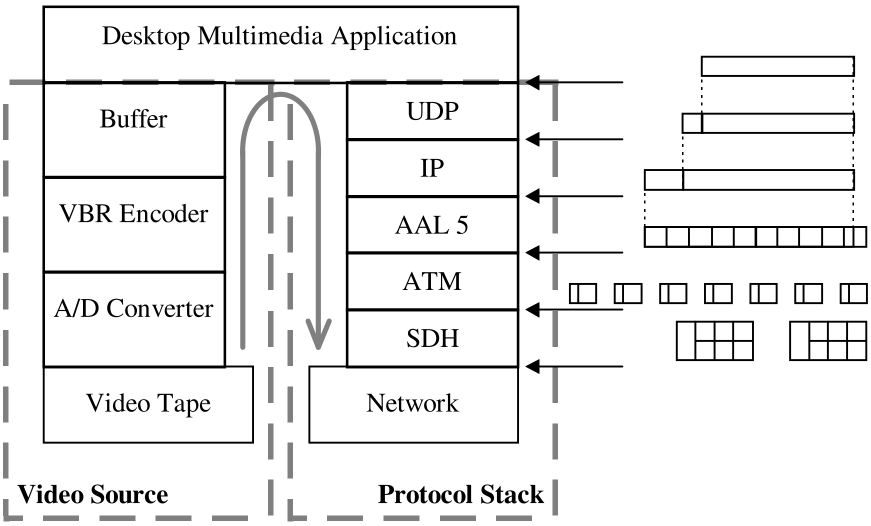
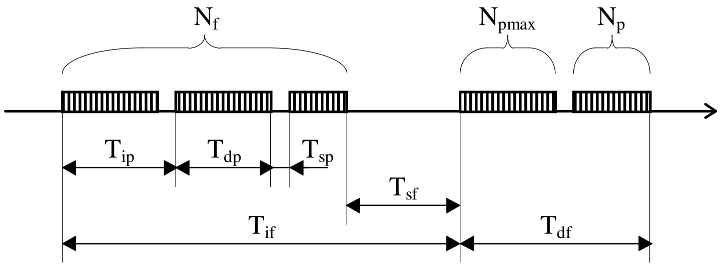


Figure 1 Traffic Generation Process

The traffic generation process and the flow of video information from the video tape to the network are depicted in Figure 1. The multimedia information is processed and transported by several hardware and software units during its way in the terminal from the input of the video and audio card to the output of the ATM board. These units are rather of a deterministic operation. Hence, a deterministic model is applicable for imitating the traffic generation process in the part denoted as *Protocol Stack* in the figure. The data transfer units of different size, which are forwarded among the layers of protocol stack, are illustrated on the left side of the figure. In spite of this quasi deterministic part, a stochastic model is necessary for describing the variable amount of compressed video data. The part containing these units is denoted as *Video Source* in Figure 1.

The amount of data produced by the VBR video encoder and stored in the buffer depends on the amount of movement in the video sequence. The CellB compression uses motion estimation and intra-frame coding and it can achieve a range of compression from about 0.75 bits per pixel to 1.5 bits per pixel for each pixel in the image [5]. Thus the typical size of a coded video frame is around 14 000 bytes in case of a resolution of 384x288 pixels. One can assume, that the application reads data from the buffer when the next video or audio frame should be transferred according to the frame rate set by the user. The frame is segmented by the application into transfer units which have a maximum length of 8192 bytes in our setting. Taking the typical frame size value estimated previously, the application transfers one transfer units of maximum length and an other, shorter one to the User Datagram Protocol (UDP) layer. For simplicity, it is supposed that there is no striping and each protocol layer sends its Protocol Data Unit (PDU) to the lower layer immediately after constructing it from the Service Data Unit (SDU) and Protocol Control Information. In this case the processing time is proportional to the length of the SDU received from the above layer. This phenomenon has an impact on the length of silence period between bursts (see Section 4.2). The UDP and IP layers insert an overhead of 8 and 20 bytes, respectively. The AAL5 layer adds an 8 bytes long trailer and a padding field of 28 bytes to get multiple of 48 octets, resulting in a packet of 172 cells on ATM level. This is the maximum size of a packet, denoted by N_{pmax} in Figure 2. This process is similar for shorter video frames and transfer units.



N_p	Packet Size	N_f	Frame Size
T_{ip}	Packet Interarrival Time	T_{if}	Frame Interarrival Time
T_{dp}	Packet Duration	T_{df}	Frame Duration
T_{sp}	Packet Silence Period	T_{sf}	Frame Silence Period

Figure 2 Multilevel Bursts in the Traffic Stream

The audio information is sent 8 times in every second. To get the constant bit rate of 64 Kbps, data units of 1024 bytes are forwarded to the protocol stack. Thus 1052 bytes arrive to the AAL5 layer which builds 23 ATM cells after segmentation and insertion of ATM headers. The cells of unshaped ATM traffic arrive on full link rate into the network, which is 155.52 Mbps because of SDH physical layer (see R_L in Table 2).

The two main parts of the traffic generation process, i.e. the Video Source and the Protocol Stack, generates traffic independently. Thus there are two major types of bursts in a hierarchical structure in the ATM cell stream of the investigated VBR video source (see Figure 2). On the higher level, frames of N_f cells can be identified, which contain the information of a compressed video frame. On the lower level, a frame consists of one or many packets of N_p cells. Figure 2 depicts the interarrival and duration time of frames (T_{if}, T_{df}) and packets (T_{ip}, T_{dp}). The silence periods between frames and between packets are given by T_{sf} and T_{sp}, respectively. The characteristics of this traffic structure are summarized in Table 2 using the notations in Figure 2.

Parameter	Value			Unit	Description
	Min.	Typical	Max.		
R_L	-	366 792	-	cell/sec	Link rate in the network (155.52 Mbps)
N_p	1	-	172	cells	Maximum packet size = (8192+8+20+8+28)/48
N_{paudio}	-	23	-	cells	Size of audio packet = (1024+8+20+8+44)/48
T_{dp}	-	-	178	cell time	Packet duration = $N_p * R_L/R_{ATM}$; R_{ATM} = 353207 cell/sec
T_{sp}	-	-	-	cell time	Linear function of N_p
N_f	-	294	-	cells	Typical frame size = 172+(5808+8+20+8+12)/48
T_{if}	-	36 679	-	cell time	Frame interarrival time at the frame rate of 10 frame/sec
T_{if}	-	14 672	-	cell time	Frame interarrival time at the frame rate of 25 frame/sec

Table 2 Traffic Parameters of the Examined ATM Cell Streams

4. Source Analysis

A detailed source analysis should be performed to verify our hypotheses about traffic generation described in Section 3 and to reveal the stochastic nature of video source. Different methods are applied on the measured data on several time scales ranging from cell level to frame level. The results of our source characterization are described in the next subsections. This analysis helps us to confirm and adjust the parameters retrieved from the knowledge of the generation process and to find the exact values or range of values.

4.1 Traffic Intensity Analysis

The traffic intensity of recorded ATM traffic traces is analyzed in different time scales, in order to investigate the structure of cell streams. Figure 3-5 show different details of a recording of the PL25 video sequence. The variation of traffic intensity shown in Figure 3 is caused by two high speed zoom periods with an intermediate period of partial motion in the picture field. Figure 4 shows the arrival pattern of seven video frames of varying size and three audio packets. Noteworthy is the deterministic property of the frame generation process. The internal structure of a frame is shown in Figure 5. At this time scale a high level of determinism can be observed in the arrival process. Finally, we show a single video packet in Figure 6 which contains 172 ATM cells arriving at practically full link rate. One can observe in Figure 4-5, that the packet silence period (T_{sp}) is shorter for the last packet in the frame. This fact confirms our presumption that a shorter packet can be processed faster by the layers of protocol stack than a larger one. This phenomenon will be investigated in detail during the packet level analysis (see Section 4.4).

The results justify the hierarchical modeling concept. The size of video frames depends on the content of video sequence while the structure of frame internal packets look very much the same for each frame.

