Modeling of multiplexed video streams in IP networks

EUGENIO COSTAMAGNA, member, IEEE, LORENZO FAVALLI, member, IEEE,
and FRANCESCO TARANTOLA

University of Pavia, via Ferrata 1, 27100 Pavia, Italy. Tel. +(39) 0382 505923. fax: +(39) 0382 422583.
e-mail {author}@ele.unipv.it

Abstract

In the recent years, it has been shown that the behavior of variable bit rate (VBR) video sources cannot be captured by the traditional Markov models since they do not exhibit a long-range dependent (LRD) characteristic which on the contrary is present in video traffic. Since then, a debate has been opened to understand the actual behavior of packet sources and the influence of this fact onto network analysis and network project. In this paper we compare two classes of video models: the first model belongs to a particular class of Markov models: the Hidden Markov Models (HMM), while the second is derived from application of chaotic attractors. The comparison is first related to their accuracy in modeling variable bit rate (VBR) video traffic out of an eventual video server, i.e. at the level of multiple aggregated streams. The two are then used to feed a simulated packet network to investigate whether they allow correct prediction of network performances.

Keywords—video modeling, traffic modeling, Hidden Markov Models, Chaos.

I. INTRODUCTION

As network traffic is expected to carry more and more video streams, the correct characterization of this type of sources is increasing its importance. The MPEG family of video coding standards [1] achieves high compression ratios by exploiting the reduction of both spatial correlation in *intra*-frame coding, using spatial transforms, and reduction of temporal correlation in *inter*-frame coding by means of motion compensation [2]. This produces a high variability in the offered load as Intra frames usually need from 2 to 5 times the number of bits necessary for inter frame coded frames (P and B frames in MPEG terminology).

Video distribution is furthermore to be seen as an aggregate of video streams coming out of a single server for multicasting over the network and a high quality of service (QoS) must be maintained for such services as video on demand since the user is expecting to receive the same quality signal he is used to receive from broadcaster or conventional cable TV.

Modeling the behavior of video sources has been the subject of countless studies and a discussion can be found in [3].

Traditionally, stochastic models (and specifically traffic models) are often implemented in the form of Markov processes as they provide a flexible framework that can be customized to fit different system properties [4]. They are also attractive since the statistical behavior of the system can be directly derived from the model. Their drawback is the definition of the state diagram and of its transition probabilities to properly describe the target system. This often leads to large and complex structures with several parameters to be tuned. Moreover, it has been shown that packet traffic exhibits long-range dependence (LRD) that cannot be captured by Markov chains as they can reproduce the autocorrelation function only in

the initial, fast decaying steps [5]. Since LRD is related to some *physical* characteristic of traffic [6], several studies have been performed to develop models that are intrinsically self-similar [7].
The discovery of LRD characteristics has raised a strong debate whether or not markovian models, known to be SRD, could be used to engineer network parameters [8].
In this work we compare the ability of a Markov based approach with that of a chaotic one in modeling a traffic mix in a simulated network environment where variable bit rate (VBR) MPEG video sources are multiplexed with other data traffic. We first describe (sections II and III) the two modeling approaches and then compare their ability to describe the statistical characteristics of the aggregated traffic in section IV. Finally 'real' and model sequences are compared in terms of network delays and packet losses under different traffic mixes.

II. HIDDEN MARKOV MODELS

Markov models have been applied since the beginning to describe the behavior of traffic events. In this large family of models we may find approaches that tend to represent some physical characteristics, binding the state space to particular values of the desired feature as well as approaches that absolutely disregard any physical meaning and 'simply' tend to emulate the *a posteriori* behavior.
In this case, there is no more correspondence between a state and a physical event and the properties are *hidden* in the model structure, from which the name of *hidden Markov models* (HMM) [9, 10].
From a mathematical point of view, an HMM can be described as a 5-ple, l=(S,V,A,B,Π) where S is the set of the states of the system, with cardinality N; *V* is the set of observable values ("the alphabet"), with cardinality M; *A* is the state transition probabilities matrix; *B* is the observable probabilities matrix and Π is the initial state vector.
Model parametrization is commonly performed using algorithms such as the method of moments or gradient or the Baum Welch (BW) procedure, [11] used in this work.
Given a set of observed samples $O=O_1 O_2 \ldots O_T$, the BW algorithm defines an iterative procedure that determines the parameters of a HMM by maximizing a *likelihood function* conditioned to the chosen model Pr(O/λ). To achieve this, the procedure defines two functions α and β as the *forward* and *backward* partial likelihood respectively, since they are determined at any given intermediate step as

$$\alpha_t(i) = P(O_1 \ldots O_t,\ q_t = S_i / \lambda)$$

and

$$\beta_t(i) = P(O_{t+1} \ldots O_T / q_t = S_i,\ \lambda).$$

Each of the partial likelihood functions is evaluated along any possible system evolution on the trellis describing the models. From these, the likelihood is finally derived as

$$P(O/\lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) \qquad t = 1, \ldots, T.$$

