MEASUREMENT AND ARMA MODEL OF VIDEO CODECS IN AN ATM ENVIRONMENT

R. Grünenfelder**, J. Cosmas*, S. Manthorpe* and A. Odinma-Okafor*
** Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne,
* Queen Mary and Westfield College, London E1

ABSTRACT - In this contribution we focus on measurement and fitting to theoretical models of ATM sources. As an example we demonstrate the method for an ATM video source. Measurements of the long term mean and the autocorrelation function of cell arrivals allow the parameter estimation of the ARMA model. Hence the video source is described by an ARMA model. These models can easily be implemented on a computer which allows to build generators for simulations.

1. INTRODUCTION

Asynchronous Transfer Mode (ATM) is a switching mechanism which has been suggested for broadband ISDN. An ATM cell stream consists of a continuous sequence of cells some of which are designated empty and others of which have been dynamically allocated to variable or fixed rate sources. The principle characteristics of Asynchronous Transfer Mode (ATM) networks are that they can absorb bit rate variations so that links are suited to cell traffic generated by variable rate sources and that they can offer high efficiencies because numerous statistically independent, encoded, variable bit rate sources can statistically share network resources.

It is expected that video will be a major source on ATM networks which will place a very large bit rate requirement on the network. Efficient utilisation of network resources is obtained by coding the video, since this reduces the video's mean output bit rate. However this affects the video codec's variance and autocovariance output bit rate characteristics. Suitable models of the video codecs are required which represent the behaviour of the arrival process, in order to study how cell arrivals from codecs affect switching networks.

This paper presents a procedure for characterising a Conditional Replenishment (CR) video codec, as described in [2], as an Autoregressive Moving Average Process (ARMA) so that subsequent dimensioning of ATM network resources may then be computed. The ARMA has been adopted because it has been shown in [1] that the quality of service (cell congestion, cell waiting time in buffer) performance measures of a queueing system are most sensitive to the autocorrelation function of the arrival process. Furthermore, the estimated autocorrelation function of the video codec gives a good indication of its functional behaviour.

The long term mean, variance and autocovariance function are measured from the cell output of the simulated video codec. These are then used to estimate the parameters of a mathematical model of an ARMA process. This means the mathematical model consists of a linear digital filter of finite order (transverse and recursive filter) and a zero memory nonlinearity (ZMNL).

Figure 1: Nonlinear Model used for VBR source modeling.

This nonlinear filter describes the cell arrival sequence which is assumed to be Wide Sense Stationary (WSS). The transfer function of the filter is calculated such that when white noise is applied to its input it produces a cell interarrival sequence which is WSS and which exhibits the same requested long term mean, variance and autocovariance function. The coefficients of the filter are determined from knowledge of the power density spectrum of the input and the required power density spectrum at the output [12].

2. VIDEO CODECS

A number of variable bit rate video codecs have been proposed in recent literature. These codecs have used a number image coding methods, either individually or in combination. These methods are interframe Differential Pulse Code Modulation (DPCM), intraframe DPCM, Conditional Replenishment (CR), Motion Compensation (MC), Transform Coding eg most notably Discrete Cosine (DCT), Run Length Coding (RLC) and Syntactic Coding (SC).

A CR teleconferencing codec, which combines interframe DPCM for areas of movement and RLC for areas with no movement, has been described in [2] in [3]. A DCT with MC codec, which sends a movement vector if transitional movement is detected and the DCT coefficients if no translational movement is detected, has been described in [4]. A CR scheme, which is combined with either interframe DPCM or intraframe DPCM for areas of movement and RLC for areas with no movement, has been described by [5]. Data which has been output from a video codec may or may not be buffered before it is presented to a celliser. Buffer sizes are normally of frame length. A CR codec with a frame length buffer is used in [6]. The output cell rate for the duration of a buffered frame is the mean cell rate for cell arrivals during the frame.

Once the video has been coded, the resultant data is presented to and transported through the ATM network in the form of cells. Cells consist of an information and a signal part. We used a cell size of 32+4 since the work has been done before CCITT recommended 48+5 bytes. When there is sufficient data to fill a cell, a full video cell is transmitted on the ATM link. When there is not sufficient data to fill a cell, an empty cell is transmitted on the ATM link.
A study of the statistical characteristics of all the video codecs which have been proposed is necessary. However the CR codec described in [2,3] was used in order to establish acceptable methodologies for statistical characterization.

<table>
<thead>
<tr>
<th>header</th>
<th>24 data bytes</th>
<th>EOL flags</th>
<th>data separation</th>
<th>error detect correct</th>
</tr>
</thead>
</table>

Figure 2: Cell organisation

Cell organisation is illustrated in figure 1. Cells are organised as 32 bytes information field and 4 bytes signalling field. The information field is subdivided as follows: 24 bytes are used for PCM and RLC data, 3 bytes for end of line/end of frame (EOL/EOF) flags, 3 bytes for PCM/RLC flags and the remaining two bytes for error detection and correction of the EOL/EOF flags. If EOL or EOF occurs then the appropriate bit is set. There is no frame buffering at the video codec output.