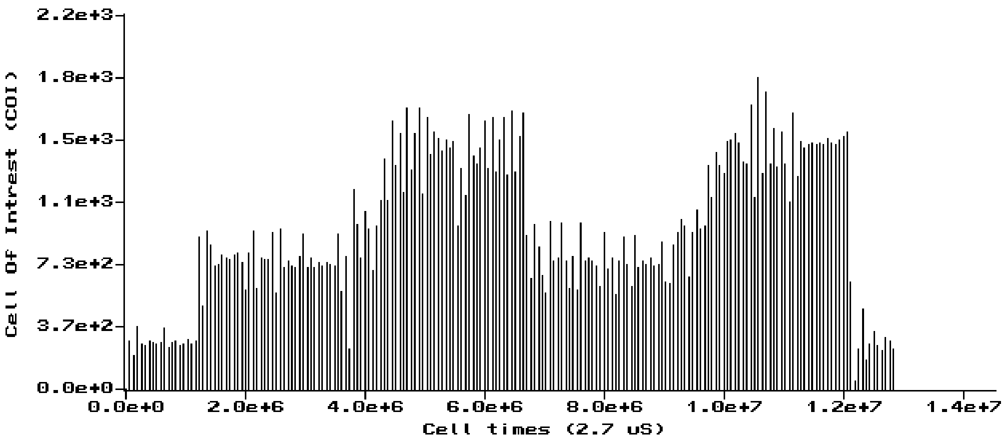


Figure 3 The Cell Arrival Intensity of the PL25 Sequence
Length of Window: 38 sec, Mean Rate in the Window: 2251 Kbps

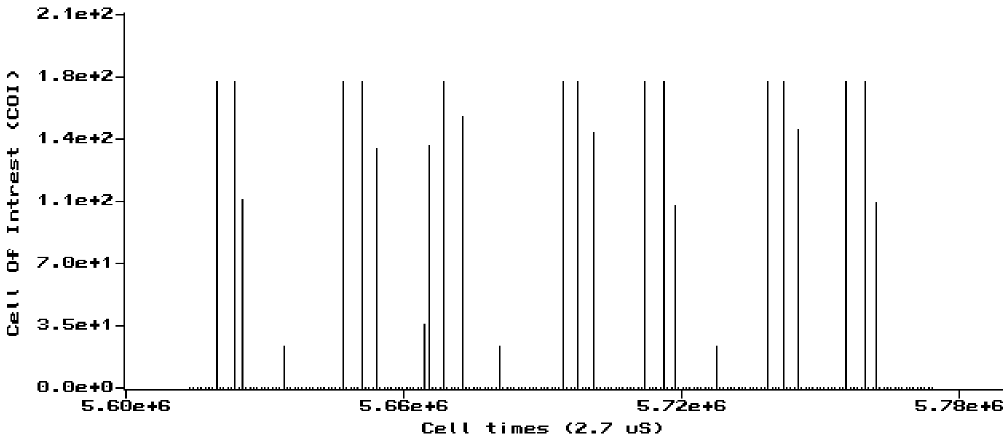


Figure 4 The cell arrival intensity of the PL25 sequence
Length of Window: 490 msec, Mean Rate in the Window: 3.26 Mbps

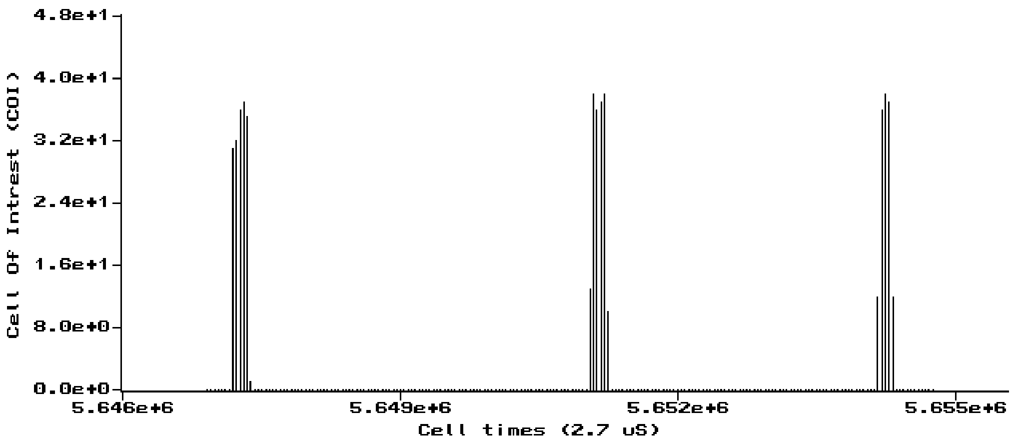


Figure 5 The cell arrival intensity of the PL25 sequence
Length of Window: 24.5 msec, Mean Rate in the Window: 9.45 Mbps

Figure 6 The cell arrival intensity of the PL25 sequence; Length of Window: 654 µsec, Mean Rate in the Window: 133.4 Mbps

4.2 Silence Period Analysis

Beside the traffic intensity analysis (i.e. analysis of busy periods) the other native way to gain information about the traffic is to calculate the probability mass function (PMF) of the Cell Interarrival Times (CIT). This latter method can be considered as analysis of the silence periods. The probabilities are estimated by counting the occurrence of CIT of different lengths and the values are smoothed by a moving average technique before drawing figures. The PMF of CITs is depicted in Figure 7 and the complementary probability density function (CPDF) in Figure 8.

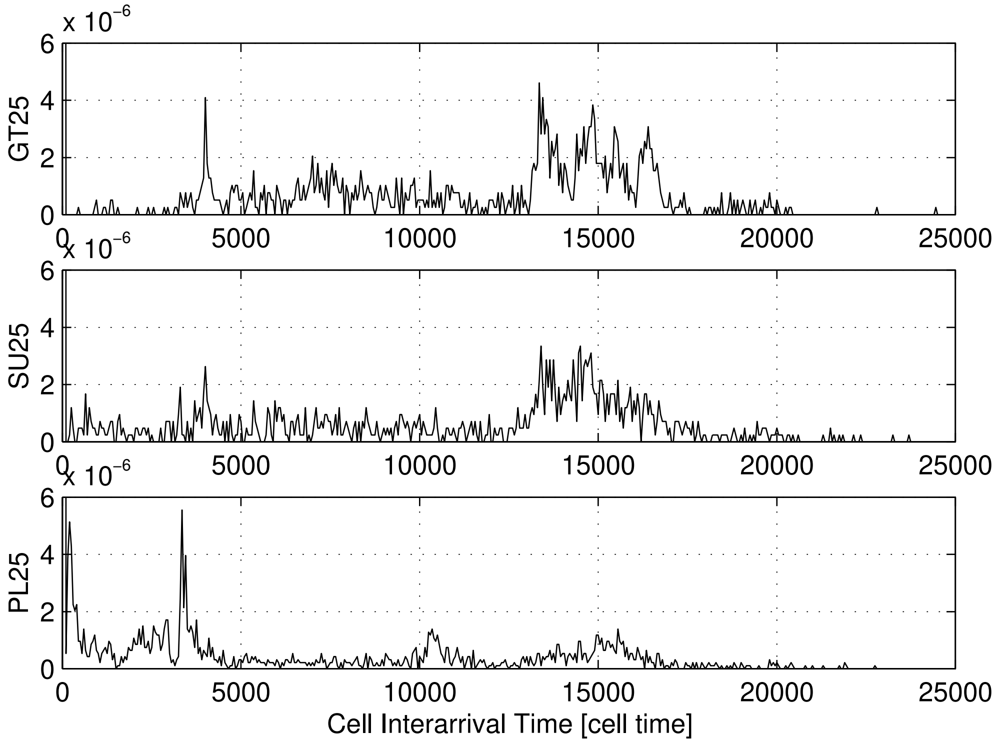


Figure 7 Probability Mass Function of Cell Interarrival Times (GT25, SU25, PL25)

