# Influence of Distortions of Key Frames on Video Transfer in Wireless Networks

E.S. Sagatov
Samara State Aerospace University
Moskovskoe sh. 34,
Samara, 443086, Russia
Email: sagatov@ya.ru

A.M. Sukhov
Samara State Aerospace University
Moskovskoe sh. 34,
Samara, 443086, Russia
Email: amskh@yandex.ru

P. Calyam
Ohio Supercomputer Center/OARnet
1224 Kinnear Road, Columbus,
OH 43212, USA
Email: pcalyam@osc.edu

*Abstract*—In this paper, it is shown that for substantial increase of video quality delivery in wireless networks, it is necessary to execute two important enhancements to existing communication schemes: (i) the video player on the receiver side should selectively discard duplicated RTP packets, and (ii) server of streaming video should duplicate the packets containing the information of key frames. Coefficients of the mathematical model to assess video quality have been found for WiFi and 3G standards-compliant wireless networks that employ MPEG-2 and MPEG-4 (DivX) codecs. We also present a novel experimental technique that enabled us to collect and process the quantitative datasets used in our modeling study.

**Keywords:** VVoIP, wireless networks, streaming video.

## I. INTRODUCTION

Mobile telecommunication solutions are increasingly being adopted in video content consumption amongst diverse user communities. Modern cellphones possess the same capabilities as that of traditional desktop computers. Also, today's laptops and netbooks provide a high degree of mobility by leveraging wireless networking capabilities. Owing to their compactness and integration with 3G, WiFi, and WiMAX network adapters, we are experiencing pervasive video communications as and when needed. According to data published by Cisco VNI [1], the Internet traffic on wireless devices has annual growth of more than 250%. By 2013 the volume of such traffic will increase by 66 times in comparison to the trends in 2008 and will make up 54% of all IP-traffic in the Internet. By 2013 various video content will make up for 64% of all traffic of wireless networks in the world.

There are several challenges for delivering scalable video services in wireless networks. The video communication quality is frequently unsatisfactory. High degradation levels due to packet losses and delay variations (i.e., network jitter) in wireless networks affect video frames transmission, and make them susceptible to distortions that adversely impact user-perceived experience. Packets of streaming video experience drastic changes in sequence numbers or are lost during transfers on network paths because of considerable packet delay variations. At the receiver-side video frames, distortions manifest as plural artifacts, lack of lip synchronization, and even frame freezing. The effects of degraded network conditions on perceived video quality due to network congestion and last-mile access-network bottlenecks are presented in [2] and [3].

The technique of certification of packet networks in recommendation IETF RFC-2544 [4] defines the following key parameters of network quality: available bandwidth, packet delay, network jitter, number of the lost packets, quantity of packages with errors. Subjective testing to assess video quality corresponding to given network quality is popularly done using the Mean Opinion Score (*MOS*) method. This method involves showing video frames received after transfer on a network path to a commission of experts who put down scores on a scale of 1 to 5 based on their impressions of video quality. The initial clip is encoded using one of codecs amongst MPEG-2, MPEG-4 or Windows Media Video 9 and is passed through a wireless network supporting standards such as WiFi, 3G or WiMAX. There are many recommendations from the International Telecommunication Union (ITU) [5] for presentation of video sequences to human subjects and gathering perceived quality estimates.

In this paper, we address the problem of adaptation of modern coding algorithms for transfer of video frames over wireless networks such as 3G, WiFi and WiMAX [6] that possess characteristics which cause distorted receiver-side video frames. In papers [7] and [8], it has been shown that subjective scope of video quality can be mapped to network parameters with gradation levels of Good, Acceptable and Poor (GAP). Such mapping [9] helps to understand the qualitative dependences arising at network translation of video frames, and to develop numerical models of user perceived video quality. By building upon findings from these prior works, we seek to find numerical dependence of video quality expressed as MOS for a given set of network quality parameters. A work that is similar to our efforts is in [10], where video distortion due to packet loss is estimated using a loss-distortion model. The loss-distortion model uses online packet loss measurements and takes into account other inputs such as video codec type, coded bit rate and packetization to estimate online relative-PSNR values. The novelty in our approach is that the specified dependence is described by a simple mathematical model that allows us to compare numerical values of co-efficients in the context of wireless networks. On the basis of similar comparisons, we find not only the most essential factors influencing quality of video, but also are able to compare them for various codecs.

