# Markovian Models for the Output of Real Word MPEG-1 Encoders

Nikolaos Doulamis, Nicolas Tsapatsoulis, Anastasios Doulamis, Athanasios Tsiodras and Stefanos Kollias

National Technical University of Athens, Computer Science Division
Zographou 15773, Athens, Greece
e-mail ntsap@softlab.ece.ntua.gr

## Abstract

In this paper we survey the output of real world MPEG-1 encoders. This output consists of packets whose sizes are multiple of 20Kbytes, while the arrival intervals are mainly gathering around 220 msec. Two markov models are proposed. The first model, which is called model A, takes into account only the packet sizes ignoring the arrival intervals. The second model (model B) quantizes the arrival intervals so as to better approximate them. Finally a combination of the two models is proposed (hybrid model) so that it can better approach the traffic behavior.

## 1 Introduction

Digital video is anticipated to be an important traffic component over broadband integrated networks. Since video applications demand very large bandwidth requirements, several coding algorithms have been proposed in order to perform efficient and effective compression while maintaining a good quality. Thus the required bandwidth to transmit coded digital video is reduced. One of the foremost coding algorithms is MPEG-1 (Moving Picture Expert Group 1) which has been standardized in 1992 and now appears in applications.

Statistical analysis and modeling of video sequences are very useful tasks since one can estimate loss probability, transmission delay, bandwidth allocation and high-speed statistical multiplexing gain. So far, the proposed models deal with the raw video sequence without taking into account the way in which encoders put the data on a transmission line [1] [2] [4].

In general, two main different modes are used for encoding any video source, namely Constant Bit Rate (CBR) and Variable Bit Rate (VBR). In a CBR scheme the video quality cannot be maintained constant for all scenes, since a resolution-reduction mechanism is activated in case of high video activity in order to achieve the desired constant bit rate. On the contrary in a VBR mode, the output bit rate is not constant but it can fluctuate depending on the video activity [5]. However, in this case the quality of the produced video stream is constant. For real world applications the CBR mode is the most frequently used; therefore it is reasonable to concentrate on this mode.

While in theory, MPEG-1 encoders produce I, P and B frames, what is actually seen at the output is packets of video data. Usually, these are constant size packets, which come out of the encoder at regular time intervals. Most of the time, though, they are not produced at an exact and constant rate. This is expected, since the rate control mechanism cannot produce an exact output rate of the requested value. It accomplishes a constant mean output rate, but at different moments, the throughput can be different. In MPEG-1 encoders which produce packets of constant size, this forces the packets to arrive at different intervals. It is exactly this process that we will emulate using the proposed models. The models have been tested with the output of a certain MPEG-1 encoder [1] with great success.

[1]This work was done in the framework of NIKA (a National project of biomedical image coding, funded by Greek Secretariat of Research and Technology

## 2 Real world MPEG-1 encoders

MPEG-1 is a well known standard used for digital video compression. To compress the digital video data, the MPEG-1 algorithm removes visual information that humans cannot perceive. It accomplishes this using the available bits to encode information at a different level of accuracy for each spatial frequency. This is due to the way the human eye works: it is much more sensitive at low spatial frequencies than at high ones. Besides encoding different spatial frequencies at different levels of accuracy, MPEG-1 attempts to remove the temporal redundancy of the video stream, by cutting video frames down to macroblocks (squares of 16 by 16 pixels) and trying to match each frame's macroblock with another one at a nearby frame. If it succeeds at this matching, it stores a vector, pointing how much the original macroblock must be moved to appear at the new location in the new frame, and encodes the error of this prediction. The first kind of coding is used in standalone frames, called I-frames (intraframes). The second one (motion estimation) is then used to produce frames that rely on other frames and motion vectors. These frames are called Predictive (P) and Bidirectional (B) respectively [5] [3]. The results of these techniques are finally entropy coded using Huffman compression.

At a first glance, the MPEG-1 algorithm does not seem to be able to produce a throughput of constant bit rate. Not only that, but the output of the algorithm is not even predictable, since the motion estimation phase can be very or hardly successful, depending on the nature of the video to be encoded. A highly active video sequence, containing a lot of movement - a music video clip, for example - will give the algorithm a hard time finding any motion vectors. Thus, the motion estimation phase will fail in a large portion of the frames, causing explicit coding of all the macroblocks that could not be matched. This will of course increase the output rate, and it is quite a problem when one considers transmitting the encoded video information through a network. This is why several feedback mechanisms have been incorporated in the algorithm to control the outcome of the encoding process. The main idea is to vary the quality of the output appropriately, so that the final throughput is almost constant.