3. EXPERIMENT

The set-up of the experiment to simulate the video codec operation is shown in the figure 2. Three main pieces of equipment are used, namely: a Harlequin image frame grabber with transputer, VP405 Philips video disc player and IBM compatible Epson host AT. The image frame grabber consists not only of frame grabbing electronic but also a single T800 transputer with a 1 Mbyte RAM memory and two 256 image buffers. The video disc player is connected to the frame grabber via a composite video link and externally controlled by the PC through an RS232 link. The host PC computer controls with the aid of development tools (such as Turbo Pascal and Transputer Development Systems (TDS2)), the transputer and the video disc player. Occam programs are developed on the TDS2 on the PC and loaded onto the transputer. The simulation is performed using a sequence of frames from three scenes, namely: a typical video phone head and shoulder scene, a selling products scene and a low flying aircraft scene.

Each frame is captured onto the local frame grabber memory and then processed frame at a time by the transputer which simulates the operation of the video codec and celliser. The monochrome intensities are sampled at pixel frequency in accordance with the algorithm being simulated. The coded and cellised interarrival time results are then passed to the PASCAL programs on the host PC for characterisation and statistical analysis.

The nature of video data is such that it recorrelates at each video frame and line interval which is at 40 msec and 156.25 μsec respectively. For lines, this is because data at one part of an image line is highly correlated with the image data on the same part of the next image line (spatial correlation). For frames, this because data at one part of an image is highly correlated with the image data on the same part of the next image (temporal correlation). For both line and frames, it is also known that this recorrelation effect decreases with increasing lag. In general, the frame recorrelation effect can be considered to decrease as a negative exponential for increasing lags. However this is highly dependent on the coincidental nature of the image sequence.

The mean and autocovariance of the number of cell counts in an interval for the output of a video codec is required to be measured in order to capture both line and frame recorrelation effects. The codec is connected to a 150 Mbit/sec or 520.833 kcell/sec ATM link for a cell size of 32+4 bytes. The maximum output bit rate of the video codec is 13.107 Mbit/sec or 68.267 kcell/sec for a cell size of 32+4 of which only 24 bytes are used for video data.

Thus the minimum interarrival distance between cells is 7.63 time slots (of 36 bytes) or every 8th ATM link cell is occupied by a video codec. Thus either only one video cell or no video cells can only ever exist in a 8 cell interval.

An interarrival of 64 ATM link cells or 122.9 μsec was selected because it is of the same order of magnitude as a line interval, namely 156.25 μsec. Thus frame recorrelation effects are captured whereas line recorrelation effects are lost.

Mean and autocovariance of the video codec’s interarrival may be estimated by two principal methods described in the following.

Method 1:

\[
\text{Estimate } \bar{E}_0 \text{ by: } \bar{X} = \frac{1}{M} \sum_{j=1}^{M} X_j
\]

\[
R(k) \text{ by: } R(k) = \frac{1}{M-k} \sum_{j=1}^{M-k} (X_{j+k} - \bar{X})(X_j - \bar{X})
\]

Both estimators are unbiased since:

\[
\bar{E}X = \frac{1}{M} \sum_{j=1}^{M} E(X_j) = EX_j = EX_0 \quad \text{and}
\]

\[
ER(k) = \frac{1}{M-k} \sum_{j=1}^{M-k} E(X_{j+k} - \bar{X})(X_{j+k} - \bar{X}) = \frac{1}{M-k} \sum_{j=1}^{M-k} E(X_{j+k} - EX_{j+k})(X_j - EX_j) = ER(k)
\]

This unbiased estimation of the autocovariance function R does not guarantee positive definiteness of R. That means the power density spectrum S is not positive over the whole
range of frequencies $\omega$, as it should be. This error is due to the finite sample size $M$. The measurements have shown, that $S(\omega)$ is positive for the most of $\omega$, as it has to be, and that the negative parts are quite small in comparison to the positive ones. For theoretical consistency we set the negative parts to zero. The problem of positive definiteness can be resolved easily by measuring in the frequency domain. That means by estimating $S$ and then performing a Fast Fourier Transform by software in order to obtain the autocovariance function. This is usual in signal processing, but in this context it is not without problem as we will see further on.

Method 2 : Estimation of the autocorrelation function by the power density spectrum $S(\omega)$.

