Network Degradation Effects on Different Codec Types and Characteristics of Video Streaming

Jaroslav FRNDA[1], *Lukas SEVCIK*[1], *Miroslav UHRINA*[2], *Miroslav VOZNAK*[1]

[1]Department of Telecommunications, Faculty of Electrical Engineering and Computer Science, VSB–Technical University of Ostrava, 17. listopadu 15, 708 00 Ostrava-Poruba, Czech Republic
[2]Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Univerzitna 8215/1, 010 07, Zilina, Slovakia

jaroslav.frnda@vsb.cz, lukas.sevcik.st1@vsb.cz, miroslav.uhrina@fel.uniza.sk, miroslav.voznak@vsb.cz

Abstract. *Nowadays, there is a quickly growing demand for the transmission of voice, video and data over an IP based network. Multimedia, whether we are talking about broadcast, audio and video transmission and others, from a global perspective is growing exponentially with time. With incoming requests from users, new technologies for data transfer are continually developing. Data must be delivered reliably and with the fewest losses at such high speed. Video quality as part of multimedia technology has a very important role nowadays. It is influenced by several factors, where each of them can have many forms and processing. Network performance is the major degradation effect that influences the quality of resulting image. Poor network performance (lack of link capacity, high network load...) causes data packet losses or different delivery time for each packet. This work focuses exactly on these network phenomena. It examines the impact of different delays and packet losses on the quality parameters of triple play services, to evaluate the results using objective methods. The aim of this work is to bring a detailed view on the performance of video streaming over IP-based networks.*

Keywords

Delay variation, packet loss, PSNR, SSIM, video quality assessment.

1. Introduction

The growth of the Internet network is using more and more resources for performance analysis. This is simulated using different models. This work compares the performance of the network from an experimental viewpoint. The objective of this work is to analyse various impacts on transmitted video, such as packet loss, jitter, reordering. The work focuses on the presence of video and data packets in the network. It compares the impact of data loss and out of order data. The results show whether it is better to get the packets in a different order or completely lost. It compares static and dynamic video; varying quality of the transmitted video. We compare the impact of the size of the transmitted data.

We deal with objective methods for the evaluation of the quality of videos in the works. There are many attributes of the video image. These can be compared; therefore, to measure the exploits of several of the most well-known methods for the evaluation of image quality. Each method has different procedures and different metric evaluation system. We apply packet loss over predetermined steps on stream. The individual objective methods will evaluate the captured streams. The aim of the paper is to evaluate the impact of loss during transmission using different compression technologies from several different perspectives.

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 network status. But on the other hand, video has become the majority 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 [1]. Analysis of video quality concentrates on the resistance of video codecs to packet loss in the network, which causes artefacts in the video [2], [3]. On the other hand, a few factors still lack, such as a complex view of video parameters on the final video quality. In our previous works [4], [7], we focused on the quality of the triple play service 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 complex 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 network situation modeling and prediction, it is good to know how it influences the final video quality.

3. Methodology

3.1. Video Processing

1) Size of Digital Image Data

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 with static images is very sensitive to memory needed saving [5].

Standard television broadcasting has a frame rate of at least 25 fps (frames per second), [6]. It is sufficient for the delay in perception of the human eye. Every second of the movie at 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 [5].

2) Video Transmission

To transfer video files, fundamentally unreliable protocols are used. The principle consists in sending and receiving data without feedback between the sender and the recipient.

Factors affecting the video transmission are:

- 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: This is the average number of packets that arrive at the destination node due to the state of the network. 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 [5].

3) Methods for Evaluating the Quality

In the work, we used the objective methods - PSNR 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.

PSNR (Peak signal-to-noise ratio) is the ratio between the maximum signal energy and noise energy. It is expressed on a logarithmic scale because different signals have different dynamic ranges. PSNR in decibels is defined by the formula [7]:

$$\mathrm{PSNR} = 10 \log_{10} \frac{\mathrm{MAX}_I^2}{\mathrm{MSE}} [\mathrm{dB}], \tag{1}$$

where MAX is the maximum value that the pixel can take (e.g. 255 for 8-bit image) and MSE is the difference between two gay-level images or video sequences. Technically, MSE reflects the diversity of the image, while PSNR expresses its identity. The strongest PSNR method is an easy and fast calculation, which is the reason why it is still used very often in scientific papers although the correlation with the human perception is worse than SSIM [6], [8].

The SSIM (Structural Similarity Index) method includes three components - the similarity of the intensity, the corresponding contrast and the corresponding structure. The combination of these three factors forms one value. This demonstrates the quality of the test video. This method differs by evaluating structural distortion and not error rate. The main reason for this difference is characteristic of the human visual system. This perceived distortion changes in the structure of the frame much better than the error rate. Since the SSIM method 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), [7].

