Predictive shaping for VBR MPEG video traffic transmission over ATM networks

L. de la Cruz, J. J. Alins and J. Mata

Department of Applied Mathematics and Telematics
Polytechnic University of Catalonia
C/ Jordi Girona, 1 i 3. Mòdul C-3, Campus Nord
08034 - Barcelona, SPAIN
E-mail : {ljcruz,juanjo,jmata}@mat.upc.es
Phone : +34-[3] 401 60 14 - Fax : +34-[3]401 59 81

Abstract

The use of smoothing techniques to remove the periodic fluctuations of the bit rate generated by the codification modes of the MPEG algorithm is very suitable in video transmission. In this way, the multiplexing gain is maximized and the resource allocation is reduced in ATM Networks. The traffic smoothing can be achieved storing the cells in a buffer. This buffer is allocated between the coder and the user-interface. To reduce the delay introduced in the storage process a new technique to forecast the VBR MPEG traffic is presented. This technique is based on the characterization of bits per frame generated by the MPEG coder as an ARIMA process. In this study the invariance of the ARIMA coefficients is verified for all coded sequences used. In addition, these coefficients are invariant also in front of the changes of the selected image quality in the coder. This characterization allows to propose a new traffic shaper scheme when forecast techniques are applied. Moreover, numerical results allows to compare the smoothing effects introduced, as well as the delays for the classic shaper and the predictive shaper.

Keywords

ATM networks, MPEG video traffic, ARIMA process, traffic shaping

Performance of Information and Communication Systems U. Körner & A. Nilsson (Eds.)

1 INTRODUCTION

Broadband Networks based on the Asynchronous Transfer Mode (ATM) will support, among others, traffic coming from variable bit rate video (VBR) coders, which are capable of maintaining a constant picture quality of the decoded image. The characterization of such VBR video sources becomes important in the analysis and design of Broadband Integrated Services Digital Networks (B-ISDN). The network architecture and its characteristics, such as cell-loss probabilities, transmission delay, statistical multiplexing gain, buffering, etc., are strongly related to the statistical properties of the sources and the coding schemes involved. Therefore, source models are useful to analyze and to dimension the network components (Nikolaidis, 1992)(Mata, 1996)(Mata, 1994).

On the other hand, a characterization of the traffic generated by a VBR source is necessary in order to allocate resources in ATM networks, as well as to keep a satisfactory quality of service (QoS). In the call establishment phase, service requirements are negotiated between the user and the network to establish a Traffic Contract. The source traffic parameters used to specify the statistical properties are the Peak Cell Rate (PCR), Sustainable Cell Rate (SCR) and Burst Tolerance (BT). The Generic Cell Rate Algorithm (GCRA) is used to provide a formal definition of the traffic conformance. This algorithm depends only on the increment parameter (I) and the limit parameter (L).

The MPEG coding algorithm was developed to achieve a high compression ratio with a good picture quality. MPEG can be used to transmit real-time variable bit rate broadcast video and it is suitable for video-on-demand in ATM networks (Pancha, 1994).

MPEG has two main coding modes: interframe mode and intraframe mode (I). In its turn, two types of frames can be distinguished by the interframe mode, predicted (P) and bidirectionally-predicted (B) frames. The Intra coded frames (I) are coded without any reference to other frames. Predictive coded frames (P) are coded using motion compensated prediction from a past I or P frame. This implies a more efficient coding. Bidirectionally-predicted coded frames (B) provide the highest degree of compression using the previous and next I or P as a reference. A video sequence of pictures (SOP) is divided into groups of N pictures (GOP). A GOP consists of subgroups of M pictures where the first is a reference picture, intra or predicted, and the rest are bidirectionally-predicted. The image quality depends on the values M, N and the selected quantizer step size (Q).

Four levels of coding can also be considered: picture, slice, macroblock and block. A picture (or frame) is a basic unit of display. The frame size in pixels depends on the application. A slice is a horizontal strip within a frame. According to the MPEG standard, each frame is divided into slices of 16 pixels of width, which implies that the frames are divided into 18 slices. A macroblock consists of four 8x8 blocks of luminance pixels and two 8x8 chrominance blocks. The smallest unit is a block which is a 8x8 matrix of pixels.

MPEG codec can be set in an open-loop mode to maintain the subjective quality with a fixed Q, and the coded variable bit-rate (VBR) output stream is delivered to the network. A suitable choice of Q, M, N parameters is important to minimize the traffic bit rate for a fixed subjective quality or for a constant signal-to-noise ratio (SNR). These parameters have to be selected to minimize the traffic rate for a constant signal to noise ratio (Mata, 1996).

The variations of the bit rate generated in the codification are produced by intrinsic and extrinsic reasons. The extrinsic ones are produced by the changes of the complexity and activity of the sequence to be coded. The intrinsic reasons are related, fundamentally, to the codification modes applied on the frames. Thus, the I frames need a higher number of bits than the frames P or B because the I frames only exploit the spatial redundancy using the DCT transform technique. In addition, the P frames tend to generate greater number of bits than B ones, since only motion compensation is applied respect to the previous reference image. Within the codification of the frames, another factor that give rise to variations of the generated bit rate is the exploitation of the entropy using run-length codes.

