Computer Communications 33 (2010) 1235–1242

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Short Communication

Modeling multiplexed traffic from H.264/AVC videoconference streams

Aggelos Lazaris [a], Polychronis Koutsakis [b,*]

[a] Department of Computer Science and Engineering, University of California, Riverside, USA
[b] Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece

ARTICLE INFO

Article history:
Received 14 December 2009
Received in revised form 9 March 2010
Accepted 10 March 2010
Available online 20 March 2010

Keywords:
Video traffic modeling
Discrete autoregressive model
H.264/AVC video

ABSTRACT

In this correspondence, we initially investigate the statistical behavior of bursty video traffic originating from single H.264/AVC streams, by modeling it with well-known distributions. Our results help to build a Discrete Autoregressive model which captures well the behavior of multiplexed H.264/AVC videoconference sources.

1. Introduction

As traffic from video services is expected to be a substantial portion of the traffic carried by emerging wired and wireless networks [1,2], 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. In the context of videoconference streaming, video packet delay requirements are strict, and this is at odds with the high burstiness of video traffic. Whenever the delay experienced by a video packet exceeds the corresponding maximum delay, the packet is dropped. Even a low video packet dropping probability may considerably deteriorate the viewer's quality of experience. In addition, errors due to packet loss in a reference frame (i.e., I frames) propagate to all of the dependent difference frames (i.e., B and P frames). This phenomenon is well-known as *propagation of errors* and has a very important impact on the quality of the received video sequence, depending on the effectiveness of the encoding/decoding scheme. 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 [3,4], discrete AR (DAR) models [3,5], Markov renewal processes (MRP) [6], MRP transform-expand-sample (TES) [7], finite-state Markov chain [8], Gamma-beta-auto-regression (GBAR) models [9,10] (which capture data-rate dynamics of VBR video conferences well but were found in [10] to not be suitable for general MPEG video sources), discrete-time semi-Markov processes (SMP) [11], wavelets [12], multifractal and fractal methods [13]. In [14,15], different approaches are proposed for MPEG-1 traffic, based on the Log-normal, Gamma, and a hybrid Gamma/Lognormal distribution model, respectively. Finally, the authors in [32] focus on H.264/AVC movies (i.e., not videoconference traffic as we do in the present work) and propose a comparison between the self-similar fractional Brownian motion (fBm) model and the two-state Markov modulated fluid flow (Mmff) model, in terms of predicting the buffer threshold exceeding probability by using the effective bandwidth approach. They showed that Mmff provides better results than the fBm model. Still, in the absence of a very accurate statistical model for each individual flow the use of the effective bandwidth approach leads to a significant overestimation of the sources' actual bandwidth requirements [33].

H.264 is the latest video coding standard of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). It has recently become the most widely accepted video coding standard since the deployment of MPEG-2 at the dawn of digital television, and it may soon overtake MPEG-2 in common use [16]. It covers all common video applications ranging from mobile services and videoconferencing to IPTV, HDTV, and HD video storage. Standard H.264 encoders generate three types of video frames: I (intra-coded), 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 estimation and compensation techniques. I frames are, on average, the largest in size, followed by P and then by B frames. The video coding layer of H.264/AVC (advanced video codec) is similar to that of other video coding standards such as MPEG-2 Video. In fact, it uses a fairly traditional approach

* Corresponding author. Tel.: +30 28210 37235; fax: +30 28210 37542.
E-mail addresses: alazaris@cs.ucr.edu (A. Lazaris), polk@telecom.tuc.gr (P. Koutsakis).

0140-3664/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2010.03.014

consisting of a hybrid of block-based temporal and spatial prediction in conjunction with block-based transform coding [16]. In 2007, the scalable video coding (SVC) extension has been added to the H.264/AVC standard. The SVC extension provides temporal scalability, coarse grain scalability (CGS), medium grain scalability (MGS), and SNR scalability in general, spatial scalability, and combined spatio-temporal-SNR scalability [17]. The study of H.264/SVC is out of the scope of this paper. In the rest of the paper, we use the term "H.264" to refer to the H.264/AVC video standard.

Similarly to our recent work on modeling H.263 and MPEG-4 videoconference traffic [18], our present work initially focuses on the accurate fitting of the marginal (stationary) distribution of video frame sizes of single H.264 video traces. More specifically, our work follows the steps of the work presented in [5], where Heyman et al. analyzed three videoconference sequences coded with a modified version of the H.261 video coding standard and two other coding schemes, similar to the H.261. The authors in [5] found that the marginal distributions for all the sequences could be described by a Gamma (or equivalently negative binomial) distribution and used this result to build a discrete autoregressive (DAR) model of order one, which works well when several sources are multiplexed. An example of multiplexed videoconference transmission is illustrated in Fig. 1. An important feature of common H.264 encoders is the manner in which frame types are generated. Typical encoders use several group-of-pictures (GOP) patterns when compressing video sequences; 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 this correspondence, we focus on the problem of modeling videoconference-type traffic from H.264 encoders (i.e., in the absence of available actual videoconference sequences we choose, as we did in [18,31] traces of content and traffic characteristics (high autocorrelation, low or moderate motion) similar to those of videoconference traffic). The modeling of H.264 traffic is a relatively new and yet open issue in the relevant literature. A much shorter version of this work has been presented in [30]. The main differences between this work and [30] are that in the current work we validate our model by conducting a queueing performance study and we extend our results by studying and modeling more video traces. In particular, our initial study was conducted using 18 video traces (i.e., "Sony Demo" and "NBC News", in several formats) while in this paper we have included also the results taken from two additional traces (i.e., "From Mars to China" and "Horizon"). The new queueing performance study validates that our modeling results provide a very good fit to the real data, in terms of packet loss ratio and packet waiting time. More details about the traces used can be found in Section 2.1. The rest of the paper is structured as follows: Section 2 presents the results for single-source H.264 traffic modeling, the evaluation of which leads us to Section 3, where the DAR (1) model is presented and the simulations results derived by its use are discussed. In Section 4, a queuing performance study is presented to further validate our proposed modeling approach. Finally, Section 5 concludes the paper.

