

CAN QOS GUARANTEES BE SUPPORTED FOR LIVE VIDEO OVER ATM NETWORKS?

Abhijit Lele and S.K.Nandy
 Supercomputer Education and Research Center,
 Indian Institute of Science
 Bangalore - 560012 INDIA
 e-mail: {abhijit,nandy}@serc.iisc.ernet.in

Abstract—Emerging *B-ISDN* networks which use Asynchronous Transfer Mode (ATM) as a backbone network, need to provide Quality of Service(QoS) to various multimedia traffic classes such as live video, high resolution images, video on demand, etc. Since live video is expected to dominate multimedia traffic, guaranteeing QoS for live video is complex because of no *a priori* knowledge of the video stream and hence requires a good understanding of such video sources. This paper investigates a complete framework for transporting live video over ATM networks with QoS guarantees. The proposed framework comprises of a source model, call admission controller and a scheduler. The source model uses *Entropy* as a measure to characterize the live video traffic. A *Markov Chain* model is developed based on this measure. The source model aids in reserving resources for the admitted sessions. The algorithm developed for call admission guarantees a minimum bandwidth throughout the duration of the session. Two new scheduling algorithms *viz* Guaranteed Bandwidth Schedule and Guaranteed Delay Schedule optimize the bandwidth utilization of all active sessions. Using this framework for transporting live video a cell loss probability of 4.87×10^{-7} and a best effort bandwidth utilization of 80% is obtained.

INTRODUCTION

Multimedia systems and ubiquitous communication is being increasingly projected as the all pervasive computing environment of the next millennium. In *B-ISDN* and computer networks, a major part of the traffic will be produced by multimedia sources like teleconferencing terminals, video-on-demand servers and pre-recorded multimedia streams. Predominantly, the live and pre-recorded video sources are MPEG streams. MPEG-4 is fast emerging as the *de facto* standard for live video encoding [2]. Characterizing live video sequences is a hot topic of current day research in source modeling because of no *a priori* knowledge of activities in live video sequences.

To ensure this QoS, current day and future local area

and wide area networks supporting real time video sequences will benefit by employing Asynchronous Transfer Mode (ATM) as a transfer medium. In order to provide QoS guarantees, the network must correctly allocate and regulate resources. Optimal resource allocation is one of the essential features of an efficient network. A source model that captures the essential features of traffic characteristics helps enhance network utilization.

Various source models proposed in the literature characterize statistical traffic in terms of parameters such as burstiness, inter-arrival time, inter-arrival correlations, etc [8].

In this paper we propose a complete framework for transporting live video sequences over the network without compromising on any of the QoS requirements. The steps involved are

1. Find an efficient source model.
2. Determine an effective call admission control policy.
3. Employ a good scheduling algorithm.

A schematic of the proposed framework is given in figure 1. The source uses entropy as a measure to capture the variations in number of bits/frame and is modeled as a markov chain. We partition the maximum bits/frame required for a live video sequence into bands of unequal extents. This is in contrast to other markov models [4] where each state is modulated by a distribution. Each band corresponds to a state in the markov chain. The coefficient of correlation between the entropies of any two frames is used to derive the state transition matrix for the markov chain (Predictor). The call admission control policy admits a call only if a certain bit/frame is guaranteed throughout the duration of the call (Call Admission Control). After a session is admitted, various parameters of the session are monitored to ensure that the negotiated QoS parameters are not violated (Resource Policing) and the network also guarantees the allotted resources (Threshold Correction). We propose two new scheduling policies *viz* *Guaranteed Bandwidth Schedule* and *Guaranteed Delay Schedule* for efficiently utilizing the net-

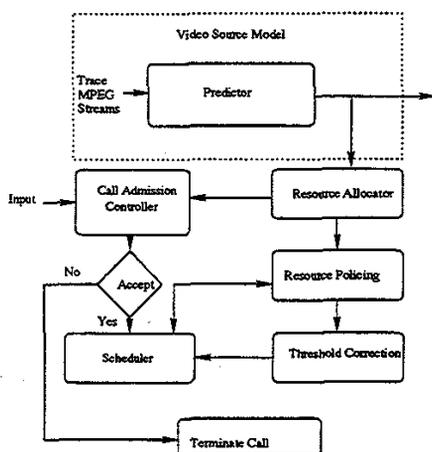


Fig. 1. Framework for transporting live video over ATM network

work resources (Scheduler). It may be noted that this source model does not replicate the actual behavior of the video stream but constructively influences call admission, scheduling mechanisms and QoS negotiation independent of the video source.

The rest of the paper is organized as follows. The second section gives the definitions of certain frequently used terms. The subsequent sections describe the mathematical framework for the proposed source model and application of this source model for QoS negotiations for benchmark live video streams. Scheduling policies are discussed in the fourth section and the subsequent section discusses the QoS reservation policy. We finally conclude in the sixth section.

