On Modeling Video Traffic from Multiplexed MPEG-4 Videoconference Streams

A. Lazaris, P. Koutsakis, and M. Paterakis

Dept. of Electronic and Computer Engineering
Information & Computer Networks Laboratory
Technical University of Crete, Chania, Greece
{alazaris, polk, pateraki}@telecom.tuc.gr

Abstract. Due to the burstiness of video traffic, video modeling is very important in order to evaluate the performance of future wired and wireless networks. In this paper, we investigate the possibility of modeling this type of traffic with well-known distributions. Our results regarding the behavior of single videoconference traces provide significant insight and help to build a Discrete Autoregressive (DAR(1)) model to capture the behavior of *multiplexed MPEG-4 videoconference movies* from VBR coders.

1 Introduction

As traffic from video services is expected to be a substantial portion of the traffic carried by emerging wired and wireless networks, statistical source models are needed for Variable Bit Rate (VBR) coded video in order to design networks which are able to guarantee the strict Quality of Service (QoS) requirements of the video traffic. Video packet delay requirements are strict, because delays are annoying to a viewer; whenever the delay experienced by a video packet exceeds the corresponding maximum delay, the packet is dropped, and the video packet dropping requirements are equally strict.

Hence, the problem of modeling video traffic, in general, and videoconferencing, in particular, has been extensively studied in the literature. VBR video models which have been proposed in the literature include first-order autoregressive (AR) models [2], discrete AR (DAR) models [1, 3], Markov renewal processes (MRP) [4], MRP transform-expand-sample (TES) [5], finite-state Markov chain [6, 7], and Gamma-beta-auto-regression (GBAR) models [8, 9]. The GBAR model, being an autoregressive model with Gamma-distributed marginals and geometric autocorrelation, captures data-rate dynamics of VBR video conferences well; however, it is not suitable for general MPEG video sources [9].

In [3] the authors show that H.261 videoconference sequences generated by different hardware coders, using different coding algorithms, have gamma marginal distributions (this result was also employed by [10], which proposes an Autoregressive Model of order one for sequences of H.261 encoding) and use this result to build a Discrete Autoregressive (DAR) model of order one, which works well when several sources are multiplexed.

Y. Koucheryavy, J. Harju, and V.B. Iversen (Eds.): NEW2AN 2006, LNCS 4003, pp. 46 – 57, 2006.

In [11-13], different approaches are proposed for MPEG-1 traffic, based on the lognormal, Gamma, and a hybrid Gamma/lognormal distribution model, respectively. Standard MPEG encoders generate three types of video frames: *I* (intracoded), *P* (predictive) and *B* (bidirectionally predictive); i.e., while *I* frames are intra-coded, the generation of *P* and *B* frames involves, in addition to intra-coding, the use of motion prediction and interpolation techniques. *I* frames are, on average, the largest in size, followed by *P* and then by *B* frames.

An important feature of common MPEG encoders (both hardware and software) is the manner in which frame types are generated. Typical encoders use a fixed Group-of-Pictures (GOP) pattern when compressing a video sequence; the GOP pattern specifies the number and temporal order of *P* and *B* frames between two successive *I* frames. A GOP pattern is defined by the distance N between *I* frames and the distance M between *P* frames. In practice, the most frequent value of M is 3 (two successive B frames) while the most frequent values of N are 6, 12, and 15, depending on the required video quality and the transmission rate.

In this work, we focus on the problem of modeling videoconference traffic from MPEG-4 encoders (the MPEG-4 standard is particularly designed for video streaming over wireless networks [14]), which is a relatively new and yet open issue in the relevant literature.

2 Videoconference Traffic Model

A. Frame-Size Histograms

We use four different long sequences of MPEG-4 encoded videos (from [15]) with low or moderate motion (i.e, traces with very similar characteristics to the ones of actual videoconference traffic), in order to derive a statistical model which fits well the real data. The length of the videos varies from 45 to 60 minutes and the data for each trace consists of a sequence of the number of cells per video frame and the type of video frame, i.e., *I, P,* or *B*. We use packets of ATM cell size throughout this work, but our modeling mechanism can be used equally well with packets of other sizes. We have investigated the possibility of modeling the four videoconference videos with quite a few well-known distributions and our results show that the best fit among these distributions is achieved for all the traces studied with the use of the Pearson type V distribution. The Pearson type V distribution (also known as the "inverted Gamma" distribution) is generally used to model the time required to perform some tasks (e.g., customer service time in a bank); other distributions which have the same general use are the exponential, gamma, weibull and lognormal distributions [20]. Since all these distributions have been often used for video traffic modeling in the literature, they have been chosen as fitting candidates in order to compare their modeling results in the case of MPEG-4 videoconferencing.

