COMMUNICATIONS

Andrew Hines - Jan Skoglund - Anil C. Kokaram - Naomi Harte *

MONITORING VOIP SPEECH QUALITY FOR CHOPPED AND CLIPPED SPEECH

Real-time monitoring of speech quality for VoIP calls is a significant challenge. This paper presents early work on a no-reference objective model for quantifying perceived speech quality in VoIP. The overall approach uses a modular design that will be able to help pinpoint the reason for degradations as well as quantifying their impact on speech quality. The model is being designed to work with narrowband and wideband signals. This initial work is focused on rating amplitude clipped or chopped speech, which are common problems in VoIP. A model sensitive to each of these degradations is presented and then tested with both synthetic and real examples of chopped and clipped speech. The results were compared with predicted MOS outputs from four objective speech quality models: ViSQOL, PESQ, POLQA and P.563. The model output showed consistent relationships between this model's clip and chop detection modules and the quality predictions from the other objective speech quality models. Further work is planned to widen the range of degradation types captured by the model, such as non-stationary background noise and speaker echo. While other components (e.g. a voice activity detector) would be necessary to deploy the model for stand-alone VoIP monitoring, the results show good potential for using the model in a realtime monitoring tool.

Keywords: *Speech Quality Model, Clip, Chop, VoIP.*

1. Introduction

As digital communication has become more pervasive, the variety of channels for human speech communication has grown. Where narrowband telephony dominated, the range of channels has expanded to include multimedia conferencing such as Google Hangouts, Skype and other voice over internet protocol (VoIP) services. Realtime assessment of the Quality of Experience (QoE) for users of these systems is a challenge as the channel has become more complex and the points of failure have expanded. Traditionally, QoE for voice communication systems is assessed in terms of speech quality. Subjective listener tests establish a mean opinion score (MOS) on a five point scale by evaluating speech samples in laboratory conditions. Aside from being time consuming and expensive, these tests are not suitable for realtime monitoring of systems.

The development of objective models that seek to emulate listener tests and predict MOS scores is an active topic of research and has resulted in a number of industry standards. Models can be categorised by application, i.e. planning, optimisation, monitoring and maintenance [1]. Full reference objective models, such as PESQ [2] and POLQA [3], predict speech quality by comparing a reference speech signal to a received signal and quantifying the difference between them. Such models can be applied to system optimisation but are constrained by the requirement to have access to the original signal, which is not always practical for realtime monitoring systems. In these scenarios, no-reference (NR) models, such as P.563 [4], LCQA [5] or ANIQUE+ [6] are more appropriate. They are sometimes referred to as single ended, or non-intrusive models, as they attempt to quantify the quality based only on evaluating the received speech signal without access to a clean reference. This restriction makes NR model design more difficult, and NR models tend to have inferior performance accuracy, when compared to full reference models [7].

This work presents the early stage development of an NR speech quality model for VoIP applications based on a modular architecture. The model will contain modules that are designed to detect and estimate the amount of degradation caused by specific issues. Ultimately the individual modules will be combined to produce an aggregate objective speech quality prediction score. The novelty of this approach over other NR models [4, 5 and 6] is that each module provides a unidimensional quality index feeding into the overall metric but can also provide diagnostic information about the cause of the degradation for narrowband

* [1,2]Andrew Hines, [3]Jan Skoglund, [3]Anil C. Kokaram, [2]Naomi Harte
[1]Dublin Institute of Technology, Ireland
[2]Trinity College Dublin, Ireland
[3]Google, Inc., Mountain View, CA, USA
E-mail: andrew.hines@dit.ie

or wideband speech. This could allow realtime remedial action to be taken to improve the overall quality of experience for the users of VoIP systems, through changing parameters such as bandwidth to adjust the quality of experience from a low quality wideband speech scenario to an alternative high quality narrowband speech scenario.

The modules proposed in this paper, as part of an overall system, are designed to work with narrowband and wideband signals. The two modules are a model sensitive to amplitude clipping and another for choppy speech. These are two common problems in VoIP. Section 2 describes these degradations and their causes. Section 3 describes the models and an experimental evaluation is outlined in section 4. Results are presented for both synthesised and real degradations. Section 5 discusses the results and compares them with the predictions of other objective metrics. The paper concludes with a description of the next stages in the overall model development.

2. Background

2.1. Amplitude Clipped Speech

Amplitude clipping is a form of distortion that limits peak amplitudes to a maximum threshold. This can be caused by analogue amplifiers where the amplification power exceeds the capabilities of the hardware. Amplitude clipping can also be caused by digital representation constraints when a signal is amplified outside the range of the digital system. If the maximum range of the signal cannot be represented using the number of quantising intervals available (number of bits per sample), the signal will be clipped. The main body of literature studying the effect of amplitude clipping on speech quality is in the field of hearing aids. For hearing aids, clipping can be can used to minimise the distortion for high level input signals [8], whereas in VoIP scenarios, clipping is generally an undesirable result of incorrect gain level settings for the speaker's hardware. The term 'clipped' is often used to describe other types of speech quality degradation, such as time clipped (choppy) or temporally clipped (front end clipping, back end clipping of words) but here it will be used to refer exclusively to amplitude clipping.

Clipping has significantly more impact on quality than intelligibility. Experiments by Licklider [9] found that word intelligibility remained over 96% when speech was clipped to 20 dB below the highest peak amplitude. To put this in perspective, the highest clipping level used in this paper was clipped to 16 dB below the highest peak amplitude and while it is fully intelligible, informal listening tests show it was perceived as very poor quality.

Examples of the clipped speech used in testing are shown in Fig. 1. The first example is clearly clipped as there is a clear threshold amplitude cut-off. The second example shows the same speech with narrowband 30 dB SNR pink noise added after clipping. This illustrates how clipping that is still apparent to the listener can be masked in the signal amplitude by other degradations.

2.2. Choppy Speech

Choppy speech describes degradation where there are gaps in the speech signal. It manifests itself as syllables appearing to be dropped or delayed. The speech is often described as a stuttering or a staccato. It is sometimes referred to as time clipped speech, or broken voice. It is generally periodic in nature, although the rate of chop and duration of chops can vary depending on the cause and on network parameters.

Choppy speech occurs for a variety of reasons such as CPU overload, low bandwidth, congestion or latency. When frames are missed or packets dropped, segments of the speech are lost. This can occur at any location within speech, but is more noticeable and has a higher impact on perceived quality when it occurs in the middle of a vowel phoneme than during a silence period. Modern speech codecs attempt to deal with some quality issues by employing jitter buffers and packet concealment methods (e.g. [10 and 11]) but do not deal with all network or codec related problems and choppy speech remains a problematic feature of VoIP systems [12].

3. Models

3.1. Amplitude Clipped Speech Detection Model

The module is a non-intrusive single ended model. It takes a short speech signal as input and bins the signal samples by amplitude into 50 bins. Two additional bins are added with values set to the minimum bin value to allow first and last bins to be evaluated as peaks. The resulting histogram for a clipped signal is illustrated in Fig. 1 where, $h[i]$, is the histogram value of peak index i. The model finds all local maxima peaks in the histogram. Local maxima peaks are constrained to a minimum height of 0.5% of the sum of the histogram and a minimum distance of 5 bins separation from other peaks. As a minimum of three bins are required to identify a peak, this constraint ensures small deviations in local maxia are not treated as new peaks. Next, all peaks are sorted into descending order yielding a set, P. Then, beginning with the largest peak, all peaks not separated by 5 or more bins are discarded. First, the centre peak and clipped peaks, illustrated in Fig. 2 are identified. The centre peak, P_c, where $P_c = h[c]$. The peak index, c, is found using auto-correlation of $h[i]$, from

$$c = \frac{1}{2} \arg\max_j R_{hh}[j] = \sum_i h[i]h[i-j] \quad (1)$$

The left peak P_l is found as the largest of the peaks to the left of the centre peak P_c, located at

$$l = \arg\max_i h[i] \;\; \forall \;\; i < c, i \in \{P\} \tag{2}$$

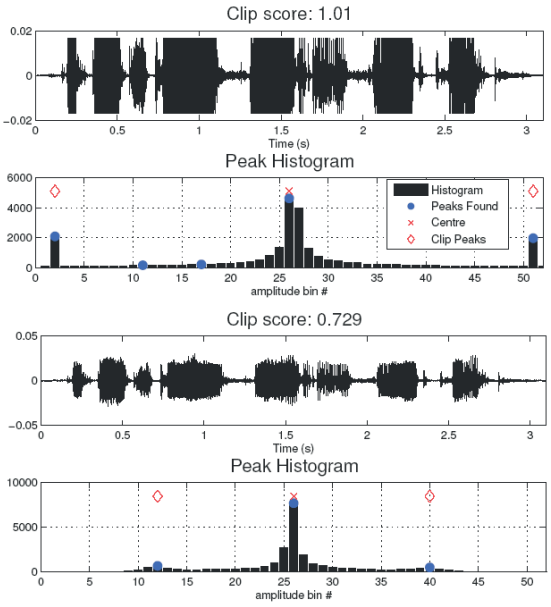


Fig. 1 Amplitude clipping signal and histogram in the time domain binned across 50 amplitude bins. Above: A signal with clipping visually apparent in the time domain. The histogram highlights the clipping with peaks in the first and last bins. Below: the clipped signal which has been further corrupted with 30 dB SNR narrowband pink noise after clipping. The clipping becomes harder to observe in the signal but the clipping peaks are still visible in the histogram.

