Frame-based Modeling of H.264 Constrained Videoconference Traffic over an IP Commercial Platform

S. Domoxoudis[1, 2], S. Kouremenos[1], A. Drigas[1] and. V. Loumos[2]

1. National Center for Scientific Research ''DEMOKRITOS''
DTE/YE Applied Technologies Department
P.O. Box 15310 Gr. Ag. Paraskevi, Attiki, Greece

2. National Technical University of Athens ''NTUA''
School of Electrical and Computer Engineering, Multimedia Technology Laboratory
P.O. Box 15780 Gr. Zographou, Attiki, Greece

E-mail
{sdomo, skourem, dr}@imm.demokritos.gr
loumos@cs.ntua.gr

Abstract

In this manuscript, we contribute measurement and modeling results for H.264 encoded videoconference traffic. The analysis of the frame-size autocorrelation function of extensive video data that were gathered during realistic experiments with two popular videoconference terminals (currently the only supporting the H.264 encoder) suggested that H.264 videoconference traffic is fully uncorrelated as a consequence of the constrained (in the loop) operation mode of the encoders. By using this fact and by statistically analyzing the sampled frame-size histogram, we conclude that H.264 traffic can be reasonably represented by a $D/G/\infty$ queue with deterministic arrivals according to the video frame rate and service times that can be fitted by a histogram-based model proposed on a previous study.

1. Introduction

The H.264 and the MPEG-4 Part 10 [1], also named Advanced Video Coding (AVC), is jointly developed by ITU and ISO. H.264/MPEG-4 supports video compression (coding) for video-conferencing and video-telephony applications. The H.264/MPEG-4 design covers a Video Coding Layer (VCL), which efficiently represents the video content, and a Network Abstraction Layer (NAL), which formats the VCL representation of the video and provides header information in a manner appropriate for conveyance by particular transport layers (such as Real Time Transport Protocol) or storage media. H.264 videoconference traffic, thanks to its improved compression algorithms which result to lower bandwidth requirements, is expected to account for large portions of the multimedia traffic in future heterogeneous networks (wireline, wireless and satellite), with the ADSL network being the most notable one. Under the above expectation, it is evident that a statistical model of H.264 videoconference traffic would be very useful to predict network usage and estimate resources. The application of a successful traffic model can lead to a more economical network usage (improved traffic policing schemes), to lower communication costs and a more affordable and of higher quality service to the end-users.

Due to the above reason, the modeling and performance evaluation of videoconference traffic, currently H.261 and H.263, have been extensively studied in literature and a wide range of modeling methods exist[1]. However, the variation of the videoconference communication parameters which in most cases are the maximum or target video bit rate,

---

[1] It must be noted here that the current study focuses on the study of the traffic generated by the H.264 encoder for videoconference applications and not for streaming applications.

the frame rate, the used encoder, the differences in the video coding algorithm implementations and finally the variety of the visual contents (low and high motion head&shoulders, sign language, telemedicine) turn accurate videoconference traffic modeling into a complex procedure. The results of relevant early studies [2]-[10], concerning the statistical analysis of variable bit rate videoconference streams being multiplexed in ATM networks, indicate that the histogram of the videoconference frame-size sequence exhibits an asymmetric bell shape and that the autocorrelation function decays approximately exponentially to zero.

In [11], a study of videoconference traffic (H.261 and H.263) measurement and simulation indicated the influence of the session parameters (codec, quality, frame rate) on the generated traffic pattern. Again, the sample histogram had a bell-shape and the autocorrelation function decayed quickly to zero. A normal mixture distribution, as a method for fitting the sample histogram, was proposed in [12], and it was proved that it performed better than the simple Gamma and Log-normal distributions.