The BW algorithm is very robust in that it always converges, but there is no guarantee that it converges to a global maximum: the final parameters may consequently be not the optimal ones in an absolute sense. Furthermore, there is no guidance in the best selection of the number of states. For what the convergence speed is concerned, it is highly influenced by the size of the alphabet of symbols *V*.
It has been shown [12] that increasing the number of states, an HMM is capable to emulate the behavior of a chaotic map.

III. CHAOS BASED MODEL

The approach used is similar to that used in [13] and for the moment is straightforwardly derived from the experience gained in simulating error gap series in mobile radio channels [15, 17] and attenuations in satellite links [14]. It stems from the observation that a chaotic attractor provides a description of a dynamical system joining a deterministic description to an unpredictable behavior that translates in a number of properties of which one of the most important is the fact that small differences in sampling time along the trajectory result in completely different evolutions. This allows the use of such mathematical instruments to generate time series of events with arbitrary statistical properties. Note that we are not assuming to describe any *physical* characteristic of the underlying system.

The first step is the selection of a chaotic attractor. After some attempts with discrete maps [15] we selected a Lorenz attractor [16] for its richness. Then, the trajectory of the Lorenz attractor is sampled at different time instants and the coordinates of the point obtained are observed. The second step involves the determination of a proper transformation of the coordinates of the selected point capable to produce meaningful values according to the specific target system.

Samples taken at different time distances will give rise to time series with different correlation levels [17]: the superposition of these different steps will enable the generation of time series of arbitrary characteristics. This procedure is thoroughly described in [17] and can be formalized as follows. Let

- $f(\bullet)$ be an exponential or polynomial function, introduced to obtain suitable length ranges;
- u_i be the selected coordinates on the chaotic equations;
- k_i be weighting coefficients to be properly determined;
- $F(\bullet)$ be a shaping function, used to match the probability distribution of the chaotic time series to that of the original sequence.

Then the number of bits for the current event is

$$l = F[\, f\,(\textstyle\sum_i (k_i u_i))],$$

The shaping function is preliminarily obtained by means of the inversion procedure for the cumulative probability function described in [18]. The sampling rates and the weights k_i are the variable optimized by means of an automatic search procedure. For the moment, only a Nelder and Mead simplex procedure [19] has been tested: this was selected because of its simplicity, and because it involves only cost function calculations, and no function derivatives. No attempt has been made for the moment to optimize computing efficiency.

To tune the model parameters, the cost functions to be optimized are normally weighted sums of three terms, where the differences between the current and the reference values are considered for the following sequence features:

1 the difference (absolute value averaged over the sequence length) between the value of any sample and the value sampled after a delay of 20 sampling times, and a similar difference, but considering a delay of 1000 sampling time;

2 the differences (absolute value averaged over the sequence length) between six moving covariance functions, computed over non overlapping intervals with reference to the interval mean value), for interval lengths of 5, 10, 20, 30, 40 and 50 samples;

3 the differences (absolute value averaged over the sequence length) between six moving histograms, locally computed for interval lengths of 5, 10, 20, 30, 40 and 50 samples, considering differences from the interval mean values and abscissa intervals having amplitudes of some 4% of the sequence maximum sample value;

4 the behavior of the Time-Variance (T-V) plot.

Table 1:Statistical properties of MPEG sequences (Kbit/s).

	Mean	Peak	s	Burst
ASTERIX	558	2,248	259	4
TERMINATOR	273	850	290	3.1

Weight coefficients have been introduced to control the ratios between terms. With three terms, the for the cost C reads:

$$C = a_1 C_1 + a_2 C_2 + a_3 C_3$$

The features selected have been found suitable to account for both short term and long term correlation between values of the source series.

IV. TRAFFIC CHARACTERIZATION

Comparison is made by considering sequences at the Group Of Picture (GOP) level instead of single frames: such a level of abstraction can be used in bandwidth allocation schemes in order to smooth out the periodic peaks due to the coding process and is commonly applied to reduce the model space complexity. We have used two MPEG sequences: ASTERIX and TERMINATOR [20]. Each sequence consists of 40,000 frames, which corresponds to approximately half an hour of video; their statistical properties (at GOP level, and considering only the luminance component) are summarized in Table 1. The two sequences are multiplexed with random starting points.