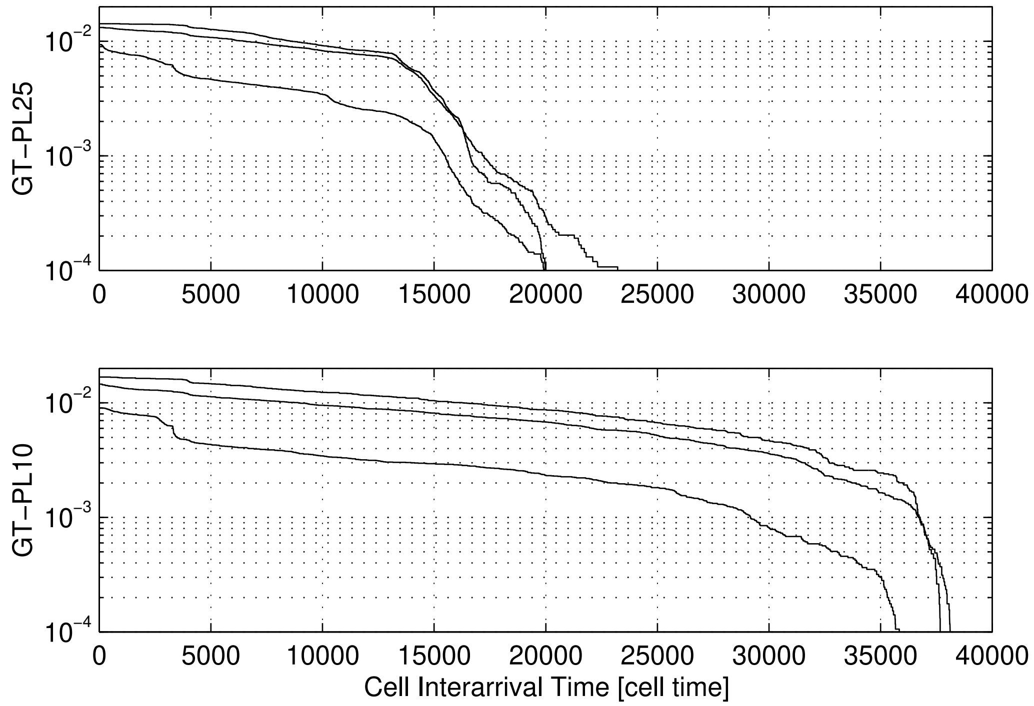


Figure 8 Complementary Distribution Function of Cell Interarrival Times (GT10, SU10, PL10, GT25, SU25, PL25)

The silence periods can be divided into three groups according to Figure 7 which depicts the PMF of GT25, SU25 and PL25 sequences. The longest periods represent the maximum silence periods in the sequence which are equivalent to the frame silence periods (T_{sf}). The medium values correspond to the silence periods within the video frames, i.e. between consecutive packets (T_{sp}) while the smallest values express the tiny, one or two cell times long gaps inside the packets. The evaluation of each group is given in the following.

In case of GT-PL10 sequences, the video frame rate is 10 frame/sec thus the theoretical frame interarrival time (T_{if}) is around 36 679 cell time (Table 2). For the GT-PL25 sequences with 25 frame/sec, the theoretical frame interarrival time (T_{if}) is 14 672 cell time. To get the frame silence period (T_{sf}), these values should be decreased by the duration of frames (T_{df}) which varies between a maximum (T_{dfmax}) and minimum (T_{dfmin}). However, even if T_{dfmin} equals to one cell time is assumed as a first approximation, it is clearly pronounced in Figure 8 that there is a certain probability above 36 679 cell times for GT-PL10 sequences and 14 672 cell times for GT-PL25 sequences, respectively. So in terms of frame rate, the terminal platform produces frames with a *somewhat worse* performance. An other phenomenon to be noticed in Figure 8 is that the maximum silence period (i.e. T_{sfmax}) is significantly shorter for the PL sequences than for the others due to the higher traffic intensity and larger size of video frames.

The silence periods, that are shorter than an estimated frame duration (see T_{df} in Table 2) but longer than a packet duration (T_{dp}), represent the packet silence period (T_{sp}). In case of SU25 sequence, the probability related to the packet silence period is less than in case of PL25. The reason is that there are more frames which consist of more than one packet in case of the more intensive PL25 sequence, thus there are more packet silence periods in the captured cell stream. The mean packet silence period read from Figure 7 is around 4000 cell time for SU25 and around 3400 cell time for PL25. The packet silence period can also be detected in Figure 8 where there is a small fall on the CPDF curve of SU25 around this value. The position of falls is almost the same but the scale is larger in case of PL25, because PL25 contains more frames that are larger than one packet.

Although ATM level data was captured in our measurements, the traces of SDH overhead information can be observed in the cell stream. Physical layer information is discarded by the measurement instrument, but it causes discontinuity within the packet. This is highlighted by the fact that the probability of very small silence periods is higher with many magnitudes than that of the longer silence periods.

The Squared Coefficient of Variation of CITs is used as a burstiness measure. This descriptor and other traditional traffic source statistics are listed in Table 3. The values underline that the burstiness of the traffic on ATM level is determined by the video source and not by the quality setting, since the same video sequences results almost the same burstiness both in case of 10 and 25 frames per second. The decrease of burstiness of PL25 sequence compared with PL10 could be caused by the saturation of terminal performance. It can be seen from the peak cell rate values that no shaping was applied.

Name of Stream	CIT mean [cell time]	CIT variance	Burstiness	Peak Cell Rate [cell/sec]	Number of captured cells
GT10	342.89	9 077 900	77.21	366 792	22 871
SU10	266.59	6 997 300	98.52	366 792	51 089
PL10	103.54	2 309 000	215.38	366 792	85 060
GT25	168.40	2 235 300	78.82	366 792	78 066
SU25	151.20	2 115 300	92.53	366 792	83 696
PL25	68.99	818 380	171.93	366 792	186 827

Table 3 Traffic Characteristics of the Captured Cell Streams

4.3 Correlation Structure Analysis

In our case, the data arriving continuously to the input of the video encoder is forwarded in small Protocol Data Units of different size through the protocol hierarchy and leaving the workstation in ATM cells. This process can be considered as a discretization of time into intervals by a slotted scheme. Therefore one may expect that every N_f^{th} and N_p^{th} CIT has almost the same length, so the sequence of CIT will exhibit high serial correlation of lags N_f and N_p and multiples thereof. The serial correlation of CITs plotted in Figure 9 confirms that suspicion. Peaks of correlation can be observed at the legs that are multiples of about 90. Therefore one can conclude that the average packet size is around 90 cells. This observation is verified in the next section (see Figure 10).

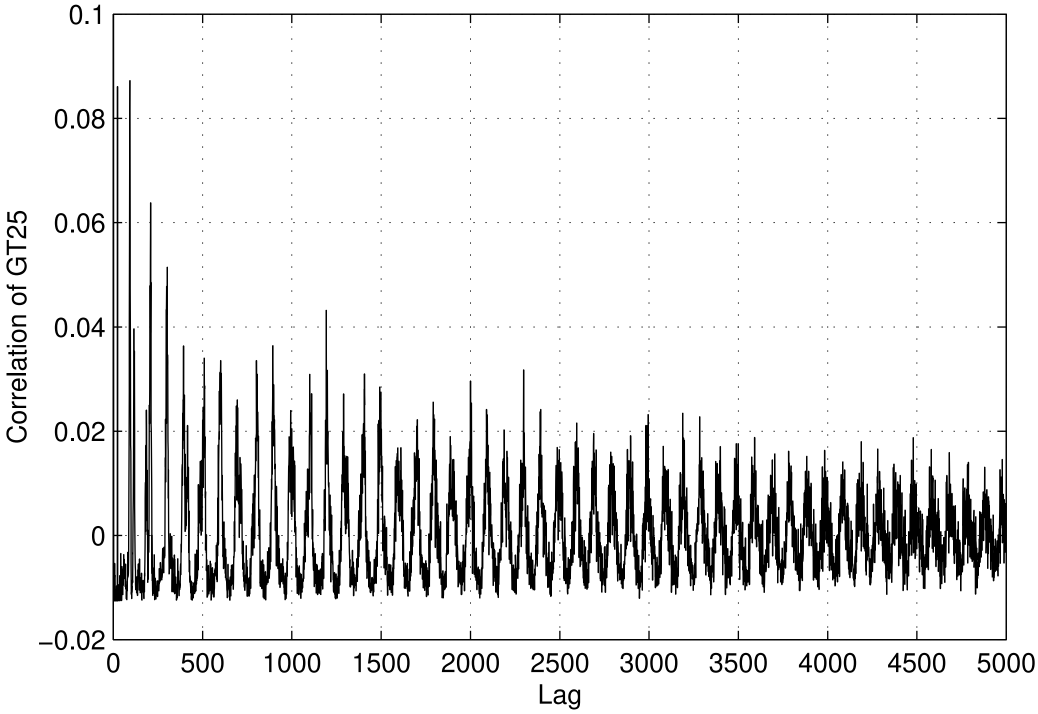


Figure 9 Autocorrelation of Cell Interarrival Time (GT25, SU25, PL25)

