A Markov Modified Model of H.264 VBR Video Traffic

Harilaos Koumaras, Charalampos Skianis, Georgios Xilouris, Georgios Gardikis, Anastasios Kourtis

***Abstract*—Video is expected to posses a large portion of the total traffic in future wireless communication networks, given that multimedia services are becoming increasingly popular. Consequently, traffic characterization of such video services is essential for efficient traffic control and resource management via the adaptation of the video properties in the design of communication and transmission networks. The new H.264/AVC standard, proposed by the ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Pictures Expert Group (MPEG), was designed satisfying two challenging tasks: Considerably higher coding efficiency and improved network adaptation compared to the already widely used standards. Hence, H.264, featuring slice partitioning, Intra refreshing, separation of picture and sequence data from slice data, provides robustness during transmission, setting it the most appropriate video coding standard for error prone environments. This paper, based on a previously published frame and layer statistical analysis of H.264 encoded sources, proposes a Markov modified model of H.264 unconstraint VBR traffic.**

***Index Terms*— H.264, VBR, Traffic Modeling**

I. Introduction

Real time multimedia transmission is considered as the most demanding application in terms of computational complexity. Among the various types of multimedia, video services (transmission of moving images and sound) are proven dominant for present and future communication networks. Regarding video transmission over wireless environments, it contains several more challenges, because transmission errors by fading signal periods cause a wide range of error conditions (i.e. from single bit errors to severe burst errors).

In 1998 the ITU-T VCEG issued a call for proposals (H.26L project), with main scope to double the coding efficiency in comparison to the already existing coding standards. In 2001, VCEG and ISO/IEC MPEG formed a Joint Video Team (JVT) in order to finalize the standard and to submit for formal approval as H.264/AVC [1].

The primary goals of the new video coding standard known as H.264/MPEG-4 Advanced Video Coding (AVC) are improved coding efficiency and improved network adaptation over a variety of network channels and conditions [2]. Some essential indicative enhancements are Variable block size support for motion compensation, Quarter-sample motion vector accuracy, extended reference frame selection for P frames, de-blocking filter within the motion-compensated prediction loop and new context-based adapted entropy coding methods.

The main target of the aforementioned enhancements is the perceived quality improvement and the high-compression efficiency. In this respect, due to the expected business models in emerging wired and wireless networks, where the end-user costs are relative to the transmitted data volume, the optimized bandwidth occupation and utilization is core goal for all the future multimedia services. For these reasons, H.264/AVC is expected to dominate in future wireless communication networks.

Today, the telecom market is facing a peculiar problem regarding the provision of audiovisual content: Although the technology has evolved and the infrastructure for both the core and the access networks are in place, the Network Operators (NOs) and Service Providers (SPs) tend to be disappointed because revenues of their investments are very low. The provision of new multimedia applications/content to the market still follow slow rates and the main reason is due to the absence of end-end QoS delivery issues.

In this respect ENTHRONE (End-to-End QoS through Integrated Management of Content, Networks and Terminals) adopts a business model in which Content Providers (CPs) and Content Consumers (CC) are considered as one entity, while NOs and SPs are other parties. In this approach, the CP will be responsible only for the generation of the content and will focus only on the business aspect of the service. A second entity (i.e., NOs) is responsible for engineering the network and for the provisioning of networking resources, required for service/content delivery.

An overlaying Integrated Management System (IMS) allows NOs to monitor, control and manage the network resources based on specific policies derived from the

Manuscript received February 10, 2006. This work is partly funded by the Information Society Technologies (IST) project ENTHRONE / FP6-507637 (End to end QoS through Integrated management of content networks and terminals).

H. Koumaras is with the N.C.S.R. Demokritos, Patriarchou Gregoriou Str., 15310 Athens Greece (e-mail: koumaras@iit.demokritos.gr).

C. Skianis is with the N.C.S.R. Demokritos, Patriarchou Gregoriou Str., Agia 15310 Athens Greece (e-mail: skianis@iit.demokritos.gr).

G. Xilouris is with the N.C.S.R. Demokritos, Patriarchou Gregoriou Str., 15310 Athens Greece (e-mail: xilouris@iit.demokritos.gr).

G. Gardikis is with the N.C.S.R. Demokritos, Patriarchou Gregoriou Str., Agia Paraskevi, Attiki, 15310 Athens Greece (e-mail: gardikis@iit.demokritos.gr).