Then the equivalent right peak P_r is the peak closest to the same distance from the centre peak as the left peak, calculated as $h[r]$ where

$$r = \arg\min_i \left| ((c-l)-(i-c)) \right| \;\; \forall \;\; i > c, i \in \{P\} \tag{3}$$

The clip score is calculated as

$$clip = \log_{10}\left[\frac{\sum_{i=l-1}^{l+1} h[i] + \sum_{i=r-1}^{r+1} h[i]}{\sum_{i=c-1}^{c+1} h[i]} \right] \tag{4}$$

Figure 2 illustrates an example histogram with the maximum peak, P_c and the clip peaks, P_l *and* P_r, as solid red bars and other candidate peaks as solid black bars.

3.2. Choppy Speech Detection Model

The chop detection model [13] uses a short-term Fourier Transform (STFT) spectrogram of the test signal to measure changes in the gradient of the mean frame power. An example is shown in Fig. 3. The STFT is created using critical bands between 150 and 8,000 Hz for wideband speech or 3,400 Hz for narrowband speech. A 256 sample, 50% overlap Hanning window is used for signals with 16 kHz sampling rate and a 128 sample window for 8 kHz sampling rate to keep frame resolution temporally consistent.

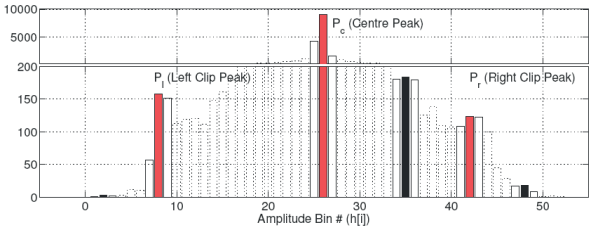


Fig. 2 Amplitude clipping algorithm. The signal is binned by amplitude into 50 bins. The peaks from the histogram are shown as solid bars. After the centre peak is found using autocorrelation, the left peak is the max peak left of the centre peak. The matching right peak is the peak closest to the same distance from the centre as the left peak. The clipped score is then calculated as a log of the sum of the clip peak bins and their adjacent bins are divided by the sum of the centre peak and adjacent bins.

A gradient of the mean power per frame is calculated, g[i], as

$$g = \nabla P = \frac{\partial P}{\partial t}. \tag{5}$$

A positive gradient signal, $g_p[i]$ and a negative gradient, $g_n[i]$ can be defined as

$$g_p[i] = \begin{cases} g[i] & \text{if } g[i] > 0 \\ 0 & \text{if } g[i] \le 0 \end{cases}$$
$$g_n[i] = \begin{cases} -g[i] & \text{if } g[i] < 0 \\ 0 & \text{if } g[i] \ge 0 \end{cases} \tag{5}$$

A cross-correlation of $g_p[i]$ and $g_n[i]$ yields the max overlap offset j as

$$\arg\max_j R_{g_n g_p}[j] = \sum_i g_n[i] g_p[i-j]. \tag{6}$$

The $g_p[i]$ and the offset $g_n[i-j]$ are summed as

$$g_c[i] = g_n[i] + g_p[i-j] \tag{7}$$

and a log ratio of the sum of values above a threshold c_T denoted c_+, to the sum below the threshold, c_-, is taken to estimate the amount of chop in the signal:

$$c_+[i] = \begin{cases} g_c[i] & \text{if } g_c[i] > c_r \\ 0 & \text{if } g_c[i] \le c_r \end{cases}$$
$$c_-[i] = \begin{cases} g_c[i] & \text{if } g_c[i] < c_r \\ 0 & \text{if } g_c[i] \ge c_r \end{cases} \quad (8)$$
$$chop = \log_{10} \frac{\Sigma_i c_+[i]}{\Sigma_i c_-[i]}$$

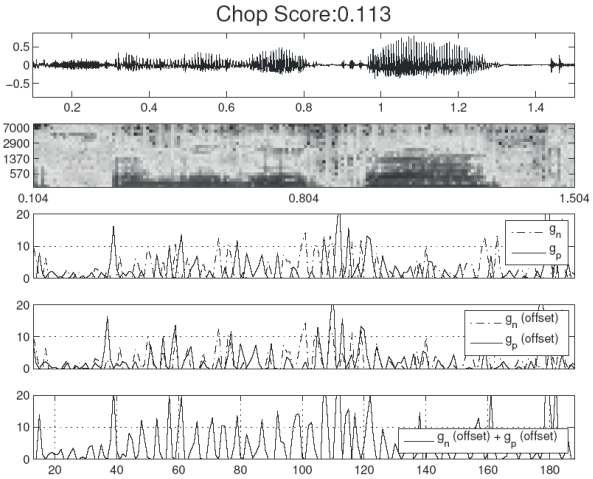


Fig. 3 Real Chop Example. This example is taken from a real recording where choppy speech occurred as a result of a codec mismatch between transmitter and receiver. The top panes show the signal and signal spectrogram, and the chop is visible as periodic white bands in the higher frequencies of the spectrogram. The gradients, gp and gn are shown in the next pane with the offset versions that have been aligned shown in the forth pane. The bottom pane shows the sum of the offset gradients. This has sharp peaks corresponding to the chop and is used to calculate the chop score, as described in section 3.

4. Model testing

4.1. Stimuli

For these experiments a test dataset was created using 30 samples from the IEEE speech corpus [14]. Ten sentences from three speakers, each of approximately 3 seconds in duration were used as source stimuli. A cursory validation with a small number of real clipped and chopped speech samples was also undertaken using wideband recordings of choppy speech caused by a codec mismatch and clipped speech recorded using a laptop microphone.

4.2. Model Comparison

The test data was evaluated using 4 other objective speech quality models: ViSQOL, PESQ, POLQA and P.563. ViSQOL is a full reference objective model developed by the authors in prior work [15, 16 and 17]. PESQ [2] is the ITU recommended standard and is still the most commonly used speech quality model although it has be superseded by a newer standard POLQA [3]. P.563 is the ITU standard no-reference model [4].

4.3. Amplitude Clipping Test

Each sentence was used to create 20 progressively degraded samples of clipped speech. For each sentence, the peak amplitude was found and the signals were clipped to a factor of the maximum peak amplitude ranging from 0.5 to 0.975 in 0.025 increments. For comparison, this is a range of 13.4 to 0.83 dB re RMS or a clipping threshold 3 dB to 16 dB below the maximum peak.

A second test used the same clipping samples but added narrowband 30 dB SNR pink noise to the signal after clipping. This was done to simulate the realities of amplitude clipping where the signal may be scaled or subjected to additional noise and or channel effects after the clipping occurred. Pink noise was chosen as it has similar spectral qualities to speech. At a 30 dB SNR level, it would not be expected to have a major impact on quality but it will mask the sharp cutoff level of the clipping, as illustrated in the signal plots of Fig. 1.

The 20 sets of stimuli created for the choppy speech detection were also used as input to test the amplitude clipping detection model. These were used to establish a minimum detection threshold boundary and to ensure that the model was only detecting the expected degradation type.

A limited test was carried out with a real recording of clipped data. A foreground speech sample spoken into a microphone over background television speech was recorded. The background speech is not clipped but the foreground speech has moderate to severe clipping. The model was used to evaluate the sentence in 1 second segments and the results are shown in Fig. 4.

