Performance Evaluation of Sign-Language Videoconference Traffic

Athanasios Drigas, Stelios Kouremenos and Dimitris Kouremenos
Net Media Lab of N.C.S.R. 'Demokritos', Agia Paraskevi, 153 10, Athens, Greece
{dr, skourem, dk}@imm.demokritos.gr

Abstract

This paper presents modeling results for sign language videoconference traces (H.261 and H.263 encoded) which were captured from realistic multipoint IP videoconference sessions between Greek signers. By comparing traffic models against simulations, it is proven that the application of the proposed theoretical parameters into a generic continuous Markovian model can lead to a conservative, though accurate, solution for the analytical treatment of this type of traffic.

1. Introduction

The increasing availability of affordable communication channels such as ADSL, together with readily available videoconferencing software (MS NetMeeting) compatible with established video coding standards (such as H.261 and H263), offer the deaf population the possibility of remote sign language communication. The new channels provide the ability to transmit at rates sufficient for video transmission, at prices low enough to be within reach of many consumers. Since sign language videoconferencing relies on the exchange of bandwidth demanding qualitative video information - minimum video bit rate of 384 kbps and frames per second are at least 15 reported in [3]-[4] although 30 is ideal, extensive deployment of this service calls for careful modeling of the associated traffic, so that the appropriate amount of resources may be anticipated by the network. Moreover, successful traffic modeling can provide valuable insights about the resulting network load and enables a theoretical assessment of the network performance. To date, most videoconferencing traffic modeling research has been tested on head & shoulders or movie sequences. Sign language, which features moving arms and hands, contains more motion than a typical head and shoulders sequence and less than a movie like Star Wars (frequently used in many studies like [5]).

Sign language videoconference traffic modeling has not been studied in literature (to the best knowledge of these authors). In [6], the author is presenting a simple method of measurement of sign language video that contributes some generic results on the requirements of sign language video transmission. The official application profile document that gives the background to the requirements of sign language (and lip-reading) successful transmission is the ITU document [7]. According to this document, the usability of sign language is reported to be good at 20 frames per second, and with some constraints 12 fps and higher.

Today, a large number of videoconference platforms exist, the majority of them over IP-based networking infrastructures and using practical implementations of the H.261 and H.263 standards for video coding. H.263 is extensively used because of its suitability for transmission over low bandwidth pipes.

Moreover, videoconference traffic modeling has been extensively studied and successful models like the DAR [2] and its continuous counterpart [2] have been proposed. However, it is of great importance to know whether the models established in literature are appropriate when sign language is used. It is a point of question whether sign language videoconference traces (H.261 or H.263 encoded) generate similar traffic so that a common model could be applied. Furthermore, videoconference service is now held through Multipoint Control Units (software or hardware) that employ a centralized management for a better quality of the sessions. In such a case, the traffic from the clients to the MCU is highly influenced by the parameters of the possible scenarios-modes of the MCU (codec used, number of participants, video bit rate, and frame rate).

Besides studies whose subject of research was videoconference traffic, all other studies (concerning vbr and MPEG video traffic modeling) use movies as the video source of their experiments (like Star Wars)

1-4244-0674-9/06/$20.00 ©2006 IEEE.

that exhibit abrupt scene changes. However, the traffic pattern generated by differential coding algorithms (like those used by H.261 and H.263) depends strongly on the variation of the visual information. Especially, for sign language, as remarked in [9], traffic sequences have different characteristics from those of typical head & shoulders sequences. As noted in the same study, the picture quality (frame size) and frame rate are both important factors in sign language perception. The resolution of the hands must be high enough to support finger spelling, while the resolution of the face is also important since facial affect may convey syntactic information. However, block-based coding standards like H.261 and H.263 demand expensive computation power to increase the compression of sign video sequences. This leads to a reduction of the frame rate and of the expected quality, in case of no centralized management by an MCU (that tends to keep the balance between algorithmic complexity and compression performance).