Different network topologies and traffic mixes have been considered although only results for the generic network in Figure 1 are given. The link between Nodes 4 and 5 is used to describe the overall network behavior so that limiting the available bandwitdh on it can account for different network loads. Transmission of the video source is performed using Real Time Protocol (RTP) while other traffic is a mix of RTP and TCP connections with the total load almost evenly splitted between the two types of sources. TCP traffic is generated following a Pareto distribution. Simulations are performed using Network Simulator 2. RTP sources start after the network has been loaded for the equivalent of a second with only TCP sources. Packet size has been set to 1024 and each node has FCFS queues of 25 packets.

Several simulations have been run initially to determine the sensitivity of the two approaches to their parameters. In particular, for the HMM, a good compromise between precision of the model and convergence time of BW procedure has been found using 10 states and quantizing the alphabet V in the model so that only 50 bit rate values are possible.

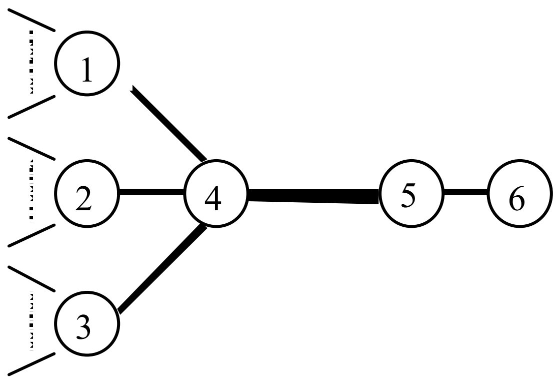

Figure 1. Generic network topology

For what the chaotic model is concerned, two different attempts have been made: In the following results obtained using the first set of parameters will be identified as ‘chaos001’ while ‘chaos002’ will be used for the second set. Both models are the superposition of nine attractors with different sampling steps and the main difference is in the speed of the ‘fastest’ attractor that is sampled about seventy times more densely in case of the chaos002 model. Furthermore, chaos001 is optimized only starting from the covariance functions that represent a short-term characteristics while the cost function to determine parameters for model chaos002 includes a term that may be related to the T-V function and consequently to the long-term behavior.

The first set of comparisons considers the stream before transmission. The time series generated by the two models (HMM and chaos002) are first compared to the original sequence (the *target*) in Figure 2. At this level of abstraction both modeling approaches seem to reproduce very well the dynamic of the target sequence although neither one is excessively fitted to it. As seen from Figure 2, both the short and the long-range model behaviors are well satisfactory, as are the moving histograms. However, some discrepancies are found at a deeper level of comparison with the target sequence, the significance of which are investigated in the following. In Figure 3, the T-V plots of the HMM and the two chaotic sequences are compared to the reference.

With respect to the curve derived from the reference sequence, the curve from the HMM is somewhat lower, whereas the chaos curves are somewhat higher. Notice that, in fact, the T-V plot was not included in the cost function leading to the chaos001 model. This allows investigating about possible differences when simulating traffic behaviors at the network level: we recall that features of the T-V plot are relevant when describing self-similarity. In particular the asymptotic slope is related to the well-known Hurst parameter (the 0.5 slope straight line is reported for the sake of comparison). All sequences are in the LRD region and this is confirmed by calculation of the Hurst parameter given in Table 3. In this table the Hurst parameter is calculated at the input of the network and after the multiplex node for different traffic loads.

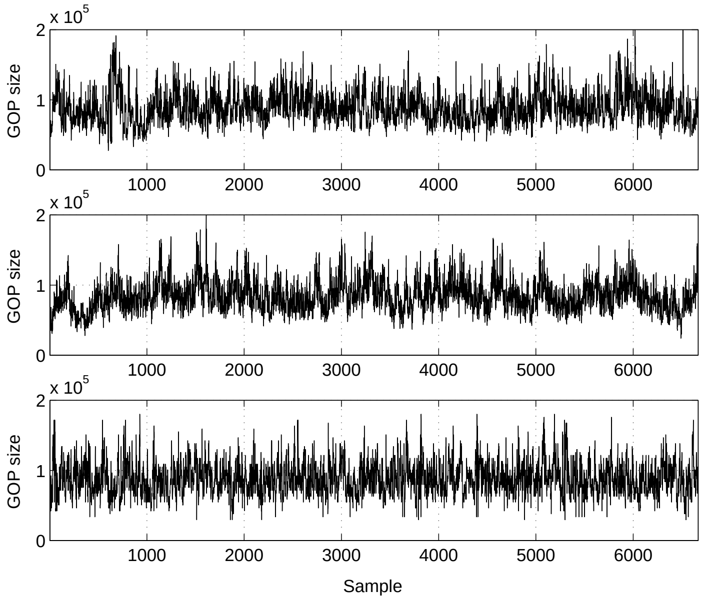

Figure 2. The original multiplexed traffic sequence (at GOP level) and two simulated sequences, from a multiple attractor model (chaos002) and from the HMM model.