The four traces under study are, respectively, a video stream extracted and analyzed from a camera showing the events happening within an office (Video Name: "Office Cam"); a video stream extracted and analyzed from a camera showing a lecture (Video Name: "Lecture Room Cam"); a video stream extracted and analyzed from a talk-show (Video Name: "N3 Talk"); a video stream extracted and analyzed from another talk-show (Video Name: "ARD Talk"). For each one of these movies we have used the high quality coding version, in which new video frames arrive every 40 msecs, in Quarter Common Intermediate Format (QCIF) resolution. The compression pattern used to encode all the examined video streams is IBBPBBPBBPBB, i.e., N=12, M=3, according to the definitions used in Section 1.

The frame-size histogram based on the complete VBR streams is shown, for all four sequences, to have the general shape of a Pearson type V distribution (this is shown in Figure 1, which presents indicatively the histogram for the lecture sequence; the other three traces have similar histograms).

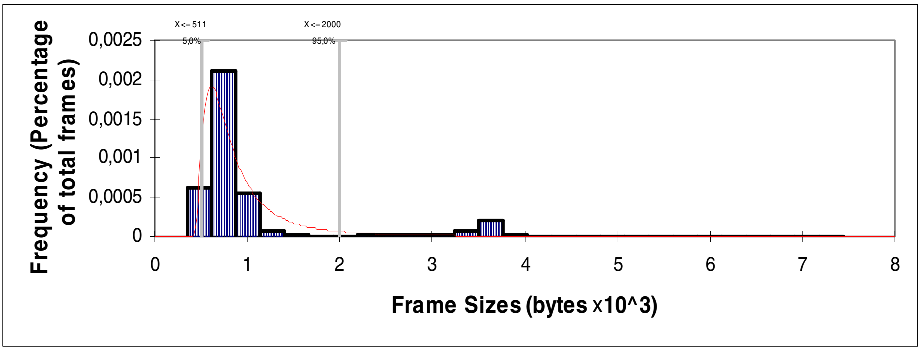

Fig. 1. Histogram for the frame size of the lecture camera trace

B. Statistical Tests and Autocorrelations

The statistical test was made with the use of Q-Q plots. The Q-Q plot is a powerful goodness-of-fit test [3, 20], which graphically compares two data sets in order to determine whether the data sets come from populations with a common distribution (if they do, the points of the plot should fall approximately along a 45-degree reference line). More specifically, a Q-Q plot is a plot of the quantiles of the data versus the quantiles of the fitted distribution (a z-quantile of X is any value x such that $P((X \leq x) = z)$.

In Figure 2, we have plotted the 0.01-, 0.02-, 0.03-,… quantiles of the actual trace versus the respective quantiles of the various distribution fits for the ARD Talk trace (the results are similar for all the traces).

The Pearson V distribution fit is shown to be the best in comparison to the gamma, weibull, lognormal and exponential distributions, which are presented here (comparisons were also made with the negative binomial and Pareto distributions, which were also worse fits than the Pearson V). However, as already mentioned, although the Pearson V was shown to be the better fit among all distributions, the fit is not perfectly accurate. This was expected, as the gross differences in the number of bits required to represent *I, P* and *B* frames impose a degree of periodicity on MPEG-encoded streams, based on the cyclic GOP formats. Any model which purports to

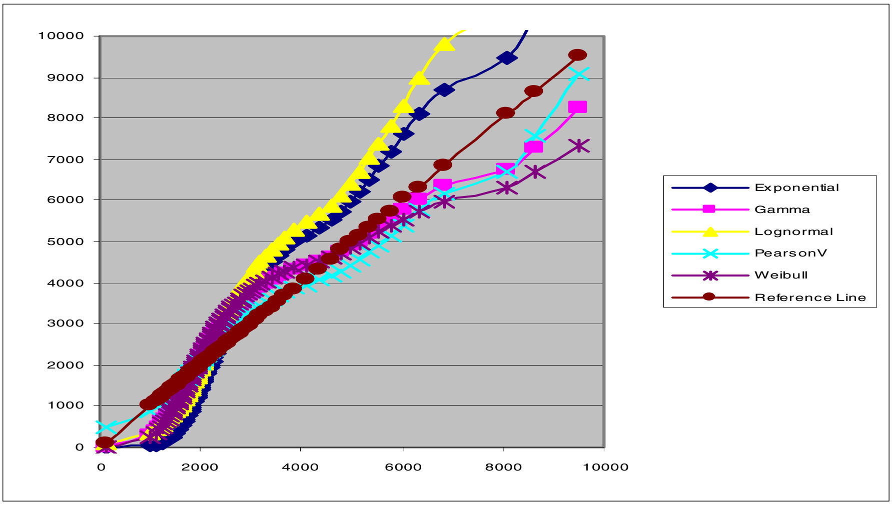

Fig. 2. Q-Q plot for the ARD TALK trace

reflect the frame-by-frame correlations of an MPEG-encoded video stream must account for GOP cyclicity, otherwise the model could produce biased estimates of cell loss rate for a network with some given traffic policing mechanism [9, 13].