PRELIMINARIES

For the sake of completeness in this section we briefly describe MPEG encoding and provide definitions for the most commonly used terms.

MPEG Video Encoding

MPEG coding is widely used for encoding video sequences. The input sequence to an MPEG encoder consists of a series of frames, each containing two dimensional array of picture elements called *pels*. The number of frames per second and number of lines per frame and the number of pels per line depend on the standard followed. For example, *PAL* standard uses 50 frames per second, 400 lines per frame and 545 pels per line. An MPEG stream consists of three types of frames:

1. **I frames** : uses intra frame coding based on DCT and entropy coding.
2. **P frames** : Uses motion compensation with respect to previous I or P frames.

3. **B frames** : Uses motion compensation with respect to previous or next I and P frames.

Typically I frames require more bits per frame than P frames, and B frames need least number of bits per frame. After coding the frames are arranged in a deterministic periodic sequence e.g. IBBPBBPBBPBB, referred to as *Group Of Pictures (GOP)*. Throughout this paper we use the following MPEG specifications.

- **GOP Pattern** : IBBPBBPBBPBB
- **Coding Scheme** : Huffman Coding followed by Run Length Encoding (RLE).
- **Transform Scheme** : 8×8 block based DCT.
- **Image Size** : 545×400
- **Frame Rate** : 50 Frames/sec.

Correspondingly the transmission rate of a frame can be expressed as

Definition 1

$$\text{Transmission Rate} = \left(\frac{\text{Bits}}{\text{Frame}} \right) \times \left(\frac{\text{Frames}}{\text{Sec}} \right) \quad (1)$$

The main focus of this paper is to provide a framework for transporting video streams over an ATM network. In the next subsection we briefly describe the transport mechanism over ATM networks.

ATM Transport Layer

ATM is a connection oriented packet switched network, with packets referred to as cells in ATM terminology. Figure 2 shows a typical ATM network, and this is the network topology used in this paper for simulation purposes. The switches used are 4×4 ATM switched and maximum capacity of each link is assumed to be 545 Mbps. The

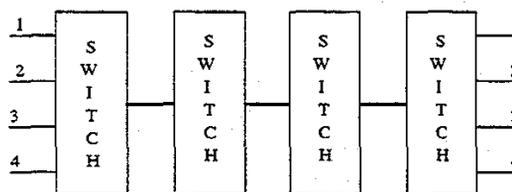


Fig. 2. Network topology used

cell size is fixed at 53 bytes (48 bytes payload and 5 bytes header). *Virtual Channel Identifier (VCI)* and/or *Virtual Path Identifier (VPI)* uniquely identify the path taken by a cell in an ATM network. Since information pertaining to error correction is provided only in the header, cells may be misrouted or lost because of network congestion. The loss of cells in ATM network is quantified by a parameter

referred to as *Cell loss Probability (CLP)* [3]. At connection establishment time the user specifies QoS parameters for the connection. QoS transactions are normally based on certain parameters such as *Peak Cell Rate (PCR)*, *Sustainable Cell Rate (SCR)*, *Delay* and *Delay Jitter* [3]. Other parameters indicative of network utilization are *Percentage Over Utilization* and *Percentage Under Utilization*.

Definition 2 *Percentage Over Utilization* : If B_i is the bandwidth associated with state i and B_{ij} utilized then

$$\text{Percentage Over Utilization} = \left(\frac{B_i - B_{ij}}{B_i} \right) \times 100 \times \rho \quad (2)$$

where ρ , the coefficient of bandwidth utilization, is the ratio of average bandwidth to total bandwidth.

Definition 3 *Percentage Under Utilization* : If B_i is the bandwidth associated with state i and B_{ij} utilized then

$$\text{Percentage Under Utilization} = \left\{ \frac{B_{ij} - B_i}{B_i} \times 100 \right\} \times \rho \quad (3)$$

where ρ , the coefficient of bandwidth utilization, is the ratio of average bandwidth to total bandwidth.

Most of the traffic carried over the network is statistical in nature and hence it is essential that a good estimate of the various statistical parameters be obtained so as to optimize the usage of various network resources. This is done with the help of a source model. Majority of the existing source models can be broadly classified as either *Markov Models* or *Autoregressive Models* [4] [8]. In the most simplistic form of the markov model [10], the maximum transmission rate is equally partitioned into N transmission rate, with each rate corresponding to a state in the markov chain. In *Autoregressive Model*, the correlation between two or more consecutive frames is used to predict the bitrate of the next frame. Details of this model are reported in [14]. More recently other types of video models based on *Self Similar* [12] and thermodynamic entropy [13] are reported in the literature. Most of the encoding in video sequences in entropy based. Hence entropy would make a good measure to capture the information content of the video sequences. Moreover scheduling policies have considerable impact on multiplexing gain and we will elaborate more on this in subsequent sections.