A. Kourtis is with the N.C.S.R. Demokritos, Patriarchou Gregoriou Str, 15310 Athens Greece (e-mail: kourtis@iit.demokritos.gr).

requested QoS and the capabilities of each network. In this context, it can be created a networking environment, which achieves provisioning and maintaining of an end-to-end agreed QoS through heterogeneous networks, owned by different NOs.

In this respect, this paper presents, based on the results of a previously published work [3] on the frame and layer level statistical analysis of the generated traffic from the JM H.264 reference encoder, a H.264 Markov Modified Model, which will be used for the configuration and engineering of a prototype heterogeneous network featuring end-to-end QoS mechanisms. Moreover, a performance study on the perceived quality of H.264 in comparison to the selected quantization scales is also presented.

The rest of the paper is organized as follows: Section II presents the results of the statistical analysis of the H.264 encoded data, which the Markov modified model of section III is based. Finally, section IV concludes the paper.

II. Statistical Analysis of the H.264 encoded data

In [3] it is performed a detailed statistical analysis of H.264 encoded traffic, based on a segment from the film "Spider-man II" as the reference signal. This segment consists of 18357 frames of YUV 4:2:0 format in 528x384 resolution, where encoding was performed using JM H.264 reference encoder, at VBR mode with constant GOP structure of the form IPBPBPBPB…. In order to study the nature of the video stream, quantization parameters are altered during the experiments. During each encoding process, video traces were captured, containing data on the type and the size of each encoded frame. Table 1 depicts the corresponding statistic of I/P/B frame sizes. The notation (x,y,z)-l is used for the quantization scales of I,B,P frames and the selected intra-frame period (i.e. GOP length, which is constant to 12 for this paper).

From Table 1, it can be derived that higher quantization parameters, which cause coarser encoding quality, result in lower mean frame sizes and variations in comparison with lower quantization parameters, which produce better encoding quality. This impact of the QP selection on the deduced perceived quality level of the encoded signal is further examined: For this purpose two different quality metrics were used: the SSIM quality metric was exploited and the well-known PSNR. The results of this process, for the various encoding schemes, appear on the Table II.

Representing the results of Table I on the Mean vs. Standard Deviation plane (Figure 1), it can be derived that the encoding schemes follow a specific pattern. Moreover, it seems that the mean frame size follows an increasing linear tendency as the selected encoding scale becomes lower.

The experimentally derived curves of Fig.1 can be successfully approximated by the power equation StdDev = $3.0056(\text{Mean})^{0.5723}$, which corresponds to a R-squared parameter equal to 0.957. R squared is the relative predictive power of a model and is a descriptive measure between 0 and 1. The closer it is to one, the better the proposed model is. Hence, the proposed power equation provides satisfactory modeling of the experimental data.

TABLE I
Frame statistics overview of the encoded signals

Quantization Settings / Frame Types	I Frames (in Kbits)		B Frames (in Kbits)		P Frames (in Kbits)	
	Mean	σ	Mean	σ	Mean	σ
(10,10,10)-12	354.47	87.29	227.65	57.67	271.34	65.67
(20,20,20)-12	148.16	58.25	43.02	33.17	67.01	41.98
(30,30,30)-12	53.91	25.68	7.86	8.71	16.33	13.93

TABLE II
The perceived quality measurements for the encoded streams

Quantization Settings	SSIM	PSNR
10-10-10-12	0.948900	30.3883
20-20-20-12	0.945567	30.3568
30-30-30-12	0.936867	30.2310