Another notable contribution of this paper is our efforts to characterize distortions which damage "key" frames as well as the "usual" video frames. Key frame is the frame which bears in itself the full information on the video image and can be restored without reinforcement with additional data. Usual frame is one that codes difference between the previous frame and the current frame. Degree of key frame compression is less than that of the usual frame, and also the key frame sizes are several times higher than the usual frame sizes. We present quantitative comparisons of the influence of the errors impacting key frames when employing MPEG-2 and MPEG-4 codecs.

This rest of this paper is organized as follows: Section 2 describes the premises for mathematical modeling of dependence of video quality from characteristizations of wireless network connections. In Section 3, we explain our experiments planning. Details of our novel method to process experiment datasets are presented in Section 4. Section 5 contains the numerical results obtained from experiment datasets analysis, and the parameters of proposed mathematical model to assess video quality in wireless networks. Section 6 concludes the paper.

## II. PREMISES FOR VIDEO QUALITY MODELING

The communication quality of video frames worsens depending on the level of degraded characteristics of network connections. In order to assess the quality of transferred video based on network parameters, we have earlier proposed the idea of an universal function $Q_{MOS}(p, j, D, B)$, which outputs video quality as per a *MOS* scale [11]. This function can be expanded in a power series on network parameters, and thus can be expressed using linear co-efficients.

For fixed speeds of video streams, it is enough to consider only a linear dependence from two parameters (losses of packets and network jitter):

$$Q_{MOS} = Q_{ideal} - \alpha p - \beta j, \qquad (1)$$

where

- $Q_{ideal}$ - maximum quality of video for a given codec; scale is from 0 to 5;
- $p$ - packet loss, %;
- $j$ - network jitter (delay variation), *sec*;
- $Q_{MOS}$ - video quality on received side, scale is from 0 to 5;
- $\alpha, \beta$ - coefficients of model which can be found experimentally.

The uniform video sequence which was compressed by codecs MPEG-4 (DivX), MPEG-2 and Windows Media video 9 with constant bitrate 256 Kbps has been developed for experimental tests.

The basic aim of our research is to reveal the influence of key frames on user perceived quality of receiver-side video. For this purpose, we separately calculate coefficients $\alpha^k$ and $\beta^k$ for a stream with damage of key frames, and also $\alpha^w$ and $\beta^w$ for sequences without damage of key frames. The $\alpha_{DivX}^k$ is the designated coefficient that characterizes deterioration of

video encoded by MPEG-4 (DivX) if a key frame is damaged. Further, the coefficients $\alpha_{DivX}^k$ and $\alpha_{DivX}^w$, as well as $\alpha_{Mpeg2}^k$ $\alpha_{Mpeg2}^w$ will need to be found to numerically compare the video deterioration levels.

## III. EXPERIMENTS PLANNING

For determining the coefficients shown in Equation (1), we have designed and conducted a number of experiments. Videos files encoded by MPEG-4 (DivX), MPEG-2 and Windows Media Video 9 were transferred to a notebook connected to wireless networks that are WiFi, WiMAX and 3G standards-compliant. We recorded received video at the notebook to files, and the network traffic at packet level was simultaneously captured using the Wireshark network packet sniffer tool. Thus, using received video files, it was possible to find video quality on a MOS scale. In addition, using the packet captures, the parameters of the network connections were determined.