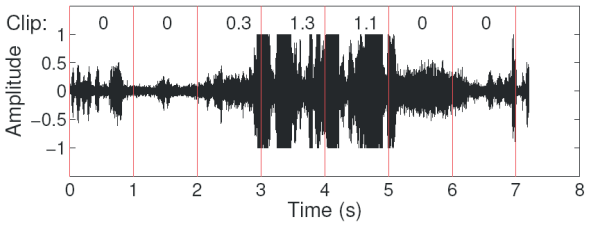


Fig. 4 Real clipped speech example. A foreground speech sample spoken into a microphone over background television speech was recorded. The background speech is not clipped but the foreground speech (from 2.9-5.1 s) has moderate to severe clipping. The model was used to evaluate the sentence in 1 second segments and the clip scores are marked above each 1 second sample.

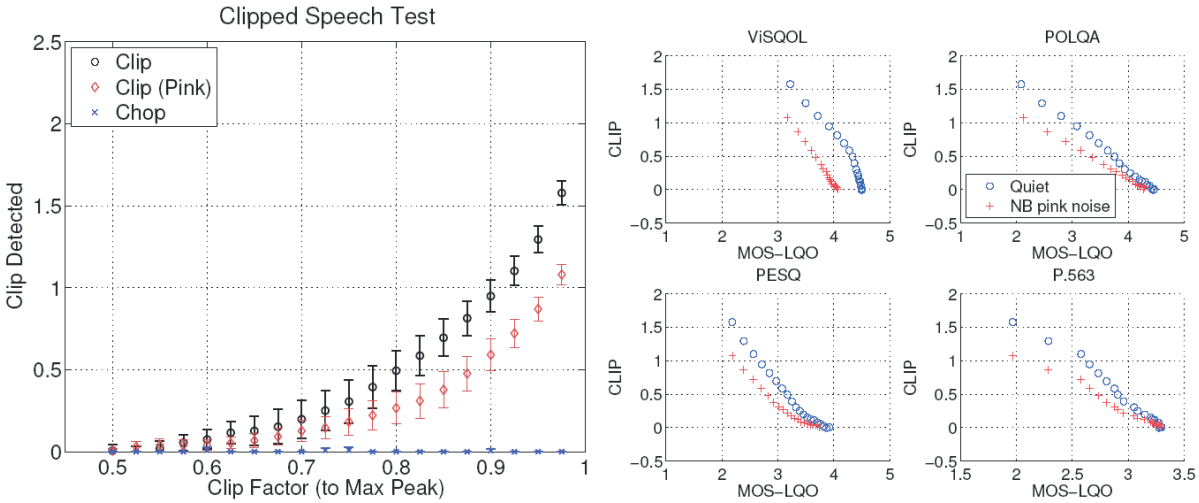


Fig. 5 Amplitude Clipping Results. Left: Results for clipped speech with the clip level plotted against the clip detected for clipping in quiet and with narrowband pink noise added after clipping. A test is also shown with 20 increasing amounts of chop to investigate the model's detection threshold and sensitivity to other degradation types. Right: Comparison with other objective metrics (both full and NR).

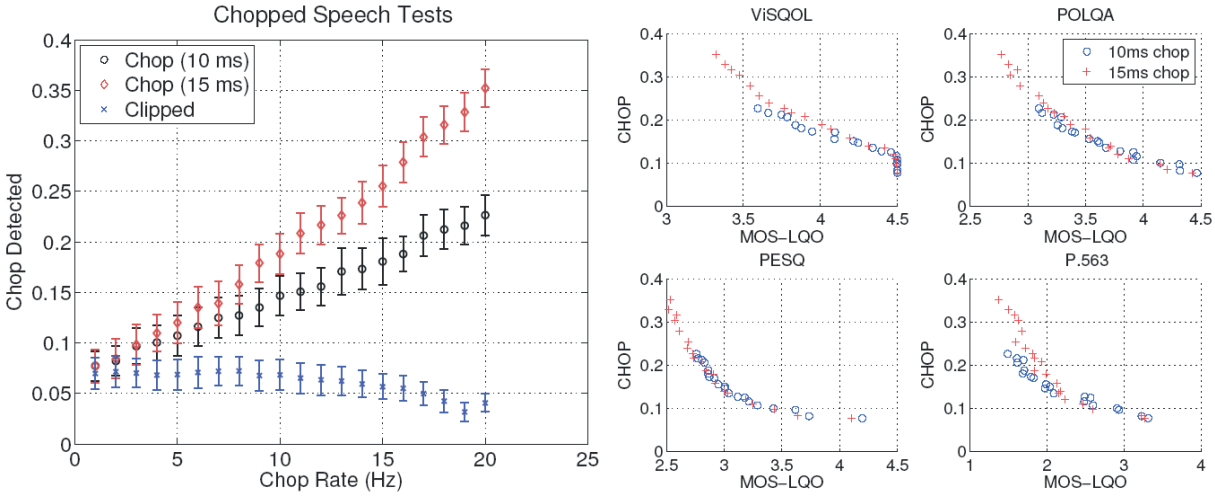


Fig. 6 Chop Detection Results. Left: Results for choppy speech with the chop rate plotted against the level of chop detected for two chop periods, 10 and 15ms. A test is also shown with 20 increasing amounts of amplitude clipping to investigate the model's detection threshold and sensitivity to other degradation types. Right: Comparison with other objective metrics.

4.4. Choppy Speech Detection Test

Two tests were carried out using chopped speech. Using the 30 source sentences, twenty degraded versions of each sentence were created using two chop frame periods of 10ms and 15ms. This simulated packet loss from 3% to 32% of the signals. The test did not simulate packet loss concealment so the samples for the chopped frames were set to zero.

As with the amplitude clipping test, the chop detection model was cross-validated with the clipped stimuli to establish a minimum detection threshold boundary and to ensure that the model was only detecting the expected degradation type.

A limited test was carried out with real choppy data. Wideband speech with a severe amount of chop was tested. The chop in the test was caused by a codec mismatch between the sender and receiver systems. A segment of the test signal is presented in Fig. 3.

5. Results and discussion

5.1. Amplitude Clipped Speech

Figure 5 presents the results for the amplitude clipping tests in quiet and pink noise. The level of clipping increases from left to right on the x-axis and the y-axis shows the model output score. The trends in both the quiet and additive pink noise show clipping begins to be detected at clip level of around 0.55 times peak amplitude. This is a 12 dB peak-to-average ratio which was reported by Kates (1994) to be the level at which clipped speech is indistinguishable from unclipped speech.

The chopped data points are shown on the same x-axis for simplicity but are not clipped in any way and represent 20 levels of progressive chop. They are reported here to highlight that the model is not sensitive to temporal or frequency degradations.

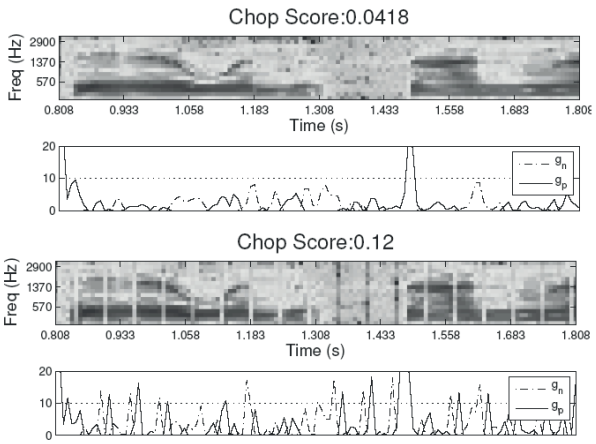


Fig. 7 Chop Example: Above: Clean speech signal with gradients gp and gn plotted below. The gradients detect the gradient in the speech with a large gradient change visible at approximately 1.5 s. Below: The same speech with chop added. The chop is visible in the spectrogram and visible in the g_p and g_n plot used to calculate the chop score.

Although the trends are similar, the range of the clip scores for the quiet and pink noise are different. This is due to the relationship between the scale and the count in the histogram bins. The difference in height between the sharp peaks seen in the quiet histogram versus the spread of peaks in the noisy histogram can been seen in Fig. 1. The use of the additional bins either side of the clip peaks and centre peaks in the ratio calculation (4) reduced the overall difference between the model estimate for a given clipping level when measured in quiet or with additive noise.