Modern MPEG-1 encoders, have this schemes implemented in hardware. However, the output is not exactly constant. It still fluctuates, but at a smaller scale. This is why a study of the output is of interest. We will present models of the output of MPEG-1 encoders, so that we can compute traffic characteristics, vital to the modern communication networks (B-ISDN). With the models we will shortly introduce, one will be able to calculate essential buffer lengths to have a specific loss probability, which is of outmost importance in quality of service assessments.

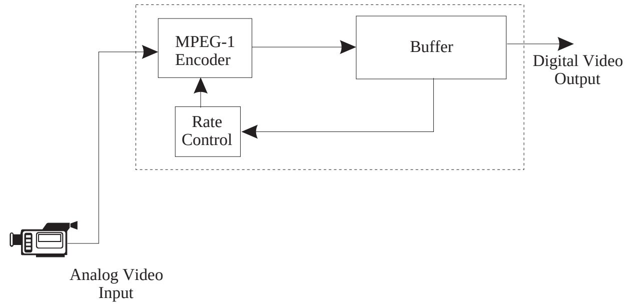


Figure 1: Rate control of MPEG-1 encoder

In real world implementations of MPEG-1 encoders, the scheme which is illustrated in figure 1 is being used. A camera or some other source of analog video is used to provide data to the encoder. The encoder is configured to produce an output at a desired bitrate. This is accomplished through two mechanisms: a rate controller and a buffer. The rate control mechanism monitors the output of the MPEG-1 encoder and dynamically reconfigures the coding parameters to keep the output rate near the requested rate [5]. Practically, this means lowering the accuracy of the DCT coefficients coding if the output bitrate exceeds the one requested. Actually, the rate controller monitors the portion of the buffer which is full of video data. Based on this value, it decides whether it is useful to reduce the image quality or not. It also works the other way: if the output bitrate falls below the one requested, the rate controller increases the image quality.

## 3 Statistical analysis for output of real world MPEG-1 encoders

Statistical models of video sequences depend on the kind of video stream which we are going to analyze. Different coding algorithms (such as with or without motion compensation, different operational modes of the encoder -like CBR or VBR- different policies that the encoder uses in order to produce the packets on a network line, constant or variable packet size) affect the statistical characteristics and the traffic behavior of the produced video stream. Since these models of video sequences rely on the statistical properties of the produced signal it is therefore anticipated that all the above parameters have a strong influence on the modeling. As a result before performing a statistical model one should understand the behavior and the basic properties of the examined signal [3].

In our study the analog video sequence that fed the MPEG-1 encoder was a 30 min ordinary TV program in which parts of film as well as advertisements were included. The bit rate was fixed at 1.5Mbits/sec. Figure 2 presents the histogram of packet arrivals for the output of the MPEG-1 encoder. All the packets which were produced had constant size and this was equal to 20Kbytes. It is observed that the most frequent intervals occur at 0 and 220 msec while the frequency of the others is very small. However there is a small variation around the value of 220 msec and an isolated part around 275 msec. The last one is somehow far from the value of 220 msec and occurs at times where the average throughput exceeds the desired constant bit rate. In this case the rate control mechanism delays to get the packet out so as to reduce the throughput. No packet arrives at the middle region of the histogram, i.e., between the time intervals around the 0 msec and 220 msec value. Due to the uncommon shape of this histogram there is no well known probability density function (pdf) that can approximate it.

The mean value of the time intervals is 109.25 msec and since the packet size is considered constant and equal to 20Kbytes, the bit rate results indeed in 1.5Mbits/sec. However, if we ignore the lower time intervals (around 0 msec) the average value is 220.81 msec. We also split the upper region into three zones, one of intervals less than 217 msec (we call this zone T-), the other of intervals above 225 msec (the T+ zone) and the last one of the rest time intervals (the T zone). The reason of this separation will be understandable at section 4 where the markov models are analyzed (the basic reason is to approximate the behavior of this region without ignoring the variation around the 220 msec value). This variation affects the traffic characterization at large buffer size. As we can notice from table 1 the average value of the T- , T+ and T zones are 212, 238, and 220 respectively. Table 1 also presents the standard deviation of the arrival intervals. The values of these deviations are rather small. Nevertheless, the deviation of the T+ zone is greater than the others due to the isolated part around 275 msec.

