

Towards an analytical model for characterising behaviour of high-speed VVoIP applications

Andrei Sukhov¹, Prasad Calyam², Warren Daly³, Alexander Ilin⁴

¹*Laboratory of Network Technologies, Samara Academy of Transport Engineering,
1st Bezymyayni per., 18, Samara, 443066, Russia
e-mail: sukhov@ssau.ru*

²*OARnet, The Ohio State University, 1224 Kinnear Road, Columbus, Ohio 43212, USA
e-mail: pcalyam@oar.net*

³*HeaNet Ltd, Crampton Ave, Shelbourne Rd, Ballsbridge, Dublin 4, Ireland
e-mail: warren.daly@heanet.ie*

⁴*Russian Institute for Public Network, Kurchatov sq., 1, Moscow, 123182, Russia
e-mail: kalend@ripn.net*

Abstract: The use of high-speed multimedia technologies such as MPEG-based interactive Videoconferencing and IPTV for delivering Television content over the Internet is becoming increasingly common. With commercial vendors and content providers envisioning products and services for delivering audiovisual content equipped with greater user-control and higher resolutions, there has arisen a significant demand to identify and quantify various bottlenecks in the Internet to better support such ventures. In this paper, we attempt to develop an analytical model that reflects the effects of various device factors and network factors that affect the performance of high-speed interactive audiovisual streams, *i.e.* the ultimate end-user perception of audiovisual quality. To develop our model, we use a set of experimental results obtained in both a LAN as well as on an Internet testbed.

Key words: videoconferencing, IPTV, network impairments, human factors

1. INTRODUCTION

The impetuous growth of traffic volumes of interactive audiovisual applications on the Internet can be attributed to the rapid advances in numerous technologies associated with both network infrastructure and Voice and Video over IP (VVoIP) applications such as Videoconferencing, Video on Demand and IPTV. With end-users demanding superior interactive voice and video communications over the Internet that offers only “best effort service”, many studies have been conducted [4, 5, 16, 18] for understanding behaviour and bottlenecks relating to such voice and video traffic to find ways to improve the Internet infrastructure, *i.e.*, the VVoIP system.

These earlier works focused on many aspects of network performance bottleneck issues such as frame rate, packet delay, jitter, loss and reordering mainly in the pur-

view of voice traffic and in some cases in the purview of video traffic. While addressing video traffic behaviour, these studies have limited their scope to low-speed interactive audiovisual streams (< 384 Kbps) and have failed to verify these findings to be applicable to high-speed interactive multimedia data streams shown in Table 1. Since high-speed VVoIP applications using technologies in Table 1 are beginning to become widespread in use [22], it has become necessary to study performance bottlenecks relating to these applications by extending earlier principles and methodologies to better understand high-speed VVoIP system bottlenecks.

The primary challenges that distinguish the study of high-speed interactive audiovisual streams compared to the study of low-speed interactive audiovisual streams include:

- Understanding the effects of device factors (*e.g.*, compression schemes, jitter buffering schemes, packet loss con-

Table 1. High-speed interactive multimedia data stream technologies

Application Video Quality	Technology	Bitrate (Kbps)
HDTV, $R \approx 1920 \times 1080$, $F_r = 60$ fps, $q = 32$	MPEG-2	2048
VCR-quality television, $R \approx 400 \times 300$, $F_r = 30$ fps	MPEG-1	1200
Videoconferencing, $R \approx \text{CIF}(352 \times 288)$, $F_r = (15-30)$ fps, $q = 16$	H.261, H.263	384-1920

cealment schemes) to maintain consistent high frame rates on the receiver side in spite of the degradation of stream quality due to network factors (*e.g.* queuing due to congestion, packet delay, loss and reordering);

- Quantifying end-user perceived experience when there are high expectations of audiovisual quality in correspondence to the effects caused by device and network factors.

In this paper, we address our above challenges by attempting to derive an analytical model that suggests the effects of the network and device factors on the resulting end-user perception (*i.e.*, Mean Opinion Scores) in any given high-speed VVoIP system. We use experimental results from a LAN and an Internet testbed to support our derived model conclusions.