4.4 Packet Level Analysis

The size, maximum silent period, duration and interarrival time of each packet of the captured cell trace are calculated in the packet level analysis. The resulted packet size (N_p) values confirm our considerations and they suit to the values in Table 2, i.e. the maximum video packet size is 172 cells and the audio packets are always 23 cells long. The duration time of packets (T_{dp}) is proportional to the packet sizes and its typical value for a packet of maximum size is around 180 cells. The content of the video sequence has a pronounced impact on the distribution of packet size as well (see Figure 10). If the peak caused by the audio packets is neglected, the PL25 sequence consists of packets of maximal size, packets shorter than an audio packet and some other packets which size is between that of the previous two groups. The latter packets have very small but equal probability. The reason of this phenomenon is that there are two mean shifts in this sequence which are quite constant (see Figure 3). Most of the packets belong to these parts of the trace, while the transient phase between the shifts happens very quickly, yielding to a limited number of packets. Since the probability of packets of maximal size is definitely higher than that of shorter ones, there are really more frames consisting of more than one packet in the PL25 sequence (see Section 4.2). The SU25 video sequence varies more around the mean packet size, but there are still many packets of maximal size. This is not the case by the GT25 sequence, where middle sized packets have much higher probability than packets of N_{pmax} cells. The constant behavior of this latter video is also visible.

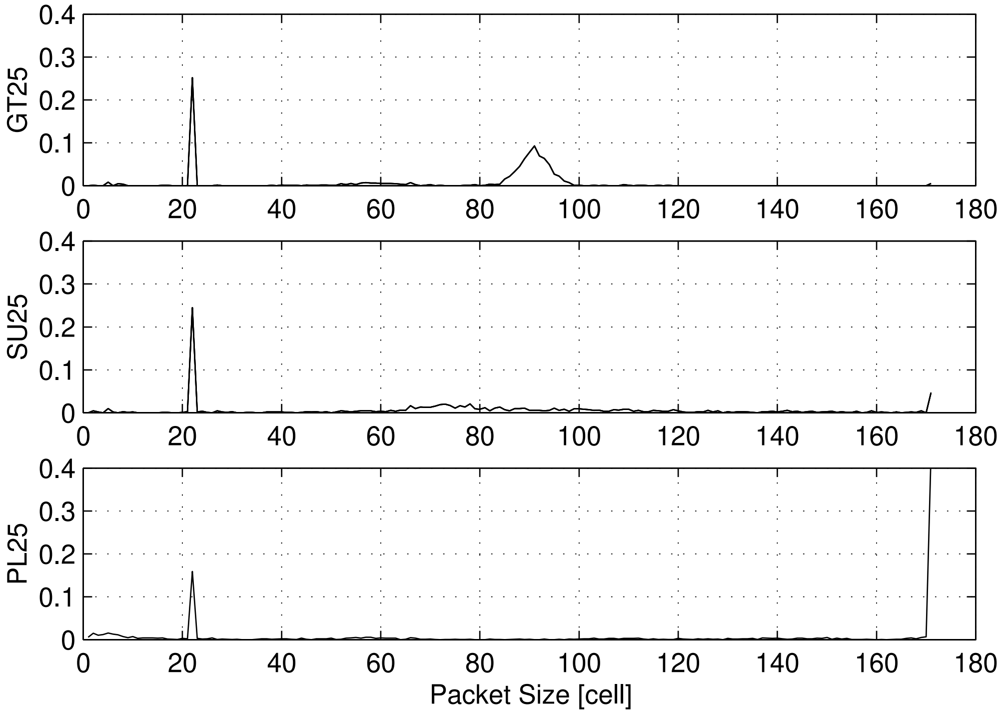


Figure 10 Probability Mass Function of Packet Size (GT25, SU25, PL25)

As an other examination on the packet level, the relationship of packet silence periods (T_{sp}) and packet sizes (N_p) is analyzed. The results confirm the presumption made in Section 3; the length of a T_{sp} period depends linearly on the size of packet generated after that period:

$$T_{sp}^{k} = \alpha N_{p}^{k+1} + \beta \qquad (4\text{-}1)$$

where k denotes the parameters of the k^{th} packet. The α and β constants can be determined by regression from the corresponding T_{sp} and N_p values. α is around 18 and β is 320 cell time.

4.5 Frame Level Analysis

The frame level behavior of the traffic is analyzed in this subsection, in order to make our understanding of traffic generation more clear. The counts of cells in video frames, the frame intararrival time and frame duration are derived from the captured cell streams by software. Consecutive frames were differentiated with the help of maximum frame silence periods determined in previous sections. The audio packets which have minor effect on the characteristics get a high importance here, since their presence had to be neglected during frame separation. It could be made by the knowledge that an audio packet consists of 23 cells (see Table 2). Therefore the software neglects every packet of 23 cells. However, the cells belonging to an audio packet are also in the sum by counting the frame size. Thus the number of cell arrivals and mean cell rate values correspond to these values of the original trace.

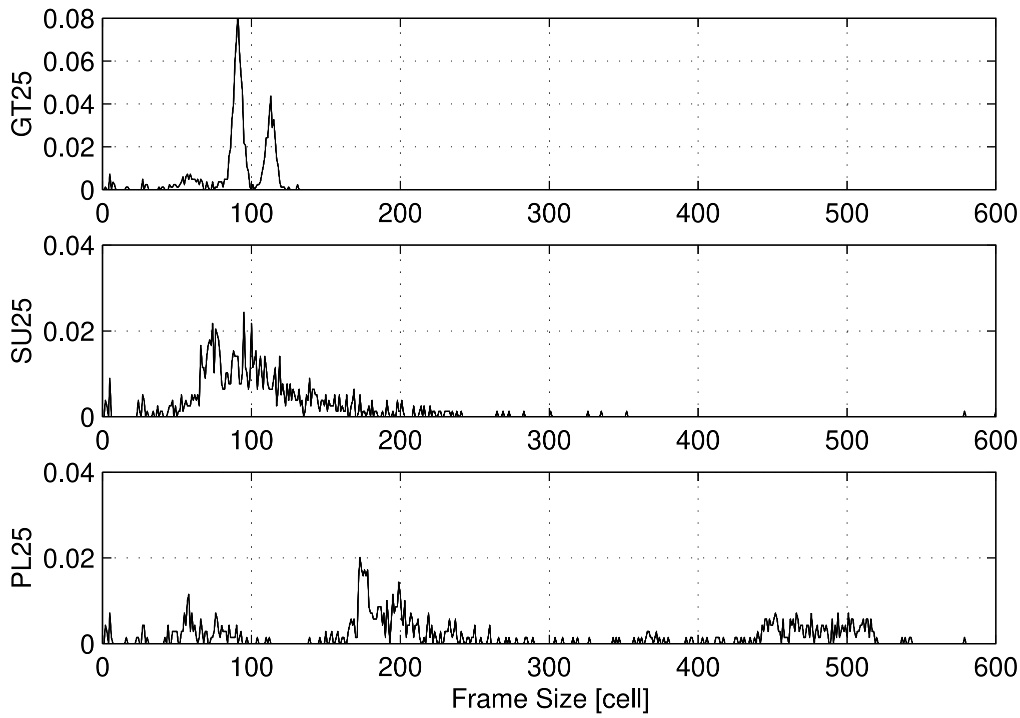


Figure 11 Probability Mass Function of Frame Sizes (GT25, SU25, PL25)