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 seconds. All the movies were measured at 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 space shuttle launch and the third video with the highest bitrate (60 Mbps) by an open source animated movie called Big Buck Bunny. The whole process of measuring is shown in the Fig. 1 and Fig. 2. To evaluate the quality we used the methods SSIM and PSNR. SSIM correlates better with the perception of the human eye [6]. 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.

This broadcast the video stream on the local computer interface. We captured and saved the stream to disk using another VLC Player. We saved this transfer video and tagged it as the original video. Our testing scenarios reflect the situation that can actually happen in the network. Especially the mobile networks capable of using IP architecture like UMTS and LTE reach high values of packet loss and delay variation [10]. For the purpose of settings of 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 re-ordering [9].

1) Packet Loss

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 %.

2) Jitter

We set that 25 % of packets will be delayed (results of our previous work [4] 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 the videos, we set the value of the de-jitter buffer to 0 in VLC, so that the delays were real.

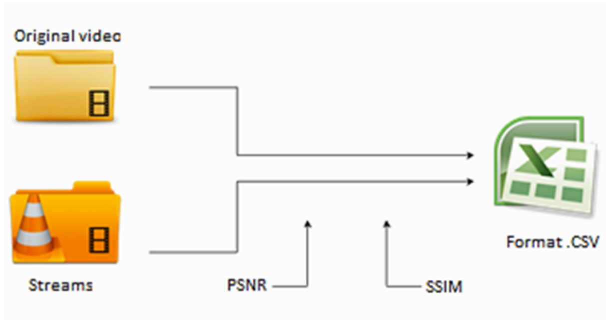

Fig. 1: Measurement procedure.

Tab. 1: Parameters for measurements.

Used codecs	MPEG-2, MPEG-4 - H.264
Video resolution [pixels]	720×576, 1280×720, 1920×1080
Evaluation methods	PSNR, SSIM

Setting packet loss on local interface:

- `#tcqdisc add dev lo root netem loss 1 %.`

Change packet loss on local interface:

- `#tcqdisc change dev lo root netem loss 2 %.`

This 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.

Setting packet delay on local interface:

- `#tcqdisc add dev eth0 root netem delay 10 ms reorder 75 % 50 %.`

In this example, 75 % of the packets (with a correlation of 50 %) will get sent immediately, the others will be delayed by 10 ms. 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 behaviour more exactly.

Change packet delay on local interface:

- `#tcqdisc change dev eth0 root netem delay 20ms reorder 75 % 50 %.`

3) Evaluation of the Results

By using the program MSU VQMT, we compared the original sample and the tested sample, which included damage caused by our settings. The program exports results into a CSV format, where we can find the value for every compared frame and the total average value for the whole video.

Fig. 2: Comparing stream with original 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 was not proven as a significant parameter of video robustness.

The most important factor was shown to be the video code type. 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, 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 [8].

The first tested dynamic video 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 significantly affects the picture distortion. That is the reason why dynamic video is generally more sensitive to data loss [5], [8].

The third video has a dynamic character, too, but a very high bitrate compared to the previous two videos. Performance of this video was very poor. High bitrate means a lot of information contained in the I, B and P frames and its loss causes significant degradation of the video quality.

For a better illustration, figures of the number 3 and 4 demonstrate the video quality results for full HD resolution. This paper follows on from our previous work [4] and extends the video prediction model that was used there. All these mentioned result were processed into the following regressive equations.

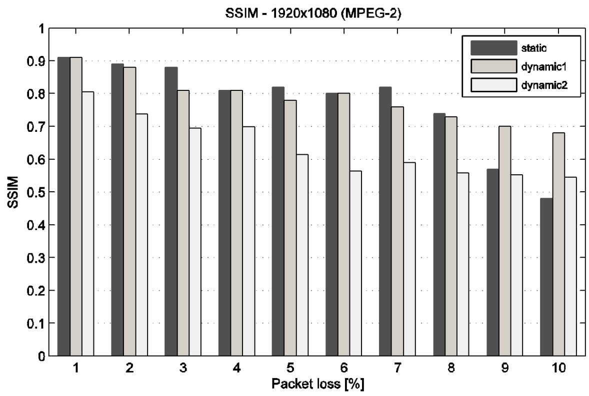

Fig. 3: SSIM results for MPEG-2.