The remainder of this paper is organized as follows: Section 2 describes various terminologies used in our model. Section 3 describes our experimental setup used for data collection. Section 4 details our analytical model and salient observations. Section 5 concludes the paper.

2. TERMINOLOGY

In this section, we define various terms used in our analytical model. The significance of a few of the terms

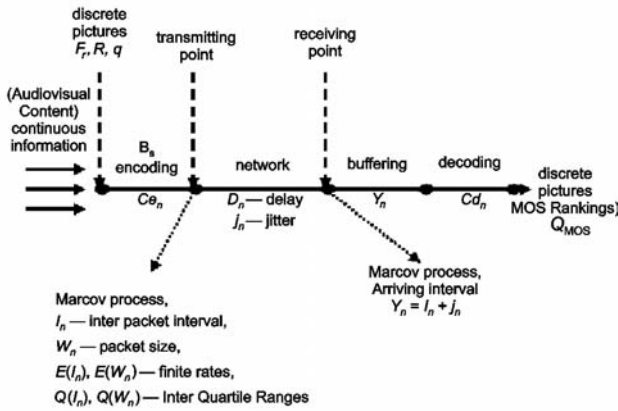


Fig. 1. End-to-end parameters for analytical model

defined in this section are shown as relevant to an actual VVoIP system in Fig. 1.

2.1. Encoding rate (B_s)

The encoding rate of the video streams is determined by a video codec with Compression factor C_{e_n} and Decompression factor C_{d_n} together with a co-efficient of compression τ . B_s value also depends on the capabilities of the source (CPU, clock rate, *etc.*), video content being encoded (A_l) and the access bandwidth to the Internet K . B_s can be written as follows

$$B_s = \tau(K)A_l(F_r, R, q) \quad (1)$$

where Frame Rate (F_r) is the number of frames per second during an audiovisual conference connection. Video Resolution (R) is the ability to discriminate closely spaced lines on the receiving screen and is measured in pixels. For example, versions of H.263 support five standardized picture formats: SQCIF (128×96), QCIF (176×144), CIF (352×288), 4CIF (704×576), and 16CIF (1408×1152). Color depth (q) is the number of bits used in a frame of video sequence.

2.2. Inter-packet times (I_n)

It can be defined as the times between successive packets at both the sender and receiver ends. Various parameters associated with inter-packet times are shown in Fig. 1. T_n depicts the transmission time at the sender (T_n^s) or arrival time at the receiver (T_n^r). I_n depicts the variation between the successive packet transmissions or variations in the received packet arrival times (Y_n). W_n depicts the packet size distribution in the packet sequences. We can also consider Z_n , which depicts the smoothed inter-packet sequence by the dejitter buffer at the receiver end-point.

2.3. Network health (N_h)

Network health [11] can be characterised as a function of Delay (D_n), Jitter (j_n) and Loss (p) as given by

$$N_h = f(D_n, j_n, p) \quad (2)$$

where Delay is the amount of time a packet takes to travel from the sender's application to reach the receiver's destination application. Jitter is defined as the variation in the delay of the packets arriving at the receiving end. Loss is defined as the percentage of transmitted packets that never reach the intended destination due to packets discarded deliberately (RED, TTL = 0) or non-deliberately by intermediate links (layer-1), nodes (layer-3) and end-systems (discard due to late arrivals at the application) along a given transmission path.

2.4. Audiovisual content (A_l)

A_l corresponds to the various low temporal (*e.g.*, casual videoconference, business presentation) or high temporal (*e.g.*, musical concert or sports broadcast) media streams generated using high-speed VVoIP applications, which need to be encoded/decoded, and transmitted through the network. The perception of the audiovisual content is quantified using the Mean Opinion Score (Q_{MOS}) metric which is used to obtain perceptual quality on a scale of 1 to 5 [15], of end-users based on the interaction response times in the system and based on the impairments observed in video streams (*e.g.*, frame freezing, tiling) and audio streams (*e.g.* echoes, dropouts) [10, 19].

