

# Impact of packet loss and delay variation on the quality of real-time video streaming

Jaroslav Frnda<sup>1</sup> · Miroslav Voznak<sup>1</sup>  · Lukas Sevcik<sup>1</sup>

Published online: 17 April 2015  
© Springer Science+Business Media New York 2015

**Abstract** The aim of this work is to bring complex view on video streaming service performance within IP-based networks. Video quality as a part of multimedia technology has a crucial role nowadays due to this increase. Since architecture of IP network has not been designed for real-time services like audio or video, there are many factors that can influence the final quality of service, especially packet loss and delay variation (also known as Jitter). The research was focused on the quality of video data delivery in many scenarios included different packet loss rate and simulating of different delay variation values in the network. Performed tests were evaluated by using of video objective methods. Based on results of these measurements, an extended QoS model for estimation of triple play services was designed. The proposed model allows us to compute the estimated objective quality parameters that describe the final quality of video service as a part of triple play services.

**Keywords** Delay variation · Packet loss · Prediction model · VQM · SSIM · Video quality assessment

## 1 Introduction

The growth of the Internet network has used more and more resources for performance analysis. This work compares the performance of the network from an experimental viewpoint. The one of the objectives of this work is to analyse the various impacts of transmitted video, such as packet loss, jitter, and packets reordering. The results should show if it is better (for video quality) to get packets in a wrong order or completely lost. Comparison includes the static and dynamic video sequences, different quality and finally the two currently most used streaming video codecs. In the work, we deal with objective methods for evaluation of the quality of video sequences. Subjective assessment methods are very demanding regarding the time and people, and comparison of each other is not a simple process.

The objective methods are commonly preferred and offer the results that can be used for immediate comparison and verification. Each used objective method has a different procedures and different metric evaluation system. The obtained results were used for an implementation of prediction model, able to compute the final quality of video service according to the network behavior. This model can be useful for internet service providers during the process of network architecture designing.

## 2 State of the art

The recently growing interest in real-time service (such as audio and video) transfer through packet networks based on IP protocol has led to analyses of these services and their behavior in such networks becoming more intensive. Logically, the greatest emphasis is being put on the transfer of voice since this service is the most sensitive to the overall net-

---

✉ Miroslav Voznak  
miroslav.voznak@vsb.cz  
Jaroslav Frnda  
jaroslav.frnda@vsb.cz  
Lukas Sevcik  
lukas.sevcik.st1@vsb.cz

<sup>1</sup> Department of Telecommunications, VSB – Technical University of Ostrava, 17. listopadu 15/2172, 70833 Ostrava, Czech Republic

work status. Packet networks based on IP protocol had not been designed to transfer delay-sensitive traffic and without any supplementary mechanisms securing the quality of service, such a transfer was not capable of providing a high-quality interactive communication similar to standard PSTN (Public Switched Telephone Network). Lack of synchronization in comparison to TDM (Time-division Multiplexing) presented in PSTN brings concerns about variable conditions on the network that cause packet loss and fluctuation in delays (jitter). Jitter causes excess packet loss on receiving buffers depending on the buffer size and delay variance [1]. These two mentioned factors (packet loss and overall delay) have significant influence on the final quality of service. The research is focused on developing new codecs and techniques that are able to eliminate the packet loss factor on voice clarity (e.g. Packet Loss Concealment in G.711), and voice flow prioritization in the network that reduces transmission time [2, 3].

But, on the other hand, video has become the major part of all data traffic sent via IP networks. In general, a video service is one-way service (except e.g. video calls) so network delay is not such an important factor as in voice service. Dominant network factors that influence the final video quality are especially packet loss, delay variation and the capacity of the transmission links [4]. Analysis of video quality concentrates on the resistance of video codecs to packet loss in the network, which causes artefacts in the video [5, 6]. The new video codecs like VP9 (Google) and H.265 (The Moving Picture Experts Group-MPEG and ITU-T) that were released in December 2012, respectively April 2013 started the process of comparison and evaluation with their predecessors. The papers brought many performed tests, which showed the suitability of codecs using. The newest video codecs offer higher efficiency compression or in other words - better quality with the same video bitrate used [7]. However, new codecs need more computing performance, and this is the main reason why they are used for a very high resolution such as Full HD ( $1920 \times 1080$ ) and 4K ( $4096 \times 2160$ ) [8]. Streaming video as a part of triple play services will still be mainly ensured by older video codecs like MPEG-2 and MPEG-4 (H.264).

Nevertheless, a few factors are still lacking, such as complex view of video parameters on the final video quality. In our previous works [2], we focused on the quality of the triple play services prediction model implementation, where one part was dedicated to the quality of the video service.

The main motivation behind this work is to extend the mentioned computational model and bring a comprehensive view of all video parameters like codec type, character and resolution, and their influence on negative network factors resistance. In addition to packet loss, we focused on another network disruption phenomenon called delay variation (also known as jitter). This phenomenon is very often overlooked

due to de-jitter buffer implementation on the receiving side, but for better process of the network situation modeling and prediction, it is good to know how it influences the final video quality.

## 3 Methodology

### 3.1 Video processing

The volume of digital video data is usually described in the terminology of bandwidth or transfer rate. Bandwidth of a classical digital video transmission without compression is up to hundreds of Mbps. The amount of data of the picture signal is higher with an increase in resolution. The volume of data is a major problem in the transmission, processing, storage and display of video information. Digital video compared to static images is very sensitive to memory needed saving [9].