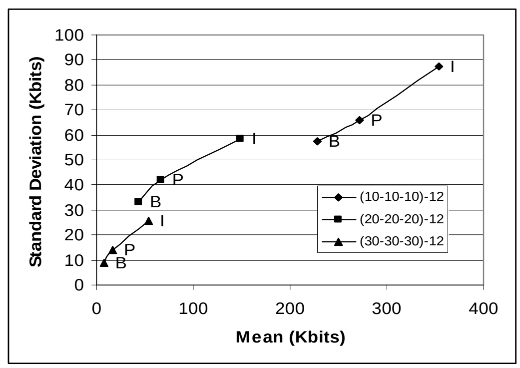

Fig. 1. I/P/B size statistics for the various coding schemes

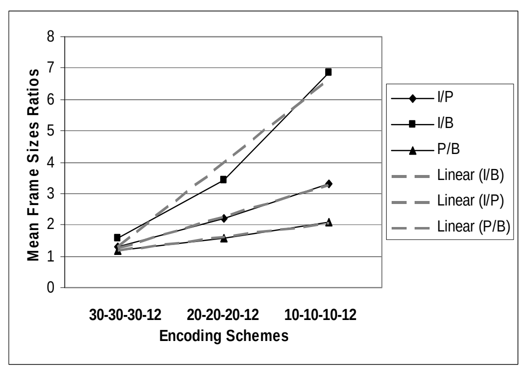

Figure 2. Mean Frame Size Ratios vs. the encoding schemes

Moreover, calculating from Figure 1 and Table I, the mean frame size ratios for the various encoding schemes and depicting them in Figure 2, it can be observed that the ratios follow a linear distribution depending on the selected encoding and quantization parameters. Thus, by exploiting this linear distribution of the frame size ratios in conjunction with the aforementioned power model of the standard deviation vs. mean frame size, it is possible to estimate the corresponding statistical properties of an encoding stream based only on an experimentally defined statistical property.

Finally, Table III contains the variation coefficients for the various quantization schemes, which represent a metric of the variation and the shape of the deduced frame size distribution. Their values are consistent with the aforementioned observations.

TABLE III
VARIATION COEFFICIENTS OF THE ENCODED SIGNALS

Quantization Settings / Frame Types	I Frames Variation Coefficient	B Frames Variation Coefficient	P Frames Variation Coefficient
(10,10,10)-12	0.2463	0.2533	0.2420
(20,20,20)-12	0.3932	0.7710	0.6265
(30,30,30)-12	0.4763	1.1081	0.8530

Moreover, in [3] it is presented that the Probability Density Functions (PDFs) for each frame type of the encoded signal at various quantization scales can be successfully approximated by the Gamma density function. Using the usual moments approach, which makes use of the fact that the Gamma distribution has mean pμ and variance $p\mu^2$, by equating to the mean and sample variance, denoted as m and v respectively, it can be deduced that μ=v/m and $p=m^2/v$. Therefore, the corresponding Gamma-fits to the sample distribution are derived in [3] and their statistical properties (i.e. p and μ values) for the various encoded schemes are depicted in Table IV.

TABLE IV
GAMMA MODEL STATISTICS OVERVIEW OF THE ENCODED SIGNALS

Quantization Settings / Frame Types	I Frames		B Frames		P Frames	
	p	μ	p	μ	p	μ
(10,10,10)-12	16.487	21499	15.584	14608	17.071	15895
(20,20,20)-12	6.468	22905	1.682	25572	2.549	26287
(30,30,30)-12	4.406	12235	0.813	7857	1.376	11874

III. THE PROPOSED H.264 MARKOV MODIFIED MODEL

We discussed in the previous section a brief overview of the statistical properties of I, B and P frame types. This detailed traffic study and mathematical quantitative approach of H.264 VBR video [3] is necessary for understanding the properties of its characteristics, which will be used for generating synthetic H.264 traffic by an appropriate video model of H.264 Unconstraint VBR traffic.

Based on the statistical analysis and quantitative mathematical approach performed in [3], it is possible to generate sequences of stationary discrete random variables with Markov properties using the discrete autoregressive process of order 1, DAR(1). Many relevant works [4-12] have been presented in the literature for older video encoding standards, but not for the latest H.264.

According to the DAR(1) model, the first order autoregressive form is given by $F_n = V_n F_{n-1} + (1 - V_n) Y_n$ for n=1,2,… where {Vn} are independent and identically distributed (iid) binary random variables with $P(V_n = 1) = 1 - P(V_n = 0) = \alpha \ with\ 0 \le \alpha < 1$ and {Yn} are iid random variables with a marginal distribution π.

In other words the model defines the current observation to be a mixture of two independent random variables: It is either the last observation with probability α, or another independent sample from the same distribution. It is a very simple and general model since π is the distribution of any random variable and the correlation structure is independent of π. The autocorrelation function (ACF) of {Fn} as defined by the aforementioned equations, is given by $\rho_F(k) = \alpha^k, k = 0,1,...$ and {Fn} is a Markov chain with transition probability matrix given by $aI + (1-a)Q$, where I is the identity matrix and Q is a matrix of whose rows are the distribution π.

This approach is used independently for the study of the I, B and P frames trying to capture the fact that a video sequence consists of various scenes with different spatial and temporal activity levels and thus a candidate video traffic model should capture this inter and intra scene state. More specifically, during a scene, the sizes of the same frame type remain typically constant, while on the contrary follow different sizes over scenes changes.