Figure 5 also presents a comparison between the model output and those of four other objective quality metrics: ViSQOL, PESQ, POLQA and P.563. For reference, the MOS-LQO predictions are presented in Fig. 8. The results for POLQA in Fig. 5 show

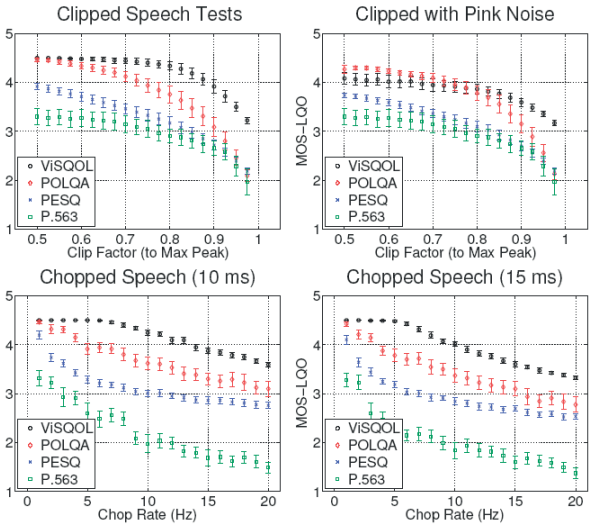


Fig. 8 Predicted MOS-LQO from objective metrics compared in Fig. 5 for clip tests and Fig. 6 for chop tests. Results have shown mean results over 30 sentences. Error bars are 95% confidence intervals.

a linear relationship between the clip scores and the objective metrics across the full range of tests while ViSQOL, PESQ and P.563 exhibit a variety of different sensitivities for the tests with low amounts of clipping, leading to nonlinear tails in the plots. It is worth noting in Fig. 8 that the addition of pink noise to the clipped signal had little effect on the POLQA results for peak clip factors from 0.50–0.6 whereas PESQ and ViSQOL results dropped by over 0.5 with the pink noise added.

5.2. Choppy Speech

Figure 6 presents the results for the chopped speech. The chop rate increases from left to right on the x-axis and the y-axis shows the model output score. The results for the amplitude clipped speech are shown on the same x-axis for simplicity but are not chopped in any way and represent 20 levels of progressive amplitude clipping. They highlight that there is a lower threshold to the chop detection. Fig. 7 shows the same speech sample with and without chop. The periodic chop is clearly visible as vertical bands across the spectrogram and in the peaks of the negative and positive gradients, gp and gn, used by the model to estimate the signal chop level. In addition to detecting chop, the natural gradients of speech are captured by the model. The natural gradient at 1.5 seconds is very apparent in Fig. 7. These speech features are responsible for the low threshold boundary of the chop detection model. The trend for both chop frame periods show chopping being detected above the threshold from a chop rate of 2 Hz. Chop at low rates are common in practice so preliminary tests (not presented here) were carried out with longer duration speech samples. They showed that better

separation between results for chop and naturally occurring gradient changes is possible. This constraint would present practical implementation challenges in a realtime monitoring implementation but should not be insurmountable.

Fig. 8 also presents a comparison between the model output and those of four other objective quality metrics: ViSQOL, PESQ, POLQA and P.563. Unlike the results for the clipping model, the chop model does not have a linear relationship with the objective model results. However, the curve is quite consistent across the different model comparisons, meaning a simple quadratic regression fitting from the chop model score to a MOS prediction may be sufficient. The 10 ms and 15 ms chop periods follow linear trends in Fig. 6 but with different slopes. When they are plotted against the objective metrics there is an overlap in the results follow the same curve. This represents a strong relationship between the chop models score and the estimated perceived quality from the objective metrics.

The real chop example tested showed that chop is detected even if the chop value is not zero and the chop frame is shorter than 10ms, as was the case in the simulated chop tests.

6. Conclusions and future work

The clip and chop measurement models for speech quality presented in this paper show promising early results and compares favourably to the other no-reference objective speech quality model. The degradation types detected are common problems for VoIP and the algorithms used are relatively low in computational complexity. These factors, combined with their applicability to both narrowband or wideband speech, mean they could be useful in applications other than full speech quality models, for example as stand-alone VoIP monitoring tools. To use the model in a realtime system, other components would be necessary. For example, the chop or clip detection will not give accurate results if the speech contains large segments of silence. This could easily be addressed with voice activity detection prior to chop and clip detection.

The models presented are still in the early stages of development. They require testing with a broader test set including a wide variety of real rather than generated degradations. Further testing with a range of wideband stimuli is also required. MOS tests on the existing data would also be beneficial as the full reference metrics disagree significantly on their MOS-LQO predictions for both the clipped and choppy speech. The correlation with quality predictions from POLQA was stronger than with the other objective models. This is seen as a positive pointer for the performance against subjective listener test results as POLQA reports better accuracy than PESQ and and has become the new benchmark standard.

References

[1] MOLLER, S., CHAN, W-Y., COTE, N., FALK, T. H., RAAKE, A., WALTERMANN, M.: Speech Quality Estimation: Models and Trends, *Signal Processing Magazine, IEEE*, vol. 28, No. 6, pp. 18-28, 2011.

[2] ITU: *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for end-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, Int. Telecomm. Union, Geneva, CH, ITU-T Rec. P.862, 2001.

[3] ITU: *Perceptual Objective Listening Quality Assessment*, Int. Telecomm. Union, Geneva, CH, ITU-T Rec. P.863, 2011.

[4] ITU: S*ingle-ended Method for Objective Speech Quality Assessment in Narrow-band telephony Applications,* Int. Telecomm. Union, Geneva, CH, ITU-T Rec. P.563, 2004.

[5] GRANCHAROV, V., ZHAO, D. Y., LINDBLOM, J., KLEIJN, W. B.: Low-complexity, Nonintrusive Speech Quality Assessment, *IEEE Audio, Speech, Language Process.,* vol. 14, No. 6, pp. 1948-1956, 2006.

[6] ANSI ATIS: 0100005-2006: *Auditory Non-intrusive Quality Estimation Plus (ANIQUE+): Perceptual Model for Non-intrusive Estimation of Narrowband Speech Quality,* American National Standards Institute, 2006.

[7] FALK, T. H., CHAN, W-Y.: Single-ended Speech Quality Measurement Using Achine Learning methods, *IEEE Audio, Speech, Language Process.,* vol. 14, No. 6, pp. 1935-1947, 2006.

[8] STELMACHOWICZ, P. G., LEWIS, D. E., HOOVER, B., KEEFE, D H.: Subjective Effects of Peak Clipping and Compression Limiting in Normal and Hearing-impaired Children and Adults, *J. Acoust Soc Am*, vol. 105, pp. 412, 1999.

[9] LICKLIDER, J. C.: Effects of Amplitude Distortion Upon the Intelligibility of Speech, *J. Acoust Soc Am,* vol. 18, No. 2, pp. 429-434, 1946.

[10] BENESTY, J., SONDHI, M. M., HUANG, Y. A. : *Springer Handbook of Speech Processing*, Springer, 2007.

[11] GOOGLE: *WebRTC NetEQ Overview*, http://www.webrtc.org/reference/architecture #TOC-NetEQ-for-Voice.

[12] RAAKE, A.: *Speech Quality of VoIP - Assessment and Prediction*, Wiley, 2006.

[13] HINES, A., SKOGLUND, J., HARTE, N., KOKARAM, A. C.: *Detection of Chopped Speech*, Patent, US20150199979 A1, 07 2015.

[14] IEEE: *IEEE Recommended Practice for Speech Quality Measurements, Audio and Electroacoustics*, IEEE Transactions on, vol. 17, No. 3, pp. 225-246, Sep 1969.

[15] HINES, A., SKOGLUND, J., KOKARAM, A. C., HARTE, N.: VISQOL: An Objective Speech Quality Model, *EURASIP J. on Audio, Speech, and Music Processing,* vol. 2015:13, May 2015.

[16] HINES, A, SKOGLUND, J, KOKARAM, A. C., HARTE, N.: *VISQOL: The Virtual Speech Quality Objective Listener,* IWAENC, 2012.

[17] HINES, A, SKOGLUND, J, KOKARAM, A. C., HARTE, N.: *Robustness of Speech Quality Metrics to Background Noise and Network Degradations: Comparing ViSQOL, PESQ and POLQA,* Acoustics, Speech, and Signal Processing, IEEE Intern. Conference on (ICASSP '13), 2013.