Standard television broadcasting has a frame rate of at least 25 fps [10] (images per second). It is sufficient for the delay in perception of the human eye. Every second of the movie at a resolution 1080p (Full HD) of uncompressed video can take up to tens of megabytes of memory. Video typically contains a large amount of redundant data. Those can be removed using the appropriate compression algorithms [9].

MPEG-2 is one of the most used compression standards. It was approved in 1994. MPEG-2 is built on the MPEG-1, and its video coding scheme is a refinement of MPEG-1 standard. The advantage of the MPEG-2 standard is that it is suitable for coding both progressive and interlaced video. Much functionality such as scalability has been introduced. MPEG-2 also defines the Profiles and Levels. The Profile describes a degree of functionality whereas the Level represents resolution and bitrates. But not all Levels are supported at all Profiles. The most important application of the MPEG-2 standard is digital television broadcasting (DVB-T, DVB-S, DVB-C) but it also specifies the format of movies and other programs that are distributed on DVDs and similar disks [5, 10].

The latest and today most used compression standard designed for a wide range of applications, ranging from mobile video to HDTV, is MPEG Part 10, also called

MPEG-4 H.264/AVC. Some of the feature enhancements in MPEG-4 H.264/AVC standard over the earlier codecs are e.g. redundant pictures, multiple reference frames, arithmetic variable-length coding or motion compensation blocks being of variable sizes. MPEG-4 H.264/AVC defines the Profiles and Levels, too, but its organization is much simpler than in MPEG-4 Part 2. There are only three Profiles currently defined (Baseline, Main, Extended) [5, 10, 11].

### 3.1.1 Group of pictures (GOP)

Very important factor that also influences the video quality is the frame type. There are three defined types of frames: I, P, B.

I (intra) frames are coded without reference to other frames (without any motion-compensated prediction), in a very similar manner to JPEG (Joint Photographic Experts Group—lossy compression for digital image), which means that they contain all the information necessary for their reconstruction by the decoder. For this reason, they are the essential entry point for access to a video sequence. An I frame is used as a reference for further predicted frames (P and B). The compression rate of I frames is relatively low.

P (predicted) frames are inter-coded using motion compensated prediction from a reference frame (the P frame or I frame preceding the current P frame). Hence, a P frame is predicted using forward prediction, and a P frame may itself be used as a reference for further predicted frames (P and B frames). The compression rate of P frames is significantly higher than of I frames.

B frames are inter-coded using motion-compensated prediction from two reference frames, the P and/or I frames before and after the current B frame. Two motion vectors are generated for each macroblock in a B frame—one pointing to a matching area in the previous reference picture (a forward vector) and one pointing to a matching area in the future reference picture (a backward vector). A motion-compensated prediction macroblock can be formed in three ways—forward prediction using forward vector, backwards prediction using backward vector or bidirectional prediction (where the prediction reference is formed by averaging forward and backward prediction references). Typically, an encoder chooses the prediction mode (forward, backward or bidirectional) that gives the lowest energy into the difference macroblock. B frames offer the highest compression rate.

All these different frame types (I, P, B) are then grouped together to a sequence (specific repeating order)—called the Group of Pictures (GOP). A GOP must always start with an I frame and can contain only I or a combination of I and P or I, P, B frames. The use and also a number of B or P frames within a GOP can be increased or decreased depending on image content, compression rate or application that the compressed video is intended for. Various GOP lengths and combinations of P and B frames can be encoded, but mostly a typical GOP pattern is used—IBBPBBPBBPBBI—where each letter represents viewing order and type of the frame. [9, 10, 12].

### 3.1.2 Video transmission

To transfer video files, whether in real or system time (Video on Demand), fundamentally unreliable protocols are used.

The principle consists of sending and receiving data without feedback between the sender and the recipient. Factors affecting the video transmission are [2, 9]:

- Latency—this is the time that elapses between sending a message from the source and adoption of the destination node.
- Packet order—variability in the packet delivery time to the destination node causes incorrect order.
- Packet loss—occurs when one or more packets of data travelling across a computer network fail to reach their destination. It is most often expressed as a percentage.
- Bandwidth—this expresses the capacity of the transmission channel.
- Delay—This is caused by overcrowding the packet queue on the outgoing interface.

Network factors like Latency and Delay are necessary for real-time services such as voice over IP (VoIP), where the communication is two-way at the same time. Video broadcasting is typically only one-way service (direction from content provider to end user), so that is the reason why we have focused mainly on packet loss and delay variation as the primary network impairments on final service quality.

### 3.1.3 Methods for evaluating the quality

In this work, we used two the objective methods—VQM and SSIM. Objective evaluation metric involves the use of the metric's computational methods, which form a “score” of the quality of the investigated video. These methods measure the physical characteristics of the video signal, such as the amplitude, timing and signal-to-noise ratio.

The SSIM (Structural Similarity Index) index is a method for measuring the similarity between two images. The SSIM index and VQM are full reference metrics; in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference. SSIM is designed to improve the traditional methods like peak signal-to-noise ratio (PSNR) and mean squared error (MSE), which have proven to be inconsistent with the human eye perception. SSIM considers image degradation as perceived change in structural information. Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close. These dependencies carry valuable information about the structure of the objects in the visual scene. This index includes three components - the similarity of the intensity, the corresponding contrast and the corresponding structure. The combination of these three factors forms one value as shown in Fig. 1.