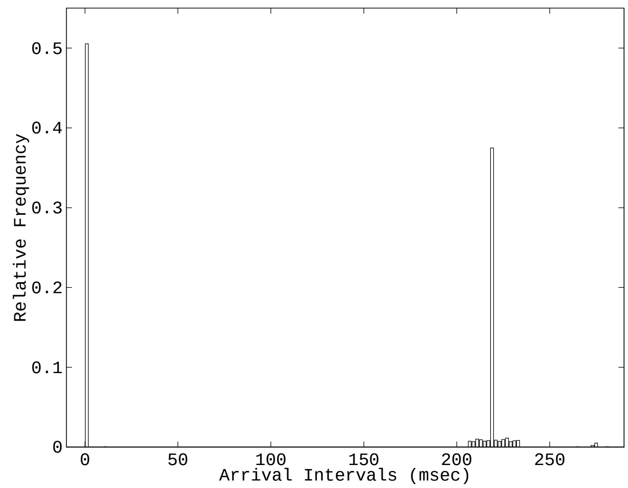


Figure 2: Histogram of arrival intervals considering constant packet size (20Kbytes)

| | T- zone | T zone | T+ zone |
|---|---|---|---|
| mean | 212 | 220 | 238 |
| std | 3.04 | 0.94 | 17.4 |

Table 1: Statistics for the three zones of the arrival intervals.

Actually the intervals around 0 msec do not exist. They occur only if the packets are double or triple or so on. Therefore if an interval which belongs to the upper region (approximately above 200 msec) is followed by a zero or so arrival interval, it means that a double packet has arrived. Two successive zeros mean a triple packet and so on. That is, there are times where the output of the MPEG-1 encoder produces packets size multiple of 20Kbytes. Figure 3 illustrates arrival intervals considering multiple size packets of 20Kbytes. Consequently, the relative frequency of this histogram is not the same as the previous one.

In figure 4 we present the packet size versus

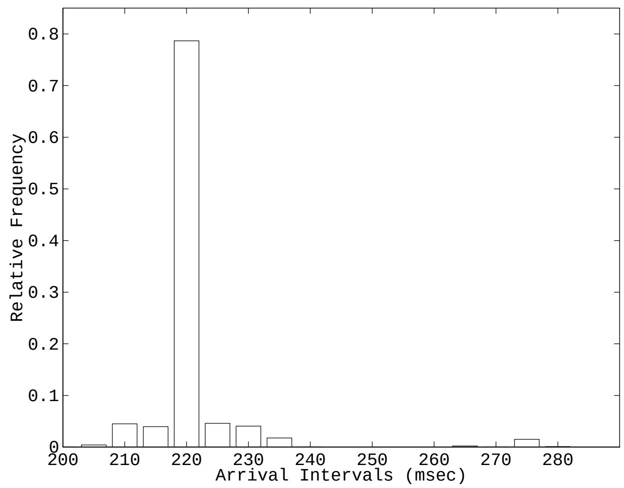


Figure 3: Histogram of the arrival intervals considering multiple size of packets

the time considering the case that time intervals around 0 msec indicate multiple size packets. The histogram of the packet size under the previous consideration is illustrated in figure 5. It is observed that the packets are only of size one, two and three times multiple of 20Kbytes for this sequence. However it is possible if the desired output rate is greater than 1.5Mbits/sec (e.g. 3Mbits/sec) for the MPEG-1 encoders to produce packets whose sizes are four or five times greater than 20Kbytes. In this paper we concentrate on MPEG-1 encoders whose output rates are about 1.5Mbits/sec. Nevertheless, the following proposed markovian models can be easily modified so as to include different output rates (see section 4.2). It is also deduced from figure 5 that the relative frequency of 40Kbytes packets is significantly greater than that of 20 or 60Kbytes packets. Thus, the video stream consists of a great deal of 40Kbytes packets which considerably affect the loss probabilities and the traffic characterization.