In [13], the authors observed that the DAR and GBAR models cannot model a variety of sources and perform unsteadily. To overcome this unsteadiness, they adjusted the DAR model by introducing a marginal matching technique and a random variable chosen from a uniform distribution. Of particular relevance to our work is the approach in [14], where an extensive study on multipoint videoconference traffic (H.261-encoded) modeling techniques was presented. In this study, the authors discussed methods for correctly matching the parameters of the modeling components to the measured H.261-encoded data derived from realistic multipoint conferences (in ''continuous presence'' mode). Two new histogram-based methods for fitting the empirical frame-size histogram, namely, LVMAX and C-LVMAX were proposed. C-LVMAX was indicated as the one that can capture accurately enough - with respect to queuing - the sample histogram due to its dominance near the tail.

Our modeling approach, in this study, is based on the recommendations towards a good traffic model that were proposed in [15]. According to them, a model must be realistic, reusable and computationally efficient. In order not to decline from the above requirements, our study was based on realistic experimental data, on-line encoded, derived from a variety of experimental conditions and on the modification of well-established and simple models proposed in literature.

Due to the advances of H.264, it is expected that it will prevail in future networks and mobile application systems, making traffic modeling and characterization of H.264 video streams a useful tool for network managers and designers. In this respect, this paper studies the output traffic from an H.264 codec at the frame level. In particular, the work focuses on video traffic that is generated by VBR H.264 coding since it offers relatively constant quality and has less bandwidth and storage capacity requirements.

The rest of the manuscript is structured as follows: section 2 describes the experiment characteristics and presents the first-order statistical quantities of the measured data. Section 3 discusses appropriate methods for parameter assessment of the encoded traffic. Finally, section 4 culminates with conclusions and pointers to further research.

2. The experimental and measurement work

The study reported in this manuscript employed measurements of the IP traffic generated by the H264 videoconference encoder. More explicitly, we measured the traffic generated by the H.264 encoder included in the following videoconference software tools: VCON Vpoint HD [18] and Polycom PVX [19], which are the only videoconference applications that support H.264. We must note that VCON Vpoint HD could not establish an H.264 connection with Polycom PVX and vide versa. This is due to the fact that the RTP payload format for the H.264 has still some open issues (media unaware fragmentation). More explicitly, the clients use different RTP payload types to communicate.

At this point, we may discuss about the basic functionality of the examined video systems which is a fundamental factor in the derived statistical features of the encoded traffic and a basic reason of the experiments' philosophy we followed. The rate control parameter (bandwidth and frame rate) sets a traffic policy, i.e. an upper bound on the encoded traffic according to the user's preference (obviously depending on his/her physical link). An encoder's conformation to the rate control of the system is commonly performed by reducing the video quality (and consequently the frame size quantity) through the dynamic modulation of the quantization level. In the case of Vpoint, and Polycom, the frame rate remains invariant and a video quality reduction is performed when the exhibited video bit rate tends to overcome the bandwidth bound. This threshold can be set through the network setting of each client. Moreover, Vpoint utilizes adaptive bandwidth adjustment (ABA).

ABA works primarily by monitoring packet loss. If the endpoint detects that packet loss exceeds a pre-defined threshold, it will automatically drop to a lower conference data rate while instructing the other conference participant's endpoint to do the same.

Exp	VCON 1		VCON 2	
Content	Low (A)	Low (B)	High (A)	High (B)
Scenario	320 Low		320 High	
Duration (sec)	3600		3600	
Video Frames	90001	89997	89911	89918
Video Bit Rate (Kbits/Sec)	**300.88**	**300.87**	**300.92**	**300.93**
Frame rate	25.000	24.999	24.975	24.977
Average Frame Size (Bytes)	1504	1504	1506	1506
Min Frame Size (Bytes)	772	260	340	260
Max Frame Size (Bytes)	2344	2328	3308	2828
Variance	234	240	820	843

Table 1a: Statistical quantities of the sample frame-size sequences of VCON Vpoint HD