This method differs by evaluating structural distortion and not error rate. The main reason for this difference is characteristic of the human visual system. Since the SSIM method

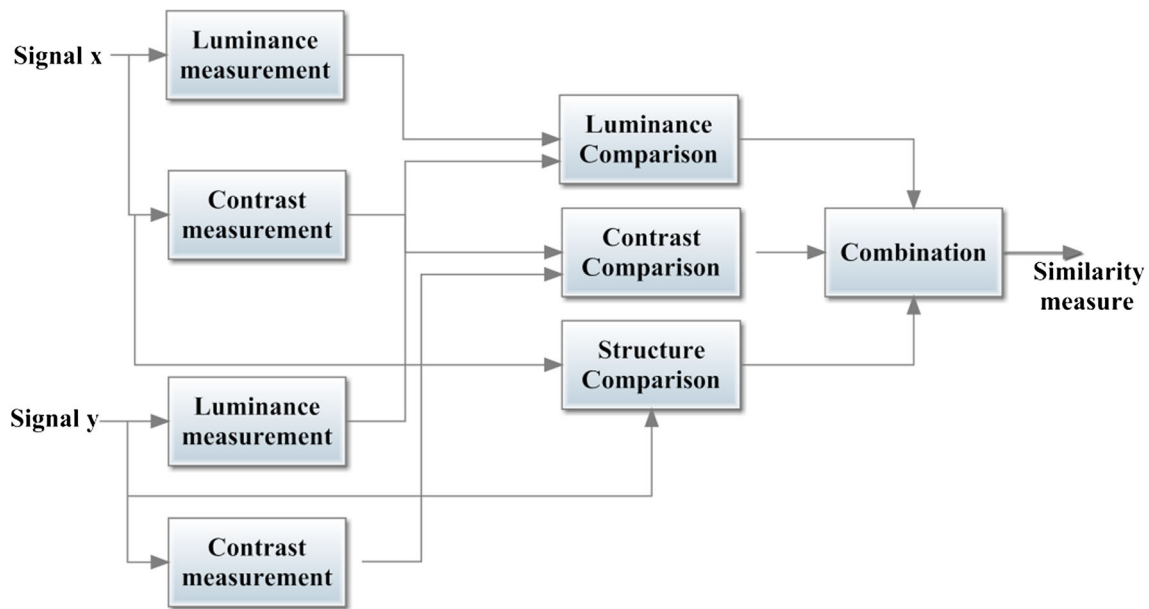


Fig. 1 The block diagram of SSIM index metric [13]

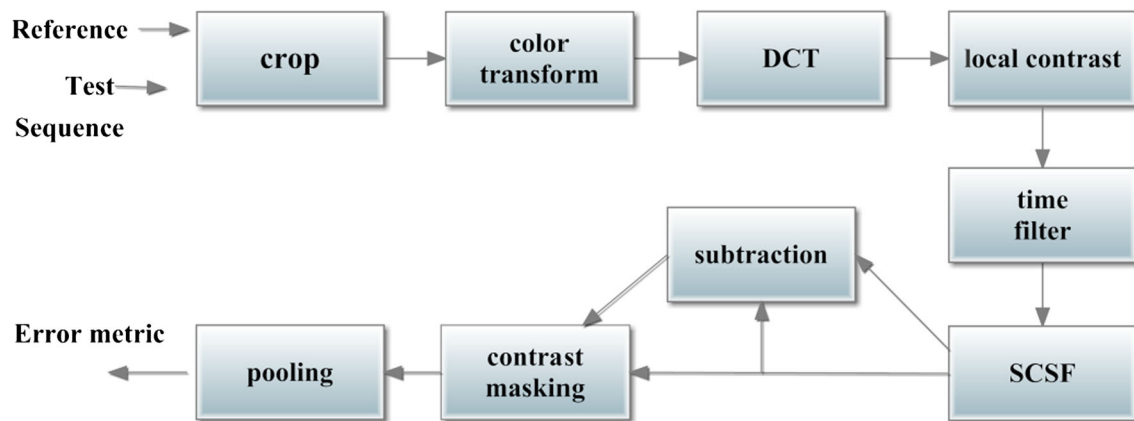


Fig. 2 The block diagram of VQM metric [12]

achieves a good correlation to the subjective impression, rating is defined in the interval [0–1], where 0 represents the worst value and 1 the best one (identity) [2]. Final SSIM value is a combination of three parameters, with original signal  $x$  and encoded signal  $y$  being defined as follows [2, 10, 13]:

$$\text{SSIM}(x,y) = [l(x,y)]^\alpha [c(x,y)]^\beta [s(x,y)]^\gamma \quad (1)$$

- Element  $l(x,y)$  compares the brightness of the signal
- Element  $c(x,y)$  compares the contrast of the signal
- Element  $s(x,y)$  measures the structure of correlation
- $\alpha > 0$ ,  $\beta > 0$ ,  $\gamma > 0$  measures the weight of individual elements

The VQM (Video Quality Measure) metric computes the visibility of artefacts expressed in the DCT domain (Discrete

Cosine Transform—express a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies). Figure 2 shows the block diagram of this metric, which can be divided into nine steps.

The input of the metric is a pair of color image sequences—the reference one and the test one. Both sequences are cropped then transformed to blocked DCT and afterwards converted to units of local contrast. In the next step, the input sequences are subjected to temporal filtering, which implements the temporal part of the contrast sensitivity function. The DCT coefficients expressed in a local contrast form, are then converted to just-noticeable-differences (JNDs) by dividing according to their respective spatial thresholds. This implements the spatial part of the contrast sensitivity function. In the next step, after the conversion to JNDs, the two sequences are subtracted to produce a difference sequence.