Hence, we proceeded to study the frame size distribution for each of the three different video frame types (*I, P, B*), in the same way we studied the frame size distribution for the whole trace. This approach was also used in [9, 19].

Another approach, similar to the above, was proposed in [11]. This scheme uses again lognormal distributions and assumes that the change of a scene alters the average size of *I* frames, but not the sizes of *P* and *B* frames. However, it is shown in [4, 12] that the average sizes of *P* and *B* frames can vary 20% and 30% (often more than that), respectively, in subsequent scenes, therefore the size changes are statistically significant.

As it will be shown from our results, none of the above choices of distribution fits are relevant to the case of *I, P* and *B* frames of MPEG-4 videoconference traffic.

The mean, peak and variance of the video frame sizes for each video frame type (*I, P* and *B*) of each movie were taken again from [15] and the Pearson type V parameters are calculated based on the formulas for the mean and variance of Pearson V (the parameters for the other fitting distributions are similarly obtained based on their respective formulas). The autocorrelation coefficient of lag-1 was also calculated for all types of video frames of all four movies, as it shows the very high degree of correlation between successive frames of the same type (it was larger than 0.7 in all the cases, and in most of the cases it was larger than 0.9). The autocorrelation coefficient of lag-1 will be used in the following Sections of this work, in order to build a Discrete Autoregressive Model for each video frame type.

From the five distributions examined (Pearson V, exponential, gamma, lognormal, weibull) the Pearson V distribution once again provided the best fitting results for 11 of the 12 cases examined, i.e., for all video frame types of the office, N3 Talk and lecture camera traces, and for the *P, B* frame types of the ARD Talk trace. The only case in which the Pearson V distribution exhibits worse fitting results than another distribution is that of the N3 Talk *P* frames, where the best fitting result is derived with the use of the lognormal distribution (still, even in this case the difference in the goodness-of-fit results is marginal).

In order to further verify the validity of our results, we performed Kolmogorov-Smirnov tests for all the 12 fitting attempts. The Kolmogorov-Smirnov test (KS-test) tries to determine if two datasets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data, i.e, it is non-parametric and distribution free. The KS-test uses the maximum vertical deviation between the two curves as its statistic *D*. The results of our KS-tests (Figure 3 is a characteristic example of these results), confirm our respective conclusions based on the Q-Q plots (i.e., the Pearson V distribution is the best fit).

Although controversy persists regarding the prevalence of Long Range Dependence (LRD) in VBR video traffic ([16, 17, 23]), in the specific case of MPEG-encoded video, research has shown that LRD is important [11, 18]. The results of our