During the experiments, two ''talking-heads'' raw-format video contents were imported in the video systems through a Virtual Camera tool [20] and then peer-to-peer sessions of one hour were employed in order to ensure a satisfactory trace length for statistical analysis. These contents were offline produced by a typical webcam in uncompressed RGB-24 format: one with mild movement and no abrupt scene changes, ''listener'', (Low) and one with higher motion activities and occasional zoom/span corresponding to the action of bringing stuff closer to the camera (e.g. to let the other party see what is on a piece of paper), ''talker'' (High). The video size was QVGA (320x240) in both cases. The target video bit rates that were imposed are shown in Table 1. In each experiment, the UDP packets were captured by a network sniffer and the collected data were further post-processed at the frame level[2] by tracing a common packet timestamp. The produced frame-size sequences were used for further statistical analysis.

Specific parameters shown in Table 1a-1b, for the High and Low traces, depend on the particular coding

---

[2] It is important to note, here, that analysis at the MacroBlock (MB), as in [16], level has been examined and found to provide only a typical smoothing in the sample data. We believe that the analysis at the frame level is simpler and offers a realistic view of the traffic.

implementation, the nature of the moving scene, and the confidence of the measured statistics.

Some primary conclusions, as supported by the experiments' results (see Table 1a-1b), arise concerning the statistical trends of the encoders' traffic patterns. For all the encoders, the use of the High content led to higher rate results (as reasonable). Similar results were observed for the mean frame size and variance quantities. In all cases, the variance quantities of the High content were higher than that of Low.

Exp	Polycom 1		Polycom 2	
Content	Low (A)	Low (B)	High (A)	High (B)
Scenario	384 Low		384 High	
Duration (sec)	3600		3600	
Video Frames	107546	94282	107515	83020
Video Bit Rate (Kbits/Sec)	**345.56**	**335.04**	**345.62**	**329.57**
Frame rate	29.874	26.189	29.865	23.061
Average Frame Size (Bytes)	1446	1599	1447	1786
Min Frame Size (Bytes)	452	692	596	612
Max Frame Size (Bytes)	3700	4228	2992	5916
Variance	11896	31959	12833	129700

Table 1b: Statistical quantities of the sample frame-size sequences of Polycom PVX

At this point, it must be noted that both implementations tend to adjust their quality in a "greedy" manner so as to use up as much of the allowed bandwidth as possible.

3. Traffic analysis and modeling assessment

The measured traffic analysis for all experimental sets confirms the general body of knowledge that literature has formed concerning videoconference traffic. Traffic analysis was employed for all experimental cases (for both high and low motion contents). More explicitly, in all cases (only High case is depicted), the frame-size sequence can be represented as a stationary stochastic process, with a frequency histogram narrow-shaped (more narrow in the case of Vpoint), see figures 1((a),(b)).
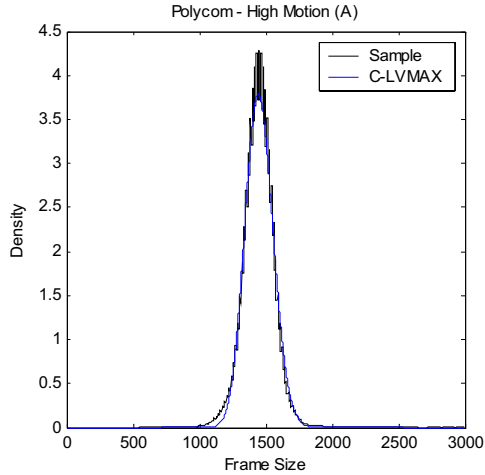
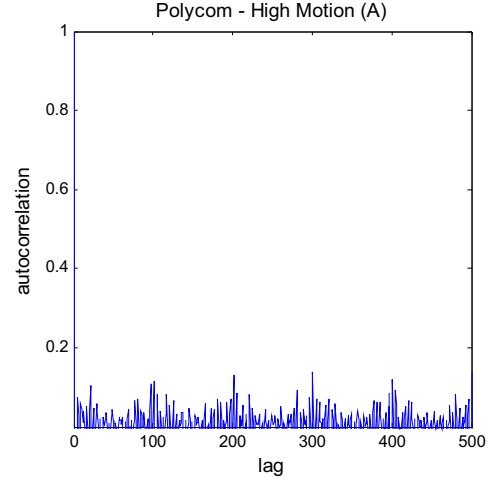**Figure 1a**: Frame-size histogram vs C-LVMAX fit
(Polycom PVX – High)