Oliwia Komperda - Hugh Melvin - Peter Pocta *

A BLACK BOX ANALYSIS OF WEBRTC MOUTH-TO-EAR DELAYS

Due to the recent shift toward cloud based computing, some of the world's leading standardization bodies have combined forces to provide guidelines and standards for native implementation of RealTime Communication (RTC) in the browsers. The World Wide Web Consortium (W3C) works on WebRTC APIs [1], while the Internet Engineering Task Force (IETF) is concerned with underlying standards [2]. Their efforts led to the development of WebRTC project by Google, Mozilla and Opera, which is an open framework that allows browsers to be RTC ready [3]. In this paper, we examine WebRTC from a Quality of Service (QoS) perspective, focusing on Mouth-to-Ear (M2E) delays under various application configurations.

Keywords: *VoIP, WebRTC, Skype, Mouth-to-ear M2E delay.*

1. Introduction

VoIP uses IP based networks (either public or private) in order to transmit digitized voice between two or more endpoints in real-time. The development of VoIP has eased the cost constraints of long-distance communications as well as completely revolutionized the way we think about real-time communications (RTC). Modern day RTC provide a rich variety of functions including, but not limited to multi-point voice calls, video conversations or data and screen sharing.

In May 2011 Google released a product called WebRTC under an open-source license. From that point, some of the world's leading standardization bodies have been combining forces to provide guidelines and standards for native VoIP support within browsers. The World Wide Web Consortium (W3C) works on WebRTC APIs, which define JavaScript APIs and HTML elements necessary to develop WebRTC communications. The Internet Engineering Task Force and their RTCweb (Real-time communications in WEB-browsers) team are concerned with underlying standards. The WebRTC project by Google, Mozilla and Opera is an open framework that allows browsers to be RTC ready. With only just a few lines of JavaScript code, developers can set-up a Video or Voice over IP (VoIP) sessions. The aim of this research is to quantify M2E delays introduced by WebRTC applications under varying application configuration settings. Firstly, baseline M2E delays are measured using standard WebRTC settings (i.e. Opus Codec with 20 ms packet size and 48000 sampling rate). Once the baseline delays are established, the impact of various Opus settings – namely packet size and encoding rate as well as other WebRTC provided codecs is examined.

The rest of the paper is organized as follows: In Section 2 the importance of mouth to ear delay (M2E) as a QoS metric is discussed. Section 3 briefly reviews WebRTC. Section 4 and 5 describe the experimental test-bed and experimental results. Finally, Section 6 concludes the paper and suggests some future work.

2. Importance of M2E delay as QoS metric

One of the most important aspects of Real-Time communications is Quality of Service (QoS). The internet was not designed with real-time capabilities in mind, therefore QoS support is required for some applications. This may be achieved in several ways including application and network configuration. Due to the time sensitivity of the voice packets, QoS is of exceptional importance when talking about VoIP applications, including WebRTC. Three main QoS indicators of the performance of VoIP applications are as follows: network jitter, packet loss rate and mouth-to-ear delay [4]. At the sender side, delays include encoding and packetization. These will depend on factors such as codec and frame size used. Other delays on sender side include operating system, sound card and NIC serialization delay. While being transmitted over the network, packets are subject to congestion delays at all intermediate routers as well as baseline propagation and serialization delay. At the receiving point packets experience similar delays to those at the sending

* [1]Oliwia Komperda, [1]Hugh Melvin, [2]Peter Pocta
[1]Discipline of Information Technology, College of Engineering & Informatics, National University of Ireland, Galway, Ireland
[2]Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia
E-mail: o.komperda1@nuigalway.ie

point, as well as extra delays caused by jitter buffer [5, 6 and 7]. The ITU Telecommunication Standardization Sector (ITU-T) provides some guidelines in the area of telecommunications. More specifically, ITU-T Recommendation G.114 [8] for mouth-to-ear delay outlines that delays of 0-150 ms are acceptable to most users, delays of 150-400 ms are acceptable, but will impact on the quality of call, while delays of over 400 ms are generally unacceptable. Those values are only applicable if relevant echo management techniques are employed.

Delays can have a drastic impact on the interactive nature and thus the QoS for VoIP. More generally, there is a growing awareness of the importance of deterministic timing for many diverse application areas, including real-time communication (RTC). To achieve such temporal determinism, research is necessary in many areas in order to create so-called Time-Aware Applications, Computers and Communication Systems (TAACCS). TAACCS is a recently established interest group that have identified areas requiring research as follows: clock and oscillator designs, time and frequency transfer methods, the use of timing in networking and communications systems, hardware and software architecture, design environments, and the design of applications. Current systems use timing mostly as performance metric. In order to provide higher QoS in the area of telecommunications, better time and timing awareness between hardware, software and the network is thus necessary. Further details on the TAACCS project can be found at [9].

3. WebRTC

WebRTC is an innovative approach to real-time communications. Its target is to implement real-time communication functionalities into all browsers making them accessible to developers through HTML and JavaScript. Some of the major international standardization communities are currently working on standards and guidelines for implementation of WebRTC into browsers [10]. Those standards are already being implemented by some of the browser vendors. WebRTC introduces a concept of peer-to-peer streaming into web browsers. In this new model two browsers are able to communicate directly between each other once the Session Description Protocol (SDP) offer has been negotiated. The Signaling Server is used to provide a signaling channel between the ends of the peer-to-peer connection. Communication is done using on-the-wire standard protocols, which normally use User Datagram Protocol (UDP) for transport. Real-Time Transport Protocol (RTP) and Real-Time Transport Control Protocol (RTCP) are used to provide more reliability when transporting time sensitive data over UDP [11].

4. Test Design

M2E delay is one of the crucial factors of QoS. Identifying the extent of delays introduced by WebRTC application could help pinpoint flaws in the current WebRTC implementation and identify areas of the project that require improvements. The questions addressed by this paper are as follows:

- How do various codecs impact on WebRTC M2E delay?
- For Opus, how does packetization size impact on WebRTC M2E delay?
- For Opus, how does encoding rate impact on WebRTC M2E delay?

A local server was used to enable a peer-to-peer WebRTC connection. Ping was used to estimate network RTT delay both in advance and during tests as our primary interest was in application performance. The SDP was used to modify the various application settings. An oscilloscope attached to both input and output acoustic interfaces ensured the most accurate results possible. The test bed is depicted in Fig. 1.

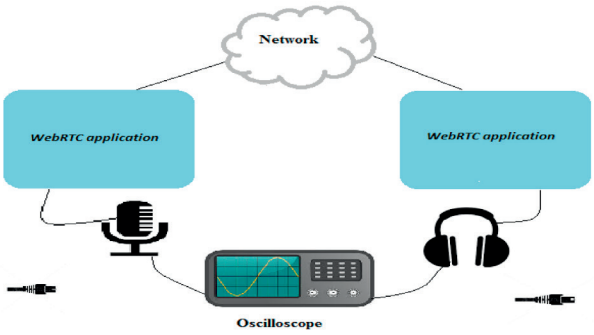


Fig. 1 Test bed

For each media/test-bed configuration, 10 VoIP calls were executed. Each call lasted between three and three and a half minutes. In each call duration, ten identifiable signals were sent between the peers, each approximately 20 seconds apart. Before and throughout each call, a ping signal was sent from one peer to the other. For each test configuration, one hundred measurements (10 x 10) and hundred ping signals (10 x 10 – 10 seconds interval) were taken for each media configuration. Two identical desktop computers were used, connected via a well provisioned (100 Mbps) wired LAN. Their specifications are as follows: Processor: AMD Athlon™ 64 X2 Dual Core Processor 4400+ 2.30 GHz. RAM: 4.00 GB, OS: Windows 7 Enterprise (64-bit). For browser, Google Canary Chrome (Version 39.0.2171.95 m) has been used.

5. Results

A. Baseline

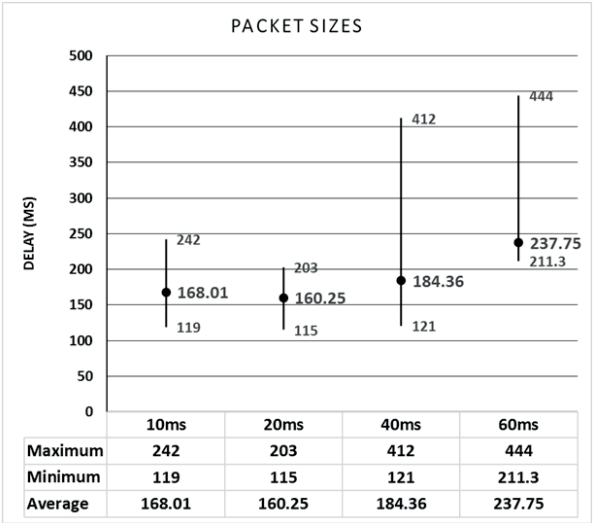
Firstly, M2E baseline delays for connections involving the default Opus codec were captured. It should be mentioned that ping results for this configuration were negligible with less than 2 ms round trip delays. Table 1 summarizes delays, outlining average and standard deviation. The experiment was performed using Voice only and then with Voice/Video call configuration. The results illustrate that in absence of significant network delays, the WebRTC default application produced M2E delays that approach the G.114 limit levels.