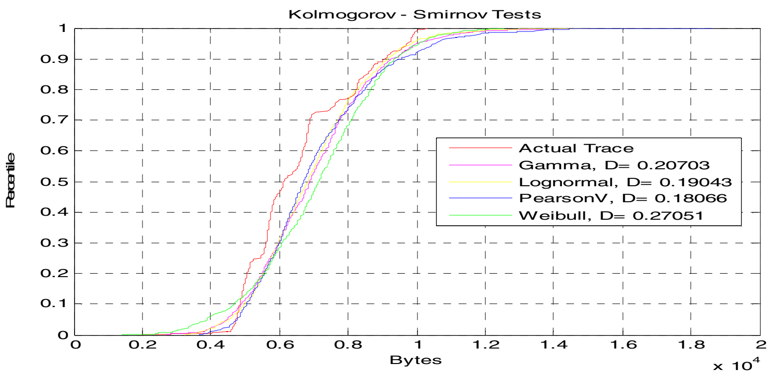

Fig. 3. KS-test (Comparison Percentile Plot) for the ARD TALK I-frames

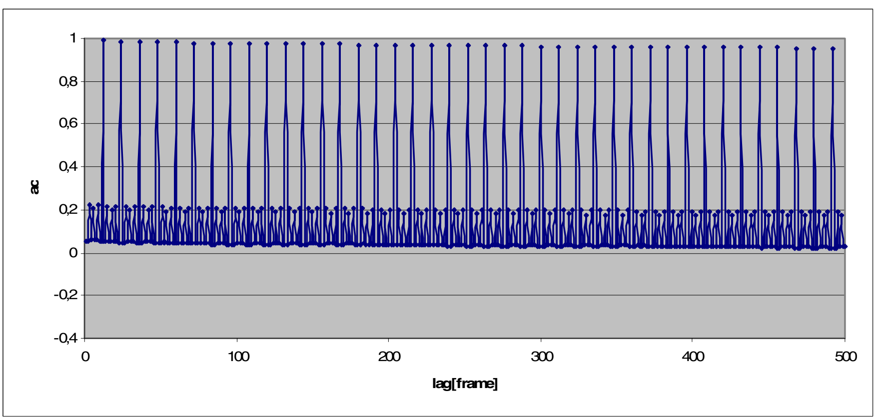

Fig. 4. Autocorrelation function of the lecture camera trace

study on single MPEG-4 videoconferencing agree with this conclusion. The autocorrelation function for the lecture camera trace is shown in Figure 4 (the respective Figures for the other three traces are similar). Two apparent periodic components are observed, one containing lags with low autocorrelation and the other lags with high autocorrelation. We observe that autocorrelation remains high even for large numbers of lags and that both components decay very slowly; both these facts are a clear indication of the importance of LRD. The existence of strong autocorrelation coefficients is due to the periodic recurrence of I, B and P frames.

Although the fitting results when modeling each video frame type separately, with the use of the Pearson V distribution, are clearly better than the results produced by modeling the whole sequence uniformly, the high autocorrelation shown in the Figure above can never be perfectly "captured" by a distribution generating frame sizes independently, according to a declared mean and standard deviation, and therefore none of the fitting attempts (including the Pearson V), as good as they might be, can achieve perfect accuracy. However, these results lead us to extend our work in order to build DAR models which inherently use the autocorrelation coefficient of lag-1 in their estimations and which will be shown to capture well the behavior of *multiplexed MPEG-4 videoconference movies*, by generating frame sizes independently for *I, P* and *B* frames.

3 The DAR (1) Model

A Discrete Autoregressive model of order *p*, denoted as DAR(*p*) [21], generates a stationary sequence of discrete random variables with an arbitrary probability distribution and with an autocorrelation structure similar to that of an Autoregressive model. DAR(1) is a special case of a DAR(p) process and it is defined as follows: let $\{V_n\}$ and $\{Y_n\}$ be two sequences of independent random variables. The random variable V_n can take two values, 0 and 1, with probabilities 1-ρ and ρ, respectively. The random variable Y_n has a discrete state space *S* and $P\{Y_n = i) = \pi(i)$. The sequence of random variables $\{X_n\}$ which is formed according to the linear model:

$$X_n = V_n X_{n-1} + (1- V_n) Y_n \tag{1}$$

is a DAR(1) process.

A DAR(1) process is a Markov chain with discrete state space *S* and a transition matrix:

$$\mathbf{P} = \rho\mathbf{I} + (1-\rho)\, \mathbf{Q} \tag{2}$$

where ρ is the autocorrelation coefficient, **I** is the identity matrix and **Q** is a matrix with $Q_{ij} = \pi(j)$ for i, j ∈ *S*.

Autocorrelations are usually plotted for a range W of lags. The autocorrelation can be calculated by the formula:

$$\rho(W)= E[(X_i - \mu)(X_{i+w} - \mu)]/\sigma^2 \tag{3}$$

where μ is the mean and σ^2 the variance of the frame size for a specific video trace.

4 DAR(1) Modeling Results and Discussion

As in [3], where a DAR(1) model with negative binomial distribution was used to model the number of cells per frame of VBR teleconferencing video, we want to build a model based only on parameters which are either known at call set-up time or can be measured without introducing much complexity in the network. DAR(1) provides an easy and practical method to compute the transition matrix and gives us a model based only on four physically meaningful parameters, i.e., the mean, peak, variance and the lag-1 autocorrelation coefficient ρ of the offered traffic (these correlations, as already explained, are typically very high for videoconference sources). According to [22], the DAR(1) model can be used with any marginal distribution.