Figure 2a: Autocorrelation Graph for Polycom PVX
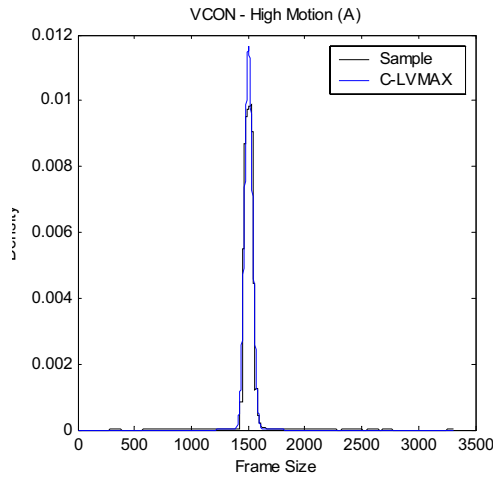(High)

Figure 1b: Frame-size histogram vs C-LVMAX fit
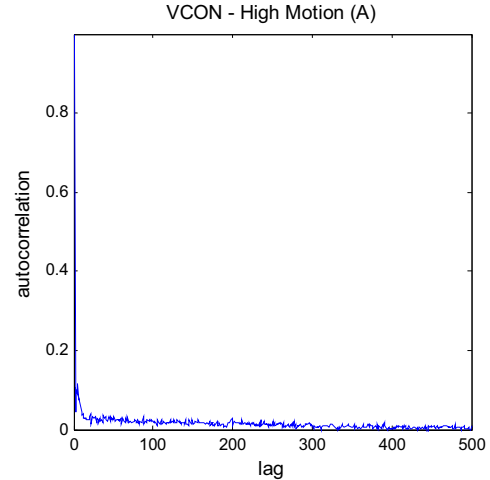(VCON Vpoint HD – High)

Figure 2b: Autocorrelation Graph for VCON Vpoint
HD (High)

The narrowness of the frame-size histograms can be easily explained by the improved compression algorithms used by the encoders that give smaller frame sizes. In fact, the case of Vpoint could be characterized as a case of CBR transmission (as it mostly sends frames with approximately the same size- around 1500 bytes). However, modeling of the total traffic, including the frame-sizes of smaller frequency could be useful for traffic policing purposes.

Furthermore, the AutoCorrelation Function (ACF) of the constrained traffic appeared to have values near zero denoting that the examined traffic is uncorrelated and exhibits no short-range dependence (see figures 2((a),(b)) as generally proved for H.261 and H.263 videoconference traffic [17].

This result is very critical in queuing as the short-term correlation parameter has been found to affect strongly buffer occupancy and overflow probabilities for videoconference traffic. As a consequence, we can reasonably say that a $D/G/\infty$ queue is a proper model in the case of the examined traffic.

This turns constrained videoconference traffic more amenable to traffic modeling than its counterpart unconstrained of H.261 and H.263 as only the

parameters of the distribution of the service times are needed to construct the proposed model. The arrivals rate can be easily assumed equal to the video frame rate (from 15 to 30 frames per second).