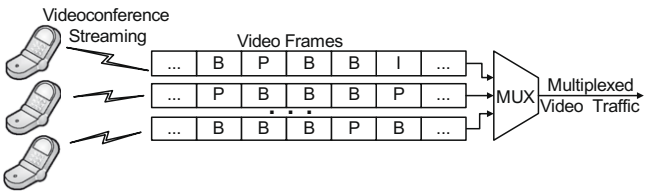

Fig. 1. An example of multiplexed videoconference transmission. Each video source generates a frame sequence based on its GoP pattern.

2. Single-source H.264 traffic modeling

The first step in our modeling approach is to analyze the statistical behavior of single-source H.264 videoconference traces. We first analyze the frame-size histograms of the individual sources and we try to fit these histograms with well-known distributions. Then, we evaluate the autocorrelation coefficients of each trace and we perform a statistical testing analysis by using three well-known statistical tests.

2.1. Frame-size histograms

In our work, we have studied four different long sequences of H.264 VBR encoded videos in 20 formats, from the publicly available Video Trace Library of [19], in order to derive a statistical model which fits well the real data. The traces used are in common intermediate format (CIF) (i.e., 352×288 pixels) and in high definition (HD) 720 and 1080 format (i.e., 1280×720p and 1920×1080i, respectively). In addition, we used several different quantization parameters (QP) for the traces under study. The four traces are: (1) a demo from the Sony Digital Video Camera, (2) an excerpt of NBC News, (3) an excerpt of KAET's documentary From Mars to China, and (4) an excerpt of KAET's news program Horizon. The length of all the videos is either 10 or 30 min. 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. Without loss of generality, we use 48-byte packets throughout this work, but our modeling approach can be used equally well with packets of other sizes. Table 1 presents the frame-size statistics for each trace. The interframe period is 33.3 ms.

We have investigated the possibility of modeling the 20 different formats of the traces 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

Table 1
Trace statistics.

Video name	[Res, G, B, F]	Mean (bits)	Peak (bits)	Standard deviation (bits)
NBC News	[CIF, 16, 1, 28]	15,816	181,096	21705.24
NBC News	[CIF, 16, 1, 48]	1197	28,032	2219.26
NBC News	[CIF, 16, 3, 28]	14,632	182,520	21631.47
NBC News	[CIF, 16, 3, 48]	1084	28,216	2237.75
NBC News	[CIF, 16, 7, 28]	15,081	186,872	21613.23
NBC News	[CIF, 16, 7, 48]	1054	29,768	2275.84
NBC News	[CIF, 16, 15*, 28]	16,624	192,272	21365.03
NBC News	[CIF, 16, 15*, 48]	1059	31,840	2290.61
Sony Demo	[CIF, 16, 1, 28]	14,067	221,664	27439.89
Sony Demo	[CIF, 16, 1, 48]	954	23,096	2166.51
Sony Demo	[CIF, 16, 3, 28]	12,801	222,888	27752.93
Sony Demo	[CIF, 16, 3, 48]	887	23,232	2203.77
Sony Demo	[CIF, 16, 7, 28]	13,129	227,680	28053.90
Sony Demo	[CIF, 16, 7, 48]	898	25,480	2289.92
Sony Demo	[CIF, 16, 15,[a] 28]	14,861	233,296	28338.22
Sony Demo	[CIF, 16, 15,[a] 48]	933	28,224	2412.26
Sony Demo	[HD,[b] 12, 2, 38]	22,513	398,544	51815.57
Sony Demo	[HD,[b] 12, 2, 48]	7618	143,408	18103.28
Mars to China	[HD,[c] 12, 2, 28]	161,656	2,615,240	241338.76
Horizon	[HD,[c] 12, 2, 28]	51,159	800,488	94292.20

Res, resolution; G, GoP size; B, number of successive B frames; F, quantization parameters.

[a] When B = 15 and G = 16 there are no P frames in the trace sequence.

[b] HD 720.

[c] HD 1080.

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 of these distributions have been often used for video traffic modeling in the literature, they have been included in this work as fitting candidates, in order to compare their modeling results in the case of H.264 videoconferencing. The frame-size histogram based on the complete VBR streams is shown, for all 20 sequences, to have the general shape of a Pearson type V distribution. Fig. 2 presents indicatively the histogram for the NBC News sequence with parameters ([CIF, G16, B7, F28]).

2.2. Statistical tests and autocorrelations

Our statistical tests were made with the use of qualitative tests like Q–Q plots [5,20], and quantitative tests like Kolmogorov–Smirnov tests [20] and Kullback–Leibler divergence tests [21]. The Q–Q plot is a powerful goodness-of-fit test, 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:

$$P((X \leqslant x) = z \tag{1}$$

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. It uses the maximum vertical deviation between the two curves as its statistic D which is defined as follows:

$$D = \sup_x \{|F(x) - G(x)|\} \tag{2}$$

where $F(x)$ and $G(x)$ are the empirical distribution function of the original data and the cumulative distribution function of the model, respectively. The Kullback–Leibler divergence test (KL-test) is a measure of the difference between two probability distributions f and g and is defined as follows:

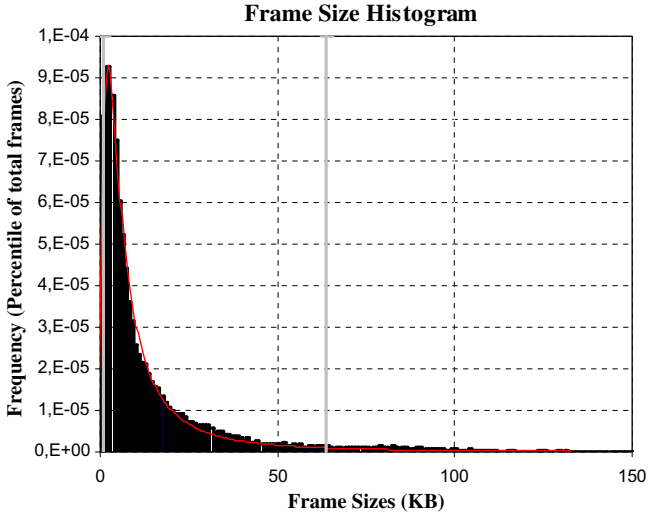

Fig. 2. Frame-size histogram for the NBC News trace with parameters: [CIF, G16, B7, F28].

$$I(f,g) = \sum_{i=1}^{k} p_i \log\left(\frac{p_i}{\pi_i}\right) \tag{3}$$

where log denotes the natural logarithm. Here, there are k possible outcomes of the underlying random variable; the true probability of the ith outcome is given by p_i, while the $\pi_1, \ldots, \pi_k$ constitute the approximating probability distribution (i.e., the approximating model). In this case, we have $0 < p_i < 1$, $0 < \pi_i < 1$, and $\sum p_i = \sum \pi_i = 1$. Hence, here f and g correspond to the p_i and π_i, respectively. The notation $I(f,g)$ denotes the information lost when g is used to approximate f or the distance from g to f.

The Pearson V distribution fit was shown to be the best in comparison to the Gamma, Weibull, Lognormal and Exponential distributions (comparisons were also made with the Negative Binomial and Pareto distributions, which were also worse fits than the Pearson V). This is the third study (after our work in [18,31] on H.263 and MPEG-4 videoconference traffic) in which the Pearson V outperforms, as a fit, the gamma distribution which has been shown in a number of studies in the past to be the best fit for H.261 and MPEG-1 traffic [5,15]. Based on the results of all three studies, *we conclude that for videoconference traffic encoded by modern encoding standards the Pearson V distribution is the best fit*. 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 H.264-encoded streams, based on the cyclic GoP formats (therefore, this case is different than the case of H.263 traffic we studied in [18], where the number of I frames was so small in each trace that the trace could be modeled as a whole). 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 [10,23]. Another approach, similar to the above, was proposed in [14]. 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 [6,15] that the average sizes of P and B frames can vary by 20% and 30% (often more than that), respectively, in subsequent scenes, therefore the size changes are statistically significant; this is also the case for H.264/AVC 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 following 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 Probability Density Function (PDF) of a Pearson type V distribution with parameters (α, β) is:

$$f(x) = \frac{x^{-(a+1)} e^{\frac{-\beta}{x}}}{\beta^{-a}\Gamma(\alpha)} \tag{4}$$

for all $x > 0$, and zero otherwise.

The mean and variance are given by the equations:

$$\text{Mean} = \frac{\beta}{\alpha - 1} \tag{5}$$

$$\text{Variance} = \frac{\beta^2}{(\alpha-1)^2(a-2)} \tag{6}$$

The autocorrelation coefficient of lag-1 was also calculated for all types of video frames of the 20 formats of the traces, as it shows the very high degree of correlation between successive frames of the same type. 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 the 60 cases (20 traces, 3 types of frames per trace) studied. In order to further verify the validity of our results, we performed Kolmogorov–Smirnov and Kullback–Leibler tests for all of the 60 fitting attempts. The results of our tests confirm our respective conclusions based on the Q–Q plots (i.e., the Pearson V distribution is the best fit). Table 2 presents indicative results from the KL-tests and Fig. 3 presents indicative results from the KS-test.

Although controversy persists regarding the prevalence of long range dependence (LRD) in VBR video traffic [22,24,25], in the specific case of H.264-encoded video, we have found that LRD is important. The autocorrelation function for the NBC News ([CIF, G16, B7, F28]) trace is shown in Fig. 4 (the respective figures for the other three traces are similar). Three apparent periodic components are observed, one containing lags with low autocorrelation, one with medium 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 were clearly better than the results produced by modeling the whole sequence uniformly, the high autocorrelation shown in Fig. 4 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 a DAR model, which inherently uses the autocorrelation coefficient of lag-1 in its estimation. The model will be shown to accurately capture the behavior of multiplexed H.264 videoconference movies, by generating frame sizes independently for I, P and B frames.

Table 2
Kullback–Leibler tests.