As explained in our work on modeling a single MPEG-4 videoconference trace, the lag-1 autocorrelation coefficient for the *I*, *P* and *B* frames of each trace is very high in all the studied cases. Therefore, we proceeded to build a DAR(1) model for each video frame type for each one of the four traces under study. More specifically, in our model the rows of the **Q** matrix consist of the Pearson type V probabilities (f_0, f_1, … f_k, F_K), where $F_K = \Sigma_{k > K} f_k$, and K is the peak rate. Each k, for $k<K$, corresponds to possible source rates less than the peak rate of K.

From the transition matrix in (2) it is evident that if the current frame has, for example, i cells, then the next frame will have i cells with probability $\rho+(1-\rho)*f_i$, and will have k cells, $k \neq i$, with probability $(1-\rho)*f_k$. Therefore the number of cells per video frame stays constant from one (*I, P* or *B*) video frame to the next (*I*, *P* or *B*) video frame, respectively, in our model with a probability slightly larger than ρ (for example, in the ARD Talk trace, with probability slightly larger than 93.23%, 77.81%, 94.49% for the *I, P* and *B* frames of the trace, respectively). This is evident in Figure 5, where we compare the actual *I* frames of the ARD Talk trace and their respective DAR(1) model and it is shown that the DAR(1) model's data produce a "pseudo-trace" with a periodically constant number of cells for a number of video frames. This causes a significant difference when comparing a segment of the

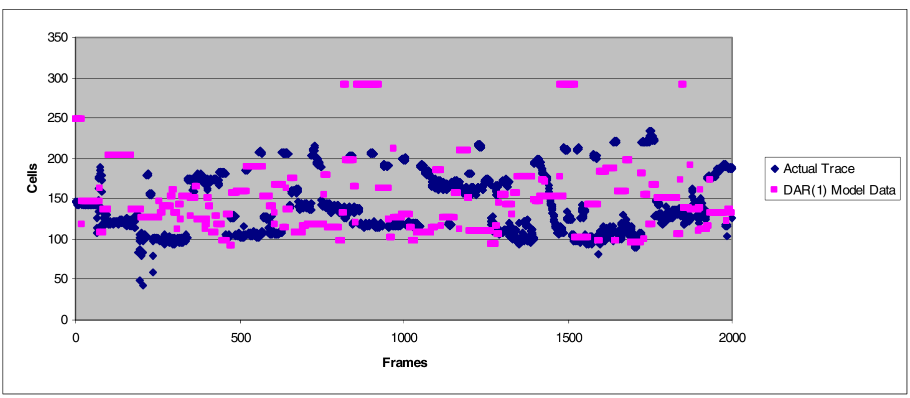

Fig. 5. Comparison for a single trace between a 2000 frame sequence of the actual *I* frames sequence of the ARD Talk trace and the respective DAR(1) model in number of cells/frame (Y-axis)

sequence of *I, P, or B* frames of the actual ARD Talk video trace and a sequence of the same length produced by our DAR(1) model. The same vast differences also appeared when we plotted the DAR(1) models versus the actual *I, P* and *B* video frames of the actual N3Talk, office camera and lecture camera traces for a single movie.

However, our results have shown that the differences presented above become small for all types of video frames and for all the examined traces *for a superposition of 5 or more sources*, and are almost completely smoothed out in most cases, as the number of sources increases (the authors in [3] have reached similar conclusions for their own DAR(1) model and they present results for a superposition of 20 traces). This is clear in Figures 6-8, which present the comparison between our DAR(1) model and the actual *I, P, B* frames' sequences of the ARD Talk video, for *a superposition* of 20 traces (the results were perfectly similar for all video frame types of the other three traces; we have used the initial trace sequences to generate traffic for 20 sources, by using different starting points in the trace). The common property of all these results (derived by using a queue to model multiplexing and processing frames in a FIFO manner) is that the DAR(1) model seems to provide very accurate fitting results for *P* and *B* frames, and relatively accurate for *I* frames.

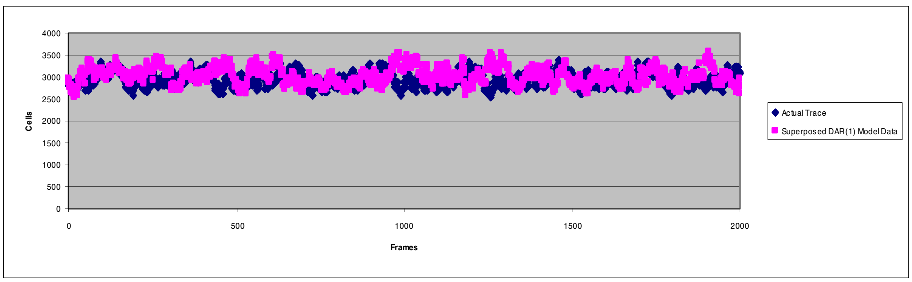

Fig. 6. Comparison for 20 superposed sources between a 2000 I frame sequence of the actual ARD Talk trace and the respective DAR(1) model in number of cells/frame (Y-axis)