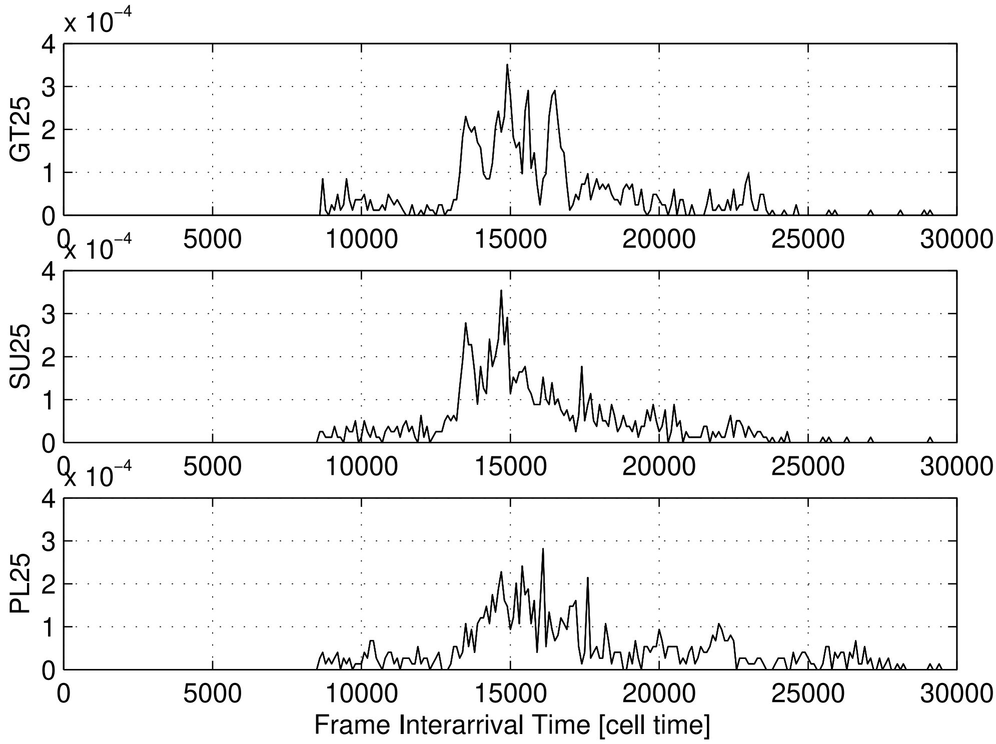


Figure 12 Probability Mass Function of Frame Interarrival Time (GT25, SU25, PL25)

It can be seen in Figure 11 that the frame sizes are more dispersed in case of the more dynamic video (PL25) than in case of the almost still picture sequence (GT25). This histogram is used for setting the parameters of our frame generator model. Figure 12 demonstrates that in reality the frame interarrival time (T_{if}) is far from being constant and above the theoretical value of 14 672 cell time (see Table 2). However, a common range can be determined from 8 000 to 28 000 cell time which contains the T_{if} values of the GT-PL25 sequences.

The difference among the frame correlation structure of the measured video sequences is shown in Figure 13. In case of the bursty PL25 video sequence, the correlation is definitely higher than for GT25.

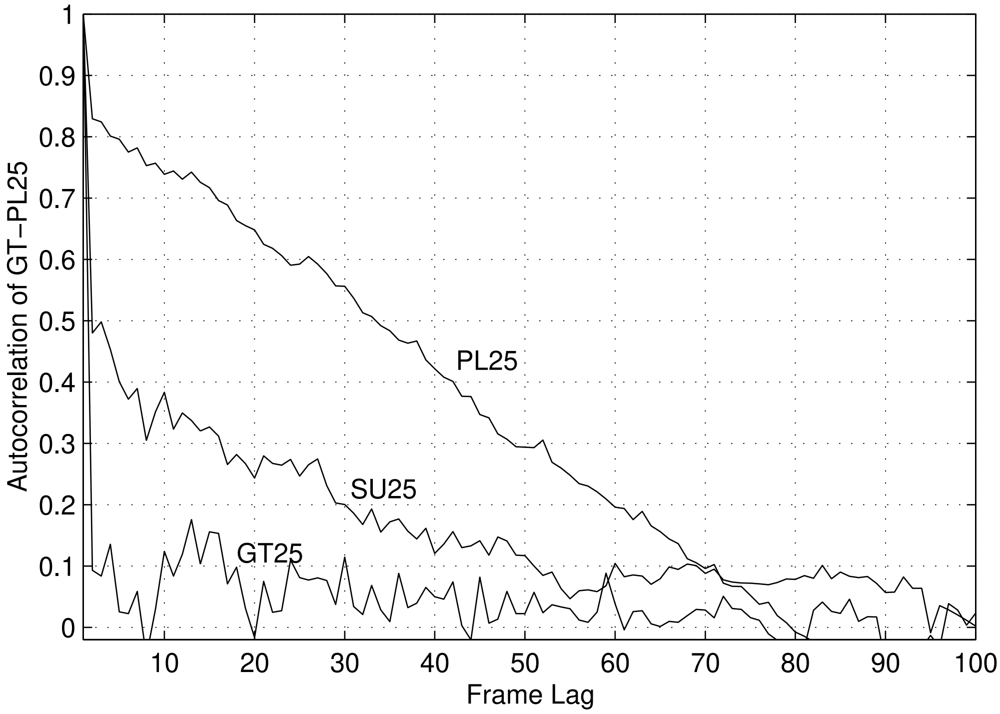


Figure 13 Autocorrelation of Frames (GT25, SU25, PL25)

4.6 Leaky Bucket Analysis

Leaky Bucket (LB) analysis is performed to check the information retrieved from previous analysis steps and get further parameters for the model. The reason for choosing this method is that this analysis is focusing on the traffic behavior that is actually tested for at the UPC. The standardized cell conformance method (Generic Cell Rate Algorithm, GCRA [17]) is based on the leaky bucket method. The LB analysis directly provides the required characteristics for the UPC (i.e. the peak cell rate and cell delay variation tolerance as well as sustainable cell rate and burst tolerance parameters). The proper settings of these parameters are needed for optimal resource allocation [3]. From network designer's point of view it is also a direct tool for dimensioning buffers or shapers [11] because it reveals the important queuing behavior of the traffic. Our aim is to get this information and to build it into our model. A detailed description of the leaky bucket traffic characterization method with applications can be found in [13].

In this paper, we use the queuing theory terminology to describe the leaky bucket method, i.e. we consider the leaky bucket as a G/D/1 queue. We are investigating the curves of the *maximum queue length* (Q) versus the *service rate* (R) by feeding the recorded cell streams to the G/D/1 queue. The curves are plotted for the three recorded sequences in Figure 14. It can be seen that the bursty PL25 sequence with higher mean cell rate has a stricter buffering requirement than the others.

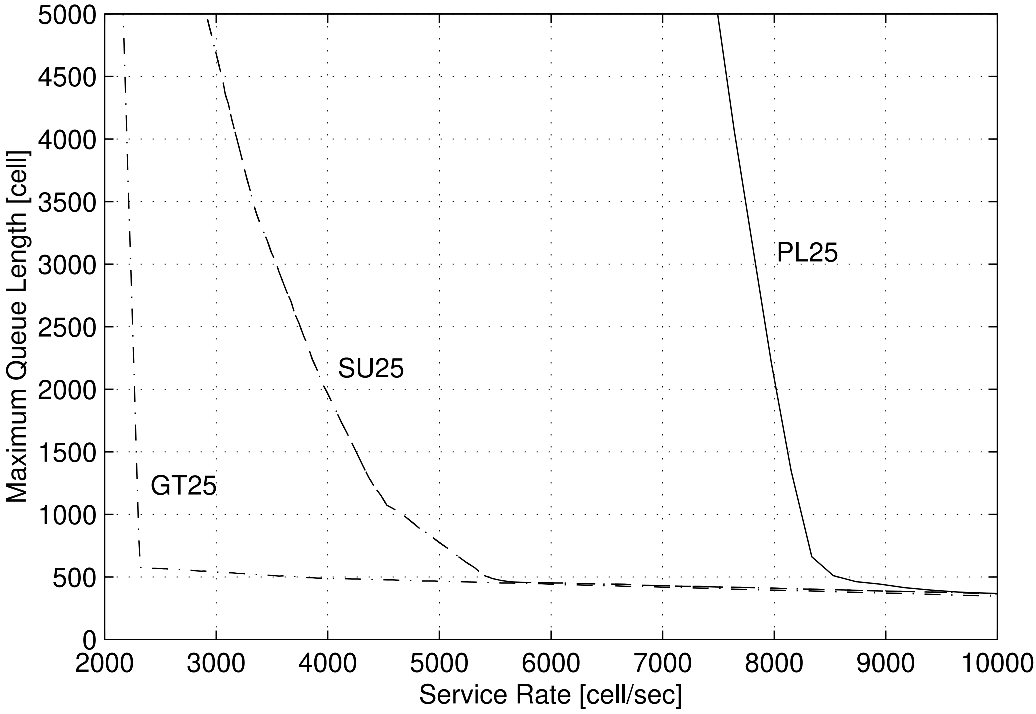


Figure 14 Leaky Bucket Curves (GT25, SU25, PL25)

Single high bit-rate periods of a video sequence increase the maximum queue length considerably as it is also proven in [16]. Therefore the points of the leaky bucket curve are determined by some worst case events and not by the average characteristics of the source, as we describe in this section.