The software packages we utilized for carrying out the data capture and analysis in our experiments are as follows:

1) VLC media player [12] was used for both the video server and the video player with inherent capability to record receiver-side video onto a file.
2) The Wireshark Network Protocol Analyzer [13] was used to capture all network traffic.
3) VirtualDub [14] was used to conduct frames analysis of the receiver-side video and for calculation of the MOS rankings.
4) AviSynth 2.5 [15] helped in the parsing of the WMV video in the VirtualDub program. The WMV codec works only using the DirectShow technology, and cannot be directly opened using VirtualDub that uses VFW (Video For Windows) technology.

We have made all the video sequences (original as well as degraded) and network packet captures collected during our experiments openly available at our Internet TV ltd website [16].

Experiments were conducted in a similar manner over the Wi-Fi, WiMAX and 3G network scenarios, with the difference being only with the corresponding network equipment that had to be used. The Wi-Fi experiments were carried out on a wireless LAN at the Samara State Aerospace University. We also used the Samara segments of all-Russian operators: the Megaphone (3G), a Beeline (3G) and Metromaks (WiMAX) in our experiments. We prepared a test video clip for our experiments featuring various types of the activity levels: static, weak movement, fast movement, and brightness change. Further, we encoded this video clip using MPEG-4 (DivX), MPEG-2 and Windows Media Video 9 codecs. Following are the video encoding configurations that were used:

- Video Resolution 320 x 240 *pixels*
- Frame Rate - 24 *Frames per second*
- Bitrate - 256 *Kbps*
- Maximal quality

## IV. NOVEL PROCESSING METHOD

Our novel method to process the receiver-side video files and the corresponding network packet traces during degraded quality scenarios is described below:

1) The Program VirtualDub is used to open a recorded video file in order to calculate the duration of video frames distortion. For example, if frame 144 is found to be the first distorted frame in a sequence, then the previous to the distorted frame i.e., 143 is searched and its display time say, 5923 millisecond *ms* is noted. Similarly, the last distorted frame is searched, say it is 171 frame with display time 7083 *ms* is noted. Then, the duration of distortion consists of 28 frames and 1160 *ms*.

2) In the network sniffer Wireshark, the corresponding logs (http://stream.ip4tv.ru/wireless/WiFi/test1/1divx.pcap) are loaded. For RTP packets, the necessary stream is selected for more detailed analysis using the built-in Wireshark RTP streams analyzer. In the analyzer, the resulting list of all packets with red labels to indicate lost packets are flagged. From column sequence numbers, we can visually see for example if packet number 30195 and packet number 30198 are contiguous and marked with red labels, based on which we can conclude that two intermediate packets have been lost.

3) One of the notable challenges is the analysis to correlate files of network packet traces with recorded video files. For this purpose, in a network packet trace, packet is searched after which there was an error, and the corresponding timestamp is used to determine the degradation point in the video sequence.

4) For reception of authentic statistical calculations it was determined that the length of sequence for the analysis should be multiple of 100 packets. Based on this idea of using 100 packet sequences, we calculate the packet loss percentage values.

5) The statistical data relating to inter-packet intervals and network jitter are recorded in the Wireshark RTP analyzer.

Other major task we encountered when processing the experimental results relates to the estimation of video quality on the MOS scale. Our algorithm of MOS estimation is as follows:

1) Labels in program VirtualDub are installed on the first and last distorted video frames.

2) The distorted video can actually be viewed in some cases and hence can be compared to the original.

3) Quality of the video at the time instant of an error on scale MOS from 1 to 5 is visually assessed.

In this paper, the assessments of MOS were taken from 4 human subjects for each error condition, and their average MOS or arithmetic mean value of assessments were calculated. We note that - while processing video files encoded by WMV9 codec, we faced some difficulties. Unfortunately, program VirtualDub does not allows us to distinguish key frames of video encoded by the WMV9 codec. Hence, results for the WMV9 codec are not presented.

## V. EXPERIMENT RESULTS ANALYSIS

Obtained data has been handled by the technique described in the previous section. All errors, both at video level, and at the network layer have been parsed. Subjective quality of video $Q_{MOS}$ depending on packet loss $p$ and network jitter $j$ have been formulated. The received values of coefficients are gathered in Tables I and II.