ENTROPY MODEL FOR VIDEO SOURCES

In the previous section we have set the motivation for exploiting entropy as a measure to predict the bits/frame for video traffic. In this section we give a more formal definition of the same.

Entropy is a measure of the information present in a data stream, and for a zero memory source is defined as:

Definition .1 *Entropy $H(s)$ is defined as*

$$H(s) = \sum_S P(S_i) \log \frac{1}{P(S_i)} \quad (4)$$

where $P(S_i)$ is the probability of the occurrence of code word S_i .

. With this we can now extend the definition of entropy to that of a MPEG stream.

Definition 2 *The entropy $E(n)$ of the n th frame of a MPEG stream is*

$$E(n) = \sum_1^M X(i) \times \log \frac{1}{X(i)} \quad (5)$$

where $X(i)$ is the probability of the occurrence of the code word i and M is the total number of code words being used in the MPEG frame.

The values of $X(i)$ are computed *a priori* from benchmark stored video sequences. We define average length of the code word as:

Definition 3 *If $S = s_1, s_2 \dots s_n$ is the set of possible code words, and l_i is the length of code word s_i , the average length L of the set of code words is defined as*

$$L = \sum_{i=1}^{i=n} P(s_i) \times l_i \quad (6)$$

With regard to MPEG encoding of a video sequence, each symbol s_i is representative of a block of size 8×8 pels. Let L denote the average length of the codes in set S in bits, and this L is used to calculate the bits/frame. If the average length is L bits and there are K symbols¹ occurring within a frame, the average bits/frame is $L \times K$ bits per frame. However it is very difficult to determine the average bits per frame. To do this we fall back on Shannon's [9] *Noiseless Coding Theorem* that derives a relationship between the average length L and entropy and is reproduced below.

Theorem .1 *For an instantaneous code with source alphabets $S = \{s_1, s_2 \dots s_q\}$ with the code length of s_i being l_i , and code alphabets $X = \{x_1, x_2, \dots x_r\}$, then the noiseless coding theorem states that*

$$H(S) \leq L \leq H(S) + 1 \quad (7)$$

and L is the average length given by equation 6.

As the average length L is bounded above, without loss of generality we let L assume the value $H(S) + 1$ for the rest of the paper and hence entropy provides an estimate for bits/frame to a high degree of accuracy. The bits/frame for each frame in the MPEG sequence can now be established as follows.

$$\text{Bits/frame} = L \times B_f \quad (8)$$

¹The words *Symbols* and *Code Words* are used interchangeably

where B_f is the total number of 8×8 blocks per frame. For a MPEG stream

$$B_f = \left\{ \begin{array}{l} (400 \frac{\text{lines}}{\text{frame}}) \times (545 \frac{\text{pels}}{\text{line}}) \times (6 \frac{\text{Blocks}}{\text{MacroBlock}}) \\ (16 \times 16 \frac{\text{pels}}{\text{MacroBlock}}) \times (\text{interlace factor}) \\ 2555 \end{array} \right. \quad (9)$$

The entropy per frame for the movie *starwars* is found to vary from 16.876 to around 88.171. In order to reserve resources on a per frame basis, we need to quantize these entropy levels. Also one of the important properties of any video source is the correlation between bits/frame across frames [8] [5]. We employ a markov chain to capture correlations across frames, where each state of the markov chain corresponds to a particular value of entropy and hence a particular bit rate. For this particular model we have selected a sixteen state markov chain. We have justified the selection of sixteen state markov chain in subsequent sections. The transition probabilities between the states of the markov chain correspond to the autocorrelation function between the entropy's of the frames.

Definition 4 The autocorrelation function $R(n)$ is defined as

$$R(n) = \frac{X(n) \times X(N-M)}{\text{Var}(X(N))} \text{ where } N \text{ and } M \text{ are integers.} \quad (10)$$

where $X(n)$ is the entropy. From equations 6,7,8 it is clear that any change in the entropy is bound to change the bits/frame. The Chapman Kolmogorov's equations for a N state model are given by

$$\pi[1] = \sum_{i=1}^N p_{1i} \dots \pi[N] = \sum_{i=1}^8 p_{Ni} \quad (11)$$

$$\sum_{i=1}^N \pi_i = 1$$

The steady state values for the above given equation are obtained by solving the Chapman Kolmogorov equations as described in [15]. The steady state values can then be used to find the transition probability matrix. This process is repeated iteratively till a quantization levels with minimum quantization error are obtained. The optimization criteria in every iteration is that the expectation of the quantization error must tend to zero. The following subsection gives the application of this mathematical framework developed so far to model live video streams.