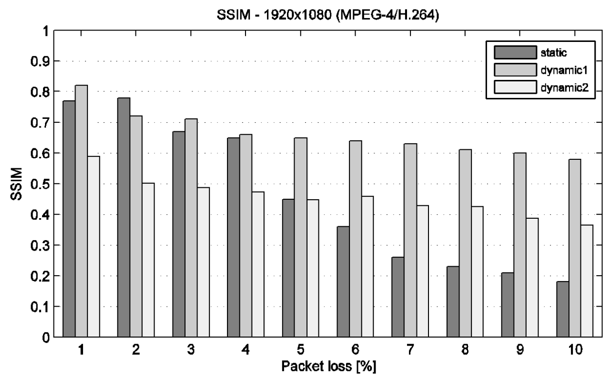

Fig. 4: SSIM results for MPEG-4(H.264).

4.1. Slow-Motion Video

1) MPEG-2

$$\begin{aligned} \mathrm{SSIM} = {} & \alpha(a + b \cdot (X^2)) + \beta(a + b \cdot \sqrt{X}) + \\ & + \gamma(a + b \cdot (X^2)). \end{aligned} \tag{2}$$

2) MPEG-4

$$\begin{aligned} \mathrm{SSIM} = {} & \alpha\left(\sqrt{a + \frac{b}{X}}\right) + \beta\left(\frac{1}{a + b \cdot X}\right) + \\ & + \gamma(exp(a + b \cdot X)). \end{aligned} \tag{3}$$

All the necessary coefficients are presented in Tab. 5.

Because measurements were performed on two dynamic videos, the following regressive equations represent a prediction model for both of them.

4.2. Dynamic Video with Ordinary and High Bitrate

1) MPEG-2

$$\mathrm{SSIM} = \alpha\left(\frac{1}{a + b \cdot X}\right) + \beta\left(a + b \cdot ln\left(X\right)\right). \tag{4}$$

Tab. 2: Static video measurements results.

Packet loss [%]	PSNR ([dB])/SSIM					
	720×576		1280×720		1920×1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	29.20/0.914	23.84/0.805	26.25/0.942	17.87/0.742	27.45/0.905	17.63/0.774
2	28.73/0.876	15.05/0.688	25.84/0.894	17.36/0.730	23.54/0.885	16.94/0.776
3	23.320.877	13.21/0.52	22.02/0.839	14.43/0.709	20.86/0.876	15.70/0.673
4	22.89/0.8	13.90/0.529	20.03/0.830	15.13/0.713	21.28/0.821	13.69/0.648
5	23.74/0.77	12.72/0.514	14.29/0.743	12.64/0.555	16.69/0.820	14.59/0.450
6	20.36/0.721	11.34/0.499	14.36/0.726	11.36/0.539	14.84/0.812	11.92/0.362
7	17.98/0.669	12.12/0.486	13.08/0.707	10.34/0.534	15.53/0.800	12.44/0.255
8	15.44/0.536	10.33/0.485	12.43/0.660	9.37/0.483	15.53/0.740	10.86/0.225
9	15.68/0.481	9.96/0.463	13.90/0.658	8.46/0.444	15.51/0.569	11.64/0.205
10	13.40/0.452	8.55/0.362	13.08/0.638	9.77/0.421	12.46/0.481	10.33/0.184

Tab. 3: PSNR and SSIM results for first tested dynamic video.

Packet loss [%]	PSNR ([dB])/SSIM					
	720×576		1280×720		1920×1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	21.22/0.864	17.71/0.777	21.75/0.859	16.29/0.736	19.00/0.915	15.90/0.817
2	19.08/0.830	16.30/0.738	19.27/0.837	15.89/0.750	17.99/0.875	13.59/0.715
3	19.08/0.746	14.89/0.692	17.90/0.789	14.73/0.745	14.96/0.809	13.25/0.715
4	18.65/0.754	14.84/0.686	18.23/0.766	14.83/0.714	15.86/0.808	11.51/0.657
5	17.38/0.716	14.73/0.683	18.37/0.752	14.69/0.699	14.41/0.781	12.24/0.649
6	16.23/0.711	14.73/0.645	16.07/0.741	15.07/0.663	14.95/0.799	12.62/0.637
7	15.02/0.706	13.83/0.644	14.83/0.717	13.21/0.582	15.29/0.764	12.97/0.632
8	15.27/0.709	12.60/0.651	14.73/0.701	12.66/0.545	14.10/0.728	12.15/0.605
9	14.29/0.693	12.46/0.622	14.83/0.694	10.57/0.502	14.13/0.703	11.46/0.604
10	14.70/0.686	11.65/0.605	14.73/0.678	11.72/0.485	13.20/0.683	12.11/0.582

Tab. 4: PSNR and SSIM results for the second dynamic tested video.