Table 2. Values of the Hurst parameter for the three models and target

	Original	hmm	chaos001	chaos002
pre-tx:	0.688	0.635	0.761	0.745
100:	0.685	0.630	0.760	0.742
130:	0.703	0.637	0.756	0.731
150:	0.737	0.709	0.731	0.710
200:	0.785	0.775	0.759	0.751

It is interesting to note that the HMM model always provides an underestimate of the parameter but generates a sequence that changes in the simulations accordingly with that of the target. The two chaotic models instead give a rather constant Hurst value irrespective of network behavior.

In Figure 4, the moving covariances obtained from the models are compared to the target sequence. Although the curves computed with GOP figures read over the shortest intervals can be hardly considered true covariances, as integration is performed over very small numbers of samples (only two for the five sample intervals), the agreement among curves is normally well satisfactory. However, different behaviors are perceived for very short covariance lags. Both the chaos covariances show a very fast decay from the maximum values for lags of 1 or 2 samples, whereas the HMM covariances exhibit a behavior more matched to the target.

Probably, this is due to the influence of model attractor sampled at high velocity in terms of curvilinear abscissa along the trajectory, which determines the very short-term behavior of the sequence, avoiding any objectionable "ringing" effect due to the Lorenz attractors but introducing a somewhat excessive pseudo-random effect.

At this point we start loading the network and determine the distribution of the interpacket delay and of the number of bits per GOP at the output node (after that the packets with delay in excess to that of a GOP size have been discarded).

The correlation of the interarrival times for different traffic loads is described in Figure 5 while Figure 6 gives the correlation curves for the number of bits. In both cases we see that the HMM model underestimates this correlation while the chaotic models are always above the target at least in the initial part of the curves while the long term behavior appears always more erratic.

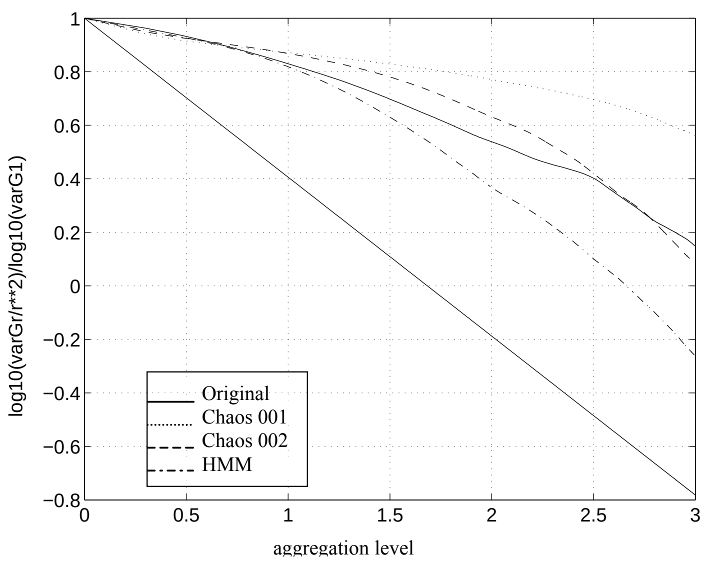

Figure 3. The Time-Variance plots derived from the HMM and the two chaotic models.