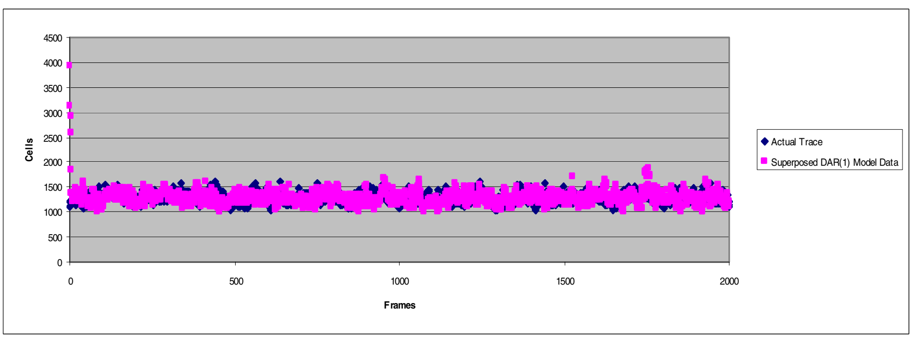

Fig. 7. Comparison for 20 superposed sources between a 2000 P frame sequence of the actual ARD Talk trace and the respective DAR(1) model in number of cells/frame (Y-axis)

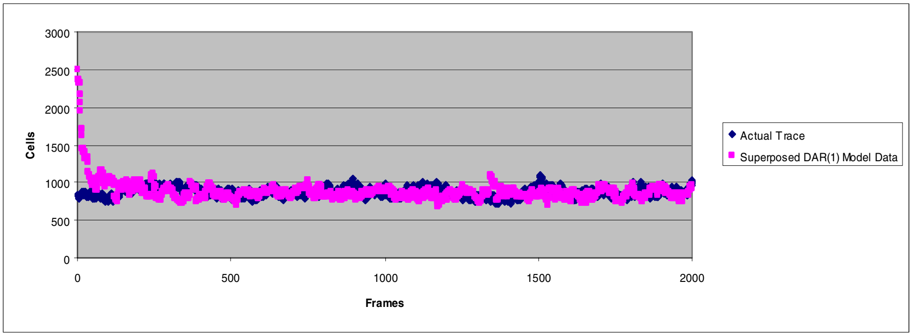

Fig. 8. Comparison for 20 superposed sources between a 2000 B frame sequence of the actual ARD Talk trace and the respective DAR(1) model in number of cells/frame (Y-axis)

However, although Figures 6-8 suggest that the DAR(1) model captures very well the behavior of the multiplexed actual traces, they do not suffice as a result. Therefore, we proceeded again with testing our model statistically in order to study whether it produces a good fit for the *I, P, B* frames for the trace superposition. For this reason we have used again Q-Q plots, and we present indicatively some of these results in Figures 9-10, where we have plotted the 0.01-, 0.02-, 0.03-,… quantiles of the actual *B* and *I* video frames' types of the N3 Talk trace versus the respective quantiles of the respective DAR(1) models, for a superposition of 20 traces. As shown in Figure 9, which presents the comparison of actual *B* frames with the respective DAR(1) models, the points of the Q-Q plot fall almost completely along the 45-degree reference line, with the exception of the first and last 3% quantiles (left- and right-hand tail), for which the DAR(1) model underestimates and overestimates, respectively, the probability of frames with a very small (large) number of cells. The very good fit shows that the superposition of the *B* frames of the actual traces can be modeled very well by a respective superposition of data produced by the DAR(1) model (similar results were derived for the superposition of *P* frames), as it was suggested in Figures 7, 8. Figure 10 presents the comparison of actual *I* frames with the respective DAR(1) model, for the N3 Talk trace. Again, the result suggested from Figure 6, i.e., that our method for modeling *I* frames of multiplexed MPEG-4 videoconference streams provides only relative accuracy, is shown to be valid with the use of the Q-Q plots. The results for all the other cases which are not presented in Figures 9-10 are similar in nature to the ones shown in the Figures.

One problem which could arise with the use of DAR(1) models is that such models take into account only short range dependence, while, as shown earlier, MPEG-4 videoconference streams show LRD. This problem is overcome by our choice of modeling *I, P* and *B* frames separately. This is shown in Figure 11. It is clear from the Figure that, even for a small number of lags, (e.g., larger than 10) the autocorrelation of the superposition of frames decreases quickly, for all the traces. Therefore, although in some cases the DAR(1) model exhibits a quicker decrease than that of the actual traces' video frames sequence, this has minimal impact on the fitting quality of the DAR(1) model. This result further supports our choice of using a first-order model.