3.1. Fitting of the frame-size frequency histograms of the traces

For the purpose of fitting the sample frame-size sequence histogram with a theoretical density (to be used as the random generator of the service times of the D/G/∞ queue), typical moments matching methods were tested for several known densities (i.e. Weibull, Parero, Lognormal, Loglogistic) and failed to be accurate in all cases. To overcome this problem, we followed the approach of histogram-based modeling. Although various histogram-based methods (e.g. [25]) have been tried in literature, as well as maximum likelihood estimations (MLE) can be used, we followed a successful and very simple previous model proposed for H.261 encoded traffic in [25], namely, C-LVMAX. This method relates the peak of the convolution function of the empirical frame-size histogram to the location at which the Gamma density (1) achieves its maximum and to the value of this maximum.

The Gamma density is given by the following equation.

Gamma: $f(x) = \dfrac{1}{\beta \Gamma(\alpha)} \left(\dfrac{x}{\beta} \right)^{\alpha-1} e^{-\frac{x}{\beta}}$ \qquad (1)

where $\Gamma(\alpha)$ is the Gamma function with parameter α. The values of the shape and scale parameters of the Gamma density via the C-LVMAX method are derived

from: $a = \dfrac{2\pi x^{*2} f^{*2} + 1}{2}$ and $b = \dfrac{1}{2\pi x^* f^{*2}}$ where

f^* is the unique maximum of the histogram's convolution density at x^*. In Figures 3((a),(b)) cumulative density values of the sample frequency histogram and the Gamma density (C-LVMAX) are compared with Quantile to Quantile plots. It is evident, by these figures, how the C-LVMAX method fit the empirical data.

4. Conclusions

The current study is a contribution of measurement and modeling practical results of H.264 traffic not examined before for talking heads communication. An extensive analysis of the measured data, a careful but simple modeling of the frame-size sequences and the

extensive evaluation of the modeling components, led us to the general conclusion that the traffic can be characterized as constrained and that the generated traffic is uncorrelated and can be accurately represented by a D/G/∞ queue.
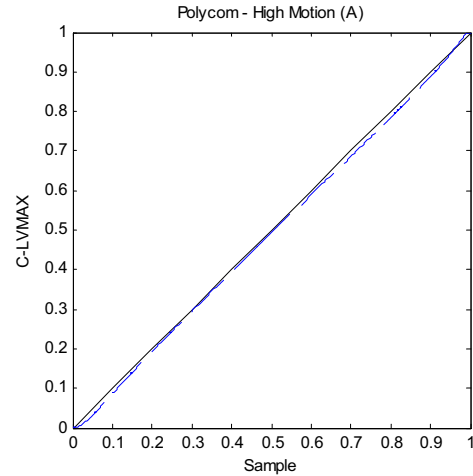

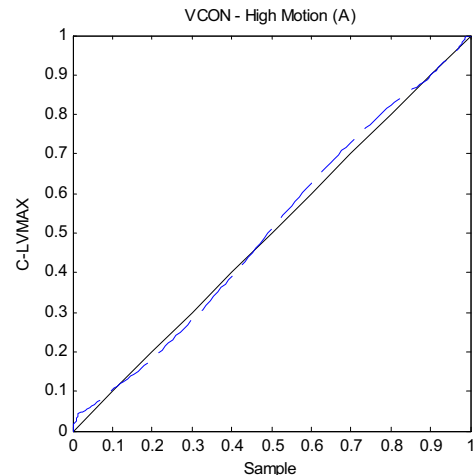**Figure 3a**: Q-Q Plots for Polycom PVX (High)

Figure 3b: Q-Q Plots for VCON Vpoint HD (High)

We used the measured data to develop a statistical traffic model for constrained traffic. This model was further validated with different videoconference contents (low motion and high motion). A statistical model for fitting the empirical distribution (C-LVMAX method) was proposed.

Another interesting assumption is that the traffic trends remain invariant when a different network constrained is selected, as evident from the Polycom traces. So, the proposed model for the constrained traffic can be applied without taking into account the specific network constraint.