Video		Expon	Gamma	Lognorm	Pearson V	Weibull
Sony Demo	I	1.149989	0.387741	0.378286	0.364346	0.452904
CIF	P	1.296492	1.027408	0.864774	0.721336	1.023750
G16,B3,F48	B	1.061298	0.542759	0.478995	0.432177	0.533222
Mars2China	I	1.027828	0.396205	0.352191	0.343299	0.429817
HD 1080	P	1.142010	0.648857	0.568190	0.523317	0.721622
G12,B2,F28	B	1.387990	1.271812	1.021952	0.869845	1.309497
Horizon	I	0.907249	0.115675	0.110301	0.107200	0.170905
HD 1080	P	0.985355	0.362541	0.324389	0.301241	0.569717
G12,B2,F28	B	1.206802	1.014664	0.781854	0.668262	0.868191

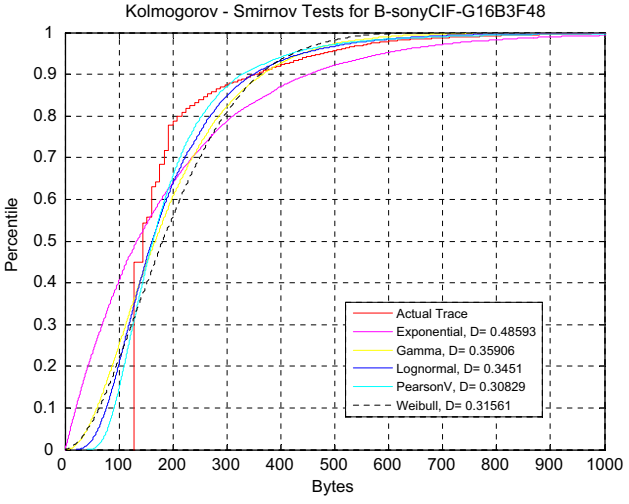

Fig. 3. KS-test (comparison percentile plot) for the Sony Demo B frames ([CIF, G16, B3, F48]).

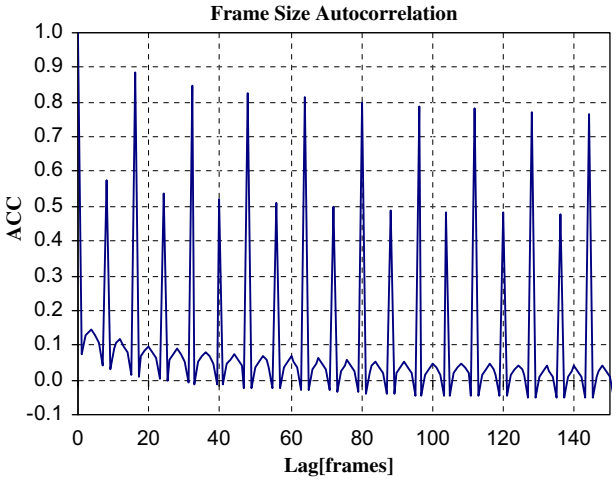

Fig. 4. Autocorrelation coefficients of the NBC News trace ([CIF, G16, B7, F28]).

3. The DAR (1) model – results and discussion

A discrete autoregressive model of order p, denoted as DAR (p) [26], 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 [26]. 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 $1-\rho$ 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: $X_n = V_n X_{n-1} + (1 - V_n)Y_n$, is a DAR (1) process. A DAR (1) process is a Markov chain with discrete state space S and a transition matrix:

$$\boldsymbol{P} = \rho \boldsymbol{I} + (1-\rho)\boldsymbol{Q} \tag{7}$$

where ρ is the autocorrelation coefficient, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{Q}$ is a matrix with $Q_{ij} = \pi(j)$ for $i, j \in S$. Autocorrelations are usually plotted for a range W of lags. The autocorrelation can be calculated by the formula:

$$\rho(W) = \frac{\mathrm{E}[(X_i - \mu)(X_{i+w} - \mu)]}{\sigma^2} \tag{8}$$

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

As in [5], 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

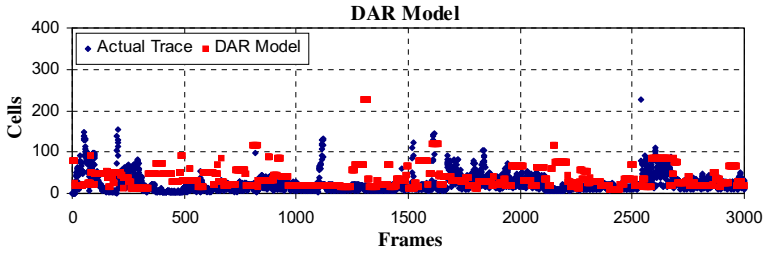

Fig. 5. Comparison for a single trace between a 10,000 frame sequence of the actual B frames sequence of the NBC News ([CIF, G16, B15, F28]) trace and the respective DAR (1) model in number of cells/frame (*Y*-axis).

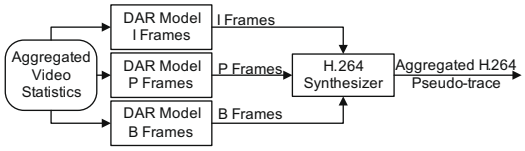

Fig. 6. The DAR (1) model used to produce an H.264 pseudo-trace for a given set of video statistical properties.

correlations, as already explained, are typically very high for video-conference sources). The DAR (1) model can be used with any marginal distribution [27].

As already explained, 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 20 traces under study. More specifically, in our model the rows of the $\mathbf{Q}$ matrix consist of the Pearson type V probabilities $(f_0, f_1, \ldots, f_k, F_K)$, where $F_K = \sum_{k>K} f_k$, 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 (7) 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. This is evident in Fig. 5, where we compare the actual B frames sequence of the NBC News ([CIF, G16, B15, F28]) 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 sequence of I, P, or B frames of the actual NBC News 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 other traces under study. 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 [5] have reached similar