The rate control mechanism affects the arrival of a packet on a transmission line. The policy of this mechanism is to maintain the output bit rate constant. Consequently, when the arrival interval of a packet is greater than 220 msec (reduction of the total bit rate) the following arrival interval is usually less than 220 msec (increasing of the total bit rate). Figure 6 presents the previous results. However, sometimes peaks appear alone in figure 6. This occurs when the rate control tries to reduce the throughput of the encoder so as to achieve the desired bit rate.

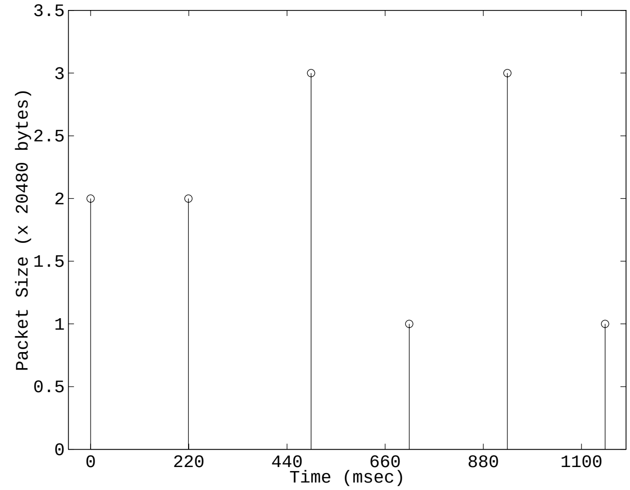


Figure 4: Data sequence considering multiple packet size

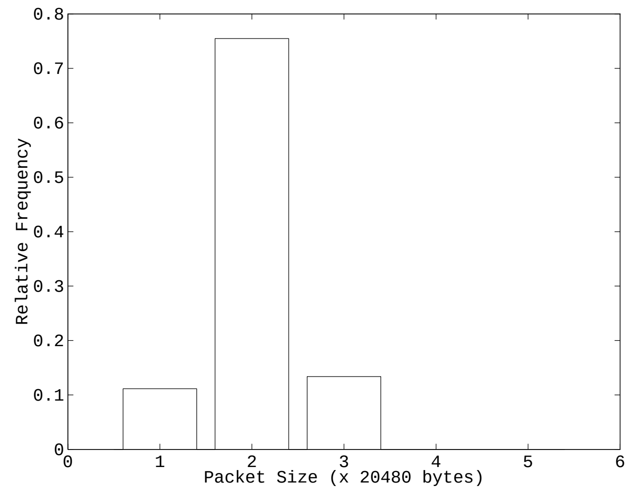


Figure 5: Histogram of the packet size

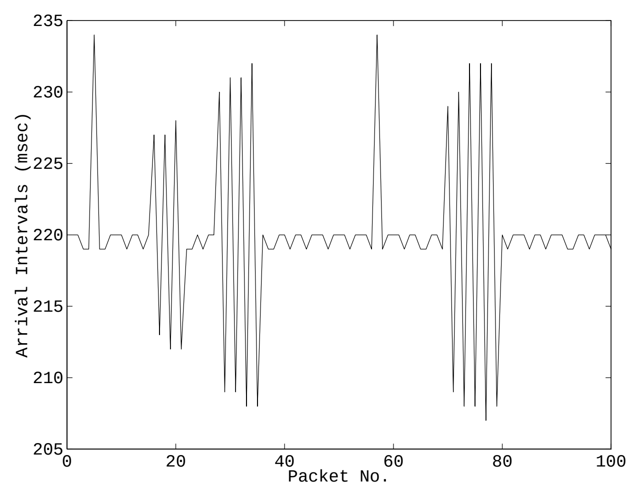


Figure 6: Time between successive packet arrivals

# 4 The proposed models

## 4.1 Background

Consider a discrete time stochastic process $X_n$. We assume that it takes values from a set of non negative integers, so the states that the process can be in are i=0,1,... We call this process a markov chain if whenever it is in state i, there is a fixed probability $P_{ij}$ that it will next be in state j regardless of the process history prior to arriving at i. That is

$$P_{ij} = P(X_{n+1} = j | X_n = i, ..., X_0 = i_0) = \\ = P(X_{n+1} = j | X_n = i) \quad (1)$$

We call $P_{ij}$ the transition probabilities and they should satisfy the conditions

$$P_{ij} \geq 0, \quad \sum_{i=0}^{i=\infty} P_{ij} = 1, \qquad \text{i=0,1,...}$$