Fig. 3 Measurement procedure

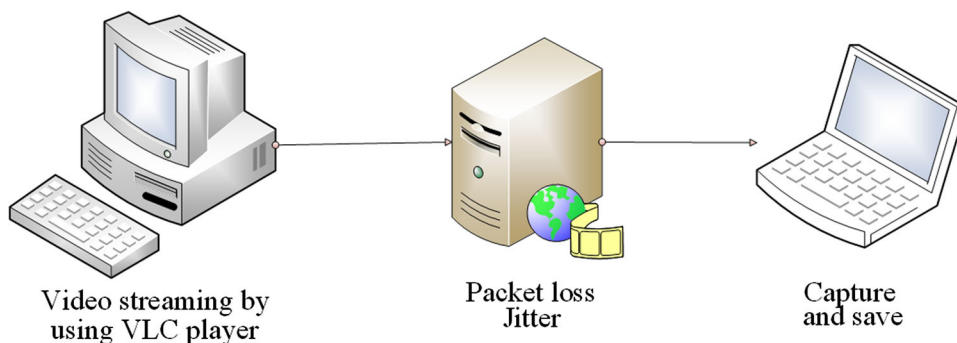


Table 1 Parameters for measurements

Used codecs	MPEG-2, MPEG-4(H.264)
Video resolution (pixels)	720 × 576, 1280 × 720, 1920 × 1080
Evaluation methods	SSIM, VQM

In the following step, the contrast masking operation to the difference sequence is performed. Finally, the masked differences are weighted and pooled over all dimensions to yield summary measures of visual error [12]. The output value of the VQM metric indicates the amount of distortion of the sequence—for no impairment the value is equal to zero and for rising level of impairment the output value rises [12].

### 3.2 Video quality evaluation

The aim of the measurement was to simulate the effect of packet loss and jitter for the video formats MPEG-2 and MPEG-4, to determine the impact on the resulting image using objective methods for measuring the quality of the video and comparing the results. We made measurements for one static and two dynamic videos of 25 s. All movies were measured in a resolution of 720 × 576 (PAL), 1280 × 720 (HD) and 1920 × 1080 (Full HD). Static video was represented by TV news (slow motion), the first dynamic video by a NASA space shuttle launch and the third video with the highest bitrate (70 Mbps in full HD) was an open source animation movie for testing purposes called Big Buck Bunny.

The whole process of measuring is shown in the Figs. 3 and 4. To evaluate the quality we used the methods SSIM and VQM. SSIM correlates better with the perception of the human eye [10, 11]. We evaluated these methods using MSU VQM Tools. As a first step, we created a stream in the VLC Player. As for the video content, streaming process RTP/UDP/IP method with MPEG2 (TS) and H.264 (MP4) was used (Table 1).

We captured the broadcast stream on the local computer interface using VLC Player. This video was saved and tagged as the original video. Our testing scenarios reflect the situations that can happen in the real network. Especially

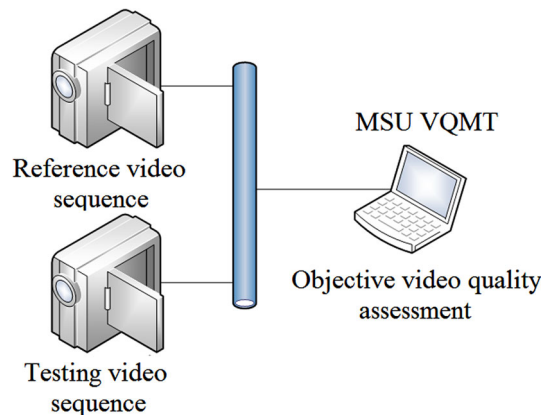


Fig. 4 Comparing tested sequence with original video

mobile networks, capable of using IP architecture like UMTS (The Universal Mobile Telecommunications System—3rd generation mobile cellular system) and LTE (Long-Term Evolution—4th generation), reach high values of packet loss and delay variation [14]. For the purpose of settings in our scenarios, we used Linux tool called Netem. Netem provides Network Emulation functionality for testing protocols by emulating the properties of wide area networks. The current version emulates variable delay, loss, duplication and reordering [15].

We set the packet loss to 1 % on the interface using Netem and then repeated the whole measurement. Then we repeated this step for packet loss in increments of 1, 2, ..., 10 %.

For the purpose of delay variation (Jitter) simulation, we set that 25 % of packets are delayed (results of our previous research [2] showed that approximately 25 % of all traffic had different one-way delay). We repeated the measurements for 10, 20, 30, 50, 75 and 100 ms delay variation. By streaming videos, we set the value of the de-jitter buffer to 0 in VLC so that the indicated delays were processed without any impact of buffering.

### 3.3 Netem settings

```
#tcqdisc add dev eth0 root netem loss 1% // add packet loss on interface
```

**Table 2** Static video measurements results

Packet loss (%)	VQM/SSIM					
	720 × 576		1280 × 720		1920 × 1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	2.62/0.914	7.82/0.805	5.53/0.942	10.14/0.742	4.87/0.905	9.96/0.774
2	4.22/0.876	9.28/0.688	7.23/0.894	10.39/0.730	8.07/0.885	9.78/0.776
3	9.20/0.877	14.70/0.52	8.29/0.839	11.10/0.709	7.77/0.876	12.11/0.673
4	11.51/0.80	15.61/0.529	9.78/0.830	13.00/0.713	8.85/0.821	14.37/0.648
5	11.30/0.77	15.99/0.514	10.61/0.743	13.64/0.555	8.26/0.820	14.49/0.450
6	12.24/0.721	16.28/0.499	10.47/0.726	15.82/0.539	8.97/0.812	15.80/0.362
7	13.33/0.669	17.45/0.486	10.40/0.707	15.52/0.534	9.64/0.800	17.91/0.255
8	13.46/0.536	18.89/0.485	11.14/0.660	17.01/0.483	9.71/0.740	18.64/0.225
9	13.41/0.481	19.70/0.463	11.10/0.658	18.61/0.444	10.14/0.569	19.13/0.205
10	14.55/0.452	21.84/0.362	13.12/0.638	20.46/0.421	12.16/0.481	20.44/0.184