The extrinsic reasons that produce fluctuations in the bit rate depend on the content of the frames to code. The frames with greater grade of detail or greater texture have a high complexity level and reduce the efficiency of the spatial redundancy exploitation. The high activity scenes with fast camera movements, zooms and plane changes, avoid the use of the predictive compression technique. In this way, these scenes increase the binary rate with respect to smaller activity sequences.

In general, the coders do not deliver directly the traffic to the user interface because, usually, a smoothing system is enabled. The smoothing is carried out through a small storage buffer. The insertion of the buffer introduces a delay in the cells delivered to the network. The use of the smoothing allows to maintain a bit rate approximately constant during a time interval. The smoothing is applied to decrease the variability of the traffic and its peak rate. Likewise, the intrinsic periodic fluctuation of rate generated by the MPEG algorithm can be removed. In this way, the VBR MPEG traffic shaping allows to reduce the allocated resources to the virtual circuit. Moreover, the effect of the periodic arrivals to the multiplexers and switch fabrics are avoided. Therefore, the employment of the traffic shaping maximize the statistical multiplexing gain.

Most of the studies are focused on the modeling and the prediction of the rate generated in a frame interval. The main reason is that the human perceptive system does not appreciate a delay less than 100 ms though it is admissible until 200 ms (Garret, 1993). Therefore, the traffic shaper can introduce a delay of only several frames. At the same time, this delay allows to use the Bidirectionally-Predicted mode in the MPEG algorithm for interactive services (Kawashima, 1993).

Depending on the temporal requirements of the service all the generated cells for the GoP can be stored. Afterwards, the cells are delivered to the network at a constant rate during an interval of the same duration. For services with more restrictive temporal requirements is essential the reduction of the storage time (about 80 ms). This reduction makes necessary the application of prediction techniques. These techniques permit the reduction of the traffic source burstiness and to satisfy the temporal constrains. The smoothing in intervals of duration one GoP allows to extract the intrinsic variations of the rate introduced by the MPEG algorithm. In this way, the rate generated only depends on the complexity and activity of the scenes.

This paper is organized as follows. In section 2 the ARIMA process is revised for digital filter theory point of view. Analyzing the coder data traces, the VBR MPEG traffic is characterized as an ARIMA process in section 3. Likewise, the perfect capture of the compressed video traffic by the ARIMA process is shown for all the long and short sequences analyzed using residual diagnostic goodness-of-fit tests. This characterization is proposed to forecast the VBR MPEG traffic in section 4. In order to evaluate the temporal response of the predictor, its behavior is also studied in sudden scene changes. The invariance of the ARIMA coefficients for all the sequences analyzed allows to introduce a new traffic shaper for VBR MPEG video in section 5. Finally, the main results of this work are discussed in section 6.

2 THE ARIMA PROCESSES

These processes have been widely studied in the literature and in their more general form are denominated autoregressive, integrative, moving average processes (ARIMA) (Box, 1994). The autoregressive models are used in the context of sources of synthetic traffic or in traffic forecast for the generation of series of rates in intervals of fixed duration (Grunenfelder, 1991)(Yegenoglu, 1993). The ARIMA(p,d,q) models are decomposed in an autoregressive component of order p, an integrative component of order d and a moving average component of order q.

The autoregressive component reflects the dependence between the current generation and the last p generations. Thus, for an AR(p) process the values generated in a time series $y=(y_0,y_1,..,y_n)$ are obtained from the p past values and an independent factor from the times series. This factor can be modelled as a process with identically independent distributed values $W=(w_0,w_1,..w_n)$. The time series W are denominated residual series. These time series are considered as the prediction error of the following generation of the process. Customarily, the values from the series W are synthesized as the realization of a gaussian variable with an average and a standard deviation directly related with the corresponding moments of the process AR to generate. So that:

$$y(n)=a_1y(n-1)+a_2y(n-2)+...+a_py(n-p)+w(n), \qquad (1)$$

where the terms a_i are constant coefficients.

The MA(q) component of the process reflects the dependency in the generation of the past values of the independent process that contributes in the obtained value. In this way, a MA(q) process would be expressed as :

$$x(n)=b_0w(n)+b_1w(n-1)+b_2w(n-2)+...+b_qw(n-q), \quad (2)$$

where the terms b_i are constant coefficients.

The integrative contribution allows to capture the non stationarity of the moments of the stochastic process. Although the integrative component can be considered within the AR component by its formulation, its synthesis depends on different factors. Thus, the integrative component also shows the dependence with past values of the series but its synthesis depends on the non stationary moments of the process. The order d of the integrative component is fixed by the order of the highest non stationary moment of the stochastic process. In general, the integrative component can be expressed:

$$z(n)=c_1z(n-1)+c_2z(n-2)+...+c_dz(n-d)+w(n), \quad (3)$$

where:

$$c_i=\binom{d}{i}(-1)^{i+1} \qquad i\in\{1,2,...,d\}. \quad (4)$$

For example, a process whose mean is non stationary and the rest of high order moments are stationaries would have an integrative component of order 1. This integrative process is the so-called "random walk". If the behavior of the variance shows a clear trend during long intervals it is convenient to apply a transformation like the Box-Cox (Box, 1994).