Packet loss [%]	PSNR ([dB])/SSIM					
	720×576		1280×720		1920×1080	
	MPEG-2	MPEG-4	MPEG-2	MPEG-4	MPEG-2	MPEG-4
1	27.0/0.923	21.32/0.695	27.63/0.927	16.10/0.530	24.72/0.806	15.56/0.589
2	26.51/0.9	17.72/0.551	22.79/0.765	15.75/0.511	22.41/0.737	15.16/0.501
3	21.59/0.714	15.36/0.446	22.56/0.745	15.30/0.422	22.0/0.694	14.42/0.487
4	21.26/0.718	15.28/0.447	21.43/0.7	14.92/0.421	22.01/0.699	13.8/0.474
5	20.73/0.705	15.33/0.441	21.10/0.664	14.04/0.428	19.84/0.614	13.37/0.446
6	19.97/0.668	15.21/0.405	21.10/0.597	12.05/0.393	18.72/0.589	13.02/0.458
7	20.07/0.618	14.28/0.338	19.15/0.566	11.46/0.372	18.22/0.564	12.17/0.429
8	19.14/0.575	14.41/0.333	17.90/0.531	10.91/0.353	17.59/0.558	12.09/0.425
9	17.97/0.562	11.08/0.303	17.65/0.528	10.91/0.352	18.28/0.552	11.09/0.387
10	17.81/0.493	10.95/0.326	17.29/0.511	10.87/0.326	18.07/0.545	10.70/0.364

Tab. 5: Coefficients for static video.

Coef.	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	−0.004924	−0.143973	−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

2) MPEG-4

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

Table 6 and Tab. 7 contain the coefficients for these two equations. All regressive models described here gained an R-square factor (R^2) higher than 90 %, which represents a high level of veracity.

The second group of measurements led to an analysis of the degradation effect of delay variation – Jitter. The results of the performed tests uncover a critical boundary of 20 ms. Above this value, a significant reduction of final video quality is observed. Due to the process of decompressing and processing the video

Tab. 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	1

Tab. 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

stream on the end user side costing some time, both codecs are tolerant for small delay variation.

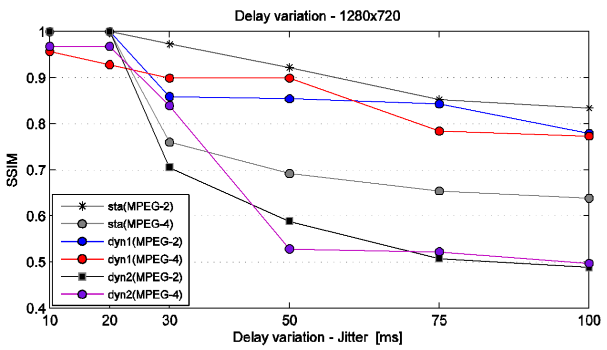

Fig. 5: Results of delay variation measurements, HD resolution.

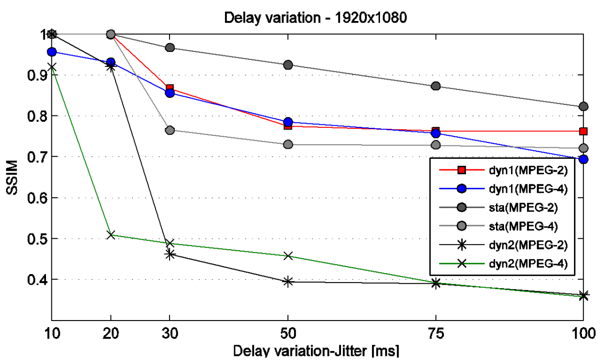

Fig. 6: Results of delay variation measurements, Full HD resolution.

The behaviour of dynamic videos is approximately on the same level, with bigger differences between the MPEG-2 and MPEG-4 codec when the static video was used.

Typically in the real world, de-jitter buffer is used for elimination of this phenomenon, but it is good to know how big a degradation effect on video quality has been caused particularly by the delay variation.

5. Conclusion

The aim of this work was to bring a detailed view of the performance of video streaming 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 helped us to create and extend our previous mathematical models of video streaming behaviour. 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.

Our future works will focus on two directions: Firstly, the new generation of video codecs such as H.265 and VP9. Due to the limitations of our evaluation MSU VQMT program, we are currently awaiting the official support for these new video codecs. The second part will be focused on analysis of the impact of security mechanisms, and on the encryption algorithm implemented to QoS parameters. Security is a highly discussed topic nowadays, and protocols such as IPsec, VPN/SSL and SRTP are becoming more and more frequently used to secure the content of voice or video, so computational mathematical models should handle this new situation.