**Table 3** VQM and SSIM results for first tested dynamic video

Packet loss (%)	VQM/SSIM					
	720 × 576		1280 × 720		1920 × 1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	5.05/0.864	7.03/0.777	8.18/0.859	8.67/0.736	8.76/0.915	8.83/0.817
2	5.28/0.830	8.60/0.738	9.33/0.837	9.53/0.750	8.67/0.875	10.26/0.715
3	6.69/0.746	9.69/0.692	9.71/0.789	10.35/0.745	9.06/0.809	12.03/0.715
4	7.33/0.754	9.61/0.686	10.30/0.766	10.46/0.714	9.33/0.808	12.11/0.657
5	7.46/0.716	9.97/0.683	11.16/0.752	10.40/0.699	9.45/0.781	13.32/0.649
6	9.06/0.711	10.45/0.645	11.18/0.741	10.62/0.663	9.85/0.799	15.16/0.637
7	9.33/0.706	11.19/0.644	12.80/0.717	10.76/0.582	9.97/0.764	15.16/0.632
8	9.30/0.709	11.88/0.651	12.88/0.701	11.08/0.545	10.19/0.728	15.36/0.605
9	10.05/0.693	12.03/0.622	13.49/0.694	11.41/0.502	10.60/0.703	15.35/0.604
10	10.53/0.686	12.02/0.605	14.08/0.678	11.79/0.485	11.54/0.683	15.80/0.582

**Table 4** VQM and SSIM results for the second tested dynamic video

Packet loss (%)	VQM/SSIM					
	720 × 576		1280 × 720		1920 × 1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	1.41/0.923	8.22/0.695	2.59/0.927	11.95/0.530	6.17/0.806	11.69/0.589
2	1.96/0.9	11.16/0.551	3.95/0.765	12.15/0.511	6.65/0.737	11.85/0.501
3	3.42/0.714	12.43/0.446	4.64/0.745	12.33/0.422	6.92/0.694	11.94/0.487
4	3.43/0.718	12.77/0.447	5.95/0.7	12.61/0.421	7.49/0.699	12.52/0.474
5	3.66/0.705	13.25/0.441	6.53/0.664	12.98/0.428	7.58/0.614	12.68/0.446
6	5.29/0.668	13.22/0.405	7.13/0.597	13.08/0.393	8.11/0.589	12.73/0.458
7	5.60/0.618	13.28/0.338	7.39/0.566	13.34/0.372	8.73/0.564	13.11/0.429
8	5.81/0.575	13.52/0.333	7.88/0.531	14.32/0.353	8.92/0.558	13.56/0.425
9	6.09/0.562	15.04/0.303	7.96/0.528	14.87/0.352	8.98/0.552	14.09/0.387
10	6.70/0.493	15.59/0.326	8.15/0.511	14.99/0.326	9.31/0.545	14.78/0.364

```
#tcqdisc change dev eth0 root netem loss 2 % // change packet loss on local interface
```

It causes that 2% (i.e., 2 out of 100) packets are randomly dropped. Videos for measurement were in these formats, so we did not set any additional transcoding by creating or capturing a stream.

```
#tcqdisc add dev eth0 root netem delay 10ms reorder 75 % 50 % // setting packet delay on local interface
```

In this example, 75% of the packets (with a correlation of 50%) are sent immediately, the others are delayed by 10ms. In our case, correlation 50% means that the delayed part of all data traffic is oscillated around a value of 25%. This setting simulates the network performance behavior more accurately.

```
#tcqdisc change dev eth0 root netem delay 20ms reorder 75 % 50 % // change packet delay on local interface
```

### 3.4 Evaluation of the results

We compared the original sample and the tested sample in tool MSU VQMT, which included a damage caused by our settings. For correct opening of video formats, it was necessary to use AviSynth plugin that is able to open many video formats. The using of this script is simple, an example of the application is listed below:

```
DirectShowSource("c : \folder\myclip.mpg")
```

This one line is written in a text editor and saved as "name.avi" is sufficient for the correct function in MSU VQMT. The DirectShow is a multimedia playback system from Microsoft. It can read most of the formats that Media Player can play, including MPEG, MP3, and some QuickTime files, as well as AVI files. MSU VQMT can work with AviSynth script that is responsible for playing the video; then MSU starts to compare every single frame with its reference. MSU VQMT offers many other objective video metrics that we have mentioned, but VQM and SSIM are the most commonly used and fully accepted by scientific community.

The tool exports results into the CSV format where we can find the value for every compared frame altogether with the total average value for the whole video.

## 4 Results

The results of the measurements verified our prediction that not only the type of video codec has a degradation impact on video quality. On the other hand, video resolution has not been proven as a significant parameter of video robustness.