The interpretation of a process ARIMA(p,d,q) can be carried out defining the delay operator z^{-1} (Proakis, 1983). So that, the general expression of an ARIMA(p,d,q) process can be expressed by its Z transform as:

$$Y(z)=[B(z)A(z)C(z)]\cdot W(z). \quad (5)$$

Understanding this expression as the relationship between the input w(n) and the output y(n) of a digital filter in a given instant n, the transfer function of the filter H(z) could be define as:

$$H(z)=\frac{Y(z)}{W(z)}=B(z)A(z)C(z). \quad (6)$$

Note that the roots of the polynomial B(z) correspond to the zeroes of the filter and the zeroes of $A^{-1}(z)$ and $C^{-1}(z)$ to the poles. According to the definition of the c_i values expressed in (4), the integrative order defines the multiplicity of the pole in z=1. This pole generates the instability of impulsional response. The rest of obtained poles (z_k) will be found in the unit circle ($|z_k|< 1$) of the Z plane. In Figure 1 a scheme of the ARIMA model is shown.

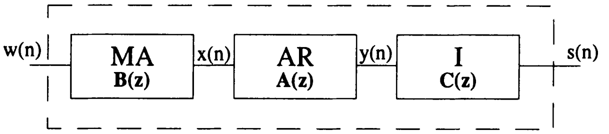

Figure 1. Components of an ARIMA process.

3 CHARACTERIZATION OF THE VIDEO TRAFFIC VBR MPEG AS AN ARIMA PROCESS

The temporal series of the VBR MPEG-I traffic presents a slow variation of the mean rate for several hundred of frames. This variation is related to the activity and complexity of the scene. The long range dependence complicates the development of a predictor because the temporal series shows an apparent non stationary mean. In order to synthesize a good predictor it is necessary to capture this long term effect. In this section, a new ARIMA model for the VBR MPEG traffic has been developed to find the predictor. Three sequences have been used to find and evaluate the predictor, "Live in Central Park" (by “América”), "Jurassic Park", and “Geografía de Catalunya” which are 34000, 174000 and 51000 frames long, respectively. The sequences have been coded using the parameters (Q =9, M= 2, N= 6), (Q= 6, M= 2, N= 6) and (Q= 6, M= 2, N= 6), respectively. The first and the second sequences presents the classic characteristics of activity and complexity, while the third has high complexity levels and short length scenes. Likewise, the results have been contrasted with the ones obtained for the sequence “Live in Central Park” coded with the set of parameters (Q =9, M= 2, N= 4).

In order to develop the ARIMA model, initially, the integrative component is found. The long term dependence produces that the mean rate varies slowly for several hundreds of frames. This variation reaches maximum and minimum levels which are very distant. However, the variance remains almost constant. This allows to conclude that the integrative component of the model should be of order 1 and its associated transfer function $C(z)=(1-z^{-1})^{-1}$.

In order to determine which are the values of the AR components and MA, it will be necessary to extract the integrative component of the actual process s(n). According to the scheme presented in Figure 1 the residual ARMA series y(n) and the real series s(n) are related as follows:

$$Y(z) = \frac{1}{C(z)} S(z) = (1 - z^{-1}) S(z) . \tag{7}$$

In this way, the temporal series y(n) will be obtained at the output of the FIR filter, whose transfer function is $(1-z^{-1})$, when it is excited with the temporal series generated by the coder. It can be checked that the temporal series y(n) is a stochastic process with mean 0 and an invariant autocorrelation coefficients. This statistical analysis has been carried out with the three sequences using blocks of 15000 frames and with autocorrelation lags of 100 units. The probability distribution function fits a gaussian distribution in all cases. The difference noted in the three temporal series is the standard deviation. This dissimilarity is related with the variability and complexity of the sequences.

The temporal series y(n) presents a seasonal behavior (Box, 1994) of period N= 4 or N= 6 according to the parameter chosen in the MPEG algorithm. Using the peaks of the autocorrelation function, which appear in multiples of N, the AR component can be synthesized. To determine the coefficients of the AR component Least Squares estimation has been employed. The order of the seasonal model found is 2.

The MA component can be analyzed when the AR component of the y(n) series is withdrawn. Using a FIR filter with transfer function $A^{-1}(z)$ the series x(n) can be obtained at the output of this filter when y(n) is applied at the input. To estimate the parameters of the MA process, least square estimation is applied to fit the partial autocovariance function of x(n). The best adjustment is obtained with a MA process with order 13.

The values of the AR and MA filter's coefficients, and a more detailed explication, can be found in (De la Cruz, 1997 b).