In Perspective : QoS for live video transaction *vis a vis* the Entropy model

A source model is obtained using five MPEG files² each of duration varying from 1 to 3.5 minutes, of which 3 MPEG

²The five MPEG files are wip.mpg token.mpg, man.mpg, H1.mpg, H2.mpg. wip.mpg has maximum explicit scene changes, man.mpg has no scene changes and background changes, token.mpg represents a high bit rate MPEG file, H1.mpg and H2.mpg are clips from movies. The first three files are benchmark files. These files can be obtained from ftp://serc.iisc.emet.in/ abhijit

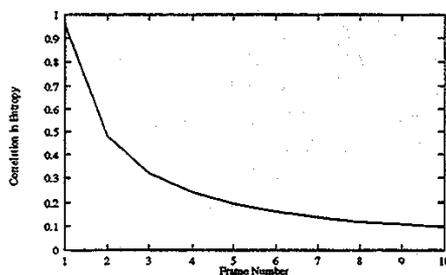


Fig. 3. Change in entropy across frames

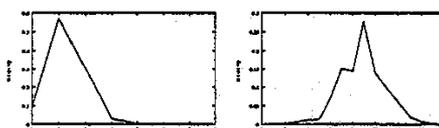


Fig. 4. State occupancy of eight state and sixteen state markov model

files are standard benchmark MPEG files used for characterizing video streams. In effect a data of about 10 minutes was used to determine the source model. The entropy per frame is found using equation 5 and the correlation in entropy across frames is shown in figure 3. The following observations can be drawn from the figure.

- The bits/frame does not change drastically across frames, indicating that there is a strong correlation in bitrates across adjacent frames.
- A sudden change in the bits/frame corresponds to a change in the scene.

To decide the order of the markov chain, we partition the maximum bits/frame for the traces into *eight* and *sixteen* equidistant states, and plot the percentage occupancy for each of the states depicted in figure 4. A preliminary observation of the bits/frame from the above graph led us to believe that quantizing the bits/frame in *eight* levels would be sufficient to track the variations in the bits/frame to a fair degree of accuracy, and hence a *eight* state markov model was derived. The transition probabilities for *eight* state markov model given in table 2 were computed using equations 10 and 11. The rows and columns in table 2 represent the current state and the next state of markov chain respectively. The associated bits/frame and the corresponding bit rate is calculated using equation 8 and equation .1 respectively for every state of the markov chain and is tabulated in table 1. The markov chain always follows the most probable path. Corresponding to the *eight* state markov model the bandwidth allocation for *starwars* trace is given in figure 5. The allocation of bandwidth according to a *eight* state markov model has an average bandwidth utilization of 66.381%, over utilization of 30.17% and an under utilization of 16.53%. Another observation is that, at some points (Typically I frames) the

State	1	2	3	4	5	6	7	8
Bits/frame in Kbps	5.8	10.2	15.03	19.38	26.71	31.33	49.61	56.01
Bit Rate in Kbps	145	255	375.75	484.5	667.75	783.25	1240	1400

Table 1. Bits/frame associated with the states in a 8 state markov model

		Next State							
		1	2	3	4	5	6	7	8
Current State	1	0.708759	0.138096	0.036514	0.004727	0.001447	0.001881	0.002122	0.003280
	2	0.386050	0.277666	0.119249	0.079410	0.039571	0.010194	0.005768	0.002683
	3	0.136121	0.180985	0.207522	0.198931	0.104811	0.062237	0.025200	0.005536
	4	0.021423	0.109539	0.254042	0.062449	0.378941	0.062652	0.040016	0.028294
	5	0.004164	0.049800	0.081113	0.387408	0.026482	0.356929	0.041139	0.018321
	6	0.007160	0.017185	0.050123	0.043576	0.550327	0.017799	0.261661	0.022095
	7	0.016257	0.015501	0.040076	0.031758	0.072968	0.640454	0.022306	0.122495
	8	0.064579	0.022505	0.035225	0.065558	0.043053	0.088063	0.510763	0.033268

Table 2. Transition probability matrix for a 8 state markov chain

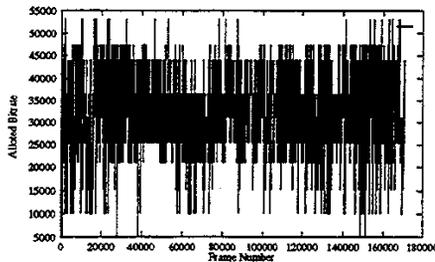


Fig. 5. Allotted bandwidth using 8 state markov model

States	Percentage Average Utilization	Percentage Over Utilization	Percentage Under Utilization
10	67.41	31.4	17.05
12	70.119	25.8	11.4
16	79.33	13.6	6.3

Table 3. Percentage utilization for 10,12,16 state markov models

allotted bandwidth is less than that required. This may lead to excessive cell loss which is not acceptable for live video transmission. One major factor contributing to the over and under utilization of the bandwidth is the non uniform distribution of bits/frame for different states of the markov chain. One of the solutions for improving the utilization is to increase the number of states of markov chain. Simulations were carried out with a markov model using 10, 12 and 16 states. The results obtained are given in table 3. The effectiveness of increasing the number of states is evident from table 3.