All measured results led to the creation of video quality prediction model. Our goal was to find regressive equations with the highest R-squared factor. R-squared is a statistical measure of how close the data are to the fitted regression line. Regression calculates an equation that minimizes the

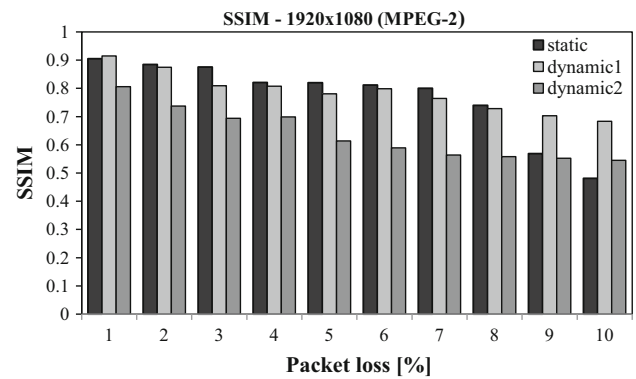


Fig. 5 SSIM results for MPEG-2

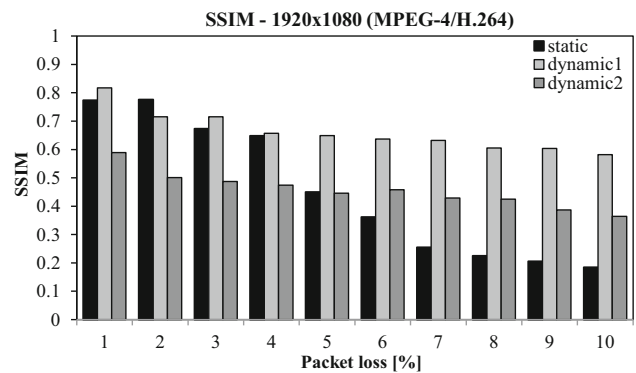


Fig. 6 SSIM results for MPEG-4(H.264)

distance between the fitted line and all of the data points. These regression equations have been calculated by statistical program Statgraphics, and as it can be seen, they have exponential character. After that, the proposal model was tested with values obtained by network simulation. Predicted values showed good correlation with measured values. The level of statistical deviation was approximately 5% and it corresponds with computed R-squared factor.

The video compression type appeared to be the most important factor. Video codec H.264 (MPEG-4 Part 10) is more prone to packet loss rate in the network infrastructure than the older MPEG-2. According to the results of the static video measurements shown in Table 2, there is no big difference between the resolutions used. We detected a slight decrease in higher resolution. During the static scene, where changes were very slow, mainly the P and B frames contained approximately the same information regardless of the resolution used [12].

Table 3 describes results of the first tested dynamic video sequence. This sequence achieved worse results than the static video. Again, the differences between the used resolutions were small. All the GOP frames contain more information, so packet loss influences more the picture reconstruction. That is the reason why dynamic video is generally more sensitive to data loss [9, 12].

**Table 5** Coefficients for static video

Coeff	MPEG-2			MPEG-4 (H.264)		
	720 × 576	1280 × 720	1920 × 1080	720 × 576	1280 × 720	1920 × 1080
a	0.89957	1.08748	0.9216	0.146704	1.1027	0.08596
b	-0.00492	-0.14397	-0.00389	0.528499	0.12312	-0.1839
α	1	0	0	1	0	0
β	0	1	0	0	1	0
γ	0	0	1	0	0	1

**Table 6** Coefficients for MPEG-2 dynamic videos

Coef.	Lower bitrate dynamic video			High bitrate dynamic video		
	720 × 576	1280 × 720	1920 × 1080	720 × 576	1280 × 720	1920 × 1080
a	0.858125	0.875076	0.927705	0.9538	0.924954	0.819482
b	-0.076882	-0.080159	-0.00389	0.094724	-0.179784	-0.1216
α	0	0	0	1	0	0
β	1	1	1	0	1	

**Table 7** Coefficients for MPEG-4(H.264) dynamic videos

Coef.	Lower bitrate dynamic video			High bitrate dynamic video		
	720 × 576	1280 × 720	1920 × 1080	720 × 576	1280 × 720	1920 × 1080
a	0.783421	1.12248	0.800991	0.678364	1.8151	00.5818
b	-0.073105	0.086301	-0.092020	-0.169198	0.122547	-0.0833
α	1	0	0	0	0	0
β	0	1	1	1	1	1

The third video has a dynamic character too, but very high bitrate when compared to the previous two videos. Obtained values from this dynamic video are listed in Table 4.

For a better illustration, Figs. 5 and 6 demonstrate the video quality results for full HD resolution.

High bitrate means a lot of information contained in I, B and P frames and its loss causes significant degradation of video quality. The conclusion of these results refers to the argument of the importance of codec type, its bitrate and character. Only better resolution does not bring significant packet loss resistance.

This paper follows on from our previous work [2] and extends the video prediction model used there. All these mentioned result were processed into the following regressive equations. As for the equations described here, X represents the packet loss in % (range 0.5–10), and all coefficients are shown in Tables 5, 6 and 7.

- Slow-motion video. MPEG-2

$$\text{SSIM} = \alpha(a + b * (X^2)) + \beta(a + b * \sqrt{X}) + \gamma(a + b * (X^2)) \quad (2)$$

MPEG-4(H.264)

$$\text{SSIM} = \alpha \left( \sqrt{a + \frac{b}{X}} \right) + \beta \left( \frac{1}{a + b * X} \right) + \gamma (\exp(a + b * X)) \quad (3)$$

All the necessary coefficients are presented in Table 5. Because measurements were performed on two dynamic videos, the following regressive equations represent the prediction model for both of them.