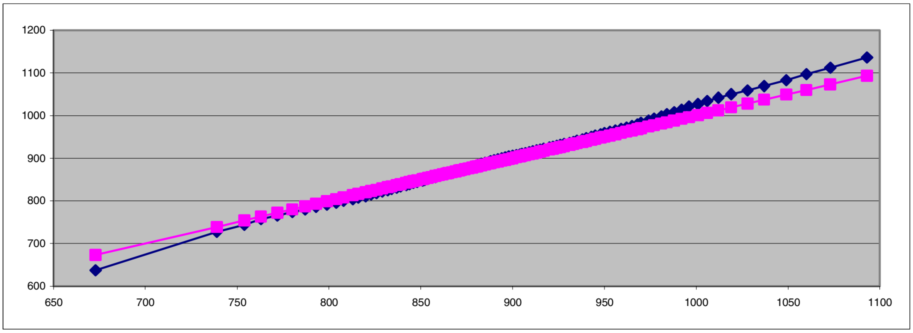

Fig. 9. Q-Q plot of the DAR(1) model versus the actual video for the B frames of the N3 Talk trace, for 20 superposed sources

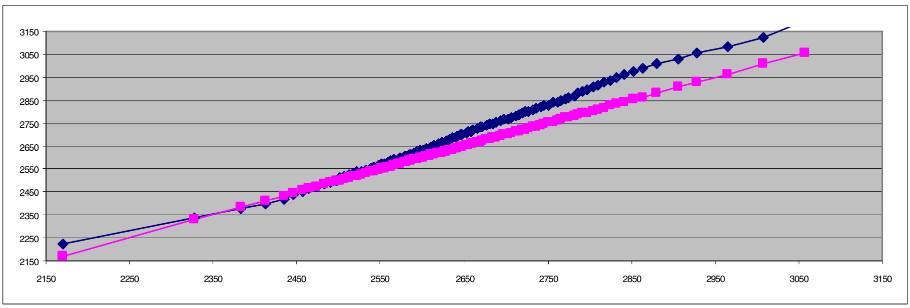

Fig. 10. Q-Q plot of the DAR(1) model versus the actual video for the I frames of the N3 Talk trace, for 20 superposed sources

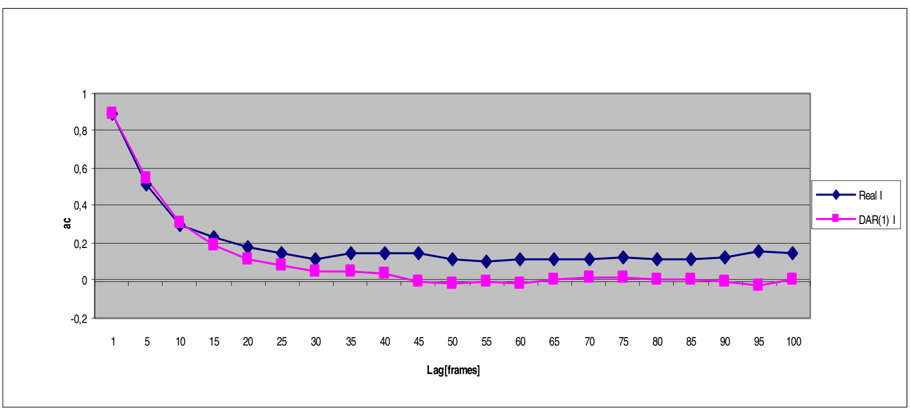

Fig. 11. Autocorrelation vs. number of lags for the I frames of the actual N3 Talk trace and the DAR(1) model, for 20 superposed sources