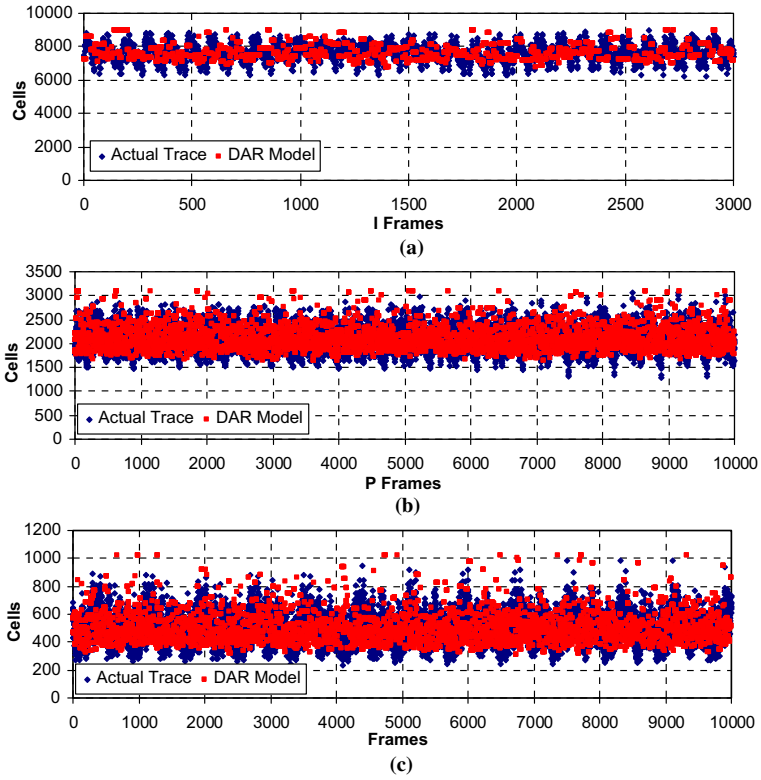

Fig. 7. Comparison for 30 superposed sources between: (a) 3000 I frames, (b) 10,000 P frames, and (c) 10,000 B frames of the actual NBC News ([CIF, G16, B1, F28)] trace, and the respective DAR (1) model in number of cells/frame (*Y*-axis).

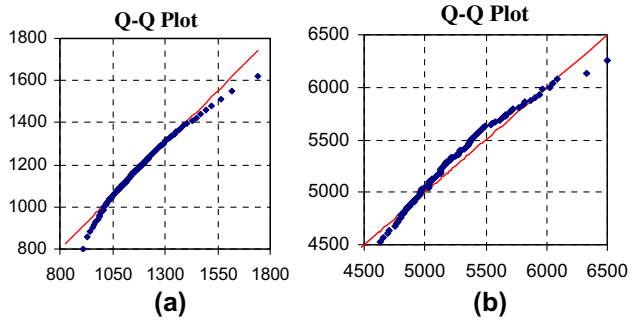

Fig. 8. Q–Q plot of the DAR (1) model versus the actual video for the: (a) P frames of NBC News ([CIF, G16, B3, F48]), and (b) I frames of NBC News ([CIF, G16, B7, F48]), for 30 superposed sources.

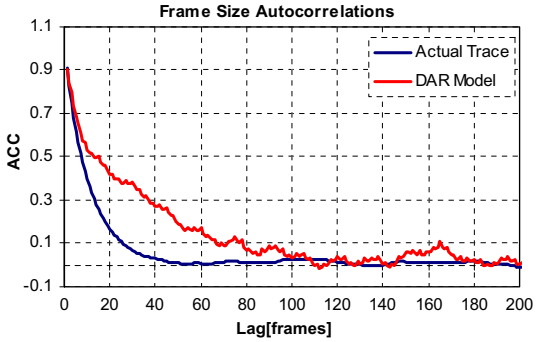

Fig. 9. Autocorrelation versus number of lags for the I frames of the actual NBC News ([CIF, G16, B15, F28]) trace and the DAR (1) model, for 30 superposed sources.

conclusions for their own DAR (1) model and they present results for a superposition of 20 traces). An overview of the proposed system's architecture is illustrated in Fig. 6. The improved fitting in this case is clear in Fig. 7, which present the comparison between our DAR (1) model and the actual I, P, B frames' sequences of the NBC News ([CIF, G16, B1, F28)], for a superposition of 30 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 30 sources, by using different starting points in the video sequence. Thus, the resulting traces are shifted in time, so they have different bandwidth requirements for a given sequence of frames, but still possess the same statistical properties (i.e., autocorrelation coeffi-