The integrative component of the obtained model has order 1. Thus, the integrative and the autoregressive components can be written together in the following way:

$$A'(z) = A(z)C(z) = A(z)\left(1 - z^{-1}\right) . \tag{8}$$

The generated series can be expressed as:

$$s(n) = b_0 w(n) + \cdots + b_q w(n-q) + a'_1 s(n-1) + \cdots + a'_{p+1} s(n-p-1) . \tag{9}$$

where the a'_i coefficients are obtained applying the inverse Z transform to A'(z).

4. VBR MPEG TRAFFIC PREDICTION

In this section, the ARIMA predicter for the bit rate generated by a VBR MPEG coder is developed. The predicter is based on the obtained ARIMA model. This prediction will be used in the next section to shape the traffic before deliver it to the network.

From (9), the (n+1) sample prediction is:

$$\hat{s}(n+1) = b_0\hat{w}(n+1) + b_1 w(n) + \cdots + b_q w(n-q+1) + a_1' s(n) + \cdots + a_{p+1}' s(n-p) \,. \tag{10}$$

Nevertheless, in the prediction context the values of the w(n) series are unknown. The predicter will have only the previous values of the s(n) series. Moreover, the $\hat{w}(n+1)$ is a future value. The forecast value of $\hat{w}(n+1)$ will be the mean value of the w(n) series. In this case, the mean value is 0. Thus, the (n+1) sample prediction can be written as:

$$\hat{s}(n+1) = b_1 w(n) + \cdots + b_q w(n-q+1) + a_1' s(n) + a_2' s(n-1) + \cdots + a_{p+1}' s(n-p) \,. \tag{11}$$

On the other hand, from (9) it is also possible to write:

$$\hat{s}(n) = b_0\hat{w}(n) + b_1 w(n-1) + \cdots + b_q w(n-q) + a_1' s(n-1) + \cdots + a_{p+1}' s(n-p-1) \,. \tag{12}$$

Subtracting (12) to (9):

$$s(n) - \hat{s}(n) = b_0\left(w(n) - \hat{w}(n)\right) . \tag{13}$$

As it has been mentioned, the forecast value of $\hat{w}(n)$ will be 0, so:

$$s(n) - \hat{s}(n) = b_0 w(n) \,. \tag{14}$$

Therefore:

$$w(n) = \frac{s(n) - \hat{s}(n)}{b_0} \,. \tag{15}$$

Replacing this value in the equation (16), the (n+1) sample prediction can be written as:

$$\begin{aligned}\hat{s}(n+1) = {} & \frac{1}{b_0}\left[b_1\left(s(n) - \hat{s}(n)\right) + \cdots + b_q\left(s(n-q+1) - \hat{s}(n-q+1)\right)\right] + \\ & + a_1' s(n) + a_2' s(n-1) + \cdots + a_{p+1}' s(n-p)\end{aligned} \,. \tag{16}$$

The derived ARIMA predicter is shown in Figure 2. The predicter supplies the estimated value for the (n+1) sample as a function of the n previous ones. This set of samples can be used also to obtain a prediction of the samples "2 ahead", "3 ahead", etc. Running the predicter with the (n+j) sample estimation, it supplies an estimated value of the (n+j+1) sample.

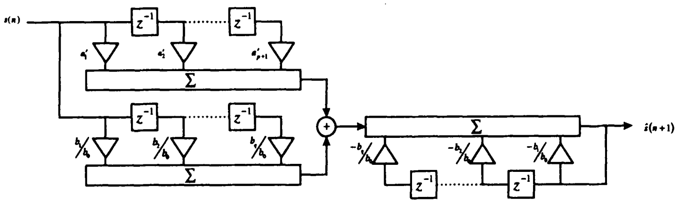

Figure 2. ARIMA Predicter.

In order to evaluate the behavior of the above transfer function, an analysis of the forecast errors has been done for all sequences. Figure 3 presents the residuals autocorrelation and the 99% confidence intervals. The residual diagnostic determines that the forecast errors are very uncorrelated. Therefore, the ARIMA model fits well the behaviour of the VBR MPEG traffic at frame level. This model could not be used to synthesize VBR MPEG traffic because the temporal series generated has a variant unbounded mean. In order to observe the temporal response of the prediction, a sudden scene change is analyzed. This response is shown in Figure 4.

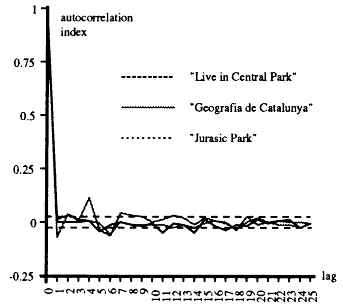

Figure 3. Autocorrelation function of the forecast errors using the ARIMA predictor.

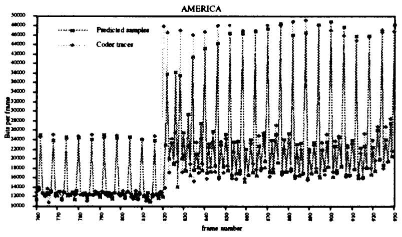

Figure 4. Unit step response of the ARIMA predictor.