3. TESTBED SETUP

Figures 2 and 3 show the setup for the experiments in the LAN and Internet environments, respectively. Both testbeds involved obtaining Ethereal traffic traces for low, medium and high temporal audiovisual content at encoding bitrates: 128, 384, 512, 768, 1024, 1472, 1920 Kbps. To extract the necessary data for determining the data distribution for inter-packet intervals I_n and packet sizes W_n from the traces, we used the RTP analysis module of the Ethereal network protocol analyser. We filtered-out the audio and video streams from the LAN traffic and decoded the audio and video packets information.

The data from the Internet experiments involved testing between RIPN, HEAnet and OARnet. The network paths between the test sites were comprehensively measured using various tools [17] such as I_{perf} , H.323 Beacon [4] and Pathping. All the tools reported excellent network path between the HEAnet and OARnet sites. For example, the H.323 Beacon Server-to-Server test indicated a MOS score of 4.33 with end-to-end jitter less than $j < 4$ ms and almost negligible loss ($p < 0.12\%$). The connection between HEAnet and the Russian site formed the best comparative site for our results, due to its poor network path characteristics. The H.323 Beacon results between the sites indicated < 4.0 MOS.

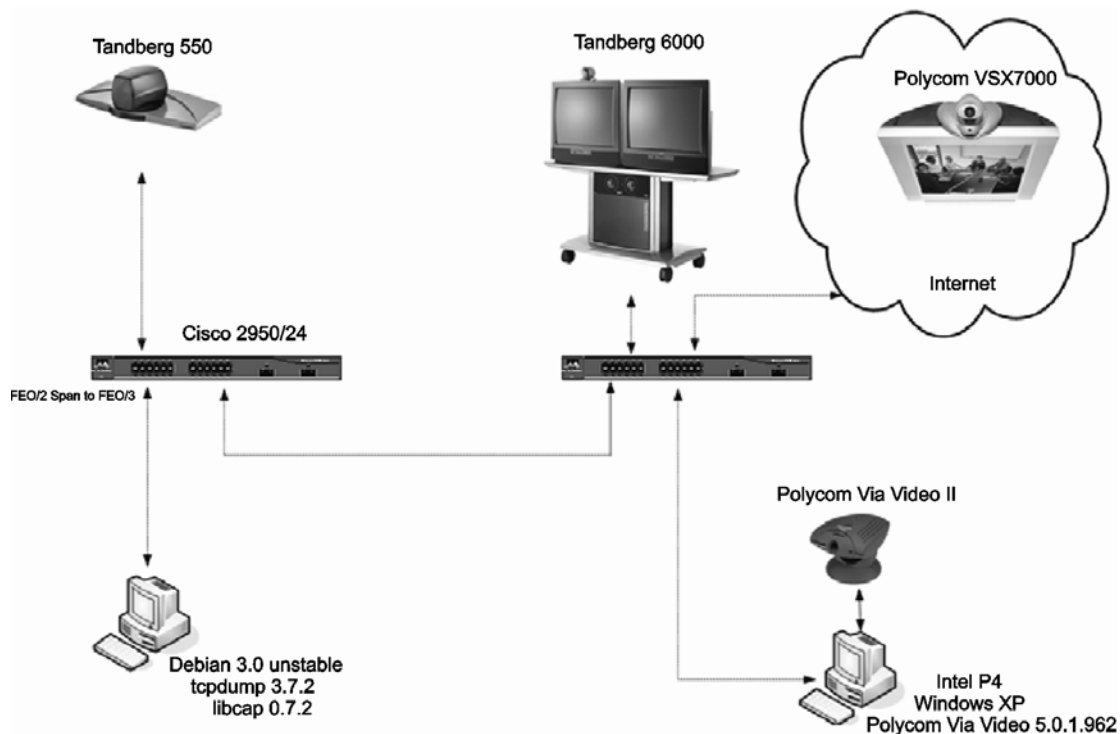


Fig. 2. LAN testbed

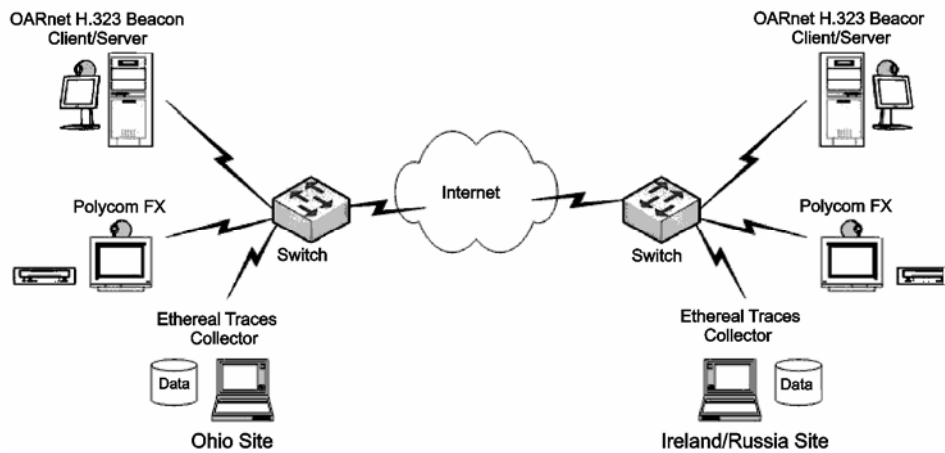


Fig. 3. Internet testbed

4. HIGH SPEED VVOIP ANALYTICAL MODEL

In this section, we describe our analytical model for a high-speed VVoIP system. We use the following assumptions in describing our analytical model [1, 2, 20]:

- A Markov process [9, 13] can be used to describe the arrival of the Audio and Video packets (RTP packets);
- The distribution of packet size W_n and sender side inter-packet intervals I_n can be characterised by using average values $E[I_n]$ and $E[W_n]$ respectively;
- A function φ from Eqn. (3)

$$\varphi(\Delta Q_{\text{MOS}}) = E[I_n]/j \quad (3)$$