The maximum queue length curve can be divided into more sections and looks like a piecewise linear curve. The breaking points of the curve are related to different worst-case events when there is a huge amount of cell arrivals within short time periods. By decreasing the service rate from the link rate, the linear part of the curve represents the situations when the queue can be emptied during a packet interarrival time (T_{ip}). This linear part breaks at the service rate when a new packet arrives and the last cell from the previous maximum size packet is still in the queue. At this rate the queue length is equal to the number of cells in this maximum size packet minus the number of served cells during this packet interarrival time:

$$R_p = \frac{R_L N_{p\max}}{T_{ip\min}} \qquad\qquad Q_p = N_{p\max}\left(1 - \frac{T_c}{T_{ip\min}}\right) \tag{4-2}$$

where Q_p and R_p are the coordinates of the breaking point, R_L is the link rate, $N_{p\max}$ is the number of cells in the largest packet, $T_{ip\min}$ is the minimum packet interarrival time for a packet of maximum size and T_c is the average cell interarrival time in a packet. By further decreasing the service rate, the second, third, etc. breaking points occur at the worst-case situations (maximum size packets with minimum packet interarrival time) when we find cells from three, four, etc. consecutive packets simultaneously in the queue.

The following significant change in the curve can be identified at the breaking point when the last cell from the previous maximum size frame with the smallest frame interarrival time is still in the queue at a new frame arrival. The calculations of the coordinates of this frame level breaking point is similar to the packet level, i.e.

$$R_f = \frac{R_L N_{f\,\max}}{T_{if\,\min}} \qquad Q_f = N_{f\,\max}\left(1 - \frac{T_{pavg}}{T_{if\,\min}}\right) \tag{4-3}$$

where Q_f and R_f are the coordinates of this frame level breaking point, R_L is the link rate, $N_{f\max}$ is the number of cells in the largest frame, $T_{if\min}$ is the minimum frame interarrival time and T_{ip} is the average packet interarrival time in the frame. By further decreasing the service rate more breaking points occur related to simultaneous queuing of cells from more consecutive frames similarly as in the packet level. The theoretical curve tends to infinity at the service rate equals to the mean rate. However, in practice we are always deal with finite data traces so the intercept of the curve related to the zero service rate equals to the total number of cells in the recorded cell sequence. The "worst cases" can be identified in the cell stream and N_f, N_p, T_{if}, T_{ip} can be determined from the leaky bucket analysis according to equations 4-2,3.

The curves presented in Figure 14 correspond to the zero cell loss requirements. However, for a specified cell loss requirement we can also use this leaky-bucket method. The curves related to different cell loss values are depicted in Figure 15. The cell loss is approximated by the overflow probability of the queue.

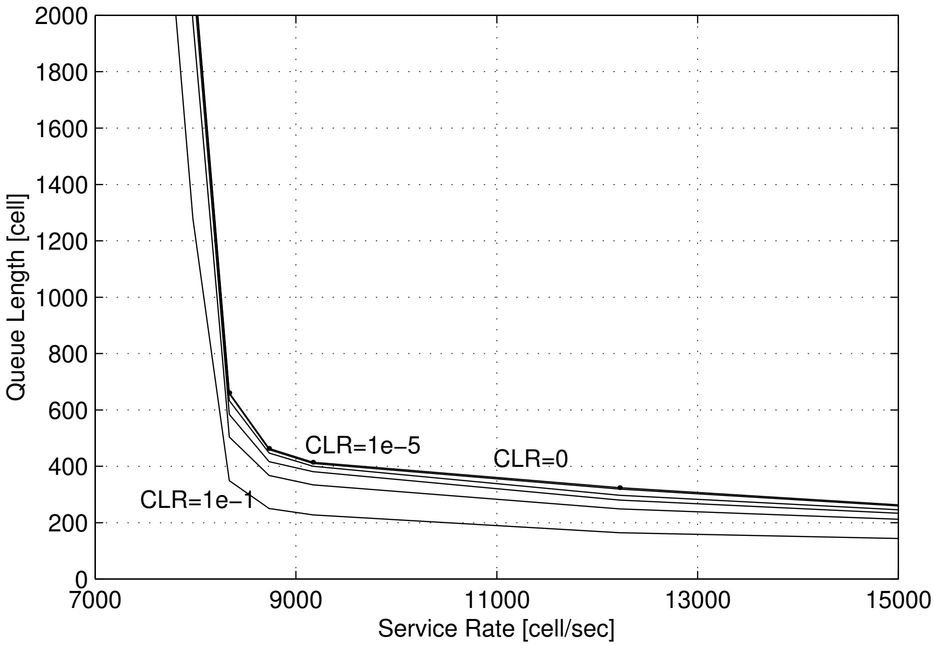


Figure 15 Leaky Bucket Curves and Estimated Cell Loss Probability (PL25)

It can be seen that for the practically interesting cell loss parameters the curves are close to the maximum queue length versus service rate curve.

5. Hierarchical Modeling Technique

Our idea is to construct a model which is built firstly on the a priori knowledge about the traffic generation process (see Section 3), secondly on results of source analysis and thirdly on LB analysis capturing the behavior of the source in the queue (see Figure 16). This construction yields to a model having the ability to imitate the queuing behavior which is of primary importance and avoids the unnecessary complexity arisen from capturing too many statistical characteristics. The parameters have to be selected which are relevant from the LB analysis point of view. As it is described in details in the previous section, these relevant parameters can be found on each time scale, ranging from cell level to frame or even scene level, which motivates our hierarchical modeling technique.

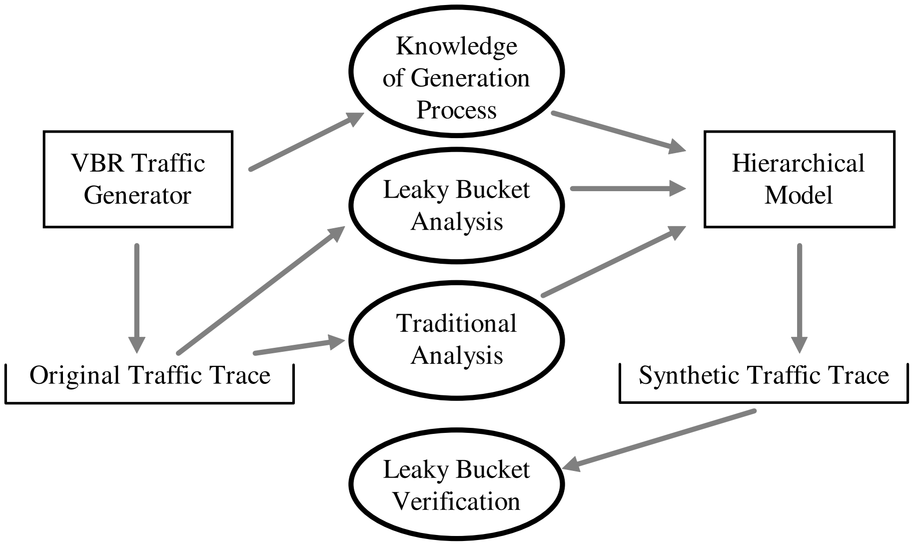


Figure 16 The Modeling Concept

The different levels of hierarchy should be modeled in different ways. The cell and packet level can simply be modeled by a deterministic traffic generator process. The packet level parameters are tuned by the information extracted from the leaky bucket analysis. This level corresponds to the part labeled Protocol Stack in Figure 1. The traffic on frame level requires a stochastic model. A simple two-stage Markov model is capable to generate frames fulfilling the LB requirements. It gets the input information from the frame histogram study. This level is labeled as Video Source in Figure 1. These models are introduced in the next subsections.

5.1 Model for Video Frame Generation

We found a simple Markov model is appropriate for modeling the frame generation process. It is because the long-term correlation does not have significant impact on cell loss [19] and the short-term correlation can be efficiently captured by Markov models. Moreover, increasing the Markovian order of the model does not improve significantly its correlation behavior [6] so there is no point of applying complex high-order Markov models. Therefore, we chose a simple two-state discrete time Markov model.