In this respect, two discrete processes for frame generation are considered:

-The first process simulates the intra-scene state, which means that the frame size, referring to frames of the same type, retains the characteristics of the previous frame.

-The second process, which models the inter-scene state, the frame size is generated using a AR(1) process of the form $x(n)=\alpha_1 x(n-1)+e(n)$, based on the size of the previous frame size, where α_1 is the autocorrelation parameter at lag-1 and e(n) a residual following the normal distribution.

Especially for the case of I frames, which are strongly responsible for the inter-GOP correlation of the video stream, which is well characterized by the ACF of the I-frames, the first process is considered as $I(n)=\alpha_1 I(n-1)$, where α_1 is the autocorrelation coefficient at lag-1, in order to better capture this phenomenon.

Although inter-GOP correlation, described by the ACF of the I frames, is an important measure, another aspect of video traffic is the correlation between I/P/B frames within the same GOP (the intra-GOP correlation). For this reason, in our case the correlation coefficients are calculated between the first neighboring I-P, I-B and P-B frames of each GOP structure. Table V shows the corresponding results for various encoding schemes, showing that there is high correlation between the neighboring first I-P frames of each GOP and lower between the P-B, and even lower between I-B.

TABLE V
CORRELATION COEFFICIENT OF INTRA/INTER FRAMES

Quantization Scale	I-P Correlation Coefficient	I-B Correlation Coefficient	P-B Correlation Coefficient
(10-10-10)-12	0.6412	0.1750	0.1721
(20-20-20)-12	0.6250	0.0653	0.0948
(30-30-30)-12	0.5180	-0.0381	0.0916

Figure 3 depicts the proposed algorithm for the generation of synthetic H.264 traffic. The proposed model, except from the use of the aforementioned DAR(1) models, it also exploits the I-P correlation coefficient and the P-B correlation coefficient for the generation of P and B streams respectively. More specifically, based on the GOP length and structure that the generated traffic has, the proposed algorithm, during a GOP initialization, propagates the first I and P frame size, multiplied by the corresponding correlation coefficient of Table V, as feed to the respective P and B DAR processes. By this way, it manages to captures the intra-GOP dependences of the actual traffic, which is an important issue for the modeling of real traffic video characteristics.

In the proposed model, the MUX component of the described block diagram is responsible for the appropriate multiplexing of I, B and P streams according to a specific GOP pattern.

For the case of (20-20-20)-12, the corresponding Q-Q plot is depicted in Figure 4, reporting the good behavior of the proposed model.

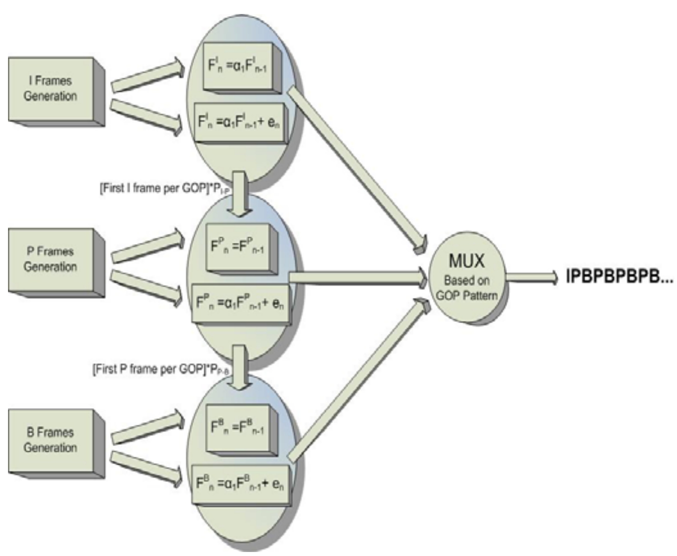

Fig. 3. Block Diagram of the proposed model

Therefore, it has been shown that the latest H.264 video encoding standard can be satisfactorily modeled by the proposed Markov Modified model.