describes the relation of network jitter j and sender-side inter-packet time $E[I_n]$ on the end-user perceived audio-visual quality QMOS.

Here $\Delta Q_{\text{MOS}} = Q_{\text{MOS}}^t - Q_{\text{MOS}}^r$ is the degradation of video quality after transmitting the data over IP network, Q_{MOS}^t and Q_{MOS}^r is the video quality at transmitting and received points correspondingly. The average value of $E[I_n]$ denotes the initial inter-packet jitter introduced by

Table 2. Values of $[W_n]$ and $[I_n]$ at different video resolutions

Resolution	B_s , Kbps	$E[W_n]$, bytes	$\Theta[W_n]$, bytes	$E[I_n]$, ms	$\Theta[I_n]$, ms
QCIF, CIF	128	600-800	600-800	70-100	50-80
CIF, 4CIF	384	900-1100	500-700	20-30	10-20
CIF, 4CIF	512	1000-1200	400-600	17-25	10-18
CIF, 4CIF	768	1000-1200	300-600	12-20	10-15
CIF, 4CIF	1024	1000-1300	300-600	10-18	8-14
4CIF	1472	1000-1300	300-600	8-15	7-14
4CIF, 16CIF	1920	1000-1300	300-600	5-10	7-12

the sender equipment before the audiovisual streams enter the network. Small values of $E[I_n]$ correspond to low jitter and thus a better quality of initial audiovisual streams (Q_{MOS}^t).

Smaller values of network jitter j correspond to higher values of user-perceived quality rankings Q_{MOS} . Thus the upper limit of Q_{MOS} [6, 7] for a given experiment with a given audiovisual content and end-to-end network topology determines the resultant I_n characteristics and consequently the inter-packet jitter distributions at the receiver end.

In order to ameliorate the distortion of initial sequence, dejitter buffering schemes are applied before the decoding process at the receiver-end. The bound of amelioration is influenced by the value Y_{max} . If the arriving interval $Y_n = I_n + j_n$ (see Fig. 1) or inter-packet Jitter after network distortion exceeds this period Y_{max} [21] then we experience a packet discard that affects Q_{MOS} :

$$Y_{\text{max}} = -\ln P(Y_n \geq Y_{\text{max}}) E[I_n], \quad (4)$$

where $P(Y_n \geq Y_{\text{max}})$ is the probability of event when arriving interval exceeds the buffering time equals.