5. VBR MPEG TRAFFIC SHAPING

The use of video compression algorithms like MPEG-I causes the variation of the generated bit rate for the images according to the coded mode I, P or B . In order to minimize the variability of the generated rate, traffic shaping is needed. This traffic shaping could be achieved for several images. To smooth the generated traffic, a number of images have to be stored in the buffer until the mean rate required is determined. This mechanism could only be applied to services which accept a transmission delay higher than the required time introduced in the stored process.

To avoid high delays for interactive services the use of prediction techniques is suitable. In this section a new VBR MPEG traffic shaper based on these techniques is introduced. It performance is studied for all the sequences under study. Moreover, it is compared with the classic storing systems. These systems are presented in the first place.

5.1 Storing systems

The most classic smoothing system is based on the storing of a number of frames. Later, the frames are delivered to the network at a constant rate. This rate will be the mean rate obtained for all the stored pictures. This kind of smoothing has been called "*ideal smoothing*" in previous works (Lam, 1996). Generally, the number of pictures used to calculate the mean rate is N, that is, the number of pictures in a GoP. Let S(n) be the n picture size. Thus, the ideal shaper will deliver the information of the previous GoP at the following rate:

$$r = \frac{S(n) + S(n+1) + \cdots + S(n+N-1)}{N\tau}, \tag{17}$$

where τ is the frame period.

The main disadvantage of this shaper is the introduced smoothing delay. Observe that for a given frame, the delay in the buffer can reach ($2N\tau$) seconds. For instance, in a system working at 25 pictures per second and N=6, a given picture can be delayed even 480 miliseconds. This delay can be admissible for broadcast services, but it is not suitable for interactive services. In Figure 5, the results of the ideal shaper for a section of the "Jurassic Park" sequence is shown. The selected section presents sudden transitions of scenes. The introduced delay is presented in Figure 6. The unit chosen to represent the delay has been the frame period.

Another kind of smoothing consist of update the obtained mean for every frame. That is, the picture n will be delivered to the network at the following rate:

$$r(n) = \frac{S(n-N) + S(n-N+1) + \cdots + S(n-1)}{N\tau}, \tag{18}$$

In this method, the rate is updated for every frame, so it is possible to call it "*sliding smoothing*". Note the main difference between the two methods presented: in the former, the window used to calculate the mean is static for every frame in a same GoP, while in the latter this window is varies for each frame. Figures about this kind of shaping can be found in (De la Cruz, 1997a).

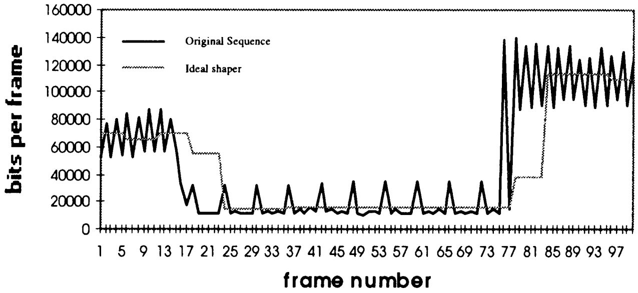

Figure 5. Ideal smoothing

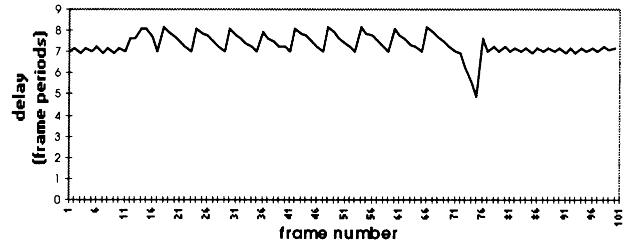

Figure 6. Delay introduced by the ideal smoothing

Some of the numerical results obtained are shown in Table 1. In this table, the two previous methods are compared. In both cases, there is a great reduction in the quadratic coefficient of variation, C_r^2, and in the burstiness, B_r. These parameters measure the variability of the rate, and they are defined as follows:

$$C_r{}^2 = \left(\frac{\sigma_r}{\bar{r}}\right) \quad ; \quad B_r = \frac{r_{peak}}{\bar{r}}. \tag{19}$$

In the worst case, B_r is reduced in a 20%. This will give rise to a considerable improvement in the resources allocation over ATM networks.

The main disadvantage for the two previous methods is the delay. For the sequences under study, the worst case presents a maximum delay of 10.2 frame periods, that is, 408 miliseconds for the systems working at 25 pictures per second. Note that the QoS will be fixed for this maximum delay.

On the other hand, the shaping goodness can be studied also in terms of the autocorrelation function. This function is shown in Figure 7 for the ideal smoothing. Results for sliding smoothing are very similar. In both cases, the extraction of the periodic peaks is verified. Therefore, both kind of shaping presents good characteristics for services without strong delay constraints. However, these shapers are not suitable for interactive services.