Summary of baseline delays Table 1

Application	Average m2e delay (MS)	Standard Deviation
WEBRTC VOICE + VIDEO	155.43	22.4235
webrtc Voice only	144.46	8.557093

B. Packetization

The influence of packetization size on M2E delay for Opus codec was then examined. Figure 2 shows average delays for 10, 20, 40 and 60 ms packets respectively, as well as the full range for each packet size. For each call, ten delay samples were taken. The Ping facility, as above, was used to measure network related delays. For all of the packet size tests, ping was negligible with values never exceeding 1ms.



	10ms	20ms	40ms	60ms
Maximum	242	203	412	444
Minimum	119	115	121	211.3
Average	168.01	160.25	184.36	237.75

Fig. 2 Packet size delays

Interestingly, a non-linear increase in M2E delays was observed as packet sizes increase, as shown in Fig. 2. Increasing packet size from 20 ms to 40 ms results in average M2E delay increase of 24 ms. This seems acceptable as an additional 20 ms of delay are added by increasing length of voice sample from 20 to 40 ms, the remaining 4 ms are possibly added due to longer processing requirements caused by encoding and transmitting of bigger packet. However – a further increase of packet size to 60 ms results in an additional 53 ms of delay - from 184 to 237 ms. Interestingly, reducing packet size from 20 ms to 10 ms did not result in M2E delay decrease. This could be caused by protocol headers overhead and the way in which WebRTC library handles 10 ms packet sizes for Opus codec. As evident from Fig. 2, a greater variation between single M2E delays for 10ms packets has been observed in comparison to that achieved for 20 ms packets.

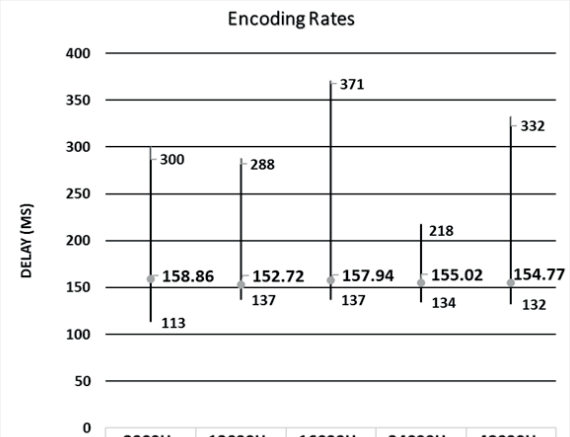
Again, in the context of G.114, it is interesting to note the non-linear impact and the extent to which application settings can greatly impact on M2E delay. In the absence of forensic under-the-hood analysis of the WebRTC codebase, the precise reasons for this non-linearity are not known. In section E below, we identify the jitter buffer behaviour as being the main contributor to this increase but also rule out sender side jitter in sending interval as a primary cause. Whilst such analysis is ongoing, we can speculate that issues such as drivers can be a contributory factor. Previous research by one of the authors has shown very unusual M2E behaviour casued by a mismatch between driver and application settings [5]. As outlined above, this reinforces and highlights the more general concerns raised by the TAACCS interest group and the consequent need for better Time Awareness.

C. Sample Rate

Besides deploying different packet sizes, Opus can also operate with a wide range of sampling rates. The sampling rate can be defined as a number of samples taken for each second of the signal. It is important to remember that Opus is a variable bit-rate codec. Bit-rate is a number of bits sent over the network in each second of the connection. Generally as sampling rate increases, greater bitrate is required to transmit data. That means that depending on network conditions and current performance, Opus can adjust encoding rate in order to maximize a quality of call [12]. The encoding rate will never exceed the maximum value as passed in SDP offer, however it may be set up to any value below it during the duration of call. Tests were performed to measure M2E delays under five encoding rates: 8000 Hz, 12000 Hz, 16000 Hz, 24000 Hz and 48000 Hz. The ping values again indicated negligible network delay with values between 0 and 3 ms.

As shown in Fig. 3, encoding rates did not have a noticeable impact on M2E delays. As expected, the bitrates present during the calls increase slightly as encoding rates rise. Opus supports bitrates of up to 510 kbps, however encoding bitrates of up to 48

kbps are mostly used for voice. During our tests, values of up to 45 kbps were observed. It is important to note that the quality of voice speech encoded at bitrates of up to 48 kbps is still very high [13].



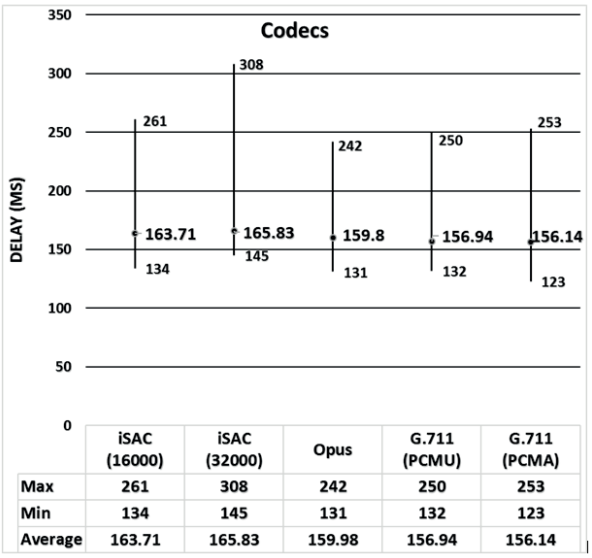
	8000Hz	12000Hz	16000Hz	24000Hz	48000Hz
Maximum	300	288	371	218	332
Minimum	113	137	137	134	132
Average	158.86	152.72	157.94	155.02	154.77

Fig. 3 Encoding Rate Delays

D. Codec

Besides Opus, WebRTC supports a number of voice codecs. Both Opus and G.711 are mandatory to implement for any WebRTC solution. This is to make sure that all the endpoints of the connection support at least one common codec. The browser generates an SDP offer contains a list of all the codecs supported on a given machine. In this way, the end points negotiate at the beginning of the communication, by comparing their SDP offers, and agreeing on a codec for voice communication. Tests were performed between two endpoints to measure M2E delays for various codecs supported by WebRTC library. Ten calls were performed for each codec and ten samples were taken during each call. Before each sample a Ping signal was sent between two endpoints to measure network delay. As before, Ping values for those tests never exceed 1 ms. In this test, the following codecs were tested: iSAC operating at 16 kHz, iSAC operating at 32 kHz, Opus operating at 48 kHz, G. 711 A-law and G.711 *μ-law operating at 8 kHz. The packet size for each codec is following: 30 ms (iSAC (16000)), 30 ms (iSAC (32000)), 20 ms (Opus), 20 ms (G.711 (PCMU)) and 20 ms (G.711 (PCMA)) respectively.*

As shown above in Fig. 4, the delays imposed by various codecs are relatively very similar, with iSAC codec shows slightly higher delays than Opus and G.711 for both sampling rates. That could be explained by the fact that iSAC operates on speech frames involving 30 ms of speech samples as oppose to speech frames of length of 20 ms used for Opus and G.711. Although the average



	iSAC (16000)	iSAC (32000)	Opus	G.711 (PCMU)	G.711 (PCMA)
Max	261	308	242	250	253
Min	134	145	131	132	123
Average	163.71	165.83	159.98	156.94	156.14

Fig. 4 Codec delays

values of M2E delay are quite similar for both Opus and iSAC, the delay range for iSAC is also greater. Opus codec achieved slightly higher average values to those reported for G. 711, however minimum and maximum values of delays for both codecs suggest that delays imposed by both of them are of the same range. Moreover it is important to note that Opus as a wideband codec is much superior to G.711 (a narrowband codec) in terms of quality of experience. Moreover G.711 performs almost identical for both versions, namely A-law and *μ-law*.