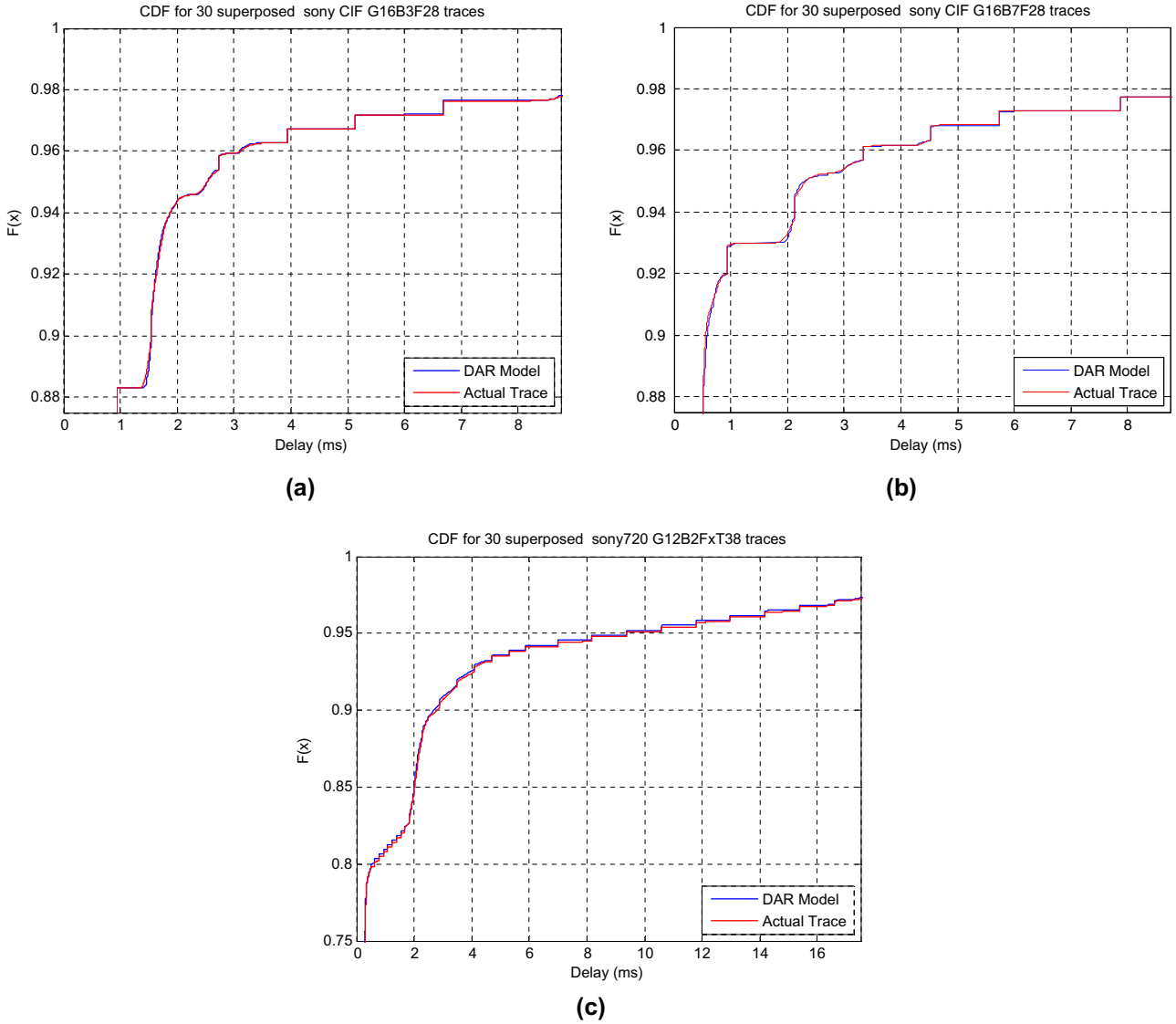

Fig. 10. Waiting time CDFs versus the respective DAR (1) model for 30 superposed sources for the: (a) NBC News trace with parameters [CIF, G16, B3, F28], (b) Sony Demo trace with parameters [CIF, G16, B7, F28], and (c) Sony Demo trace with parameters [HD, G12, B2, F38] (unlimited buffer size).

cients, mean and variance). 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 provides accurate fitting results for P and B frames, and relatively accurate for I frames. However, although Fig. 7 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 Fig. 8, where we have plotted the 0.01-, 0.02-, 0.03-,... quantiles of the actual *P* and *I* video frames' types of the NBC News trace versus the respective quantiles of the respective DAR (1) models, for a superposition of 30 traces. As shown in Fig. 8(a), which presents the comparison of actual P frames with the respective DAR (1) models for the NBC News ([CIF, G16, B3, F48]) trace, 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 the probability of frames with a very small and very large, respectively, number of cells. The very good fit shows that the superposition of the P 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 B frames), as it was suggested in Fig. 7. Fig. 8(b) presents the comparison of actual I frames with the respective DAR (1) model, for the NBC News ([CIF, G16, B7, F48]) trace. Again, the result suggested from Fig. 7(a), i.e., that our method for modeling I frames of multiplexed H.264 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 Fig. 8 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, H.264 videoconference streams show LRD. This problem is overcome by our choice of modeling I, P and B frames separately. This is shown in Fig. 9. 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 (similar results were found for all the traces and all types of frames in our study). Therefore, although in some cases the DAR (1) model exhibits a slower 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.

4. Queuing performance study

We have also conducted a queuing performance study similar to the one presented in [28], in order to complete our statistical analysis and further validate the quality of our results. More specifically, we fed a discrete-time queuing system (representing a downlink channel) with unlimited buffer size, for a 20 Mbps channel transmission rate. The transmission slot had a 33.3 ms duration (equal to the inverse of the video frame rate). We assumed, in our simulation, that up to 10 packets (this value is taken from [28]) of length equal to 48 bytes, may be served during each transmission slot (i.e., we consider a TD/CDMA channel frame with a duration of 12 ms [29] and 625 slots/frame). We studied the packet waiting time and the packet loss ratio to validate the model for various load factors, given the delay constraints of real time video streams (packets of a video frame need to be transmitted before the arrival of the next video frame, i.e., within 33.3 ms, otherwise they are dropped; we set an upper bound of just 0.01% for videoconference packet dropping [29]). A load of 0.3, e.g., corresponds to a load of 0.3×20 Mbps = 6 Mbps. As shown in Fig. 6, in our queuing performance study we do not compare separately the I, P and B frames sequences of the actual traces and those created by our models, but instead we use our separate models for I, P and B frames in order to generate a whole "pseudo-trace" based on the actual traces' GOP parameters. It is this