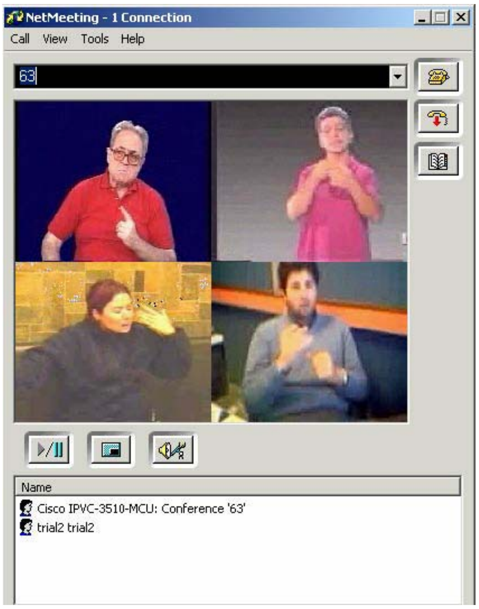

NetMeeting - 1 Connection
Call View Tools Help
63
Name
Cisco IPVC-3510-MCU: Conference '63'
trial2 trial2

Figure 1. In continuous presence mode, four QCIF H.261 videos are combined into one CIF H.261 video

On the above basis, the research reported in this paper undertook measurements of the IP traffic generated during videoconference sessions (at continuous and switched presence mode) between four videoconference clients (MS NetMeeting) used by Greek signers (three interpreters and one deaf user). At switched presence mode, the MCU sends to all terminals the output from one participant (QCIF video), designated as ‘‘currently active’’ while at continuous presence, the MCU combines the signal from all terminals and sends back the output to all the participants (CIF video, see Figure 1).

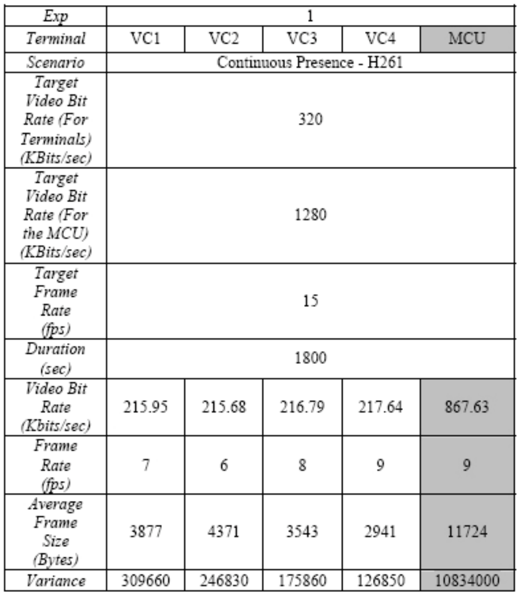
The rest of the paper is structured as follows: section 2 discusses the videoconferencing platform employed for experimentation. Section 3 proposes methods for the modeling of the generated traffic for all cases and presents the analytical theoretical against the simulation results. Finally, section 4 culminates with conclusions and pointers to further research.

2. Description of the videoconference experiments

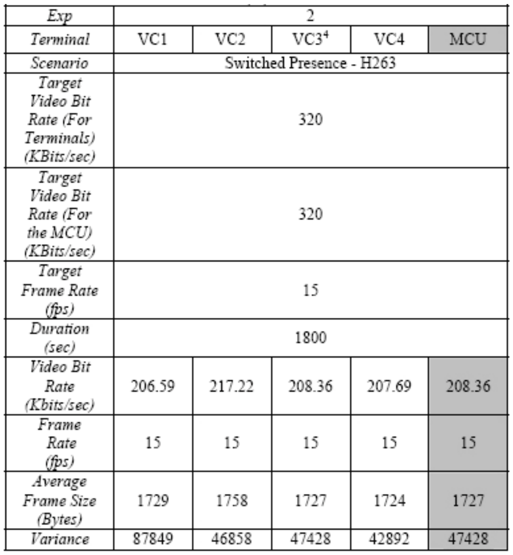
Concerning the experimental work, two experiments were held at two different high qualities (with an acceptable for sign language peak rate equal to 320 Kbits/sec) modes of CISCO MCU 3510 (hardware MCU): continuous presence mode - H.261 coding and switched presence mode - H.263 coding (see Table 1 for more details). In both cases, to ensure the quality of the sign sequences, MS NetMeeting clients were configured with the same video parameters (High Quality - QCIF). Furthermore, the same sign language Video Contents VC_{1-4} (signers conversing) were used in both cases for reasons of statistical comparison between H.261 and H.263 traffic.