The transition probabilities formulate the transition probability matrix which is denoted as

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & . & . & . \\ P_{10} & P_{11} & P_{12} & . & . & . \\ P_{20} & P_{21} & P_{22} & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \end{bmatrix} \quad (2)$$

We concentrate on finite number of states, however similar results can be deducted in the case of infinite number of states. If we symbolize with $P^n$ the n-step transition probabilities that is

$$P_{ij}^n = P(X_{n+m} = j | X_m = i), \; n, m \geq 0, \; i, j \geq 0.$$

The Chapman Kolmogorov equations provide a method for calculating $P_{ij}^n$, since a markov chain satisfies the condition

$$P_{ij}^{n+m} = \sum_{k=0}^{k=\infty} P_{ik}^n P_{kj}^m, \quad (3)$$

with non negative n,m,i,j.

It can be seen that $P_{ij}^n$ are the elements of the matrix $P^n$.

If the markov chain is irreducible and aperiodic then the following equation which is called global balanced equation holds and estimates the probabilities in the steady state $p_i$.

$$p_j = p_j \sum_{i=0}^{i=\infty} P_{ji} = \sum_{i=0}^{i=\infty} p_i P_{ij} \quad (4)$$

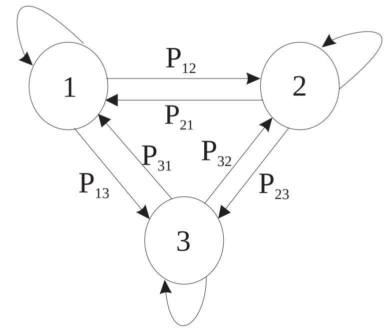


Figure 7: Proposed model A

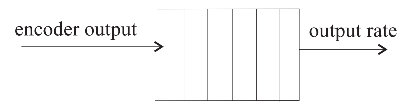


Figure 8: Buffer's subsystem

More details about Markov chains can be found in [6]

An estimation of the transition probabilities is given by [1] [4]:

$$\hat{P_{ij}} = \frac{number \;\; of \;\; tranitions \;\; i \;\; to \;\; j}{number \;\; of \;\; tranitions \;\; out \;\; of \;\; i} \quad (5)$$

## 4.2 Markov model A

In following when we refer to "buffer" we mean a FIFO subsystem located after the output of the MPEG-1 encoder. The output rate of this subsystem should be constant and greater than the average input rate. For our simulation results the input rate was 0.97 of the output one (therefore utilization is 0.97). If a packet could not totally insert to the buffer, the last 20Kbytes are lost [4]. A scheme of this subsystem is shown in figure 8.

Model A assumes that the time arrivals of the output signal are constant. The small variation of the time around the average value of 220 msec is ignored. In this consideration the packet size is not constant but multiple of 20Kbytes. The states of the markovian chain correspond to the packet size. However, the probability of a packet size being greater than 3x20Kbytes is negligible. Therefore, only three states for the markovian chain are used. State 1 corresponds to packet size of 20Kbytes, state 2 to packet size of 40Kbytes and state 3 to packet size of 60 Kbytes. In figure 7, the proposed model A is illustrated.

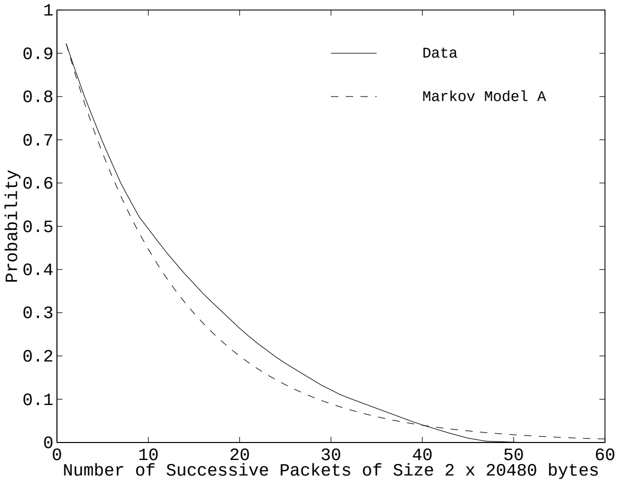


Figure 9: Probability distribution of n successive packets of size 2x20Kbytes for data and model A