Future work includes the integration of the proposed models in dynamic traffic policy schemes in real diffserv IP environments in order to measure the bandwidth gains from application of the proposed model.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", *IEEE Trans. Circuits Syst. Video Technol.*, vol 13, July 2003, pp. 560-576.

[2] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. D. Robbins, "Performance models of statistical multiplexing in packet video communications", *IEEE Trans. Commun.* (1988), 36(7): pp. 834–843.

[3] R. Kishimoto, Y. Ogata, and F. Inumaru, "Generation interval distribution characteristics of packetized variable rate video coding data streams in an ATM network", *IEEE JSAC* (1989), 7(5): pp. 833–841.

[4] H. S. Chin, J. W. Goodge, R. Griffiths and D. J. Parish, "Statistics of video signals for viewphone-type pictures", *IEEE JSAC* (1989), 7(5): pp. 826–832.

[5] M. Nomura, T. Fujii and N. Ohta, "Basic characteristics of variable rate video coding in ATM environment", *IEEE JSAC* (1989), 7(5): pp. 752–760.

[6] D. P. Heyman, A. Tabatabai and T. V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks", *IEEE Trans. Circuits Syst. Video Technol.* (1992), 2(1): pp. 49–59.

[7] D. M. Cohen and D. P. Heyman, "Performance modelling of video teleconferencing in ATM networks", *IEEE Trans. Circuits Syst. Video Technol.* (1993), 3(6): pp. 408–422.

[8] D.P. Heyman, T.V. Lakshman, "Modelling Teleconference Traffic from VBR Video Coders", *IEEE ICC* (1994), pp.1744-1748.

[9] D.M. Lucantoni, M.F. Neuts, "Methods for Performance Evaluation of VBR Video Traffic Models", *IEEE/ACM Transactions on Networking* (1994), Vol.2, No.2, pp. 176-180.

[10] S. Domoxoudis, S. Kouremenos, V. Loumos and A. Drigas, "Measurement, Modelling and Simulation of Videoconference Traffic from VBR Video Encoders", *Second International Working Conference, HET-NETs '04, Performance Modelling and Evaluation of Heterogeneous Networks* (2004), Ilkley, West Yorkshire, U.K., 26 - 28 July.

[11] K. Dolzer, W. Payer, "On aggregation strategies for multimedia traffic", *Proceedings of the 1st Polish-German Teletraffic Symposium* (2000), Dresden

[12] W.C. Poon and K.T. Lo, "A refined version of M/G/∞ processes for modelling VBR video traffic", *Computer Communications* (2001), vol.24, no.11, pp.1105-1114.

[13] L. Yan-ling, W. Peng, W. Wei-ling, "A Steady Source Model for VBR Video Conferences", International Conference on Information Technology: Computers and Communications (2003), Las Vegas, Nevada, p. 268.

[14] C. Skianis, K. Kontovasilis, A. Drigas and M. Moatsos, "Measurement and Statistical Analysis of Asymmetric Multipoint Videoconference Traffic in IP Networks", *Telecommunications Systems* (2003), Volume 23, Issue 1. pp. 95-122.

[15] B. Ryu, "Modeling and Simulation of Broadband Satellite Networks: Part II-Traffic Modelling", *IEEE Communications Magazine* (1999).

[16] G. Sisodia, L. Guan, M. Hedley, S. De, "A New Modeling Approach of H.263+ VBR Coded Video Sources in ATM Networks", RealTimeImg(6) (2000), No. 5, pp. 347-357.

[17] A. Drigas, S. Kouremenos, Y. Bakopoulos and V. Loumos, "A study of H.263 Traffic Modeling in Multipoint Videoconference Sessions over IP Networks", *Computer Communications*, In Press, Corrected Proof, Available online 13 May 2005.

[18] VCON Vpoint HD, *http://www.vcon.com*.

[19] Polycom PVX, *http://www.polycom.com*.

[20] MorningSound, *http://www.soundmorning.com*.