We found that video quality depends not only on percent of packet losses and network jitter, but also from frame type on which there was an error. Key frame carries in itself the complete information on the video and can be restored without using additional data, and usual frame is one which encodes the difference between the previous frame and current frame progressively. Accordingly, if the error damages a key frame, video quality worsens more strongly in comparison with a similar error in a usual frame. Therefore in Tables I and II, we specially selected two types of coefficients - with losses on key frames and without them.

In Tables I and II, $\alpha^k$ and $\beta^k$ are coefficients of our model with losses of packets on key frames, $\alpha^w$ and $\beta^w$ are coefficients for the intact key frames, and $Q_{ideal}$ is an estimation on scale *MOS* for a source file (before transfer on a network).

Initially for the video encoded by MPEG-4 (DivX), quality is above the value that of video encoded by MPEG-2. However, in cases of deterioration of characteristics of network quality, it decreases considerably. And, at the high levels of network interference, MPEG-4 (DivX) quality becomes similar to the quality of MPEG-2 encoded video. In case of damage of a key frame, the video quality when encoded using MPEG-4 (DivX) will fall by more than 2 times at the same characteristics of a network, whereas MPEG-2 falls by 3.75 times.

Thus, for substantial increase of video quality during transmission on a wireless network, it is necessary to execute two important enhancements to existing communication schemes:

1) The video player on the receiver side should selectively discard duplicated RTP packets

2) The server of streaming video should duplicate the packets containing the information of key frames

Further, we observed that the period between key frames cannot exceed 2 seconds (optimally 1 second) to avoid video quality degradation.

These simple measures will lead to even poor grade [7] networks delivering video quality with MOS estimates higher than 3.5. It is necessary to notice that for codecs MPEG-4 (DivX) and MPEG-2, the transferred size of the traffic will increase by 7-10%, and quality of video grows not less than 2 times. Some of the video players we researched playback MPEG-TS MPEG-4 streams by automatically discarding duplicate packets.

The experiments on WiMAX network have shown good grade [7] performance in terms of network health. WiMAX

TABLE I
VALUES OF COEFFICIENTS OF MODEL FOR CODECS MPEG-2, DIVX IN WIFI NETWORKS

| $N$ | Codec | $Q_{ideal}$ | $\alpha^k$ | $\beta^k$ | $\alpha^w$ | $\beta^w$ |
|---|---|---|---|---|---|---|
| 1 | MPEG-2 | 4.2±0.2 | 0.15±0.03 | 0.011±0.002 | 0.04±0.01 | 0.003±0.001 |
| 2 | DivX | 4.7±0.2 | 0.27±0.05 | 0.013±0.003 | 0.13±0.02 | 0.01±0.002 |

TABLE II
VALUES OF COEFFICIENTS OF MODEL FOR CODECS MPEG-2, DIVX IN 3G NETWORKS

| $N$ | Codec | $Q_{ideal}$ | $\alpha^k$ | $\beta^k$ | $\alpha^w$ | $\beta^w$ |
|---|---|---|---|---|---|---|
| 1 | MPEG-2 | 4.2±0.2 | 0.005±0.002 | 0.005±0.002 | 0.004±0.001 | 0.003±0.001 |
| 2 | DivX | 4.7±0.2 | 0.01±0.003 | 0.003±0.001 | 0.002±0.0005 | 0.002±0.0008 |

networks tend to have the characteristics that are comparable to wired networks such as Ethernet. It is very difficult to find network errors in practice since the loss percent in all our tests was found to be near 0%, and a variation of delay was on the order 20 *ms* even in the tests that were run during large amounts of competing traffic.