Estimate $E X_0$ by : $$\mathcal{X} = \frac{1}{M} \sum_{j=1}^{M} X_j$$

Define $u_N(t) = \sum_{k=-N/2}^{N/2} (X_k - \mathcal{X}) \delta(t-ka)$

apply the Fourier Transform $\mathcal{F}$ to $u_N(t)$, which yields:

$$U_N(\omega) = \mathcal{F} \{u_N(t)\} = \sum_{k=-N/2}^{N/2} \mathcal{F} \{\delta(t-ka)\} = \sum_{k=-N/2}^{N/2} (X_k - \mathcal{X}) e^{-j\omega k a}$$

The power density spectrum $S(\omega)$ is given by :

$$S(\omega) = \lim_{N \to \infty} \frac{E |U_N(\omega)|^2}{N} \quad \text{for } \omega \in [-\pi, \pi]$$

The autocovariance function can be obtained by inverse Fourier Transform of the power density spectrum. The problem of negative power spectra is omitted, but the autocovariance function is truncated by the method of measurement itself, that means it becomes zero for lags greater than $N$. This may be a significant loss of information. Since we are interested in the autocovariance function for all lags, we have chosen method 1. This allows the measurements of all the autocorrelations. Nevertheless theoretical studies [7] justify method 2 since they show that the influence of the correlation of the queueing time becomes small enough for large lags $k$. Hence the choice of $N = 1024$ assures that the autocorrelation of lags higher than 1024 does no more influence the queueing delay.

4. SOURCE MODELS

Cell traffic transported on an ATM network are categorised in [8], according to their time resolution into four levels, namely: connection, dialogue, action and transmission levels. The connection level describes the behaviour of a traffic source on a virtual call basis. These are the call setup and clear events which delimit a call duration. The dialogue level describes the interaction between subscribers at both ends of the connection. Four situations are possible: both subscribers are transmitting, both subscribers are not transmitting and either one or the other of the subscribers are transmitting. The action level describes the behaviour of a transmitting subscriber. Its statistical behaviour includes the mean, probability distribution and autocovariance. The transmission level describes the behaviour of the cell generation at the lowest level. This is the minimum and maximum interarrival times. The purpose of this study is to derive a source model which successfully represents the variable rate video codec cell output at the action and transmission levels. A number of source models at the action level for variable bit rate video codecs have been proposed in recent literature.

A discrete time, discrete state space, Markov chain source model is suggested in [8] for modelling variable rate video on ATM links. This is a modification of a continuous time, discrete state space, Markov Process source model proposed in [9]. A thorough study of Markov chains and processes is presented in [10].

A continuous time, discrete state Markov model is presented in [6]. This is used to model a CR codec with frame buffering at its output. Cells are output at the fixed mean cell arrival rate for each buffered frame. Therefore recorrelations from frame to frame do not occur except through pure coincidence and can be fitted to the autocovariance of the continuous time, discrete state Markov model which is a negative exponential function. The parameters of the model are obtained by matching the sample means and autocovariance with theoretical derivations of the mean and autocovariance. A continuous state autoregressive model is also presented in [6] and used to model the same CR codec with frame buffering at its output. The autocovariance of this model is also a negative exponential function. The parameters of the model are obtained by matching the sample means and autocovariance with theoretical derivations of the means and autocovariance.

Source models whose autocovariance exhibit recorrelation are:

- Periodic State Models
- Autoregressive Moving Average Models (ARMA)

A description of periodic state models is given in [10]. In [11] the CR video codec is fitted to a periodic state model. The ARMA model exhibits a recorrelation which closely matches the sample recorrelation. The cell stream from a single video source is modeled by the number of arriving video cells within a predefined interval of 64 time slots of 36 bytes. The intervals are hopped on the time axes. In particular the video cell arrival process is considered to be WSS. A sequence of random variables $(X_n)$ with $E |X_n|^2 < \infty$, when $n$ is an integer, is WSS if for all $n$ in the following two equalities are fulfilled.

$$E X_n = E X_0$$

$$\text{cov} (X_{n+k}, X_{n}) = \text{cov} (X_k, X_0) = R(k)$$

WSS processes of Gaussian sequences can as usually known in signal processing be described by an ARMA model.