Using I_n and W_n data, we can determine the average values: $E[I_n]$ and $E[W_n]$, as well as their Inter Quartile Range (IQR) values: $\Theta[I_n]$, $\Theta[W_n]$. The values observed in our experiments to study the characteristics of W_n and I_n are shown in Table 2 and in Figs. 4-6.

Figure 4 shows the cumulative number N of RTP (H.261) packets that arrive at the receiver end-point during a 60 seconds interval. The arrival rate $\lambda = 1/E[I_n]$ remains mostly constant throughout the 60 seconds interval. Figure 5 shows typical example of a distribution density of inter-packet intervals I_n for a video sequence. The local peaks on Fig. 5 corresponds approximately to 0.5, 10, 20, 30, 40 ms, etc at connection speeds of 512 Kbps and CIF resolution. The local peaks compare to the Gaussian curves, therefore the distribution density $p^I(x)$ can be written as described in [12]

$$p^I(x) = k \sum_i A_i \exp\left(-\frac{(x-I_i)^2}{2\sigma_i^2}\right) \quad (5)$$

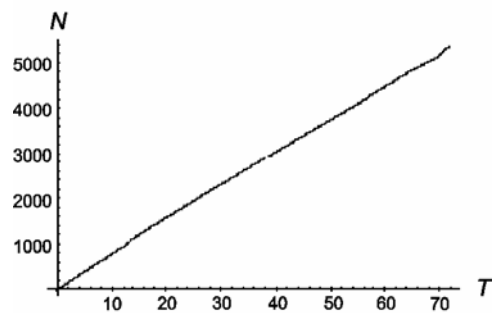


Fig. 4. RTP (H.261) packet arrivals during 60 seconds interval at 768 Kbps

Here

- k is the normalizing coefficient, so that $\int_1^\infty p^I(x) dx = 1$.
- Values I_i describe the local peaks of $p^I(x)$; A_i and σ_i characterises their respective amplitudes and dispersions.

It should be noted that a voice sequence (Ex., ITU-T G.722) consists of packets of constant packet sizes of

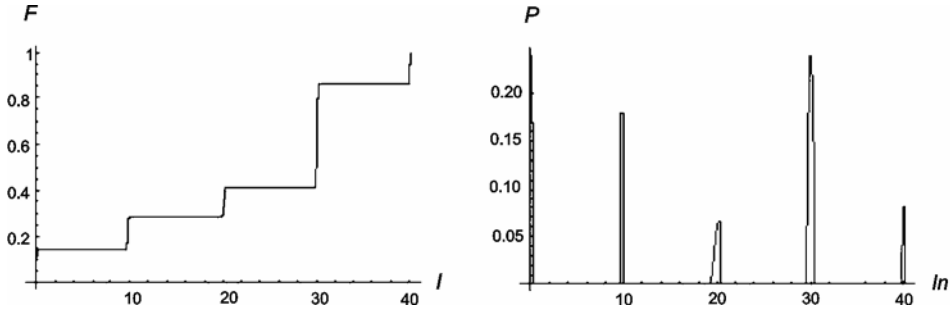


Fig. 5. Typical distribution F and distribution density P of inter-packet intervals I_n at 512 Kbps
 NOTE: For data shown in Fig. 5, the average size of inter-packet interval $E^{512}[I_n] = 23.0$ ms, IQR value $\Theta^{512}[I_n] = 29.7$ ms and quartiles are $\Theta_{1/4}^{512}[I_n] = 10.4$ ms, $\Theta_{3/4}^{512}[I_n] = 30.1$ ms