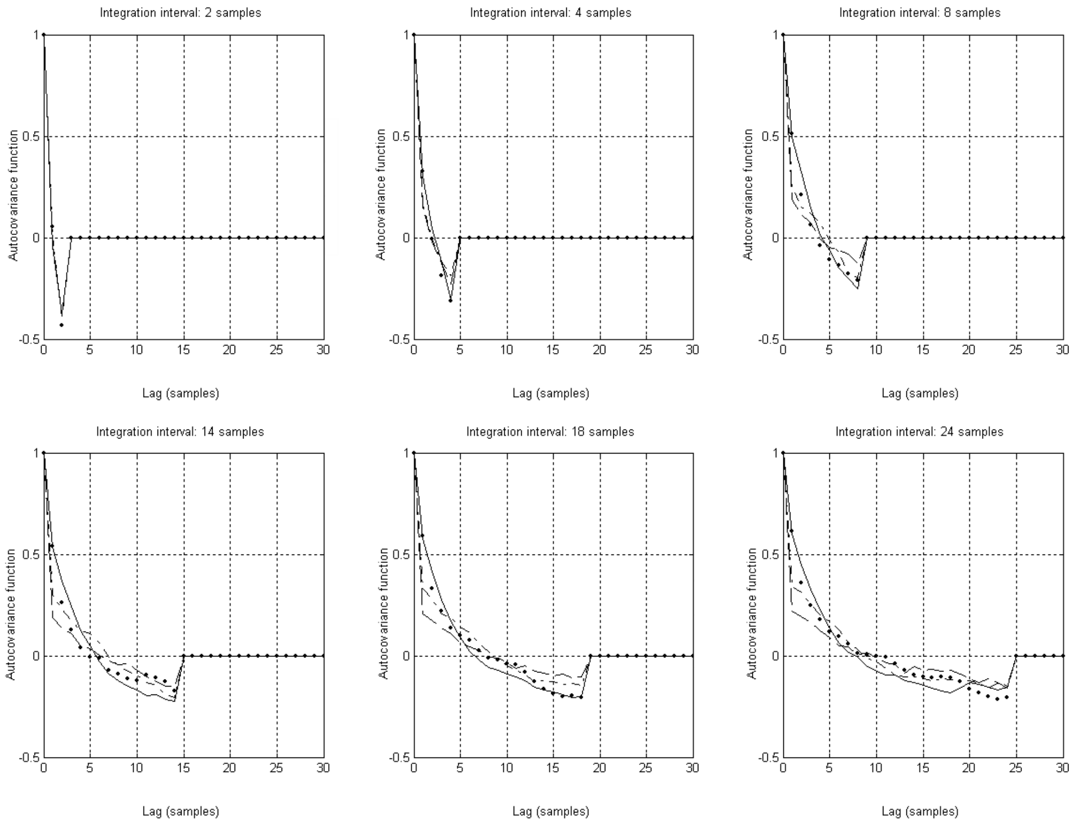

Figure 4. Moving covariances for the three models and different windows. Dots represent the target sequence.

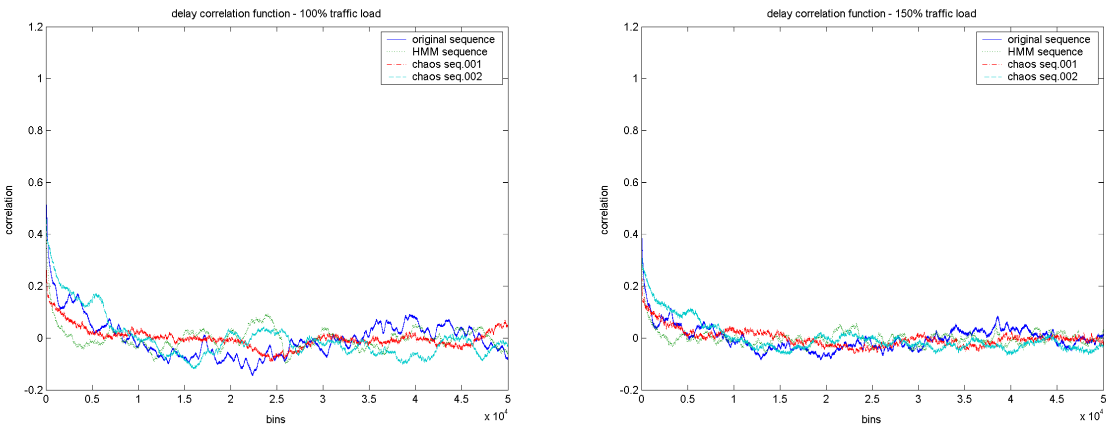

Figure 5. Delay correlation function at different traffic loads (right to left: 100%, 150%).