Even though the computational complexity in forming a sixteen state markov model is large, it contributes to the

startup cost alone. Once the model is ready, only a lookup table to determine the most probable path is sufficient. As the bandwidth utilization is near optimal for a sixteen state markov model, all further simulations were carried out using this model. The details of this model are given in the following subsection.

Sixteen State Markov Model

From the observations in previous section, we conclude that it is advantageous to use a 16 state markov model. The transition probability matrix and the corresponding bitrates is given in table 4 and table 5 respectively. The rows and columns of the matrix represent the next state and the current state of the markov chain. This Bits/Frame and the bitrate is calculated using equation 8 and equation .1 respectively. This transition probability matrix is obtained from equation 10 and by setting $N = 16$ in equations 11. In addition to the five MPEG files used to develop the entropy based *Markov Model*, this model is used on traces of five MPEG files, viz *Starwars*, *Soccer*, *Lambs*, *Term and Bean*. The actual bandwidth requirement and the allotted or the predicted bandwidth per frame for the movie starwars is shown in figure 6. The over utilization is 13.6% and the under utilization is 6.3%. A CLP of 6.27×10^{-7} is obtained. The average utilization of bandwidth in case of a 16 state markov model is around 81% , indicating that this model is able to track the changes in the bandwidth sufficiently well. For the I frames, the bandwidth is over allocated so that cell loss does not occur because of lack of resources. Over allocation of bandwidth is essential as losses in the I frame may lead to unacceptable picture quality. Apart from this, the CLP is monitored for every frame. If the CLP value exceeds a threshold value for three consecutive frames, then a corrective measure is taken in the form of re-entering

		Next State							
		1	2	3	4	5	6	7	8
Current State	1	0.608975	0.052724	0.027410	0.024780	0.010198	0.019649	0.023625	0.014795
	2	0.354018	0.016098	0.063016	0.025316	0.024491	0.050908	0.032196	0.010732
	3	0.283796	0.122874	0.010072	0.098030	0.006714	0.018129	0.042525	0.023724
	4	0.267527	0.055988	0.114167	0.005842	0.104430	0.020448	0.016796	0.038462
	5	0.180054	0.070625	0.015134	0.175786	0.002328	0.201009	0.007761	0.012418
	6	0.196996	0.077959	0.022747	0.020760	0.123012	0.003313	0.244700	0.013251
	7	0.198776	0.040691	0.028088	0.015484	0.008642	0.231725	0.002521	0.295823
	8	0.126456	0.014975	0.020521	0.024589	0.007210	0.013866	0.350712	0.000740
	9	0.221373	0.014615	0.013543	0.031912	0.022392	0.025342	0.038348	0.294985
	10	0.227039	0.022384	0.008172	0.009238	0.010481	0.016699	0.012969	0.032866
	11	0.223870	0.050466	0.008610	0.006458	0.006697	0.018656	0.010524	0.011002
	12	0.277924	0.111403	0.026247	0.013415	0.007874	0.021289	0.018664	0.004666
	13	0.300613	0.212533	0.075811	0.023225	0.008326	0.017967	0.027607	0.014461
	14	0.373898	0.209877	0.153439	0.083951	0.027513	0.019400	0.020106	0.017989
	15	0.548597	0.086621	0.119561	0.101261	0.039447	0.029687	0.013420	0.007320
	16	0.373003	0.153896	0.102706	0.103358	0.055103	0.093903	0.048419	0.014183
		9	10	11	12	13	14	15	16
1	0.035406	0.027901	0.020653	0.020525	0.014795	0.022022	0.028136	0.048405	
2	0.014309	0.013071	0.024904	0.054210	0.070171	0.086544	0.026417	0.133599	
3	0.024396	0.010072	0.007386	0.017681	0.033348	0.111907	0.066473	0.122874	
4	0.055988	0.010954	0.006329	0.010711	0.014606	0.051120	0.073515	0.153116	
5	0.074893	0.022507	0.012806	0.009313	0.005821	0.021731	0.039969	0.147846	
6	0.047482	0.032023	0.027385	0.021201	0.005521	0.010822	0.014576	0.138251	
7	0.030969	0.022866	0.020166	0.021966	0.017285	0.009363	0.004681	0.050954	
8	0.343132	0.021815	0.017194	0.012757	0.013866	0.011647	0.003143	0.017378	
9	0.000402	0.246581	0.029767	0.021453	0.012604	0.009654	0.003352	0.013677	
10	0.391721	0.000000	0.214781	0.026292	0.009416	0.003908	0.006751	0.007284	
11	0.091844	0.351591	0.000000	0.183449	0.015307	0.003109	0.007893	0.010524	
12	0.032663	0.079615	0.293963	0.000292	0.085156	0.006124	0.005249	0.015456	
13	0.017528	0.006135	0.042945	0.201139	0.000000	0.029360	0.002629	0.019720	
14	0.019048	0.005996	0.005996	0.008466	0.034921	0.000353	0.004233	0.014815	
15	0.010167	0.012607	0.007320	0.011387	0.003253	0.004880	0.000000	0.004473	
16	0.016303	0.007825	0.006684	0.008803	0.008151	0.006358	0.001304	0.000000	