3G networks are very sensitive to external interferences and even at good grade performance levels of signals, there is considerable packet loss in the ranges of poor grade, and delay variations of the order of 40 *ms* that falls in the acceptable grade. Also, it has been found in the case of one of the 3G providers, the equipment frequently duplicated outgoing packages that created repeated video artifacts at the receiver-side, which existing video players are unable to compensate.

## VI. CONCLUSION

There is an increasing demand for voice and video applications on wireless devices due to the recent developments in smart phones and tablet PCs. To cater to this demand efficiently and reliably, it is important for the application developers and service providers to characterize and tune the performance of RTP streams that deliver the content. In addition, the area of video performance measurement is in its early stages, and developing effective techniques to measure video quality is vital. In this paper, we have addressed both the above dimensions of requirements. We show scenarios where RTP packets are duplicated and evaluate how the video player should handle these duplicated packets. We also show a novel experimental technique to identify key frames in video, and evaluate how redundancy of those key frame packets can improve video quality.

A significant contribution of our work is the development of a mathematical model for estimating video quality of codecs MPEG-2 and MPEG-4 (DivX) codecs used in open-source VLC on WiFi and 3G standards compliant wireless networks. Our model development has been done with due considerations given to the handling of duplicate packets, and adding redundancy to key video frames.

The video quality measurement experimental technique and research findings presented in this paper have been implemented in our Internet video broadcasting efforts at SSAU, Togliatty branch SSAU and Internet TV service. Owing to these implementations, we have observed that our Internet video broadcasting service offering has become more automated, predictable and our service has seen notably lowered operation costs.

## REFERENCES

[1] Cisco Systems, Cisco Visual Networking Index: Forecast and Methodology, 2008-2013, 9 June 2009 (http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf)
[2] X. Hei, C. Liang, and et. al., A Measurement Study of a Large-Scale P2P IPTV System, IEEE Trans. on Multimedia, vol. 9, 8, 2007.
[3] Y. Won and M. Choi, End-User IPTV Traffic Measurement of Residential Broadband Access Networks, in IEEE NOMS, 2008.
[4] Bradner S., McQuaid J., RFC2544 - Benchmarking Methodology for Network Interconnect Devices, 1999.
[5] International Telecommunications Union, Methodology for the subjective assessment of the quality of television pictures (ITU-R BT.500-11), 2002.
[6] Haghani, E. De, S. Ansari, N., VoIP Traffic Scheduling in WiMAX Networks, in: IEEE Global Telecommunications Conference, 2008.
[7] Calyam P., Sridharan M., Mandrawa W., Schopis P., Performance Measurement and Analysis of H.323 Traffic, Passive and Active Measurement Workshop (PAM), 2004.
[8] Claypool M., Tanner J., The effects of jitter on the peceptual quality of video, Proceedings of the seventh ACM international conference on Multimedia (Part 2), p.115-118, 1999.
[9] T. Silverston, O. Fourmaux, A. Botta, A. Dainotti, A. Pescape, G. Ventre, K. Salamatian, Traffic analysis of peer-to-peer IPTV communities, Computer Networks, Vol. 53, pp. 470-484, 2009.
[10] S. Tao, J. Apostolopoulos, R. Guerin, Real-Time Monitoring of Video Quality in IP Networks, IEEE Trans. on Networking, Vol. 16, No. 5, pp. 1052-1065, 2008.
[11] Sukhov A., Calyam P., Daly W., Ilin A., Towards an analytical model for characterizing behavior of high-speed VVoIP applications, Computational Methods in Science and Technology, 11(2), pp. 161-167, 2005.
[12] VideoLAN team, VideoLAN, Free streaming and multimedia solutions for all OS! (http://www.videolan.org/)
[13] Wireshark Packet Capture Tool, (http://www.wireshark.org/)
[14] Avery Lee, Virtualdub.org, (http://www.virtualdub.org/)
[15] Main Page - Avisynth. (http://avisynth.org)
[16] Traces of video streams in WiMAX, 3G, and WiFi standards-compliant wireless networks, (http://www.ip4tv.ru/stati/aaa.html)