E. Jitter Buffer

The non-linear increase in M2E delay for different packet sizes outlined above encouraged the authors to investigate the jitter buffer behaviour during those calls. It is worth noting here that the jitter buffer values can be observed using chrome://webrtc-internals/. Our observations showed that when packet size is changed from the default value of 20 ms, the jitter buffer starts introducing additional delays, see Figs. 5 - 8 for more detail.

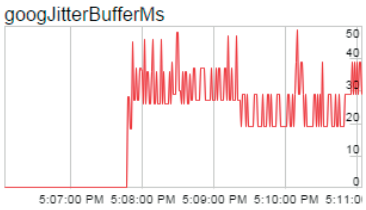


Fig. 5 Jitter buffer for 10 ms packet size

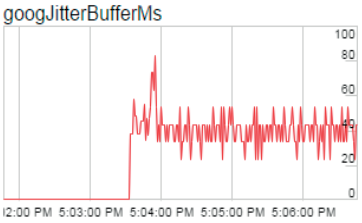


Fig. 6 Jitter buffer for 20 ms packet size

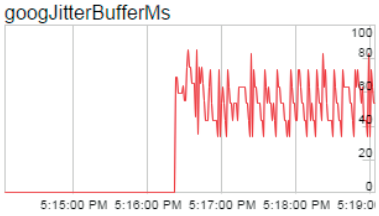


Fig. 7 Jitter buffer for 40 ms packet size

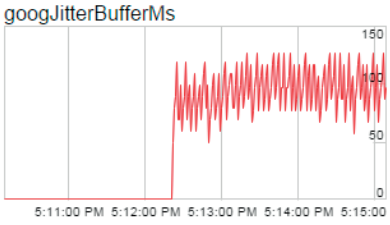


Fig. 8 Jitter Buffer for 60 ms packet size

We initially speculated that one possible reason for such jitter buffer behaviour could be irregular transmission of packets at the sending side caused by sender side non-determinism. The authors investigated the sender-side time intervals for each packet size using the Wireshark network analyzer. We concluded that with exception of a few out of place packets, packets are being transmitted relatively regularly from the sending side, regardless of packetization. There is some jitter in the data - however as the issue occurs for all the packet sizes, it is clear that sending time intervals are not the source, and thus the cause of jitter buffer behaviour lies elsewhere. We plan to investigate this issue as a further step in this research.

6. Conclusions and further work

Overall, some very interesting results in terms of M2E delays of WebRTC have been presented in this paper. The results firstly showed that baseline delays for WebRTC in presence of minimal network delay and jitter (treated as negligible) resulted in delays very close to or greater than those defined in ITU-T Rec. G.114. This is of significant concern for QoS/QoE as network delay can easily introduce 10-100 ms baseline increase of delay with jitter adding further due to jitter buffer behaviour. Regarding other tests, the sampling rate and in turn bitrate for Opus codec has no significant impact on M2E delays that occur during WebRTC calls. However, the experiments with packet sizes showed its significant impact on M2E delay. The packet size of 20 ms has been identified as the preferable choice, as it resulted in the lowest M2E delays. For a packet size increase to 60 ms, the corresponding M2E delay increase is of particular note. Moreover the operation of jitter buffer has been identified as a source of additional delay. While the jitter buffer for 20 ms packets tends to stay within 30-50 ms range, once the packet size increases to 60 ms jitter buffer values increase to between 60 and 130 ms. The source of such behaviour was not evident from an analysis of sender side packets, and further research is needed.

The study also showed that codec choice had little impact on M2E delay. Delays introduced by iSAC codec for both sampling rates were slightly higher than those achieved for Opus as well as G.711 though packetisation size is the obvious reason for this.

In conclusion, the tests emphasize the need for RTC application developers to ensure Time Awareness in the design and implementation of RTC applications. More broadly, these test results highlight the essential message of the TAACCS project whereby the full chain of hardware and software need to be better integrated and Time Aware so as to provide better guarantees and temporal determinism.

Acknowledgement

This work has been partially supported by the ICT COST Action IC1304 - Autonomous Control for a Reliable Internet of Services (ACROSS), November 14, 2013 - November 13, 2017, funded by European Union.

References

[1] W3C, World Wide Web Consortium, no date, accessed on 04/12/14 from http://www.w3.org/

[2] IETF, International Engineering Task Force, no date, accessed on 03/12/14 from https://www.ietf.org/

[3] WebRTC.org, WebRTC project, accessed on 09/10/14 from www.webrtc.org

[4] KARAPANTAZIS, S., PAVLIDOU, F-N.: VoIP: A Comprehensive Survey on Promising Technology, *Computer Networks*, 2009, vol. 53(12), pp.2050-2090, accessed on 18/10/14 from: http://www.sciencedirect.com.libgate.library.nuigalway.ie/science/article/pii/ S1389128 609001200

[5] MELVIN, H.: *The Use of Synchronized Time in Voice over Internet Protocol (VoIP) Applications,* Unpublished doctoral dissertation, University College Dublin, accessed on 16/11/15 from http://www.csi.ucd.ie/staff/jmurphy/publications/thesis-hugh.pdf, 2004

[6] MARKOPOULOU, A. P., TOBAGI, F. A., KARAM, M. J.: *Assessment of VoIP Quality over Internet Backbones,* Proc. 2002, vol. 1, pp. 150-159, accessed on 18/01/15 from http://ieeexplore.ieee.org.libgate.library.nuigalway.ie/xpl/articleDetails.jsp?arnumber=1019256

[7] CISCO.COM: Understanding Delay in Packet Voice Networks, February 2006, accessed on 24/11/2014 from http://www.cisco.com/c/en/us/support/docs/voice/voice-quality/5125-delay-details.html

[8] ITU-T, Recommendation G.114: One way transmission time, 2003, Accessed on 07/11/14 from http://www.itu.int/rec/T-REC-G.114-200305-I/en

[9] WEISS, M., EIDSON, J., BARRY, CH., BROMAN, D., GOLDIN, L., BOB IANNUCCI, LEE, E. A., STANTON, K.: *Time-Aware Applications, Computers, and Communication Systems (TAACCS),* NIST Technical Note 1867. NIST, National Institute of Standards and Technology, http://dx.doi.org/10.6028/NIST.TN.1867, U.S. Department of Commerce, February 2015.

[10] ALVESTRAND, H.: *Overview: Real Time Protocols for Browser-based Applications Draft-ietf-rtcweb-ovieview-13*, November 28 2014, accessed on 12/02/15 from https://tools.ietf.org/html/draft-ietf-rtcweb-overview-13

[11] JOHNSTON, A. B., BURNETT, D. C: *WebRTC, APIs and RTCWEB. Protocols of the HTML5,* Real-Time Web, Digital Codex LLC, St. Louis, 2014

[12] VOS, K., VALIN, J.: RTP *Payload Format for Opus Speech and Audio Codec Draft-ietf-payload-rtp-opus-11*, April 2015, accessed on 20/05/15 from https://tools.ietf.org/ html/draft-ietf-payload-rtp-opus-11

[13] VALIN, J.-M., TERRIBERRY, T. B, MONTGOMERY, C., MAXWELL, G.: A High Quality Speech and Audio Codec with Less Than 10ms Delay, *IEEE Transaction on Audio, Speech, and Language Processing*, vol. 18, No. 1, 2010, accessed on 01/07/15 from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4926218.

Yusuf Cinar - Hugh Melvin - Peter Pocta *

A BLACK-BOX ANALYSIS OF THE EXTENT OF TIME-SCALE MODIFICATION INTRODUCED BY WEBRTC ADAPTIVE JITTER BUFFER AND ITS IMPACT ON LISTENING SPEECH QUALITY

WebRTC is an open-source platform for real-time communications over the web and has been experiencing widespread adoption in recent years. WebRTC clients employ the technique of time scaling of packets to cope with the impact of network jitter and/or clock skew. A black-box study presented in this paper focuses on two aspects, namely time scale modification behaviour under different packet arrival interval and its impact on the listening quality perceived by the end user. Specifically, we examine the MOS scores predicted by the POLQA speech quality prediction model. Our tests involve both iSAC and Opus codecs, two of the widely used WebRTC codecs. In the experiment, a speech file played from one client, is directed through a network simulation before reaching the receiving client. Our results surprisingly show that the extent of time scaling is consistently higher for Opus producing shorter speech files. Regarding the consequent impact on quality, we also find that there are many cases where POLQA is reporting MOS predictions that contradict expert listener assessments.