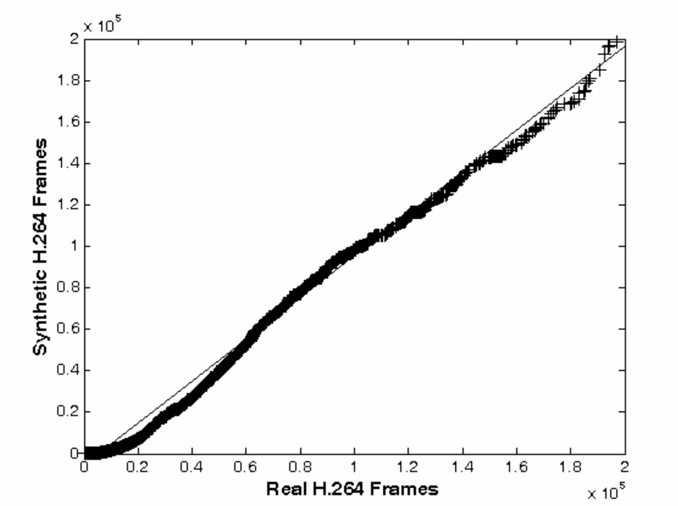

Fig. 4. Q-Q Plot of real and generated H.264 traffic

All the new features of the H.264 standard aim at increasing compression efficiency in conjunction with improvement of the deduced perceived quality level for a specific encoding bit rate. Thus, H.264 by featuring variable block size support down to 4x4, quarter-sample motion vector accuracy, extended reference frame selection for P frames, de-blocking filter within the motion-compensated prediction loop and new context-based adapted entropy coding methods: CAVLC and CABAC, it succeeds better compression efficiency in comparison to the previous standards, without altering the distribution characteristics of the generated stream and the encoding scheme, which remains similar to the previous MPEG standards.

IV. CONCLUSION

This paper reports on an experimental study of H.264 encoded video streams where additional statistical analysis established general results about the video traffic. The experiments covered cases with different quantization scales, showing that the derived data can be expressed as superimposition of three discrete frame contributions. Finally, a novel model for generating and simulating H.264 VBR traffic is presented and evaluated by comparing real and generated video traffic.

ACKNOWLEDGMENT

This work is partly funded by the Information Society Technologies (IST) project ENTHRONE / FP6-507637 (End to end QoS through Integrated management of content networks and terminals).

REFERENCES

[1] Wiegand T., Sullivan G., Bjontegaard G. and Luthra A. (2003), ‘‘Overview of the H.264/AVC Video Coding Standard’’, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue in H.264.

[2] Sullivan G. (2005), “The H.264/MPEG-4 AVC vdeo coding standard and its deployment status”, Proceedings of VCIP 2005, Beijing, China.

[3] Koumaras H., Skianis C., Gardikis G., Kourtis A. (2005), "Analysis of H.264 video encoded traffic", INC 2005 Fifth International Network Conference, Samos Island, Greece, July 2005

[4] Alheraish A. (2004), "Autoregressive video conference models", International Journal of Network Management, Vol. 14, pp329-337

[5] Chin H.S., Goodge J.W., Griffiths R. and Parish D.J. (1989), "Statistics of video signals for viewphone-type pictures'', IEEE Journal on Selected Areas in Communications, Vol.7, No.5, pp826–832.

[6] Cohen D.M. and Heyman D.P. (1993), ''Performance modeling of video teleconferencing in ATM networks'', IEEE Transactions on Circuits Systems Video Technology, Vol.3, No.6, pp408–422.

[7] Cox D.R. and Miller H.D., The Theory of Stochastic Processes, Chapman & Hall, London, 1965

[8] Dalgic I. and Tobagi F.A., Performance evaluation of ATM networks carrying constant and variable bit-rate video traffic, IEEE Journal on Selected Areas in Communications 15(6) (1997) 1115–1131.

[9] Doulamis N.D., Doulamis A.D., Konstantoulakis G.E. and Stassinopoulos G.I. (2000), ''Efficient Modeling of VBR MPEG-1 Coded Video Sources'', IEEE Transactions on Circuits Systems Video Technology Vol.10, No.1, pp 93–112.

[10] Grunenfelder R, Cosmas JP, Manthorpe S, Odinma-Okafor A (1991) Characterization of Video Codecs as Autoregressive Moving Average processes and Related Queuing System Performance. IEEE J Sel Areas Commun 9(3): 284–293

[11] Haskell B.G. (1972), ''Buffer and channel sharing by several interframe picturephone coders'', Bell Systems Technical Journal, Vol.51, No.1, pp261–289.

[12] Heyman D.P., Tabatabai A. and Lakshman T.V. (1992), ''Statistical analysis and simulation study of video teleconference traffic in ATM networks'', IEEE Transactions on Circuits Systems Video Technology, Vol.2, No.1, pp49–59.