Table 4. Transition probability matrix for a 16 state markov chain

the markov chain at the exact state corresponding to the bits/frame for the frame in error. This step is essential, as not taking a corrective measure would lead to accumulations of errors resulting in large cell loss. As observed from figure 6, the bandwidth requirement in the 82961th frame of the Starwars movie is around 131Mbps. This is the only place where the bandwidth requirement is high. The maximum allowed bits/frame in this particular model is around 80Mbps. Thus a large cell loss occurs during this frame. Due to lack of space, the bandwidth allocation plots for the other four traces are not shown here. However, these plots also show an average utilization of 81% and a *CLP* of 6.18×10^{-7} .

Switching in a network is equivalent to multiplexing of multiple video streams, and performance evaluation after multiplexing will give a realistic estimate of the effective-

ness of source model. It is assumed that the destinations addresses requested by the video streams are uniformly distributed. The network performance under such multiplexing is given by the multiplexing gain, and is described in the next subsection.

Multiplexing Gain

Multiplexing gain is indicative of the network utilization, and depends on the source model being used. Each of the five test traces mentioned in section are multiplexed according to the following policies.

- *Round Robin (RR)* : In this scheme, samples from each of the five traces are multiplexed in a round robin fashion.

States	1	2	3	4	5	6	7	8
Bits/frame in Kbps	5.3	10.87	12.37	14.18	16.09	17.381	20.66	24.19
Bit Rate in Kbps	132.5	271.75	309.25	354.5	402.25	434.52	516.5	604
States	9	10	11	12	13	14	15	16
Bits/frame in Kbps	26.81	30.75	33.68	36.66	41.22	46.5	58.93	70.8
Bit rate in Kbps	670.25	768.75	842	916.5	1030.5	1162.5	1473.25	1770

Table 5. Bit rates for a 16 state markov model

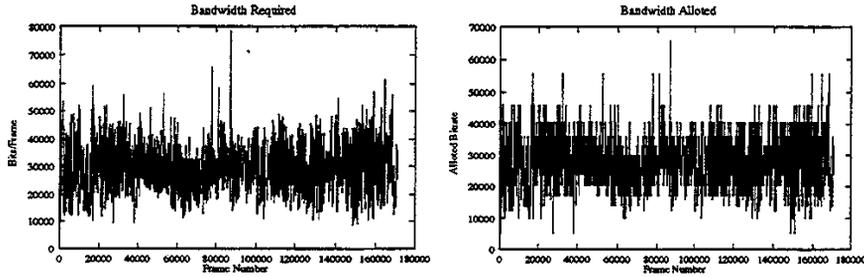


Fig. 6. Bandwidth required and bandwidth allotted for Star Wars trace

Policy	Multiplexing Gain
Round Robin	21.008
Random	22.43

Table 6. Multiplexing gain obtained at first stage

- *Random Multiplexing (RM)* : In this scheme, one of the five traces is randomly picked in a uniform manner, and a frame from that particular trace is selected.

The multiplexing gain as determined at the output of the first stage of the network is given in table 6. From table 6 we find that the scheduling policies considerably influences the multiplexing gain and hence the usage of the available bandwidth. This necessitates to find out a proper scheduling policy so as to optimize the bandwidth utilization. In the next section we formally define the different scheduling policies used and the multiplexing gain obtained for the various scheduling policies

SCHEDULING POLICIES

A proper scheduling policy helps to improve the overall network utilization. In this section we formally define the various scheduling policies.

Markov Scheduling Policy

Consider the switch with N input ports. In this paper it is assumed that the traffic at the input of each of the ports is video traffic, and hence the markov model can be applied to each active session. We propose two scheduling policies referred to as the *Guaranteed Delay Schedule (GDS)* and the *Guaranteed Bandwidth Schedule (GBS)*. Every

active session follows a particular transition graph in the markov chain. Thus at any given instance of time, every active session is in a particular state of the markov chain. Denote p_{ij}^k to be the transition probability from state i to j for the K th active session.