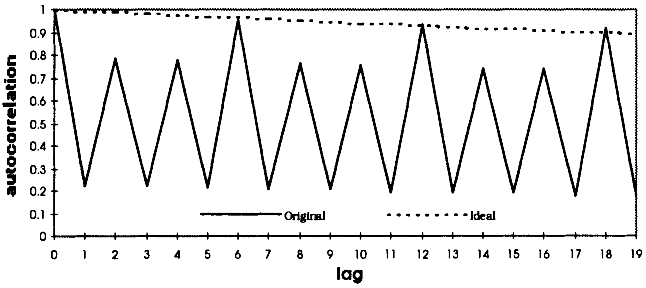

Figure 7. Ideal shaping autocorrelation

Table 1 Shaping methods comparison

Shaping	C_r^2	B_r	*Min. delay*	*Max. delay*	*Mean delay*
None	0.52	5.96	1	1	1
Ideal	0.27	4.73	4.5	9.6	7.4
Sliding	0.28	4.77	3.0	10.2	5.6
Predictive	0.28	4.77	1	3	2.03

5.2 Predictive shaping

In the previous section, two shaping methods have been studied. They are based on the stored samples. In this section, a predictive shaper scheme is presented. The main goal of the new scheme is to achieve a small and bounded delay. The predictive shaper is based on the developed ARIMA model to forecast the future samples.

In Figure 8 the structure of a generic traffic shaper is presented. This structure contains a buffer to temporarily storage the information generated by the coder. This information will be packetized and delivered to the network at a given rate. This rate is calculated for the controller. In the methods studied in the previous section, this controller is mainly an averager. The mean of the N previous pictures is used to obtain the output rate. With the ideal smoothing, this rate will be constant for all frames in the same GoP. When the sliding shaping is applied, the output rate is calculated for each frame.

The controller for the new predictive scheme is presented in Figure 9. In this case, the controller use K samples to determine the output bit rate. The K samples are distributed in the following way: L_1 samples from the past and L_2 samples forecast by the ARIMA predicter. The current sample is included in L_1, so $K=L_1+L_2$. With these K samples, the controller calculates the mean value. This value will be the output rate supplied to the buffer.

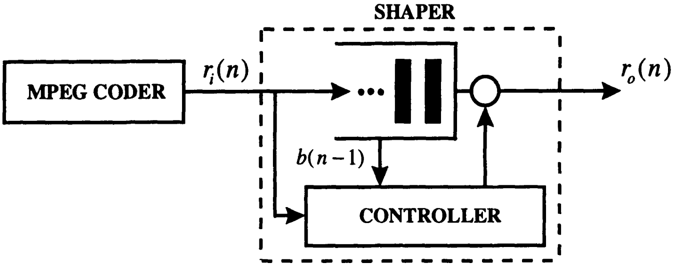

Figure 8. Traffic shaper.

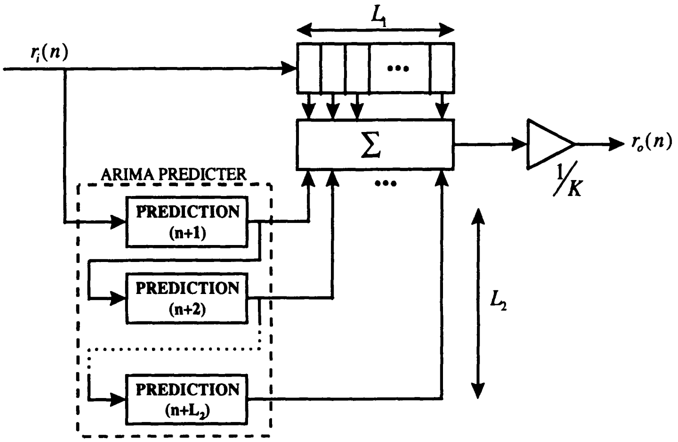

Figure 9. Predictive controller.

In particular, in the presented controller it will not be necessary a number L_1 greater than one. All the needed information about the past is included in the current sample and in the buffer fullness b(n-1). The controller will use this quantity in order to calculate the minimum rates necessaries to keep the delay under a given value.

On the other hand, the minimum number needed of future samples will be the necessary to complete a GoP, that is, (N-1) samples. Thus, the controller has the information about all the I, P and B frames in a GoP, in order to calculate the mean rate. In previous works (Lam, 1996), the possibility of working with a number of predicted samples greater than N has been refused. Here it is possible to corroborate this assumption, since the prediction of a second GoP will be practically the same than the first one. Thus, it will not introduce any difference in the output rate. Nevertheless, the prediction chosen in (Lam, 1996) for a given picture is the bite rate of the equivalent picture in the previous GoP. That is, the bit rate estimated for the (n+1) picture is the bit rate of the (n-N+1) one. This prediction is very accurate when there is not scene changes. However, when a sudden transition occurs, the prediction is not accurate. The difference is shown in Figure 10.