In each case, the IP packets exchanged between the terminals and the MCU were captured by the traffic monitoring software ‘‘Ethereal’’. The collected data were further post-processed at the ‘‘frame’’ level by tracing a common packet timestamp. The produced sequences were used for further analysis. Some first conclusions, as supported by the experiments’ results (Table 1), arise concerning the statistical trends of sign language videoconference traffic: It is indicated that the first-order statistical characteristics of H.261 sign language traffic are similar to those of [1] where head & shoulders content was used. This is due to the centralized management of the MCU that keeps a balance between frame rate and frame size to acquire a constant video bit rate (MS NetMeeting appears to try to remain at the 75% of the target video bit rate). However, H.323 traffic reaches 15 fps in most cases, reducing the average frame size.

Table 1. The experiments’ two scenarios and some first-order statistical characteristics of the generated frame sequences



Exp	1				
Terminal	VC1	VC2	VC3	VC4	MCU
Scenario	Continuous Presence - H261				
Target Video Bit Rate (For Terminals) (KBits/sec)	320				
Target Video Bit Rate (For the MCU) (KBits/sec)	1280				
Target Frame Rate (fps)	15				
Duration (sec)	1800				
Video Bit Rate (Kbits/sec)	215.95	215.68	216.79	217.64	867.63
Frame Rate (fps)	7	6	8	9	9
Average Frame Size (Bytes)	3877	4371	3543	2941	11724
Variance	309660	246830	175860	126850	10834000



Exp	2				
Terminal	VC1	VC2	VC3[4]	VC4	MCU
Scenario	Switched Presence - H263				
Target Video Bit Rate (For Terminals) (KBits/sec)	320				
Target Video Bit Rate (For the MCU) (KBits/sec)	320				
Target Frame Rate (fps)	15				
Duration (sec)	1800				
Video Bit Rate (Kbits/sec)	206.59	217.22	208.36	207.69	208.36
Frame Rate (fps)	15	15	15	15	15
Average Frame Size (Bytes)	1729	1758	1727	1724	1727
Variance	87849	46858	47428	42892	47428

3. Analysis of the video traffic

The sequence of the frame sizes from a terminal can be represented as a stationary stochastic process, with an AutoCorrelation Function (ACF) quickly decaying to zero and a marginal frame-size distribution of approximately Gamma form. These general characteristics remain invariant for both experiments. Analysis of the traffic from the MCU to the clients, in experiment 1, confirms the results of a previous study [1], where the Probability Density Function (PDF) of the MCU in the continuous presence mode had the form of a weighted sum of four Gamma components (equal to the number of the conferring terminals, see Figure 2). In experiment 2, the MCU PDF has the form of a simple Gamma Distribution, same with terminal VC3 (that of the currently active user). This being the case, analysis of the MCU trace in experiment 2 is included in the analysis of the VC3 trace.

The analysis of the MCU trace in experiment 1 will not be included in the present study as it demands a separate methodology than that of the terminals' analysis. Thus, the rest of our work will concentrate on the analysis of the frame sequences generated by the terminals.

Our work is based on the analytical theoretical model C-DAR (1) proposed in [2] for videoconference performance analysis. The C-DAR (1) model combines an approach utilizing a discrete-time Markov chain with a continuous-time Markov chain. More explicitly, this model produces a sequence of frame sizes according to the transitions of a continuous time Markov Chain, of the form:

$$\begin{aligned} &P_{cdar} = f\,(P_{dar} - I) \\ &where \rightarrow f = \frac{\ln \rho}{(\rho - 1)} T\,, P_{dar} = \rho I + (1 - \rho) Q \end{aligned} \qquad (1)$$

from DAR (1) of D.P. Heyman [10] where *T* is the frame rate of the videoconference traffic (easily obtained from Table 1), *I* is the identity matrix, *ρ* is the autocorrelation coefficient at lag-1 and *Q* is a rank-one stochastic matrix with all rows equal to the probabilities resulting from the negative binomial density corresponding to the Gamma fit for the frame size distribution.