- Dynamic video with ordinary and high bitrate. MPEG-2

$$\text{SSIM} = \alpha \left( \frac{1}{a + b * X} \right) + \beta(a + b * \ln(X)) \quad (4)$$

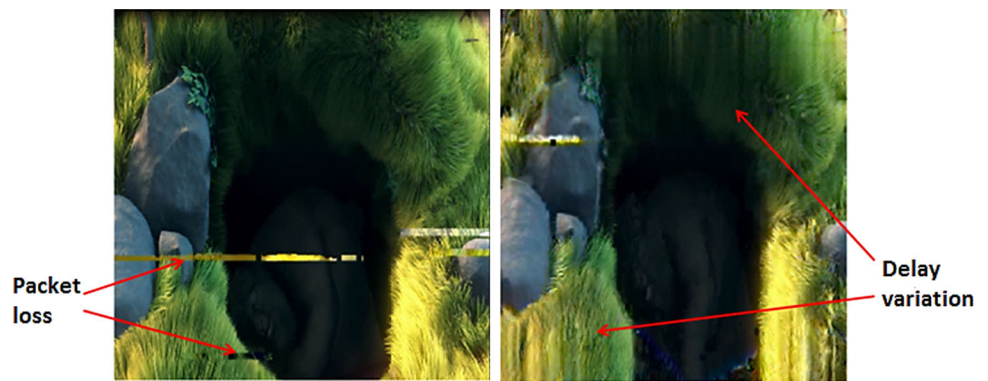
MPEG-4(H.264)

$$\text{SSIM} = \alpha(a + b * \ln(X)) + \beta \left( \frac{1}{a + b * X} \right). \quad (5)$$

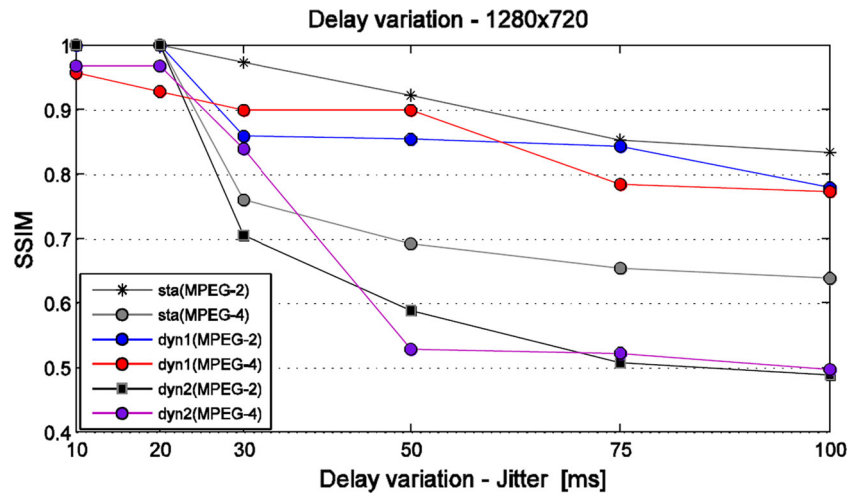
The Tables 6 and 7 contain the coefficients for these two equations. All regressive models described here have gained an R-square factor ( $R^2$ ) higher than 92 %, which represents a high level of veracity.



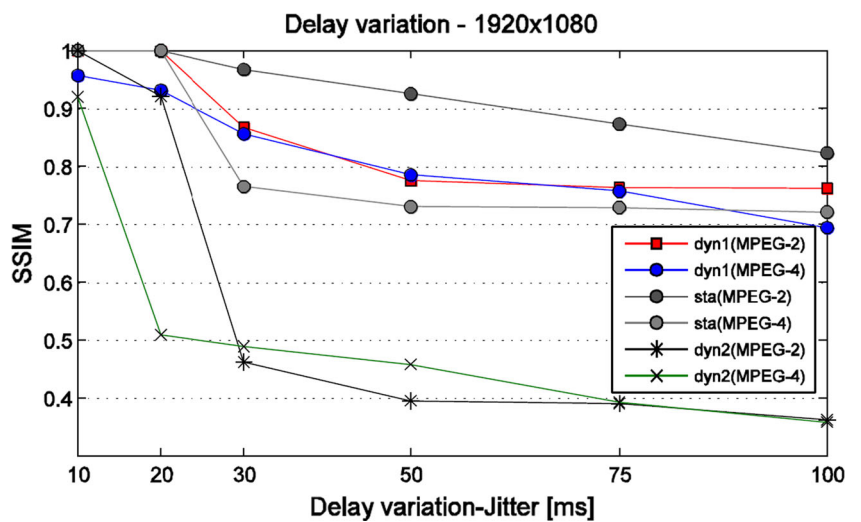
**Fig. 7** Comparison of packet loss and Jitter effect on video quality



**Fig. 8** Results of delay variation measurements—HD resolution



**Fig. 9** Results of delay variation measurements—full HD resolution



The second group of measurements led to an analysis of the degradation effect of delay variation—Jitter. The results of the performed tests showed in the Figs. 8 and 9, uncover a critical boundary of 20 ms. Above this value, a significant reduction of final video quality is observed. This boundary indicates, that if a network provider is able to ensure the Jitter not exceeding the value of 20 ms, the using of de-jitter buffer on receive side would not be necessary.

The value of 20 ms points to the typical interval of RTP transmitted audio/video datagrams. Due to the process of decompressing and video stream processing which requires a particular time, both codecs are tolerant of a small delay variation.

As can be seen in Fig. 7, packet loss causes artefacts in the image. On the other hand, delay variation influences the overall image sharpness. The video, affected by delay vari-

ation is more blurred in comparison to the video with some packets missing.

The behavior of dynamic videos is approximately on the same level, with bigger differences between MPEG-2 and MPEG-4 if the static video is used. Typically in the real operation, de-jitter buffer is used for elimination of this phenomenon. Nevertheless, our experiments showed the significant degradation impact on video quality in case of a particular delay variation.