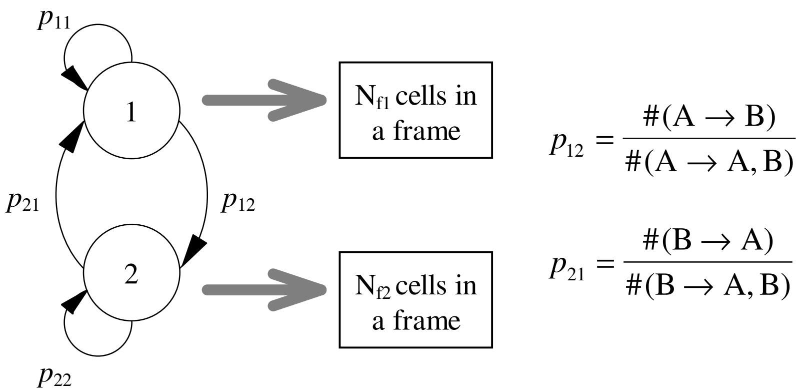


Figure 17 Frame Generation Model

The parameters of the Markov model (Figure 17) are set according to the PMF of frame sizes. The unit time of the Markov model is the frame interarrival time (T_{if}). State A is assigned to a given quantile of the probability mass toward to the maximum frame size as denoted in Figure 18. In this state frames of N_{f1} cells are generated, where N_{f1} is the mean of the shadowed quantile. The frame size in state B (N_{f2}) is determined such a way that the long-term average cell rate of the traffic should be fitted. The transition probabilities of the states (p_{12}, p_{21}) are estimated from calculating the number of the given frame size changes (A→B, or B→A) divided by the total number of frame size changes (see Figure 18).

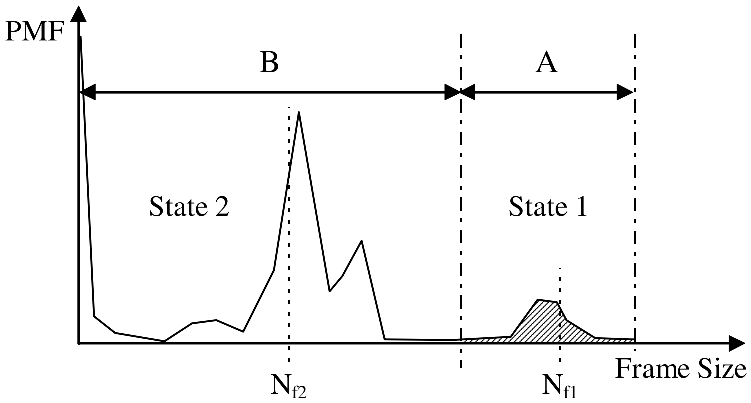


Figure 18 Frame Size Histogram

5.2 Model for Protocol Stack

The operation of the deterministic part of the traffic generator can be emulated by a simple routine based on considerations in Section 3 (see Figure 19).

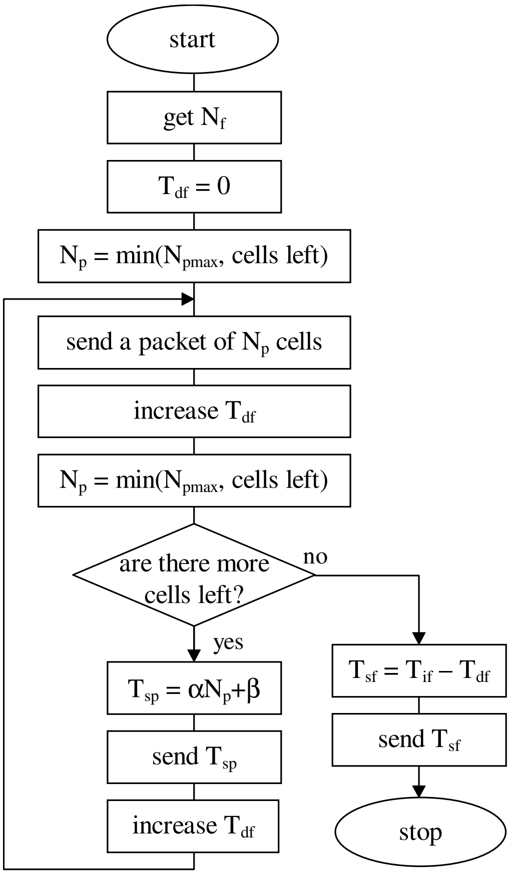


Figure 19 Model for the Deterministic Part

The parameters α, β, N_{pmax} and T_{if} are retrieved from the traditional and LB analysis described in Section 4. The frame size (N_f) is determined by the Markov model. Since there are ranges determined for this parameters instead of exact values, some iterative steps should be made until the parameter set is found which yields to the best model verification.

5.3 Model Verification

Leaky bucket analysis is performed both on the original sequences and the synthetic traces generated by our hierarchical model. The results are plotted for GT25 and PL25 sequences in Figure 20 and 21, respectively.

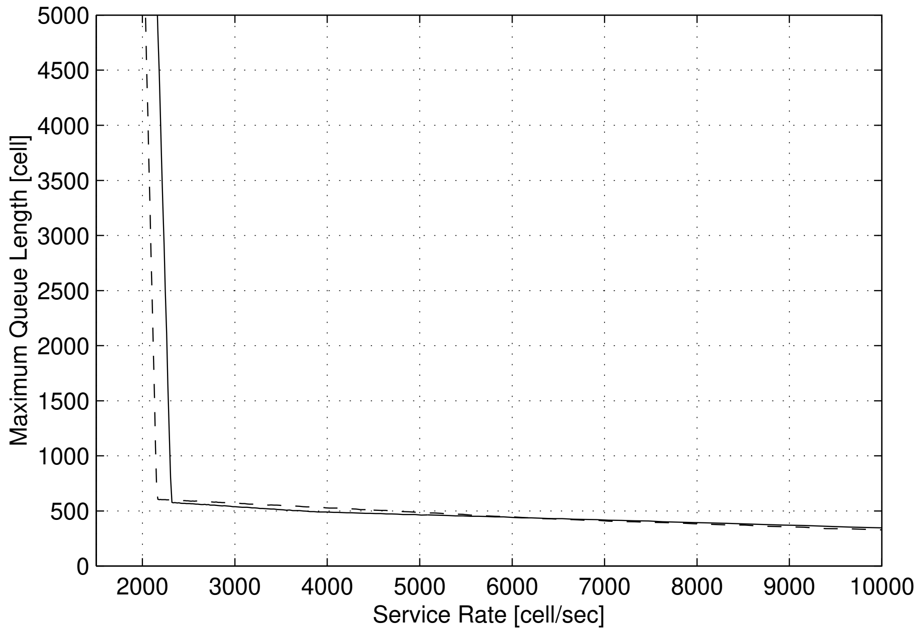


Figure 20 Leaky Bucket Curves for GT25 and the Model

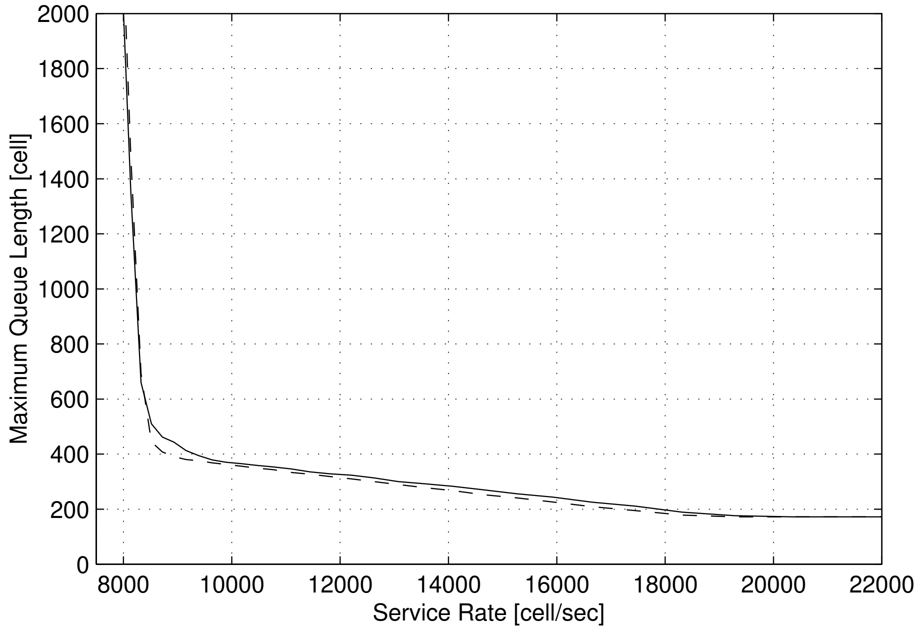


Figure 21 Leaky Bucket Curves for PL25 and the Model

We can see in these figures that the important breaking points are accurately captured by our model. The LB curves depend on scene changes in range of service rate from 0 to 2200 cell/sec for GT25 and 0 to 8300 cell/sec for PL25 sequence, since there are many frames in the queue. Although it was not our goal to have a good curve fitting in this range, the results are promising even in this range.