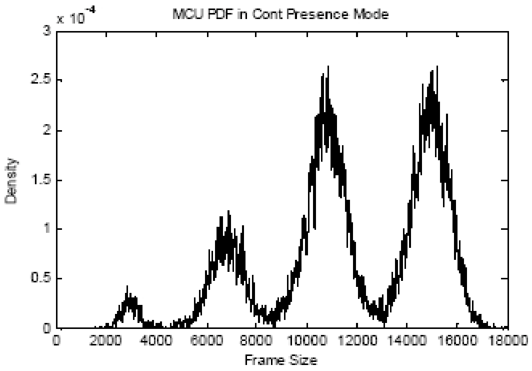


Figure 2. MCU PDF in Continuous Presence Mode

The rest of our work will concentrate on the calculation of the correlation coefficient *ρ* and the stochastic probability matrix *Q* of the eight videoconference traces of Table 1. This will be performed with theoretical fits on the ACF and PDF of the frame sizes sequences correspondingly.

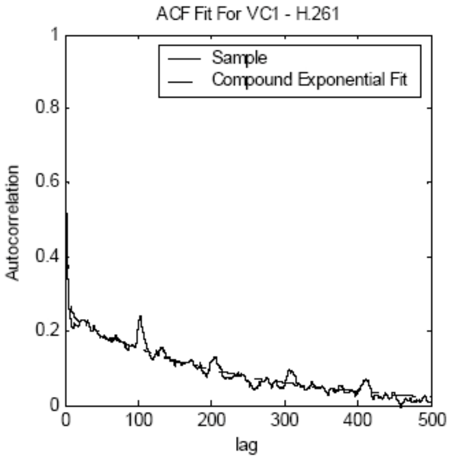
3.1. Calculation of the autocorrelation coefficient *ρ*

To find the most accurate fitting model for the ACF of sign language traffic, the Compound Exponential Fit method proposed in our previous study [1] was used.

$$\rho_{\kappa} = w\lambda_1^{\kappa} + (1-w)\lambda_2^{\kappa}, with|\lambda_2| < |\lambda_1| < 1 \qquad (2)$$

This method was tested with a least squares fit to the autocorrelation samples for the first 500 lags. In fact, what matters is the correlation coefficient λ_1 that has been calculated around 0.98 for videoconference traffic in [10] and has been applied in C-DAR(1) analytical model in [2]. However, our approximation of calculating the correlation coefficient is more conservative than that of the AR(1) method of the C-DAR(1) model as we pay attention to the long-term trend of the ACF[5]. This is further confirmed by our results where λ_1 has values larger than 0.99 (see Table 2).

Figure 3 reflects the fact that the ACF of sign language videoconference traffic has very strong correlations (especially H.263 traffic). This is due to the motion compensated techniques used by the two differential coding algorithms. It seems that the bursty nature of sign language force the algorithms to send regularly Intra frames in order to improve the image quality. In the case of Figure 3, the periodicity of the ACF is attributed to the similarities (temporal redundancy) that exist between sequential H.263 video frames. However, the reflection of these characteristics to queuing is minimal as only the value of the correlation coefficient matters.



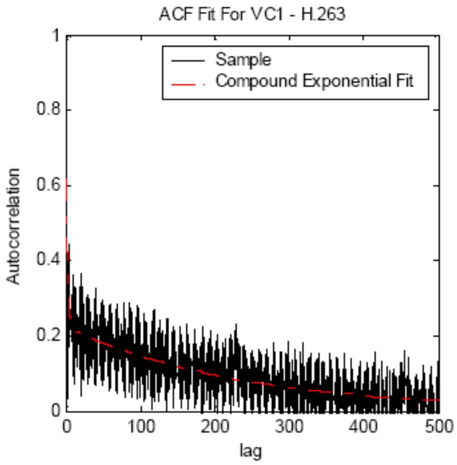


Figure3. The ACF fits for H.261 and H.263 sign language traces

3.2. Calculation of the stochastic matrix *Q*

Now, we may proceed to the analysis of the PDF of the traffic patterns' frame sizes which is one of the main points of the current study. In actuality, all density distributions seem to fit a gamma-like shape. In the analysis to follow, the Gamma density function will be used to fit the empirical PDFs.