Acknowledgment

This work was supported by Grant of the SGS No. SP2014/72 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) and by the Development of human resources in research and the development of latest soft computing methods and their application in practice project (CZ.1.07/2.3.00/20.0072) funded by the Operational Programme Education for Competitiveness.

References

[1] CHIKKERUR, S., V. SUNDARAM, M. REISSLEIN and L. J. KARAM. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison. *IEEE Transactions on Broadcasting*. 2011, vol. 57, iss. 2, pp. 165–182. ISSN 0018-9316. DOI: 10.1109/TBC.2011.2104671.

[2] YIM, C., A. C. BOVIK, M. REISSLEIN and L. J. KARAM. Evaluation of temporal variation of video quality in packet loss networks: A Classification, Review, and Performance Comparison. *Signal Processing: Image Communication*. 2011, vol. 26, iss. 1, pp. 24–38. ISSN 0923-5965. DOI: 10.1016/j.image.2010.11.002.

[3] FEAMSTER, N. and H. BALAKRISHNAN. Packet Loss Recovery for Streaming Video. In: *12th International Packet Video Workshop*. Pittsburgh: ACM, 2002.

[4] FRNDA, J., M. VOZNAK, J. ROZHON and M. MEHIC. Prediction model of QoS for Triple play services. In: *2013 21st Telecommunications Forum Telfor (TELFOR)*. Belgrade: IEEE, 2013, pp. 733–736. ISBN 978-1-4799-1419-7. DOI: 10.1109/TELFOR.2013.6716334.

[5] BOVIK A. *Handbook of image and video processing*. 2nd ed. Burlington: Elsevier Academic Press, 2005. ISBN 01-211-9792-1.

[6] UHRINA, M., J. HLUBIK and M. VACULIK. Correlation Between Objective and Subjective Methods Used forVideo Quality Evaluation. *Advances in Electrical and Electronic Engineering*. 2013, vol. 11, no. 2, pp. 135–146. ISSN 1804-3119. DOI: 10.15598/aeee.v11i2.775.

[7] FRNDA, J., M. VOZNAK and L. SEVCIK. Network Performance QoS Prediction. In: *Proceeding of the First Euro-China Conference on Intelligent Data Analysis and Applications*. Shenzen: IEEE, 2012, pp. 165–174. ISBN 978-3-319-07776-5. DOI: 10.1007/978-3-319-07776-5_18.

[8] UHRINA, M., J. HLUBIK and M. VACULIK. Impact of Compression on the Video Quality. *Advances in Electrical and Electronic Engineering*. 2012, vol. 10, no. 4, pp. 251–258. ISSN 1804-3119. DOI: 10.15598/aeee.v10i4.737.

[9] Netem. *Linux Foundation* [online]. 2009. Available at: `http://www.linuxfoundation.org/collaborate/\workgroups/networking/netem`.

[10] CHAN, M. C. and R. RAMJEE. TCP/IP performance over 3G wireless links with rate and delay variation. In: *Proceedings of the 8th annual international conference on Mobile computing and networking - MobiCom '02*. New York: ACM Press, 2002, pp. 71–82. ISBN 1-58113-486-X. DOI: 10.1145/570645.570655.

About Authors

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, and now he is continuing his Ph.D. study at the same place. His research interests include Quality of Triple play services and IP networks.

Lukas SEVCIK was born in 1989 in Cadca, Slovakia. He received his M.Sc in Informatics from the Faculty of Management Science and Informatics, University of Zilina, in 2013. Now, he is continuing his Ph.D. study at the Department of Telecommunications, VSB–Technical University of Ostrava. His research interests include Quality of Triple play services and IP networks.

Miroslav UHRINA was born in 1984 in Zilina, Slovakia. He received his M.Sc and Ph.D. degrees in Telecommunications from the Department of Telecommunications and Multimedia, University of Zilina, in 2008 and 2012, respectively. Nowadays, he is an assistant lecturer at the mentioned Department of Telecommunications and Multimedia. His research interests include audio and video compression, TV broadcasting (IPTV. DVB-T, DVB-H) and IP networks.

Miroslav VOZNAK is an associate professor at the Department of Telecommunications, VSB–Technical University of Ostrava. He received his Ph.D. in telecommunications, and his dissertation thesis was entitled "Voice traffic optimization with regard to speech quality in networks with VoIP technology" in 2002. Topics of his research interests are Next Generation Networks, IP telephony, speech quality and network security.