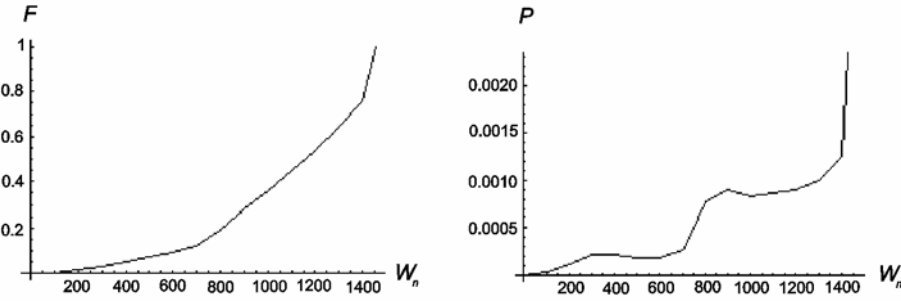


Fig. 6. Typical distribution F and distribution density p of inter-packet intervals W_n at 512 Kbps
 NOTE: For data shown in Fig. 6, the average size of inter-packet interval $E^{1920}[W_n] = 1086$ bytes, IQR value $\Theta^{1920}[W_n] = 533$ bytes and quartiles are $\Theta_{1/4}^{1920}[W_n] = 860$ bytes, $\Theta_{3/4}^{1920}[W_n] = 1393$ bytes

374 bytes whereas video sequence (Ex., ITU-T H.261) consists of packet sizes varying between 65 bytes to 1518 bytes.

A typical packet size density distribution for H.261 traffic can be studied using Fig. 6. The x axis corresponds to the size of W_n in bytes, y axis reflects the relative frequency of such packets in our data. It can be seen that about 25% of the small packet sizes do not exceed 860 bytes. Thus the video sequence is quite different from a voice sequence in terms of packet sizes and inter-arrival jitter values.

We can express the objective MOS ranking for voice as follows:

$$\Delta Q_{\text{MOS}}^{\text{voice}} = 0.012\sigma(D_{n+1} - D_n) + 0.24p \quad (6)$$

where $\sigma(D_{n+1} - D_n)$ is the quadratic mean and p is packet loss value measured in percentage of packets received as result of a ping sequence [8] of more than 100 packets:

ping -s [value >100] [IP address of another end point]

Since there is no conventional test method for determining objective MOS ranking for video connections, our coefficients in formula for video are purely subjective [3]. Thus the function $\varphi(\Delta Q_{\text{MOS}})$ is given by the equation:

$$\Delta Q_{\text{MOS}}^{\text{video}} = 0.015J + 0.008I - 0.0006JI \quad (7)$$

where the $J = \sigma(D_{n+1} - D_n) + 20p$ is the cumulative variable described Network Factor and σ, p are the results of a ping sequence of more than 100 packets:

ping -s [value >100] [IP address of another end point]

The value of I is given from the Table 2.

During signal processing of the audiovisual streams, the transformation of continuous picture in discrete sequence causes an unavoidable deterioration of quality of technical picture in comparison to the original generated streams. Therefore the maximal value Q_{MOS} reaches upto only 4.41 even in the most ideal conditions. When the connection bitrate becomes smaller than the quantity required for encoding or decoding of the picture, a frame drop is noticed. This fact explains the term βI in Eq. (8) when $J = 0$. The network induced degradation described by variable J causes drop in the connection video quality more intensely at high bit rates. In other words the variable $\Delta Q_{\text{MOS}}^{\text{video}}$ must have a higher magnitude at higher bit rates. Also, since J and I can be ameliorated at receiver endpoints using packet loss concealment schemes and dejitter buffering schemes, we can express the overall VVoIP model in general as follows:

Hence, we have

$$Q_{\text{MOS}}^{\text{video}} = Q_{\text{ideal}} - \alpha J - \beta I + \gamma JI \quad (8)$$

where:

- Q_{ideal} is maximum value of multimedia quality obtainable at the sender end-point before ingress into the network. Its value depends on the video and audio codecs used and the audiovisual content generated at the end-point side;
- $J = \text{IPDV} + \zeta p$ is variable that quantifies Network induced degradation. Here, IPDV is the inter-packet delay variance and p is percent of packet loss in the audiovisual stream;
- I is Inter-packet jitter before network distortion caused by the initial encoding device at the sender side;
- $\alpha, \beta, \gamma, \zeta$ are the numerical coefficients that must be found experimentally for any given path.