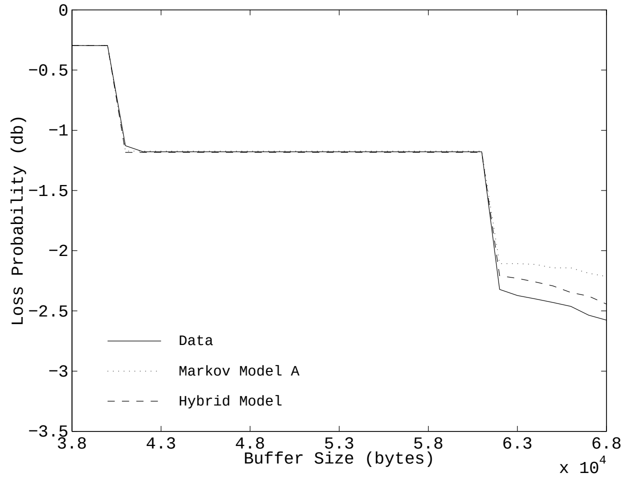


Figure 10: Loss probability for data, model A and hybrid model

Notice that bit rates higher than 1.5 Mbits/sec can be modeled using more states since in this case packets exist of size greater than 3x20Kbytes. However, 1.5 Mbits/sec is a typical bit rate for MPEG-1 encoders operating in CBR mode.

The transition matrix P for the Model A, calculated using the input video sequence and the definition given in equation 5 is as follows:

$$P = \begin{bmatrix} 0 & 0.150 & 0.850 \\ 0.025 & 0.923 & 0.052 \\ 0.688 & 0.312 & 0 \end{bmatrix} \tag{6}$$

Apart from the transition matrix, of great interest is the probability $p_j$ of state j (j=1,2,3) [6]. If $\mathbf{\Pi}$ is the column vector of the steady state probabilities $p_j$, i.e., $\mathbf{\Pi} = [p_1 \quad p_2 \quad p_3]^T$ then:

$$\mathbf{\Pi} = \mathbf{P}' * \mathbf{\Pi} \tag{7}$$

subject to $\sum_{i=0}^{i=3} p_i = 1$. The solution of the above equation gives:

$$[p_1\ p_2\ p_3]\text{=}[0.111\ 0.756\ 0.133]$$

and approximates perfectly the state histogram of the input data (see figure 5).

From the values of $p_j$ (j=1,2,3) it is observed that, by far, the most probable packet size is that of 2x20 Kbytes. In addition, from matrix P arises that the most probable transition, from state 2, is to state 2 with probability $P_{22}$. In the real data $P_{22}$ expresses the probability of the next packet being of size 2x20 Kbytes while the current is of the same size. A crucial point is whether the probability $s_2(n)$ of n successive packets of size 2x20 Kbytes, has exponential distribution, i.e., $s_2(n) = a^n$, as the model assumes. In figure 9 it is illustrated that $s_2(n)$ can be indeed approximated by an exponential distribution $s_2(n) \simeq P_{22}^n = 0.923^n$ [6]. Due to the small probabilities $P_{11}$ and $P_{33}$, a similar analysis for the state 1 and state 3 is of no interest.

The more important test for the efficiency of Model A is the simulation test. In this test the Model A produces packets of different sizes, starting from an arbitrary initial state, using the transition matrix P. In figure 10 the loss probability is shown (in db) as a function of the buffer size. This function is expected to be a step one for small buffer sizes, due to the packetized output of the MPEG-1 encoder. For example a buffer size of 43 Kbytes is smaller than a packet of size 3x20Kbytes so all packets of this size will be damaged. Model A approximates perfectly the loss probability for buffer sizes smaller than the maximum packet size, i.e., 3x20 Kbytes. For buffer size larger than the maximum packet size the loss probability is due to the arrival intervals distribution of the packets (figure 3). Model A assumes constant arrival intervals and is not expected to approximate this loss probability; actually overestimates it. Therefore a model for the arrival intervals distribution must be introduced so as to approximate the loss probability for large buffer sizes.