For the video cell arrival autocorrelation function, the recorrelation peak occurs every 40 ms, that means at all integer multiples of the cell lag $m = 40 \text{ msec} / 122.9 \mu\text{sec} = 325$ (see Figures 4, 5 and 6). This periodic occurrence will
be modeled by the autoregressive part of the ARMA model. The parameter $\alpha$ is an attenuation parameter of the recorrelation peaks. Figures 4, 5 and 6 show that the "amplitude" of the $i$th peak, $i \neq 0$, is almost the same as for the $i+1$th. Replacing $R(0)$ by $R(1) + \text{Var}\,v$, with $v$ an independent and identical (iid) distributed random variable, it becomes possible to choose the parameter $\alpha$ nearly one, since $R(1)$ is approximatively equal to $R(m)$ for each measured scene. The lags from -162 up to 162 will be modelled by the aid of the moving average part. This yields together with the autoregressive part the periodic behaviour of the autocorrelation function of the video cell stream. The white noise $v$ can be added to the output of the ARMA model, since $v$ is iid. The expected values of the ARMA filter's output sequence $\{Z_i\}$ is zero, since a white noise sequence $\{\epsilon_i\}$ with expected value zero will be applied to the filter's input. Expected value and variance of the Gaussian random sequence $v_i$ are $\text{E}\,v_i = 0$ and $\text{Var}\,v_i = R(0) - R(1)$ for all integers $i$. The variance of the white noise $\epsilon$ is chosen so that the variance of $V$ (ARMA filter output plus noise $v$) becomes 1. The ARMA filter output $Z$ is followed by a zero-memory nonlinearity (ZMNL) $g(V_i) = \max(0, aV_i + b)$. The coefficients $a$ and $b$ are given in Table I for the three considered scenes:

<table>
<thead>
<tr>
<th>scene</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>head/shoulder</td>
<td>1.3477</td>
<td>0.0206</td>
</tr>
<tr>
<td>selling products</td>
<td>1.615</td>
<td>1.588</td>
</tr>
<tr>
<td>flying aircraft</td>
<td>1.808</td>
<td>2.525</td>
</tr>
</tbody>
</table>

Table I: Coefficients $a$ and $b$ of the ZMNL.

It should be pointed out that the presented methodology is not restricted to video sources. All the mathematics of the ARMA+ZMNL model is developed in [12], where it is shown that the sequence $\{V_i\}$ has the following autocovariance function $R(.)$. Furthermore is shown that the autocovariance function of $g(.)$ is closely approximated by the digital filter's one. Hence

$$R(0) = \frac{1}{1 - \alpha^2} \text{Var} \sum_{\mu = -m/2}^{m/2} h_{\mu}^2 + \text{Var} v$$

$$R(m+r) = \frac{\alpha^{\lceil j \rceil}}{1 - \alpha^2} \text{Var} \sum_{\mu = -m/2}^{m/2} h_{\mu} h_{r - \mu}$$

if $0 \leq r \leq m/2, j m + r \neq 0, j \in \mathbb{Z}$ and $m, r \in \mathbb{N}$

where $h_{i} = 0 \, \forall \, |i| > q$, $\{\epsilon_i\}$ and $\{V_i\}$ white noise sequences. The Fourier coefficients (filter taps coefficients) of the function $f$

$$f(\omega) = \left( \frac{1}{\text{Var} \epsilon} \left( R(0) + 2 \sum_{k=1}^{m/2} R(k) \cos(\omega k) \right) \right)^{1/2}$$

are listed below. They have to be multiplied by $\sqrt{1-\alpha^2}$ in order to be the transverse filter's coefficients.

The sample autocorrelation recorrelated at lags of unit frame intervals. Line correlation was not observed because of the size of the interval. The autocorrelation function spans value from 1 to near 0, without approaching -1. This means that the successive cell rates from the video codec changes gradually rather than abruptly across cell rate extremes and that successive rates are grouped by size (there are groups of large rates and groups of low rates which keep separate). The recorrelation peaks decay very slowly, and this corresponds with the parameter $\alpha = 0.99$ for the recursive filter. The difference between measured and fitted autocovariance function is exceptionally small.

Using stochastic optimization techniques e.g. simulated annealing allows to characterize the codec with less than 15 parameters. In this case the linear digital filter is an ARMA filter characterized by 15 coefficients.

5. CONCLUSIONS

A methodology has been developed for characterising video codes in an ATM environment, as an ARMA process fully taking into account its correlation behaviour. The characteristics of a video codec are measured with methods well known in signal processing. ARMA processes can be applied to any type of ATM source, with any type of correlation function. A ZMNL transforms the linear digital filter's output to strictly positive random variables. The chosen ZMNL does not influence a lot the autocovariance function of the process. Hence it can directly be fitted to the ARMA filter. The smaller the coefficient of variation of the process the lesser the ZMNL influences the autocovariance function. It is demonstrated that the ARMA process is suited to video codecs and ATM traffic in general. An ATM hardware demonstrator requires sources which incorporate all the statistical characteristics using simple hardware implementation. The ARMA satisfies both requirements. Even for software simulations ARMA is suitable since on a VAX 8530 100'000 events are simulated within 3 minutes 10 seconds.

The methodology is suited to characterize variable bit rate and compound ATM sources with less than 15 parameters.
6. ACKNOWLEDGEMENT

The authors gratefully acknowledge support from the CEC RACE programme, project 1022 "Technology for ATD".

REFERENCES


Figure 4: Autocorrelation Function of the Head and Shoulder Scene

Figure 5: Autocorrelation Function of the Selling Scene

Figure 6: Autocorrelation Function of the Flying Aircraft Scene