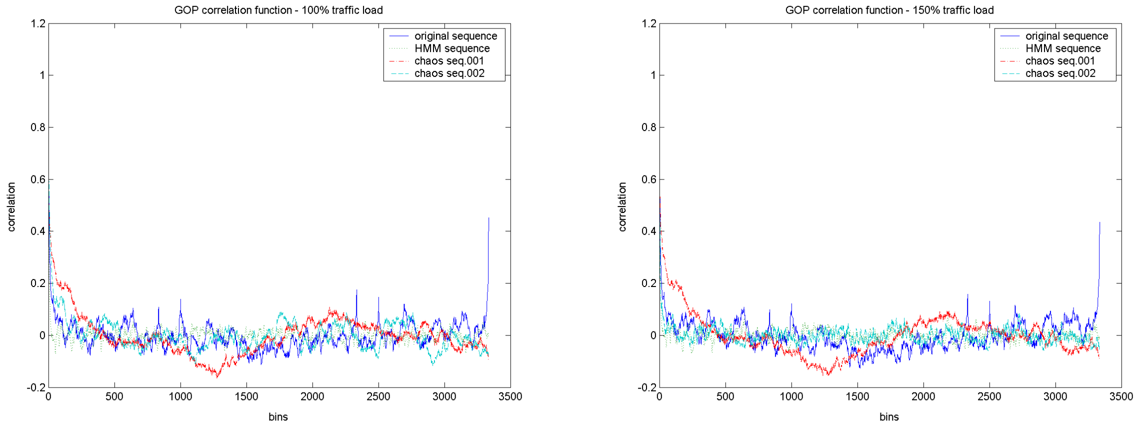

Figure 6. Bit per GOP correlation function at different traffic loads (right to left: 100%, 150%).

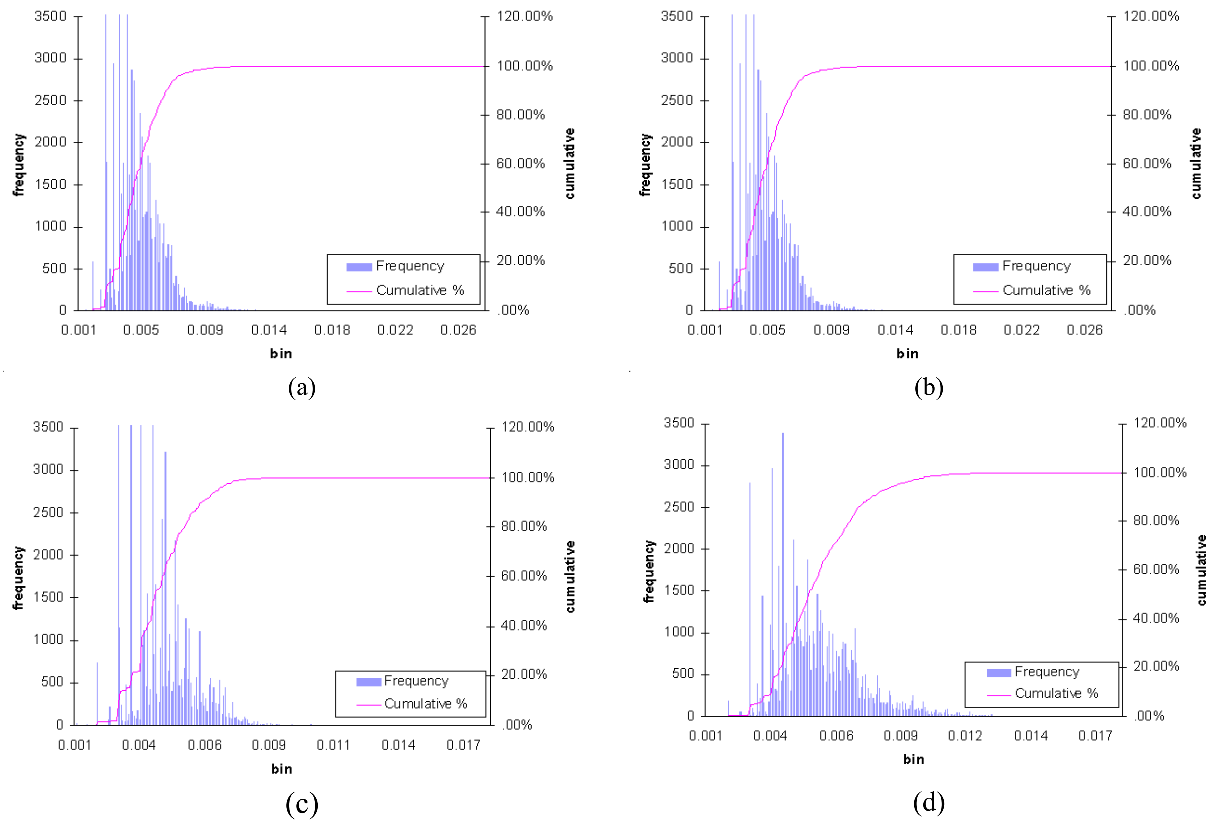

Figure 7. Delay statistics at 100% load: (a)original, (b) HMM, (c) chaos 001, (d) chaos 002.