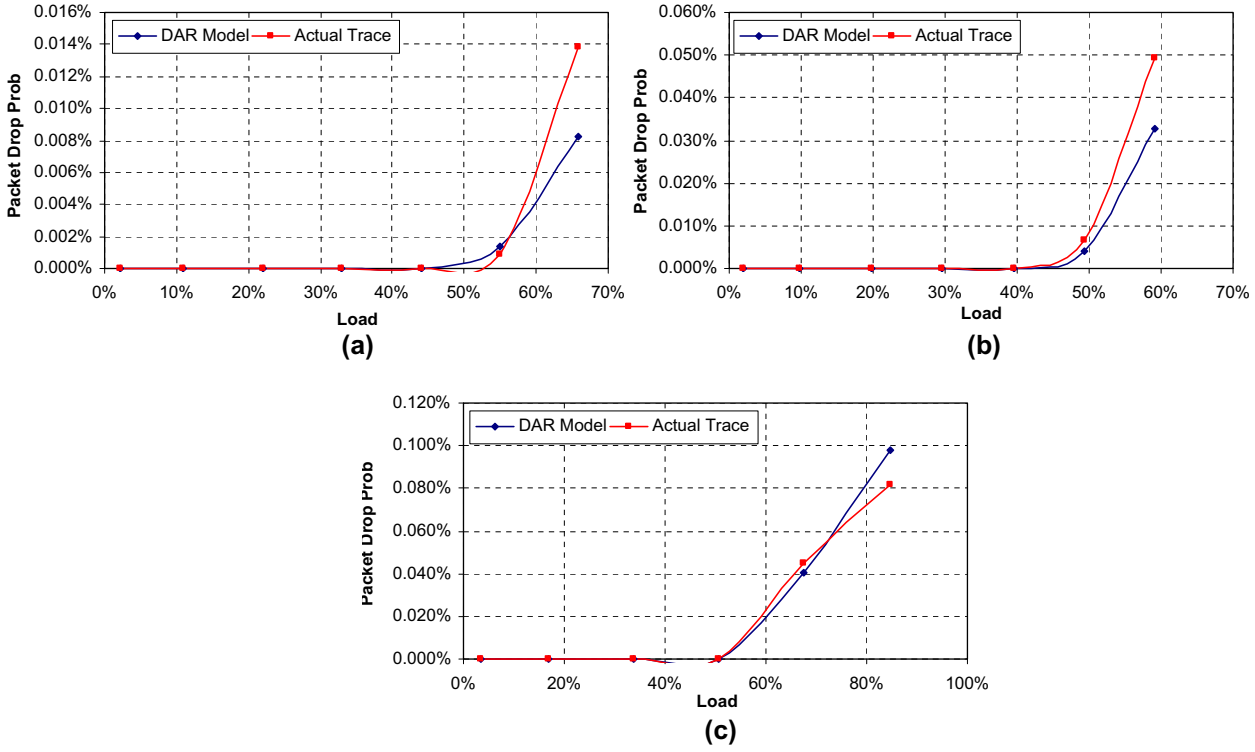

Fig. 11. Video packet dropping ratio versus offered load from various superposed sources of: (a) NBC News traces with parameters [CIF, G16, B3, F28], (b) Sony Demo traces with parameters [CIF, G16, B7, F28], and (c) Sony Demo traces with parameters [HD, G12, B2, F38] (unlimited buffer size).

"pseudo-trace" that we compare with the actual trace. We have derived results for all the traces under study; we present here selectively some of them, for brevity reasons. The results for all cases not shown below are of very similar nature to the ones presented. Fig. 10 presents a comparison of the waiting time Cumulative Density Functions (CDFs) for superposed NBC News and Sony Demo traces and the proposed models; it is clear from Fig. 10(a)–(c) that our modeling approach allows a very accurate characterization of the waiting time experienced by the packets of the multiplexed H.264 videoconference sources, for various realistic load factors (5–85%). Fig. 11 presents a comparison of the packet dropping probability for various numbers of actual NBC News and Sony Demo superposed sources (i.e., several channel loads) and the respective DAR models. This result, combined with the results presented in Fig. 10, verifies the validity of our modeling approach, hence showing that a first-order model is a competent candidate for modeling H.264 videoconference traffic from multiplexed video sources. However, as shown in Fig. 11, when the load increases over 50% or 60%, depending on the trace, it becomes too large for the system to handle (both for the real traces and our pseudo-traces) without the 0.01% upper bound on packet dropping probability being violated. Fig. 11 also shows that, due to the fact that the modeling of I frames sizes is not perfectly accurate, there is a small difference in video packet dropping between our models and the actual traces, but this difference exists only for higher loads, where the upper bound on packet dropping probability has already been violated.

5. Conclusions

We have shown that the Pearson type V distribution provides the relatively best fit for modeling videoconference traffic from single H.264 sources. Then, we proceeded to use this result in order to build an accurate and simple Discrete Autoregressive model (separately for I, B, and P frames) to capture the behavior of multiplexed H.264 videoconference sources. Based on the very accurate results of our study in modeling P- and B-frames' sizes, and the low complexity of our first-order model, we believe that our approach is very promising for modeling this type of traffic, possibly combined with the use of wavelet modeling for modeling I frames' sizes. In recent work [12], which focuses on single traces (not multiplexed traffic as in our work) and on video traces (i.e., traces with low autocorrelation and high motion, contrary to the videoconference-type traces used in our study) it was shown that for modeling I frames especially, wavelet modeling is a very competent solution. The work in [12] develops a model at the frame-size level, as we do in our work, not at the slice-level or block-level as in other relevant studies in the literature (e.g., [15]). Corroborating our conclusion that a different approach should be used for I frames than that of P and B frames, Dai et al. [12] also use a simpler model for P and B frames than the model used for I frames. In addition, based on the present work and other recent work of our group, that for videoconference traffic encoded by all major modern encoding standards the Pearson V distribution is the best fit.

Finally, the implementation of our proposed scheme in a wireless testbed (e.g., IEEE 802.11 b/g) in order to allocate bandwidth to video-streaming clients and the consecutive measurement of QoE would be of great interest and is considered as future work.