In order to show the performance of buffering, the CPDF for both the PL25 traffic and the model generated traffic are depicted in Figure 22. We illustrated three cases corresponding to load 0.43, 0.64 and 0.81. The related service rates are shown in the figure. The curves of the model are illustrated by dotted lines.

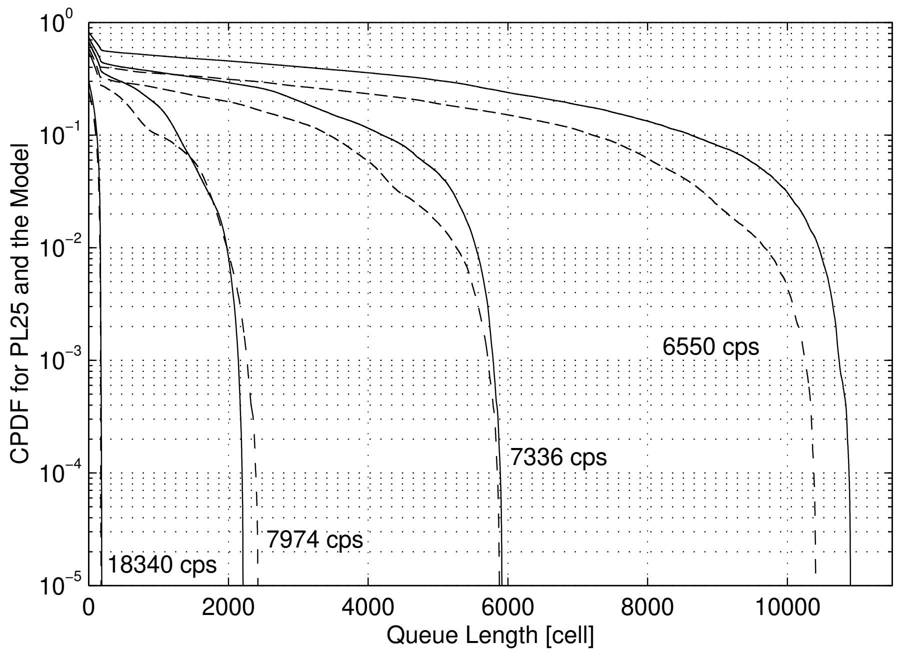


Figure 22 Complementary Distribution Function of Queue Length (PL25 and the Model)

The results show that the queuing performance is successfully reproduced by the synthetic traffic generated by our model. We think that this model and our general modeling approach is a good candidate for building video traffic models in practice.

6. Conclusion

In this paper, we propose a "white-box" modeling technique which is based on the a priori knowledge of the protocol stack and the queuing characteristics of the measured traffic obtained by the leaky bucket analysis method. Using this approach we introduce a simple and easily practicable model which consists of a two-state Markov process for video frame generation and a deterministic model imitating the mechanisms in the layers of the protocol stack. Our point is to focus on the final goal of modeling, namely the performance evaluation of traffic in the network. It means that we directly capture the important queuing behavior instead of describing the cell arrival process itself with a usually rather complex queuing problem. This methodology allows us to use merely the parameters that are important from queuing point of view.

Our measurement study with source analysis is described in the paper. We also discuss in detail the leaky bucket analysis and the information that can be extracted from the results. The model construction and verification results are also presented.

We may conclude the general validity of this modeling technique on the basis of our study performed on a relatively large number of video sequences and performance settings, from which six examples are presented. This concept can be easily adopted even to other hardware and software platforms, because of the similar operation of other terminal architectures and VBR coding techniques. We have ongoing work on the generalization of our Hierarchical VBR Video Source Model.

Acknowledgments

We wish to thank Urban Hansson and Aziz Miah from Telia Research for their help during the preparation and fulfillment of the measurements and our colleagues from High Speed Networks Laboratory for their valuable comments on this paper.

References

[1] J. Beran, R. Sherman, M. S. Taqqu, W. Willinger, Variable-Bit-Rate Video Traffic and Long Range Dependence, IEEE Transactions on Communications, Vol. 43, No. 2/3/4 (1995) 1566-1579.

[2] N. Björkman, A. Latour-Henner, U. Hansson, O. Pers, A. Miah, Practical ATM Resource Dimensioning Based on Real-Time Traffic Measurements and Analysis, GLOBECOM '95, Vol 1. (1995) 399-403.

[3] S. Blaabjerg, S. Molnár, Methods for UPC Dimensioning of a CDV Perturbated Cell Stream, RACE BRAVE Workshop, Milano, Italy, (June 1995)

[4] C. Blondia, O. Casals, Statistical Multiplexing of VBR Sources: A Matrix Analytic Approach, Performance Evaluation, Vol. 16, No. 5 (1992) 5-20.

[5] *The Complete Guide to ShowMe 2.0.1*, SunSolutions, (Sun Microsystems, Inc. 1994)

[6] COST 242: *Methods for the Performance Evolution and Design of Broadband Multiservice Networks*, Final Report, (1996)

[7] I. Cselényi, S. Molnár, N. Björkman, On Measurements of Multimedia Traffic in ATM Networks, Int. Conference on Multimedia Technology and Digital Telecommunication Services, Budapest, Hungary, (October 1996) 191-198.

[8] V.S. Frost, B. Melamed, Traffic Modeling For Telecommunications Networks, IEEE Communications Magazine, (March 1994) 70-81.

[9] P.E. Heegaard, B.E. Helvik, Establishing ATM Source Models for Traffic Measurement, 11'th Nordic Teletraffic Seminar, Stockholm, (August 1993)

[10] D.P. Heyman, A. Tabatabai, T.V. Lakshman, Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 2, No. 1 (1992)

[11] A. Latour-Henner, N. Björkman, I. Cselényi, ATM Traffic Shaping at the Multimedia Terminal, International Conf. On Telecommunications, Melbourne, (April 1997)

[12] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, J.D. Robbins, Performance Models of Statistical Multiplexing in Packet Video Communications, IEEE Transactions on Communications, Vol. 36, No. 7 (1988) 834-844.

[13] S. Molnár, I. Cselényi, N. Björkman, ATM Traffic Characterization and Modeling Based on the Leaky Bucket Algorithm, IEEE Singapore International Conference on Communication Systems, Singapore, (November 1996)

[14] 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 1996) 237-250.

[15] G. Ramamurthy, B. Sengupta, Modeling and Analysis of Variable Bit Rate Video Multiplexer, INFOCOM '92, (1992) 6.C.1.1-11

[16] E.P. Rathgeb, Policing of Realistic VBR Video Traffic in an ATM Network, Int. Journal of Digital and Analog Communication Systems, Vol. 6 (1993) 213-226.

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

[18] A. Romanow, S. Floyd, Dynamics of TCP Traffic over ATM Networks, IEEE Journal on Selected Areas in Communications, Vol. 13, No. 4 (May 1995) 633-641.

[19] B.K. Ryu, A. Elwalid, The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities, Proc. ACM SIGCOMM '96, (1996) 3-17.

[20] G.D. Stamoulis, M.E. Anagnostou, A.D. Georgantas, Traffic Source Models For ATM Networks: A Survey, Computer Communications, Vol. 17, No 6 (1994) 428-438.

[21] A. Wolman, G. Voelker, C.A. Thekkath, Latency Analysis of TCP on an ATM Network, 1994 Winter USENIX, San Francisco, CA, (January 1994) 167-179.

List of Figures