Finally, Figure 7 provides in detail the distribution and the cumulative function of the interarrival times. It is possible to see that histograms of the two chaotic models have a different shape, with values more dispersed than the original one.
Figure 8 summarizes the results of the cumulative distributions at 100% traffic while a 150% overload situation is in Figure 9. We may note that all models differ of a similar quantity at 100% load with a slight overestimate of the delay for both HMM and chaos_001. The HMM model is close to the original sequence also at the higher load of Figure 9.

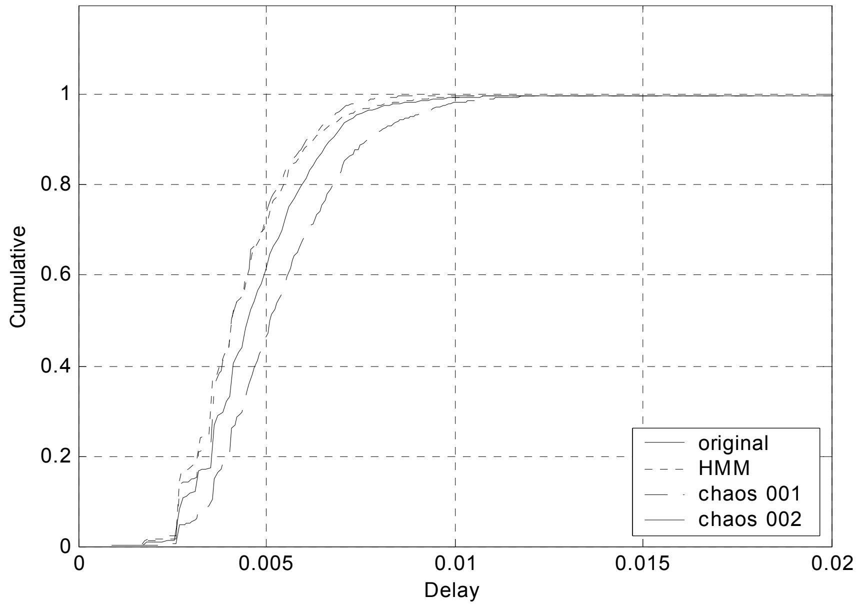

Figure 8. Cumulative distribution at100% load (see also figure 7).

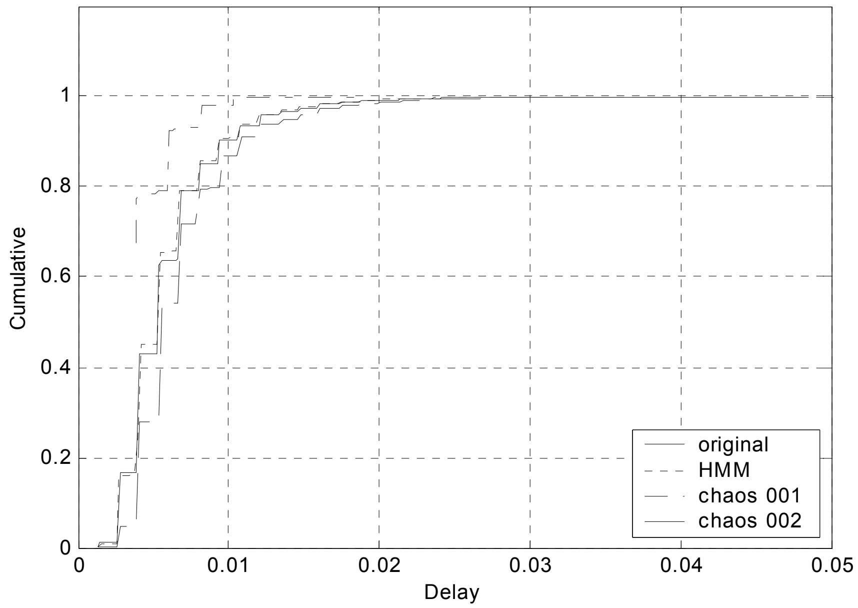

Figure 9. Cumulative distribution at150% load.

The two chaotic models, on the contrary, don't show a clear behavior: in Figure 8 in particular they give opposite indications, but are close enough to the behavior of the original sequence. If we increase the overall traffic as in Figure 9, the sequence chaos002 becomes undistiguishable from the original and HMM traces while the sequence chaos001 still provides an overestimate of the initial part of the curve. This indeed leads to an underestimate of the probability of having higher interarrival times.
A final remark is devoted to the complexity of the two models. Both are able to generate data at run time and they are also comparable in terms of time required to tune the parameters: convergence time is about five minutes on a 1.4G Pentium IV pc with 256MB memory.