References

[1] S.M. Cherry, Fiber to the home, IEEE Spectrum 41 (1) (2004) 42–43.
[2] M. Etoh, T. Yoshimura, Advances in wireless video delivery, Proceedings of the IEEE 93 (1) (2005) 111–122.
[3] D.M. Lucantoni, M.F. Neuts, A.R. Reibman, Methods for performance evaluation of VBR video traffic models, IEEE/ACM Transactions on Networking 2 (1994) 176–180.
[4] M. Nomura, T. Fuji, N. Ohta, Basic characteristics of variable rate video coding in ATM environment, IEEE Journal on Selected Areas in Communications 7 (5) (1989) 752–760.
[5] 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 2 (1) (1992) 49–59.
[6] A.M. Dawood, M. Ghanbari, Content-based MPEG video traffic modeling, IEEE Transactions on Multimedia 1 (1) (1999).
[7] B. Melamed, D.E. Pendarakis, Modeling full-length VBR video using Markov-renewal modulated TES models, IEEE Journal on Selected Areas in Communications 16 (5) (1998) 600–611.
[8] K. Chandra, A.R. Reibman, Modeling one- and two-layer variable bit rate video, IEEE/ACM Transactions on Networking 7 (3) (1999) 398–413.
[9] D.P. Heyman, The GBAR source model for VBR videoconferences, IEEE/ACM Transactions on Networking 5 (4) (1997).
[10] M. Frey, S. Ngyuyen-Quang, A gamma-based framework for modeling variable-rate video sources: the GOP GBAR model, IEEE/ACM Transactions on Networking 8 (6) (2000) 710–719.
[11] S. Kempken, W. Luther, Modeling of H.264 high definition video traffic using discrete-time semi-markov processes, in: Proceedings of the ITC 2007, Lecture Notes in Computer Science (LNCS), vol. 4516, 2007, pp. 42–53.
[12] M. Dai, Y. Zhang, D. Loguinov, A unified traffic model for MPEG-4 and H.264 video traces, IEEE Transactions on Multimedia 11 (5) (2009) 1010–1023.
[13] I. Reljin, A. Samcovic, B. Reljin, H.264/AVC video compressed traces: multifractal and fractal analysis, EURASIP Journal on Applied Signal Processing 2006 (1) (2006) 1–12.
[14] M. Krunz, S.K. Tripathi, On the characterization of VBR MPEG streams, in: Proceedings of the ACM SIGMETRICS, vol. 25, June 1997.
[15] U.K. Sarkar, S. Ramakrishnan, D. Sarkar, Modeling full-length video using Markov-modulated gamma-based framework, IEEE/ACM Transactions on Networking 11 (4) (2003) 638–649.
[16] D. Marpe, T. Wiegand, G. Sullivan, The H.264/MPEG4 advanced video coding standard and its applications, IEEE Communications Magazine 44 (8) (2006) 134–143.
[17] G. Van der Auwera, P. David, M. Reisslein, Traffic and quality characterization of single-layer video streams encoded with the H.264/MPEG-4 advanced video coding standard and scalable video coding extension, IEEE Transactions on Broadcasting 54 (3) (2008) 698–718.
[18] P. Koutsakis, A new model for multiplexed VBR H.263 videoconference traffic, in: Proceedings of the IEEE GLOBECOM, 2006.
[19] Available from: <http://trace.eas.asu.edu/hd/index.html>.
[20] A.M. Law, W.D. Kelton, Simulation Modeling & Analysis, second ed., McGraw-Hill, New York, 1991.
[21] K.P. Burnham, D.R. Anderson, Model Selection and Multi-Model Inference, Springer, New York, 2002.
[22] K. Park, W. Willinger (Eds.), Self-Similar Network Traffic and Performance Evaluation, Wiley, New York, 2000.
[23] M. Krunz, H. Hughes, A traffic model for MPEG-coded VBR streams, in: Proceedings of the ACM SIGMETRICSm 1995, pp. 47–55.
[24] D.P. Heyman, T.V. Lakshman, What are the implications of long-range dependence for VBR-video traffic engineering, IEEE/ACM Transactions on Networking 4 (3) (1996) 301–317.
[25] 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, pp. 3–14.
[26] P.A. Jacobs, P.A.W. Lewis, Time series generated by mixtures, Journal of Time Series Analysis 4 (1) (1983) 19–36.
[27] T.V. Lakshman, A. Ortega, A.R. Reibman, VBR video: trade-offs and potentials, Proceedings of the IEEE 86 (5) (1998).
[28] S. Baey, Modeling MPEG4 video traffic based on a customization of the DBMAP, in: Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'04), 2004, pp. 705–714.
[29] D.A. Dyson, Z.J. Haas, A dynamic packet reservation multiple access scheme for wireless ATM, Mobile Networks and Applications (MONET) Journal 4 (2) (1999) 87–99.
[30] A. Lazaris, P. Koutsakis, Modeling video traffic from multiplexed H.264 videoconference streams, in: Proceedings of the IEEE GLOBECOM, 2008.
[31] A. Lazaris, P. Koutsakis, M. Paterakis, A new model for video traffic originating from multiplexed MPEG-4 videoconference streams, Performance Evaluation Journal 65 (1) (2008) 51–70.
[32] Z. Avramova, D. De Vleeschauwer, K. Laevens, S. Wittevrongel, H. Bruneel, Modelling H.264/AVC VBR video traffic: comparison of a Markov and a self-similar source model, Telecommunication Systems 39 (2) (2008) 91–102.
[33] R. Guerin, H. Ahmadi, M. Naghsineh, Equivalent capacity and its application to bandwidth allocation in high-speed networks, IEEE Journal on Selected Areas in Communications 9 (7) (1991) 968–981.