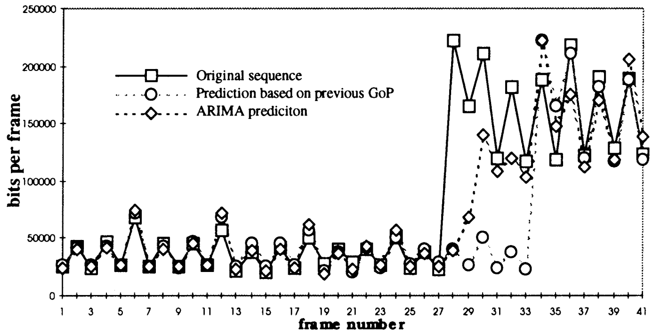

Figure 10. ARIMA prediction and. "previous GoP" prediction.

As it has been mentioned, the buffer fullness information will be employed by the controller. This information is necessary when the shaper has to offer a given QoS, that is, a maximum smoothing delay. Using the buffer fullness, the shaper can obtain a minimum rate for each frame in order to keep its delay under a minimum. This minimum rate will be calculated as a function of the picture size S(n), the buffer fullness b(n-1), and the maximum delay allowed D:

$$r_{min}(n) = \frac{S(n) + b(n-1)}{D\tau}, \qquad (20)$$

where D is expressed in frame periods.

The controller will work normally in the same manner explained previously. Nevertheless, it will have in memory a minimum output rate for every picture which has not been yet completely extracted of the buffer. Therefore, the minimum output rate will be the maximum of these bounds. This mechanism can be called "rate corrector".

A section of the sequence "Jurassic Park", and its predictive smoothing with D=4, D=5 and D=6, are presented in Figure 11. Note that if the maximum delay allowed is D=1, the shaped sequence is the same than the original one. The section chosen for both figures presents a sudden negative transition. It is possible to observe that the smoothing is better when the allowed delay D is greater. The introduced delay in every case is presented in Figure 12. In this figure, it is shown that maximum delay is not exceeded in any case.

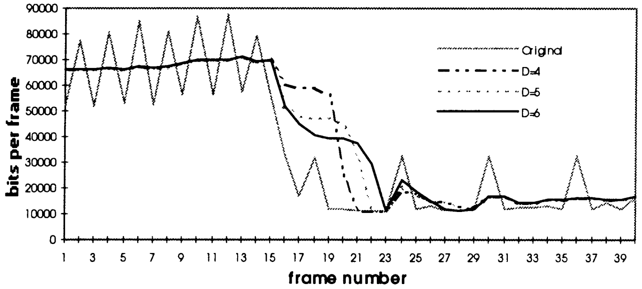

Figure 11. Predictive shaping in a negative transition.

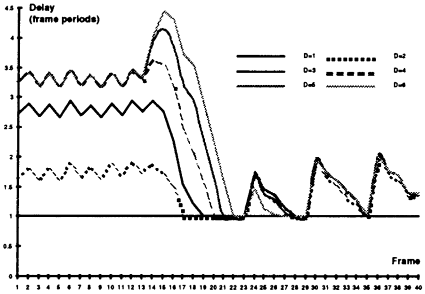

Figure 12. Delay in a negative transition.

Analyzing the previous results, the worst cases appears when sudden negative transitions occurs. This is due to the following reason. In these moments, the predicter will calculate a low output rate, according to the low new rate coming from the coder. However, there is still in the buffer a lot of information. It is the remaining information from de previous frames. These frames were coded with a great number of bits, but now the shaper tries to extract them out of the buffer at a low bit rate. At these moments it becomes necessary to correct the output rate, with the "rate corrector" mentioned above.

The numerical results obtained with this new shaper are presented at the end of Table 1, for the sequence "Jurassic Park" and for D=3. As with the previous kinds of shapers, the quadratic coefficient of variation and the burstiness are strongly reduced. However, the introduced delay is smaller and perfectly bounded. This characteristic permits the use of this kind of shaping with all the services.

The final test to determine the goodness of the new scheme consist of checking the autocorrelation function, in order to observe if the periodicity of the sequences is extracted. In Figure 13, this function is presented for the cases D=1 (that is, the original sequence), D=2 and D=3. It shows that for D=1, the sequence has a great periodicity. This periodicity is strongly reduced for D=2, but it is still possible to observe some peaks every (iN+1) lags. This peaks are completely extracted for D=3. The obtained functions for D greater than 3 are not presented because they are very similar with the case D=3. From this figure, this analysis permits to conclude that the maximum decorrelation of the output sequence is achieved for values of D greater or equal than 3. This will give rise to an improvement in the statistical multiplexing gain over ATM networks. For stronger delay constraints, the shaper could be used with D=2. In this case, the improvement in the resources allocation is reduced.

The different tests has been carried out also with the sequences "America" ("Live in Central Park") and "Geografia de Catalunya", with very similar results.