V. CONCLUSIONS

The work presented in this paper discusses the ability of different approaches to describe the behavior of VBR video traces transmitted over an IP network. The results are somehow interlocutory as they show that the HMM performs better than the chaotic models but chaotic models catch up when traffic increases. Furthermore, there is an indication that LRD is indeed relevant in assessing the parameters of the models since of the two chaotic models provided, the one with better performances is the one whose parameters have been optimized considering the long term statistics. Work is currently in progress to achieve a better understanding of the statistics of interest in more realistic scenarios with larger buffers and different traffic mixes.

REFERENCES

[1] ITU-T Recommendation H.262, also ISO/IEC 13818-2, Information Technology - generic coding of moving pictures and associated information, 1994.

[2] A. N. Netravali, and B.G. Haskell, *Digital pictures. Representation, Compression, and Standards.* Second Edition. Plenum Press, New York, 1995.

[3] N. Ohta, *Packet video. Modeling and Signal Processing.* Artech House, Boston, 1994.

[4] D. Heyman, and T.V. Lakshman: "Source models for VBR broadcast-video traffic," *IEEE Trans. on Networking*, Vol. 4, No. 1, pp. 40-48, Feb. 1996

[5] J. Beran, R. Sherman, M. S. Taqqu, W. Willinger: "Long-range dependence in variable bit rate video traffic," *IEEE Trans. Comm.*, Vol. 43, No. 2,3.4, pp. 1566-1578, Feb.-Apr. 1995

[6] W. Willinger, M.S.Taqqu, R.Sherman, and D.V. Wilson, "Self-Similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at Source Level," (Extended Version), *IEEE/ACM Transactions on Networking* Vol. 5, N. 1, pp.1-16, February 1997

[7] K. Park, and W. Willinger, eds., *Self-similar Network Traffic and Performance Evaluation.* J. Wiley and Sons, 2000.

[8] D.P. Heyman, and T. V. Lakshman, "What are the implications of long range dependance for VBR-video traffic engineering?," *IEEE/ACM Tr. On Networking*, vol. 4, N. 3, pp. 301-317, June 1996.

[9] L.R Rabiner and B.H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, pp. 4-16, Jan. 1986.

[10] G. Lindgren, and U. Holst, "Recursive Estimation of Parameters in Markov-Modulated Poisson Processes," *IEEE Transactions On Communications*, Vol. 43, No. Ii, November 1995 pp. 2812-2819.

[11] W. Turin, and R van Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE JSAC*, vol. 16, n.9 pp. 1809-1817, Dec. 1998.

[12] D.H. Kil and F.B. Shin, *Pattern recognition and prediction with application to signal characterization*, American Institute of Physics Press, Woodbury, New York, 1996.

[13] E. Costamagna, L. Favalli, P. Gamba, and G. Iacovoni, “A simple model for VBR video traffic based on chaotic maps: validation through evaluation of ATM multiplexer QoS parameters,” *IEEE - ICC’98*, Atlanta (GA), June 7-11, 1998, pp. 568-572

[14] E. Costamagna, L. Favalli, P. Gamba, “Ka band fading channel models derived from chaotic dynamical systems, ” *Proc of the IEEE 54th Vehicular Technology Conference* VTC Fall 2001, 7-11 October 20001, Atlantic City, NJ, USA.

[15] E. Costamagna and A. Schirru, “Channel error models derived from chaos equations,” in *Proc. 1994 IEEE Int. Symp. on Systems, Man, and Cybernetics*, San Antonio, TX, 1994, vol. 1, pp. 577-581.

[16] H. O. Peitgen, H. Jurgens, D. Saupe, *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, 1992.

[17] E. Costamagna, L. Favalli, P. Gamba, and P. Savazzi, “Block-error probabilities for mobile radio channels derived from chaos equations,” *IEEE Communications Letters*, vol. 3, pp. 66-68, March 1999.

[18] A. Papoulis, Probability, Random Variables, and Stochastic Processes 3rd ed., McGraw-Hill, Toronto: 1991.

[19] W. H. Press, S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery, "Numerical Recipes in FORTRAN, the art of scientific computing," Cambridge University Press, 1992.

[20] O. Rose, *Statistical properties of MPEG video traffic and their impact on traffic modelling in ATM system*, University of Wuerzburg, Institute of Computer Science Research, Report Series, Report No. 101, Feb. 1995.