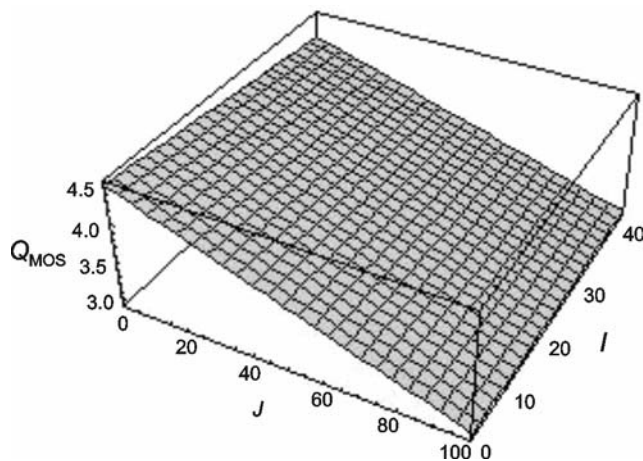


Fig. 7. 3D representation of VVoIP model surface

Our VVoIP model ultimately can be visualized as a 3D surface whose surface characteristics are determined by the values of the numerical coefficients α, β and γ explained above. Figure 7 shows a typical 3D representation of our VVoIP model with axes –inter-packet jitter before network distortion (I), network induced degradation (J) and the final perceived Q_{MOS} . Thus, we can imagine a unique surface that overall characterises the human perceived audiovisual quality by taking into account the degradation of the audiovisual streams due to both intermediate encoding/decoding device and network factors and any amelioration that the audiovisual streams can be subjected to, in a given VVoIP system.

5. CONCLUSION

In this paper, we presented our measurement methodology to develop a basic analytical model that features the impact and inter-dependencies of various network and

device features on the end-user perception of audiovisual quality when using interactive high-speed multimedia applications. We are planning to conduct further tests with more extensive experiments whose results could further strengthen the validity of our analytical model and also could lead us in the direction of extending our proposed analytical model to be more comprehensive in nature.

References

- [1] Ch. Barakat, P. Thiran, G. Iannaccone, Ch. Diot, Ph. Owezarski, *A flow-based model for Internet backbone traffic*, IEEE Transactions on Signal Processing – Special Issue on Signal Processing in Networking, 51, 8, 2111-2124, August 2003.
- [2] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, J. Roberts, *Statistical Bandwidth Sharing: A Study of Congestion at Flow Level*, ACM SIGCOMM, August 2001.
- [3] S. Bonaventura, F. Baldi, M. Alfano, M. Mirabelli, J. Antkowiak, A. Dehnel, *Testing user perceived multimedia communication quality using realistic tasks*, EURESCOM, Project P807: JUPITER2 – Joint Usability, Performability and Interoperability Trials in Europe, 1999.
- [4] P. Calyam, W. Mandrawa, M. Sridharan, A. Khan, P. Schopis, H.323 Beacon: *An H.323 application related end-to-end performance troubleshooting tool*, Proceedings of ACM SIGCOMM Network Troubleshooting Workshop (NetTs' 04), Portland, October 2004.
- [5] P. Calyam, M. Sridharan, W. Mandrawa, P. Schopis, *Performance Measurement and Analysis of H.323 Traffic*, Passive and Active Measurement Workshop (PAM), 2004.
- [6] A. Clark, *Modelling the effects of burst packet loss and recency on subjective voice quality*, 2001.
- [7] M. Claypool, J. Tanner, *The Effects of Jitter on the Perceptual Quality of Video*, ACM Multimedia", 1999
- [8] L. Cottrell, W. Matthews, C. Logg, *Tutorial on Internet Monitoring \& PingER at SLAC*, 2001, <http://www.slac.stanford.edu/comp/net/wan-mon/tutorial.html>
- [9] Chen-Nee Chuah and Randy H. Katz, *Characterizing Packet Audio Streams from Internet Multimedia Applications*, ICC 2002.
- [10] R. Finger, A. W. Davis, *Measuring Video Quality in Videoconferencing Systems*.
- [11] F. Fluckiger, *Understanding Networked Multimedia*, Prentice Hall, Inc. 1995.
- [12] B. V. Gnedenko, *Theory of Probability*, NY, Gordon and Breach Science Publishers, 1999.
- [13] W. Jiang and H. Schulzrinne, *Modeling of packet loss and delay and their effect on real-time multimedia service quality*, Proc. of NOSSDAV, 2000.
- [14] ITU-T Recommendation G.107, *The Emodel, a computational model for use in transmission planning*, 1998.
- [15] ITU-T Recommendation P.911, *Subjective audiovisual quality assessment methods for multimedia applications*, 1998
- [16] M. Karam, F. Tobagi, *Analysis of the Delay and Jitter of Voice Traffic over the Internet*, IEEE INFOCOM, 2001.
- [17] S. Leinen, M. Przybylski, V. Reijts, Sz. Trocha, *Testing of Traffic Measurement Tools*, GEANT Report GEA-01-113, October 2001.
- [18] A. Markopoulou, F. Tobagi, M. Karam, *Assessment of VoIP quality over Internet backbones*, IEEE INFOCOM, 2002.
- [19] J. Mullin, L. Smallwood, A. Watson, G. Wilson, *New techniques for assessing audio and video quality in real-time interactive communications*, IHM-HCI Tutorial, 2001.
- [20] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, *Modeling TCP Throughput: A Simple Model and its Empirical Validation*,