## 5 Conclusion

The article brings detailed view of the video streaming performance over an IP-based network. The measured results showed the relation between the video codec type and bitrate to the final video quality. These results have helped us to create and extend our previous mathematical models of video streaming behavior. The second part of the measurements was dedicated to another adverse network impact on video quality called Jitter. The results proved the importance of De-jitter buffer implementation not only for voice services but also for video streaming services.

The extended prediction model of video quality described in the paper is very easy to be implemented in any programming language. Then, this model will become a very simple and fast tool for prediction of video behavior in the network. Constant network monitoring, along with network performance intervening as needed, seems to be a method for securing at least minimal QoS level in the packet network. The application aims to be a helpful tool for designing the network infrastructure with regard to achieving at least minimal QoS level.

Our future works will focus on two directions: Firstly, the new generation of video codecs such as h.265 and VP9. The used tools do not support these new codecs yet. Therefore, we will include them into our model immediately after their release as soon as possible. The second part will be focused on analysis of the impact of encryption algorithms on QoS. Security is a highly discussed topic nowadays, and protocols such as IPsec, VPN/SSL and SRTP are becoming more and more frequently used for securing of the content of voice or video. Therefore, computational mathematical models should handle this new situation.

**Acknowledgments** This work was supported by Grant of the SGS No. SP2015/82 and was partially supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070).

## References

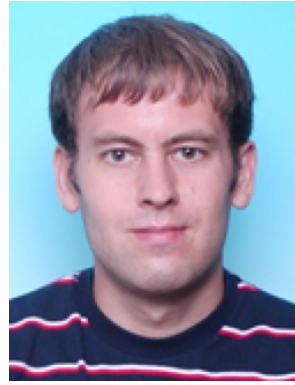
1. Voznak, M., Kovac, A., & Halas, M. (2012). Effective packet loss estimation on VoIP jitter buffer. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (Vol. 7291 LNCS, pp. 157–162).
2. Sevcik, L., Voznak, M., & Frnda, J. (2014). QoE prediction model for multimedia services in IP network applying queuing policy. In *Proceedings of the 2014 international symposium on performance evaluation of computer and telecommunication systems, SPECTS 2014—part of summerSim 2014* Multiconference, art. no. 6879998, (pp. 593–598).
3. International Telecommunications Union, ITU-T G.711, Appendix I. (1999). A high quality low-complexity algorithm for packet loss concealment with G.711.
4. Syvampasad, Ch., Vitaj, S., Reisslein, M., & Karam, L. J. (2011). Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting*, 57(2), 165–182.
5. Changhoon, Y., & Bovik, A. C. (2011). Evaluation of temporal variation of video quality in packet loss networks. *Signal Processing: Image Communication*, 26, 34–38.
6. Feamster, N., & Balakrishnan, H. (2002). Packet loss recovery for streaming video. In *12th international packet video workshop, Pittsburgh, PA*.
7. Uhrina, M., Frnda, J., Sevcik, L., & Vaculik, J. (2014). Impact of H.264/AVC and H.265/HEVC compression standards on the video quality for 4K resolution. *Advances in Electrical and Electronic Engineering*, 12(4), 368–376.
8. Kim, I. K., Lee, S., Piao, Y., & Chen, J. (2014). Coding efficiency comparison of new video coding standards: HEVC vs VP9 vs AVS2 video. In *International conference on multimedia and expo workshops (ICMEW)* (pp. 1–6).
9. Bovik, A. (2005). *Handbook of image and video processing* (2nd ed.). Burlington: Academic Press.
10. Uhrina, M., Hlubik, J., & Vaculik, M. (2013). Correlation between objective and subjective methods used for video quality evaluation. *Advances in Electrical and Electronic Engineering*, 11(2), 135–146.
11. Loke, H.M., Ong, P.E., Lin, W., Lu, Z., & Yao, S. (2006). Comparison of video quality metrics on multimedia videos. In *IEEE international conference on image processing* (pp. 457–460).
12. Uhrina, M., Hlubik, J., & Vaculik, M. (2012). Impact of compression on the video quality. *Advances in Electrical and Electronic Engineering*, 10(4), 251–258.
13. Zhou, W., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 12(4), 1057–1149.
14. Chan, M. C., & Ramjee, R. (2005). TCP/IP performance over 3G wireless links with rate and delay variation. *Wireless Networks*, 11(1–2), 81–97.
15. Netem, <http://www.linuxfoundation.org/collaborate/workgroups/networking>.



**Jaroslav Frnda** was born in 1989 in Martin, Slovakia. He received his M.Sc. from the VSB-Technical University of Ostrava, Department of Telecommunications in 2013, now he continues with his Ph.D. study at the same place. His research interests include Quality of Triple play services and IP networks.



**Miroslav Voznak** (born in 1971) is an Associate Professor with Dept. of Telecommunications, Technical University of Ostrava, Czech Republic. He received his M.S. and Ph.D. degrees in telecommunications, dissertation thesis “Voice traffic optimization with regard to speech quality in network with VoIP technology” from the Technical University of Ostrava, in 2002. Topics of his research interests are Next Generation Networks, IP telephony, speech quality and network security.



**Lukas Sevcik** was born in 1989 in Cadca, Slovakia. He received his M.Sc. in Informatics at the Faculty of Management Science and Informatics, the University of Zilina in 2013. Now, he continues with his Ph.D. study at the Department of Telecommunication, VSB-Technical University of Ostrava. His research interests include Quality of Triple play services and IP networks.