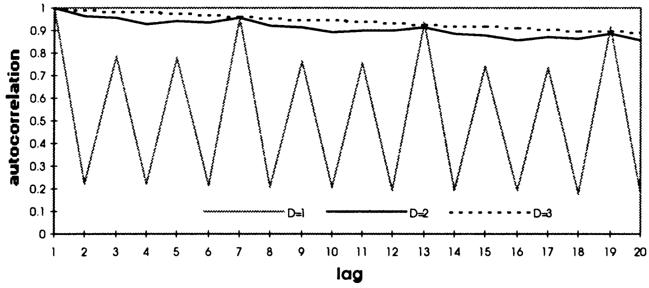

Figure 13. Predictive shaping output autocorrelation

6. CONCLUSIONS

In this paper, a new VBR MPEG traffic shaper is presented. This shaper use prediction techniques to smooth the traffic. In this case, the applied technique is based on the characterization of real traffic as an ARIMA process. The long range dependence of the VBR MPEG traffic can be approximated with the integrative component of the ARIMA process. In this way, the predictor is very accurate in a short temporal range. This short temporal dependence is captured from the autoregressive and the moving average components of the ARIMA predictor. A result of this work is the invariance of the ARIMA coefficients obtained for all the real data sequences analyzed. Moreover, the invariance is also presented from the quantizer step set on the MPEG VBR coder. This invariance allows to apply the VBR MPEG traffic shaper to coders which vary the image quality according to the congestion level in the ATM networks.

The new shaper scheme has been compared with the classics storing systems. The main disadvantage of these systems is the delay introduced in each frame. Nevertheless, the new scheme has a smaller and bounded delay. This characteristic allows to employ this shaper for interactive services, where small delays are needed.

In order to synthesize artificially VBR MPEG traffic an integrative component as $(1-z^{-1})^{H-1/2}$ can be applied. In this way, the Hurst parameter (H) define the persistence of the process. This persistence or self-similarity is associated to the long range dependence of the process. Future works will be focused in this sense using fractional ARIMA process to obtain a new model with bounded mean.

7 REFERENCES

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis. Prentice Hall, New Jersey.

De la Cruz, L., Alins, J., Pallarés, E., Fernández, M. and Mata, J. (1997a) Conformación predictiva de tráfico de vídeo VBR MPEG a partir de su caracterización como proceso ARIMA, Jornadas de Ingeniería Telemática JITEL'97, 197-208.

De la Cruz, L., Alins, J. and Mata, J. (1997b) Prediction Techniques for VBR MPEG Traffic Shaping. *Proceedings of IEEE GLOBECOM'97*, **3**, 1434,39.

Garret, M.W. (1993) Contributions Toward Real-Time Services on Packet-Switched Networks. Ph.D. Dissertation CU/CTR/TR 340-93-20, Columbia University, New York.

Grunenfelder, Cosmas, Manthrope and Odinma-Okafor (1991) Characterization of Video Codecs as Autoregressive Moving Average Processes and Related Queueing System Performance. *IEEE J. on Selected Areas in Communications,* **9**, no. 3, 284-92.

Kawashima, K., Chen, C., Jeng, F. and Singhal, S. (1993) Adaptation of MPEG Video Coding Algorithm to Networks Applications. *IEEE Transactions on Circuits and Systems for Video Technology*, **3**, no. 4, 261-9.

Lam, S.S., Chow, S. and Yau, D.K.Y. (1996) A Lossless Smoothing Algorithm for Compressed Video. *IEEE/ACM Transactions on Networking,* **4**, no. 5, 697-708.

Mata, J. and Sallent, S. (1996) Source Traffic Descriptor for VBR MPEG in ATM Netwoks, in *Multimedia Communications and Video Coding* (ed. Plennum Press).

Mata, J., Pagán, G. and Sallent, S. (1996) Multiplexing and Resource Allocation of VBR MPEG Traffic on ATM Netwoks. *Proceedings of the IEEE ICC'96*, **3**, 1401–5.

Mata, J., Sallent, S., Balsells, J., Zamora, J. and Van der Kolk, A. (1994) Statistical Models for MPEG Video Standard. *Proceedings of IEE EUSIPCO'94*, **1**, 624-7.

Nikolaidis, I. and Akyildiz, I.F. (1992) Source Characterization and Statistical Multiplexing in ATM Networks. *Paper of the College of Computing*. Georgia Institute of Technology, Atlanta, USA.

Pancha,P., and El Zarki, M. (1994) MPEG Coding For Variable Bit Rate Video Transmission. *IEEE Communications Magazine,* **32**, no.5, 54-66.

Proakis. (1983) Digital Communications. McGraw-Hill, New York.

Yegenoglu, F., Jabbari, B. and Zhang, Y. (1993) Motion Classified Autoregresive Modeling of Variable Bit Rate Video. *IEEE Transactions on Circuits and Systems for Video Technology*, **3**, no. 1, 42-53.

ACKNOWLEDGEMENTS

This work has been developed using the Global ATM Standard Simulator (GLASS) and supported by the SIGLA (GLASS) Project [Spain CICYT, TEL 96-1452].