Proc. SIGCOMM Symp. Communications Architectures and Protocols Aug. 1998, 304-314.

- [21] A. Sukhov, P. Calyam, W. Daly, A. Iliin, *Network requirements for high-speed real-time multimedia application*, III IPv6 global Summit (Internet. New Generation -IPv6), Moscow, November 2004, p. 2833.

- [22] Y. Zhu, *A Survey of Network Requirements to Support Current and Future Data Streams*, CS 7001 Introduction to Graduate Studies Mini-project 1, Fall 2000.



Andrei Sukhov is an Associate Professor, Head of Laboratory of Network Technologies, Samara State Academy of Transport Engineering, Russia and was awarded a PhD in Moscow, in Physics and Mathematics in 1993. Over the last 10 years he has been involved in acting as an investigator for more than 10 telecommunication projects supported by the Russian government, INTAS, NATO, ESA, US Information Agency, etc. These are the construction of Samara Regional Network for Science and Education - the first Russian provincial network providing a digital connection to Moscow, the regional transport network at level STM16, etc. He is a founder of the biggest Regional ISP and an expert in the area of telecommunication investments in Russia.



Prasad Calyam is a Systems Developer/Engineer at OARnet, A Division of The Ohio State University. He obtained his M.S. in Electrical and Computer Engineering from The Ohio State University in 2002 and is currently enrolled in the Ph.D. program in the same department. His expertise is in the design and development of software systems comprising of advanced architectures and network-based performance analysis techniques and tools. He has previously worked on projects for companies such as Mphasis, Samsung Electronics Corporation, Korea and Eaton Corporation, USA in the commercial sector and for research groups at the Indian Institute of Science, The Ohio State University, OARnet and Internet2 in the academic sector.



Warren Daly works as a Network Security Expert for HEAnet Ltd. He has been an IT professional for 8 years and holds a BSc in Business and Information Technology from Trinity College Dublin and a Diploma in Electronic Systems from Dublin Institute of Technology. He has successfully deployed a PKI infrastructure for the Irish Academic & Research community and has extensive knowledge in Information Security Management Systems (ISO17799). He is also a member of the EU National Research Networks Policy Committee Review panel.



Alexander Iliin works as Network Engineer at Russian Institute for Public Networks. He graduated Moscow Institute of Physics and Technology at 1998 as engineer-physicist with specialization Telecommunication Network and Systems. He is primary working under building internetworks and monitoring systems. He is participating in different network projects for educational and commercial needs in Russia. Some of them: RBNET (Russian Backbone Network), that has been established to provide Internet service for science and high school community at main regions of Russia, RELARN (association of scientific and educational organizations) and others. He is working at IT business over 10 years and have a great experience of building and supporting global networks. He is also certified specialist Juniper, Cisco and Unix administration.