References

1. D. M. Lucantoni, M. F. Neuts, A. R. Reibman, "Methods for Performance Evaluation of VBR Video Traffic Models," *IEEE/ACM Trans.Networking,* Vol. 2, 1994, pp. 176-180.
2. M. Nomura, T. Fuji, N. Ohta, "Basic Characteristics of Variable Rate Video Coding in ATM Environment", *IEEE Journal on Selected Areas in Communications,* Vol. 7, No. 5, 1989, pp. 752-760.
3. D. P. Heyman, A. Tabatabai, T. V. Lakshman, "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks", *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 2, No. 1, 1992, pp. 49-59.
4. A. M. Dawood, M. Ghanbari, "Content-based MPEG Video Traffic Modeling", *IEEE Transactions on Multimedia,* Vol. 1, No. 1, 1999, pp. 77-87.
5. B. Melamed and D. E. Pendarakis, "Modeling Full-Length VBR Video Using Markov-Renewal Modulated TES Models", *IEEE Journal on Selected Areas in Communications,* Vol. 16, No. 5, 1998, pp. 600-611.
6. K. Chandra, A. R. Reibman, "Modeling One- and Two-Layer Variable Bit Rate Video", *IEEE/ACM Transactions on Networking*, Vol. 7, No. 3, 1999, pp. 398-413.
7. Q. Ren, H. Kobayashi, "Diffusion Approximation Modeling for Markov Modulated Bursty Traffic and its Applications to Bandwidth Allocation in ATM Networks", *IEEE Journal on Selected Areas in Communications,* Vol. 16, No. 5, 1998, pp. 679-691.
8. D. P. Heyman, "The GBAR Source Model for VBR Videoconferences", IEEE/ACM Transactions on Networking (1997), Vol. 5, No. 4, 1997, pp. 554-560.
9. M. Frey, S. Ngyuyen-Quang, "A Gamma-Based Framework for Modeling Variable-Rate Video Sources: The GOP GBAR Model", *IEEE/ACM Trans. on Networking*, Vol. 8, No. 6, 2000, pp. 710-719.
10. S. Xu, Z. Huang, "A Gamma autoregressive video model on ATM networks", *IEEE Transactions on Circuits and Systems Video Technology,* Vol. 8, No. 2, 1998, pp. 138–142.
11. M. Krunz, S. K. Tripathi, "On the Characterization of VBR MPEG Streams," *in Proceedings of ACM SIGMETRICS*, Vol. 25, June 1997.
12. U. K. Sarkar, S. Ramakrishnan, D. Sarkar, "Modeling Full-Length Video Using Markov-Modulated Gamma-Based Framework," *IEEE/ACM Trans. on Networking*, Vol. 11, No.4, 2003, pp. 638-649.
13. O. Rose, "Statistical Properties of MPEG Video Traffic and Their Impact on Traffic Modeling in ATM Systems," *in Proceedings of the 20th Annual Conference on Local Computer Networks*, October 1995.
14. D. Arifler, B. L. Evans, "Modeling the Self-Similar Behavior of Packetized MPEG-4 Video Using Wavelet-Based Methods", *in Proceedings of the IEEE International Conference on Image Processing,* Rochester, New York, USA, 2002, pp. 848-851.
15. F. H. P. Fitzek, M. Reisslein, "MPEG-4 and H.263 Video Traces for Network Performance Evaluation", *IEEE Network,* Vol. 15, No. 6, 2001, pp. 40-54.
16. D. P. Heyman, T. V. Lakshman, "What are the Implications of Long-Range Dependence for VBR-Video Traffic Engineering", *IEEE/ACM Transactions on Networking*, Vol. 4, No.3, 1996, pp. 301-317.
17. B. K. Ryu, A. Elwalid, "The Importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities", *in Proceedings of the ACM SIGCOMM 1996,* Stanford, CA, USA, pp. 3-14.

18. J. Beran, R. Sherman, M. S. Taqqu, W. Willinger, "Long-Range Dependence in Variable Bit-Rate Video Traffic", *IEEE Transactions on Communications*, Vol. 43, No. 2/3/4, 1995, pp. 1566-1579.
19. M. Krunz, H. Hughes, "A Traffic Model for MPEG-coded VBR Streams", *in Proceedings of the ACM SIGMETRICS,* Ottawa, Canada, 1995, pp. 47-55.
20. A. M. Law, W. D. Kelton, "Simulation Modeling & Analysis", 2nd Ed., McGraw Hill Inc., 1991.
21. P. A. Jacobs, P. A. W. Lewis, "Time Series Generated by Mixtures", *Journal of Time Series Analysis*, Vol. 4, No. 1, 1983, pp. 19-36.
22. T. V. Lakshman, A. Ortega and A. R. Reibman, "VBR Video: Trade-offs and potentials", *Proceedings of the IEEE,* Vol. 86, No. 5, 1998, pp. 952-973.
23. K. Park, W. Willinger (editors), "Self-Similar Network Traffic and Performance Evaluation", John Wiley & Sons, Inc., 2000.
24. R. Schafer, "MPEG-4: A Multimedia Compression Standard for Interactive Applications and Services", *IEE Electronics and Communications Engineering Journal,* Vol. 10, No. 6, 1998, pp. 253-262.