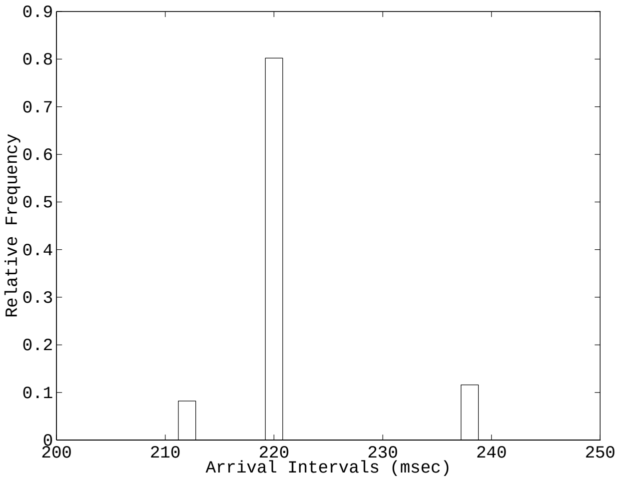


Figure 11: Histogram of the arrival intervals produced from model B

## 4.3 Markov model B

In model B (figure 12), a three state markovian chain is proposed so as to approximate the arrival intervals distribution. The states of the markovian chain correspond to the time in which the next packet will arrive. Actually, only three kinds of time intervals are considered, the time intervals between 217 and 225 msec (state T), the time intervals which are greater than 225 msec (T+ state) and the time intervals which are smaller than 217 msec (T- state) respectively. Thus, the state T represents the case in which the next packet arrives after T msec where $T \in$[217 225] msec. The T+ state represents the case in which the next packet arrives after time greater than 225msec and the T- state after time smaller than 217 msec. If a packet arrival occurs, the markovian chain will define the time of the next arrival. In this way, the model can approximate the arrival times regardless of the exact value of the packet size. The transition matrix P for the model B, calculated using the input video sequence and the equation 5, is as follows:

$$P = \begin{bmatrix} 0 & 0.291 & 0.709 \\ 0.009 & 0.931 & 0.060 \\ 0.632 & 0.304 & 0.064 \end{bmatrix} \qquad (8)$$

The probabilities $p_T$, $p_{T+}$ and $p_{T-}$, of the tree states are calculated using equation 6 and produce a histogram of the arrival intervals which is shown in figure 11. A comparison with the real data is given in Table 2.

The most probable state is state T with probability $p_T$=0.812 while the $p_{T+}$, $p_{T-}$, are 0.111 and 0.077 respectively. In state T the transition with the highest probability is towards state T.

|  | Arrival Interv. in T- zone (%) | Arrival Interv. in T zone (%) | Arrival Interv. in T+ zone (%) |
|---|---|---|---|
| Real Data | 7.55 | 81.51 | 10.94 |
| Model B | 7.71 | 81.23 | 11.06 |

Table 2: Relative frequencies for real data and simulation results of model B

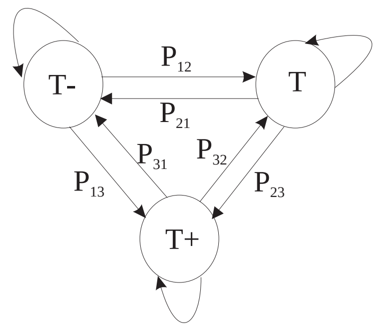


Figure 12: Proposed model B

This transition denotes successive arrivals in the interval [217 225] msec. As it can be seen in figure 13 the probability $s_T(n)$ of n successive arrivals in [217 225] msec has exponential distribution as Model B assumes. Since the state probability $p_T$ is by far the highest and the transition probabilities $P_{11}$ ( T- to T-) and $P_{33}$ (T+ to T+) are very small, there is no necessity to examine the behavior of the probabilities $s_{T-}(n)$ and $s_{T+}(n)$ to evaluate the efficiency of Model B [6].