$$f(x) = \frac{1}{\Gamma(p)}\frac{1}{\mu}(\frac{x}{\mu})^{p-1} e^{-\frac{x}{\mu}}, \mu, p > 0, x \geq 0, \qquad (3)$$
$$\Gamma(p) = \int_0^{\infty} u^{p-1} e^{-u} du$$

The method used is the widely used MOM method (method of MOMents) successfully applied for traffic modeling in previous studies. When the mean, *m*, and the variance, *v*, of the data sample are known (in our case these values are obtained from Table 1), the method of moments produces estimates for the shape and scale parameters of the Gamma distribution:

$$p = \frac{m^2}{v}, and, \mu = \frac{v}{m} \qquad (4)$$

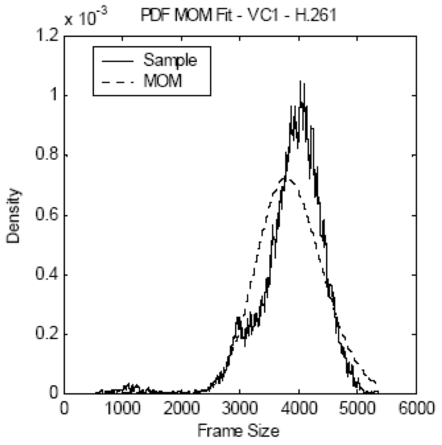
Table 2. The correlation coefficients and gamma parameters of the traces

	H.261				
	Compound Exponential Fit			Method of Moments	
	w	λ_1	λ_2	p	μ
VC1	0.7532	0.9952	0.6791	48.53	79.88
VC2	0.936	0.9903	0.7171	77.40	56.47
VC3	0.7849	0.9977	0.586	71.39	49.63
VC4	0.2288	0.9987	0.5131	68.19	43.13
	H.263				
VC1	0.7772	0.9957	0.5119	34.03	50.81
VC2	0.9528	0.9947	0.6881	69.84	56.79
VC3	0.7943	0.9979	0.486	62.91	27.46
VC4	0.3176	0.998	0.5326	69.25	24.89

Indicatively, for the VC1 trace, we show the fitted PDF with the MOM method (see Figure 4) in both experiments. In all cases, the fit of the MOM method was satisfactory and no other models needed to be tested. The values of p and μ parameters for all sign traces are given in Table 2 (to be used in the formation of the Q Matrix).

3.2. Analysis results for the C-DAR model

Since the C-DAR model is a continuous-time Markov chain model, it may be looked at as a Markov modulated rate process [11], and therefore suitable for theoretical analysis using the fluid flow method. In the paragraph to follow, we present experimental results comparing the analytical C-DAR model (via the fluid flow method as described in [2]), versus a simulation using actual video traces. In order to perform the simulation, we used a discrete-time-event simulator (ns-2) [12].



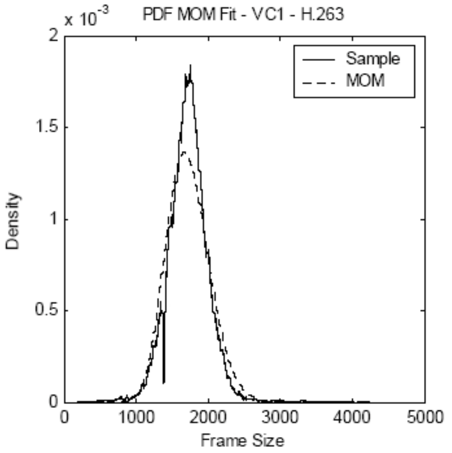
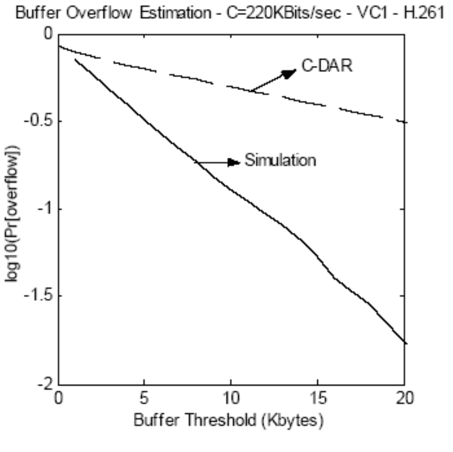