1. Guaranteed Delay Schedule

GDS is a weighted round robin schedule. In this round robin schedule, the sessions at the input port 1 is serviced first and the sessions from 2, 3, ..., N are serviced in order.

Let d^k be the maximum delay jitter and D^k the maximum delay associated with session K . Let each of the sessions be in a state i and let S_p be the set of next possible state of the markov chain.

$$\text{Let } B_{max}^j = \max_{i \in S_p} \{B_i\}$$

$$W = \{B_{max}^1, B_{max}^2, \dots, B_{max}^N\}$$

$$T_{min}^k = B_{max}^k \times (D^k - d^k) \quad \forall B_{max}^k \in W$$

$$T_{max}^k = B_{max}^k \times (D^k + d^k) \quad \forall B_{max}^k \in W$$

$$T_{max} = \max_{\forall k} \{T_{max}^k\} \quad \exists B_{max}^k \in W \quad (12)$$

$$T_{min} = \max_{\forall k} \{T_{min}^k\} \quad \exists B_{max}^k \in W \quad (13)$$

Let P be the port number associated with T_{max} given by equation 13. The GDS corresponds to a round robin starting from port number P . The calculation for T_{max} and T_{min} are done at every clock cycle, thus changing the starting point of the round robin schedule in every cycle.

2. Guaranteed Bandwidth Schedule

Let each of the session be in some state i , and let j denote the possible next state of the markov chain. Let S_α denote the set of all active sessions and B_i and B_j denote the bandwidth being utilized in the present state and the bandwidth required in state j respectively.

$$B_{max}^k = \max_{k \in S_\alpha} \{B_j^k\} \exists p_{ij}^k = \max_{k \in S_\alpha} p_{ij}^k \quad (14)$$

The GBS now corresponds to a round robin schedule that starts from the queue associated with the session K , and continues with the same cycle five times. The value of B_{max}^k is then recalculated at the end of the fifth cycle. The recalculation is done to accommodate any new session that may have arrived.

The multiplexing gain for each of these policies at different stages of the network are given in table 7. The best

Policy	Stage	Multiplexing Gain
GDS	1	16.2
GBS	1	17.381
GDS	2	12.18
GBS	2	10.4
GDS	3	6.38
GBS	3	6.91
GDS	4	4.57
GBS	4	4.1

Table 7. Multiplexing gain for various Scheduling policies

multiplexing gain is obtained for the GBS. Now that the scheduling policies are defined, we go on to define the QoS transaction phase and the Call Admission Control policies

QOS RESERVATION POLICIES

Having obtained a source model for the incoming traffic, the reservation of resources per Virtual Connection is done based on the source model. Every switching node runs a dynamic bandwidth allocation policy based on the markov model. At the start of every session, a bandwidth equivalent to sustainable cell rate is allocated to the session for duration of one frame. This is the bandwidth that is guaranteed throughout the duration of the session. Subsequent frames are allotted a bandwidth dynamically depending on the transition probabilities between frames. The algorithm for QoS reservation policies is given in pseudo code given in table 8.

Steps 1 to 3 of the pseudo code form the CAC policy. An incoming session is accepted if and only if the SCR can be guaranteed throughout the duration of the session. The set S_α is the index set of all the active sessions and S_{low} is the

index set of all under allocated sessions. After every active session, the most probable bit rate B_i of the next frame is determined using the transition matrix. If the sum of all these new bit rates is less than the total bandwidth(TBW), every active session is allocated these new bit rates. If the sum is found to be more than the total bandwidth (step 11), we start from the session taking up the maximum bandwidth, and reduce it's bit rate to the corresponding sustainable bit rate (step 17 to 20). The above given algorithm is bound to allocate at least the sustainable cell rate to every active session per frame, and is stated as theorem 1.

Theorem 1 Every sessions is allotted a bandwidth at least equal to the sustainable cell rate for the complete duration of the session.

Following the algorithm in table 8 a maximum of 14 live video sessions could be admitted. The bandwidth utilization and the CLP for the maximum number of active sessions and for the various scheduling policies is summarized in table 9. Minimum cell loss probability is obtained

Policy	Cell Loss Probability ($\times 10^{-7}$)	Percentage Bandwidth Utilization
GDS	12.1	79.3
GBS	4.87	74.17

Table 9. CLP and bandwidth utilization for various scheduling policies

for GBS and the optimal bandwidth utilization is obtained for GDS.

CONCLUSION

In this paper we have proposed Entropy as a new measure to model VBR traffic, and a 16 state markov model is developed. The simulation results based on this model, carried out on benchmark video streams and traces indicate a Cell Loss Probability (CLP) of 6.27×10^{-7} and an average bandwidth utilization of 79.33% for a maximum of 11 active video sessions. The balance 21% bandwidth can be easily allotted for ABR and UBR traffic giving a near 100% utilization.