## 4.4 Hybrid model

The hybrid model (figure 14) combines Model A and Model B so as to estimate both the size and the arrival time of the next packet. In this model the Model A and Model B work independently, i.e., the estimation of the next packet size is based only on Model A and the estimation of the arrival time is based on Model B. A fully connected nine-state markov model has slightly better performance but much more complexity. In the simulation test the representative arrival intervals for the states T-, T and T+ are considered the mean values of the respective zones (see section 3)

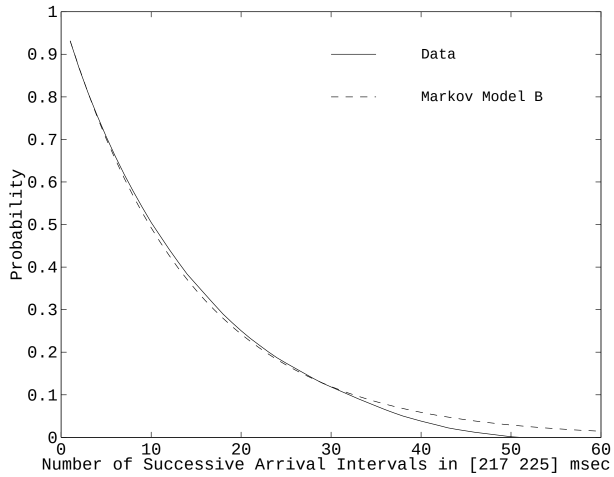


Figure 13: Probability distribution of n successive arrivals in [217 225] msec for the data and model B

The results of the simulation test for the hybrid model are also illustrated in figure 10. As it can be observed, while the hybrid model approximates the loss probability for large buffer sizes much better than the Model A, the performance of the two systems is the same for small buffer sizes. It is clear that the loss probability for small buffer sizes is mainly due to the packetized output of the MPEG-1 encoder . On the other hand for large buffer sizes the loss probability is mainly due to the arrival intervals distribution of the packets.

# 5 Conclusion

Two markovian models are proposed for the traffic characterization of real MPEG- 1 encoders. These models simulate the real data, approximating the statistical properties of the video signal as well as the loss probability. The first model (model A) assumes that the arrival intervals are fixed and thus it gives an over estimate at large buffer sizes. In order to take into account the variation of t he arrival intervals a second markov model is proposed (model B). Better results for the loss probability at large buffer sizes are deducted using a hybrid mode l which combines the previous ones. To reduce the number of the estimated parameters in the hybrid model, the two mo dels (model A and B) are considered to be independent. As a result only twelve parameters are required (six for model A and six for model B). A simple model t hat correlates arrival intervals and packets sizes is a field for further work, so as to achieve more accurate results.

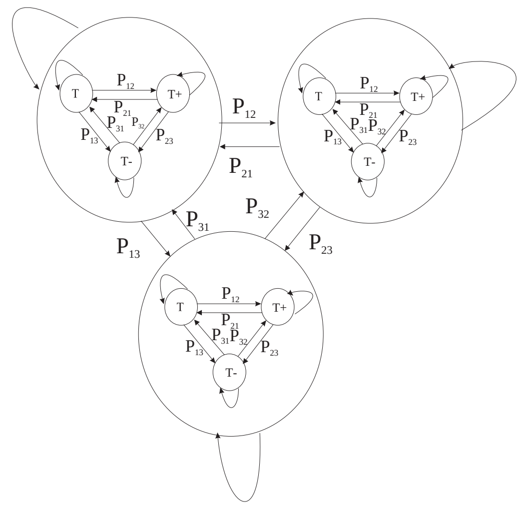


Figure 14: Hybrid model

# References


[1] Heyman, A. Tabatabai and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks," *IEEE Trans. On Circ. and Syst. for Video Technol.* , vol. 2, pp. 49-59, Mar. 1992.

[2] Daniel P. Heyman, Ali Tabatabai, and T. V. Laskhman, "Statistical Analysis of MPEG2-Coded VBR Video Traffic, " *PV'94 B2.1*

[3] Pramod Pancha and Magda El Zarki, "Bandwidth-Allocation Schemes for Variable - Bit - Rate MPEG Sources in ATM Networks," *Trans. On. Circ. and Syst. for Video Technol.*, vol. 3, pp.1991-19 8, Jun. 1993

[4] A. D. Doulamis, N. D. Doulamis, G. E. Konstantoulakis, G. I. Stassinopoulos, "Traffic Characterisation and Modelling of VBR Coded MPEG Sources," *IFIP 4th Intern. Work. on Performance Modelling and Evaluation of ATM Network*, U.K., July 1996

[5] ISO/IEC 13818-2, " Generic Coding of Moving Pictures and Associated Audio," *Recommendation H.262,* Committee Draft May 1994.

[6] Dimitri Bertsekas, Robert Gallager, Data Networks Prentice-Hall Inter. Editions