Figure 4. The PDF plots together with the MOM fit

Running ns-2, we simulated a queuing system with trace-driven arrivals of the actual data gathered during our experiments. By recording the queue length, we estimated the complementary buffer overflow estimation for a given server capacity C (always a little larger than the mean rate of the sample). The results (Figure 5) prove the conservativeness of the C-DAR model mainly due to the calculation method of the correlation coefficient ρ. A less conservative choice of the value of ρ (i.e. fitting the ACF at the first 100 lags) leads to more tight results. It is clearly shown that our model is a simple - only three parameters (mean, variance and the correlation coefficient) are demanded - and robust method for performance analysis of sign language videoconference traffic. Again, plots are given for VC1 contents in both experiments.



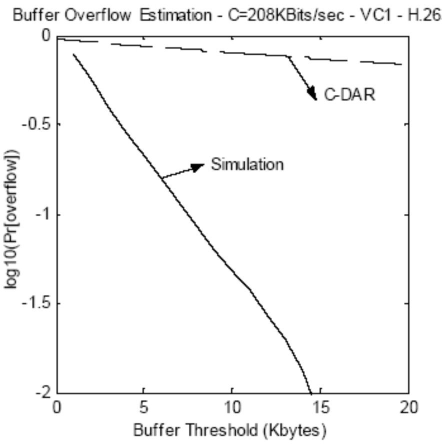


Figure5. The complementary distributions plots of buffer overflow estimation for a given server capacity *C*

4. Conclusion

This manuscript was a modeling assessment of H.261 and H.263 encoded sign language traffic in multipoint videoconference sessions over IP Networks. The modeling results showed that the sign language terminal traffic was stationary and seemed to possess a rapidly decaying strongly correlated autocorrelation function and a Gamma-formed marginal distribution well fitted with the MOM method. Although the correlations of generated traffic are more complex than a simple geometric term, careful choice of the decay rate allows the construction of a conservative, accurate though, approximation. Our generalization of the C-DAR(1) model is proven to be a robust solution for theoretical studies on videoconferencing for deaf people. A simple calculation of the mean and variance of the frame sizes sequence and a choice of a correlation coefficient near 0.99 are the demanded input parameters of our proposed model.

Further work will include analysis of the MCU generated traffic in continuous presence mode and studies of statistical multiplexing of sign language sources into IP links.

6. References

[1] W. C. Skianis, K. Kontovasilis, A. Drigas, and M. Moatsos, "Measurement and statistical analysis of asymmetric multipoint videoconference traffic in IP networks", *Telecommunications Systems*, vol. 23, Issue 1, pp. 95-122, 2000.

[2] W. S. Xu, Z. Huang, and Y. Yao, "An analytically tractable model for video conference traffic", *IEEE Transactions Circuits Systems Video Technologies*, 10(1):63–67, 2000.

[3] G. Hellström, Delevert, Revelius, "Quality requirements on videotelephony for sign language", *Swedish National Association of the Deaf*, 1997.

[4] H.W. Frowein, "Improved speech reception through videotelephony", IEEE Journal on Selected Areas in Communication, 1991.

[5] F. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation", *IEEE Network*, vol. 15, no. 6, pp. 40-54, 2001.

[6] G. Hellstrφm, "Quality measurement on video communication for sign language", http://www.omnitor.se/english/qualityrequirements.html

[7] ITU-T, Series H–Sup1, Application profile - sign language and lip-reading real-time conversation using low bit-rate video communication

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

[9] R. P. Schumeyer, K.E. Barner, "Color-based content coding with applications to sign language video communications", *IEEE Transactions on Circuits and Systems for Video Technology*, 1997.

[10] 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.*, 2(1):49–59. 1992.

[11] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer", *Adv. Appl. Prob.*, Vol. 20, 646-676, 1988.

[12] The Network Simulator – ns-2,
http://www.isi.edu/nsnam/ns/.