Scheduling policies also play an important role and have considerable effect on CLP. We have proposed two new scheduling policies, viz. Guaranteed Delay Schedule (GDS) and Guaranteed Bandwidth Schedule (GBS). The average bandwidth utilization of around 75 percent and a CLP of 4.87×10^{-7} is obtained for GBS. The corresponding values with GDS are 12.1×10^{-7} and 80% for CLP and average bandwidth utilization respectively.

A logical extension of this work is to model the multiplexed traffic based on this source model and fine tune the various scheduling policies being used so that optimal bandwidth utilization is obtained.

```

1  Wait for New Session (State 1);
2  On New Session ( $S_n$ ) Arrival with Sustainable Bandwidth ( $SBW_n$ );
3  If  $\sum_{j=1}^{j=n-1} (SBW)_j + SBW_n \leq TBW$  ;/* TBW is Total Sustainable Bandwidth */
4     Check BW Requirement Based on Qos and Accept Session,  $S_\alpha = S_\alpha \cup \{n\}$  ;
5  Else
6     Reject Session ;
7  For every active session
8     do
9         On Completion of frame transmission for session  $i$  ;
10        Find bit rate ( $B_i$ ) of next frame using transition table matrix;
11        if  $BW = \sum_{j=1}^{j=N} B_j \leq TBW$  ;
12            Allocate New Bandwidth;
13        Else
14            Set  $S_\alpha = \{1, 2, 3 \dots N\}$  ;
15            while( $BW \geq TBW$ )
16                do
17                    Find L such that  $B_L = \max_{i \in S} (B_i)$  ;
18                    Change  $B_L$  to Sustainable Cell Rate  $B_{sl}$  ; /*  $B_{sl}$  is Sustainable Bandwidth */
19                    Set  $S_{low} = S_{low} \cup \{L\}$  ; /*  $S_{low}$  = Set of under allocated Bandwidth Sessions */
20                end
21        end

```

Table 8. Pseudo code for call admission and bandwidth reservation

References

- [1] "ATM Forum Standards", 1996 (<http://www.atmforum.com/standards/accept/>)
- [2] "Mpeg Forum Standards Document", 1997 (<http://www.mpeg.org/standards/>)
- [3] Martin de Prynker, "Asynchronous Transfer Mode, Solutions for Broadband ISDN", 1992 Publisher Ellis Horwood.
- [4] Abdelnaser Adas, "Traffic Models in Broadband Networks" *IEEE Communications Magazine*, July 1997, pp. 82-89.
- [5] Marwan Krunz and Satish K. Tripathi, "Modeling Bit Rare Variations in MPEG Sources", *Technical Report, Computer Science Dept. University of Maryland*.
- [6] Sanjeev Rampal, D.S.Reeves and Dharma P. Agrawal, "End-to-end guaranteed QoS with statistical multiplexing for ATM networks" *Technical Report, North Carolina State University*. (http://www.ece.ncsu.edu/cacc/tech_reports).
- [7] Harry G. Perros and Khaled M. Elsayed, "Call Admission Control Schemes a Review", *Technical Report, North Carolina State University*. (http://www.ece.ncsu.edu/cacc/tech_reports).
- [8] Daniel P. Heyman and T.V Lakshman, "Source Models for VBR Broadcast Video Traffic" *IEEE/ACM Trans. on Networking*, Vol.4, No.8, Feb.1996, pp.40-49.
- [9] H. Shannon, "Information Theory and Coding", 1980 *Springer Verlag*.
- [10] Daniel P. Heyman Ali Tabatabai and T.V Lakshman, "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.2, No.1, Mar.1992, pp. 49-59.
- [11] David M. Lucantoni, Marcel F. Neuts and A.R. Reibman, "Methods for Performance Evaluation of VBR Video Traffic Models", *IEEE/ACM Trans. on Networking*, Vol.2, No.1, April.1994, pp.176-180.
- [12] Will E. Leland, Walter Willinger et al, "On Self Similar Nature of Ethernet Traffic", *IEEE/ACM Trans. on Networking*, Vol.2, No.1, Feb.1994, pp.1-15.
- [13] N.G.Duffield, J.T.Lewis et al, "Entropy of ATM Traffic streams: A Tool for Estimating Qos Parameters" *IEEE Journal Of Selected Areas in Comm.* Vol. 13, No.6, August 1995, pp. 981-990.
- [14] Ferit Yegenoglu, B. Jabbari et al, "Motion-Classified Autoregressive Modeling of Variable Bit Rate Video", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.3, No.1, Feb.1993, pp-42-53.
- [15] L. Klienrock, "Queuing Theory and Applications", *McGraw Hill.*, 1980.