Keywords: *WebRTC, jitter, adaptive playout, time-scale modification.*

1. Introduction

WebRTC is one of the latest developments in the area of multimedia real-time communications (RTC) and it is a set of standards from WC3 [1] and IETF [2] that enables real-time communication on the web. WebRTC has drawn significant interest from not only browser vendors but also application and web developers due to the potential new services it can offer [3]. There is an open source project, with the same name WebRTC [4], which implements the standards and is used by the browser vendors and application developers. Traditional VoIP components, such as audio coding modules, jitter buffer, play-out decision and codec implementations are integral parts of WebRTC project.

There are numerous dynamics that can influence the quality experienced by the end user for a voice call session. The non-deterministic nature of best-effort Internet causes many network impairments, such as network delay variations (jitter) and consequent packet bursts, as well as packet loss. In particular, jitter results in voice packets arriving at irregular intervals to the receiver. In order to maintain the speech intelligibility and quality for the listener, the voice stream must be reconstructed in a similar manner to which it was created. Due to the presence of network jitter and congestion, the voice packets are typically held in the receiver jitter buffer before they are played out in a way that sustains the conversational and listening quality at a certain level. WebRTC has a component called NetEQ for this purpose [4] describes NetEQ as: "A dynamic jitter buffer and error concealment algorithm used for concealing the negative effects of network jitter and packet loss. It aims to keep latency as low as possible while maintaining the highest voice quality." As a by-product, it also deals with clock skew issues between sender and receiver clients.

When playing out the speech from the jitter buffer, the receiver's playout strategy tries to adapt to the changing network conditions. Adaptive playout techniques are grouped in two categories, per packet and per talkspurt, as mostly referenced in the literature [5] and [6]. Per talkspurt techniques apply adjustments only to silence periods between talkspurts. However, the latter technique, which is also referred to as time scale modification, applies compression or expansion to all the packets regardless of silence or voice segment.

Literature to date had focused mostly on the impact of the degradations caused by per-talkspurt playout strategies on the listening quality. A comprehensive study was conducted by some of the authors in [5], and has shown that small playout

* [1]Yusuf Cinar, [1]Hugh Melvin, [2]Peter Pocta
[1]Discipline of Information Technology, College of Engineering & Informatics, National University of Ireland, Galway, Ireland
[2]Dept. of Telecommunications and Multimedia, Faculty of Electrotechnical Engineering, University of Zilina, Slovakia
E-mail: cinar.yusuf@gmail.com

delay adjustments of 30 ms or less introduced to silence periods is negligible according to subjective listening quality scores and POLQA predictions. On the other hand, the PESQ model predicts contradicting scores to the scores obtained from the subjective test; hence they conclude PESQ fails to correctly predict quality scores with those adjustments. It is worth noting here that [5] does not study the impacts of per packet time-scaling modifications.

The fundamental idea of adaptive playout mechanism via time scale modification was first introduced separately by Liu et al. [7] and Liang et al. [8] in 2001, according to [9]. A further study by Liu et al. [9] in 2002 investigated the stretching-ratio transition effect on perceived audio quality by measuring the objective PESQ MOS, and found that PESQ, which was the most up to date ITU standard in the area of objective speech quality assessment at the time, does not provide a good objective quality measure for packet-based time-scale modified speech signals. However, this study was carried out more than a decade ago and doesn't cover more recent methods such as POLQA.

In [10], a more recent study has shown via extensive experiments that POLQA can predict the quality with high accuracy for different sampling rate adjustments, (referred to as time scale modification in [11]), if the total range of sample rate deviations is +/-3% of the nominal sample rate. It is also claimed here that POLQA predictions start failing after 6% but no data was presented to support that. There is also no detailed description of how the time scale modification was applied to the speech signals. Firstly, although not explicitly mentioned, it is implicitly indicated that a constant rate between -3% and 2.9% of sampling rate difference is applied to the entire speech which in a real world scenario is not likely. For instance, our experiments show that WebRTC applies a variable rate of time scaling per speech frame. Liu et al. [9] use the term 'stretching dynamic' to describe how flexible the time scale modification ratio can change from one segment to the next, within constraint that audio quality is not sacrificed. Secondly, a simple change in the sampling rate doesn't reflect the true nature of time scale modification that is typically employed in VoIP as can be seen in [7] and [8]. In our experiments, we study simulated WebRTC VoIP calls, where variable rates of time scale modification are applied by the playout algorithm, which better reflects the reality.

It is also worth noting that subjective listening tests [12] or objective methods such as PESQ [13] and POLQA [10] do not consider the effect of mouth-to-ear delay. Hence, the ITU-T E-model [14] is also an important method of validation since it provides a direct link to perceived conversational speech quality by estimating user satisfaction from the combined effect of information loss, delay and echo. This method has been widely used by researchers including one of the authors [11] to evaluate playout buffer schemes.

Based on the above summary, we conclude that regarding per-packet algorithms, aforementioned gaps exist. Therefore, our objective here is to examine the quality impacts of WebRTC's time scaling algorithm, under different codec settings and network conditions.

1.1. Research motivation

The literature to date has not examined the performance of time scale modification of adaptive jitter buffer algorithm employed by the NetEQ component of WebRTC project. This prompted us to undertake this study as WebRTC is available to billions of devices [15].

In advance of the research, there was little published about the internal workings of WebRTC' adaptive jitter strategy. Although several points were made by Hines et al. [16], such as that it accommodates time scale modifications, we believed more characterisation would help further studies.

Furthermore, because WebRTC supports many voice codecs, we wished to see the quality implications of choosing one codec over another when used with same time scale modification technique at the adaptive jitter buffer algorithm. To our knowledge, a comparative performance analysis of playout buffer algorithm coupled with codec has not yet been conducted.

In the context of research motivation above, the following key research questions are identified:

1. What are the core characteristics of the time scaling adaptive jitter buffer algorithm employed in WebRTC?
2. What impact does time scale modification have on the perceived listening quality – both via objective and expert subjective evaluations?

This paper is organised as follows. Section 2 outlines the experiment methodology. Results are presented in Section 3. Section 4 concludes the paper and suggests future work.

2. Methodology

2.1. Testbed Description

The test-bed, illustrated in Fig. 1, involves a sender and receiver side of WebRTC communication chain and an emulated network channel in between, all hosted on a single host machine. Sender represents a speaker of a virtual conversation while receiver represents a listener. The channel emulates the IP network carrying one-way voice packets from the sender to receiver. Our objective was to simulate severe network jitter leading to packets bursts arriving at the receiver, and examine the associated jitter buffer response.

An important note is that the default WebRTC jitter buffer can hold up to 50 packets, i.e. a maximum buffer size at the receiver is hardcoded to 50. Since our simulation results in extreme packet bursts, the buffer reaches its threshold quickly and starts dropping packets. Hence we modified WebRTC's jitter buffer maximum capacity to 500 packets to avoid dropping packets. This allows us to focus on time scaling only.

As illustrated in Fig. 1, the sender reads a PCM file, which contains speech ranging from 8 to 10 seconds including silent periods, and encodes and emits the data to the network channel in the form of RTP packets. The experiment is designed to accumulate all the packets generated in the network channel before dispatching them to the receiver.

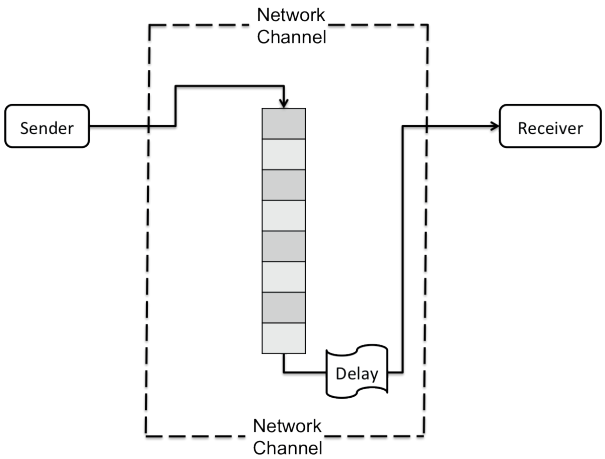


Fig. 1 Simulation Design

Once all the packets are in the channel, i.e. the representation of network in our application, we then dispatch each packet according to the delay profiles of the experiment. This is done by a Delay component displayed in Fig. 1. For instance, one delay profile is to dispatch each packet to the receiver at an interval of 15 ms.

The test is automated end to end and there are several configurations, i.e. test conditions, which can be controlled. As shown in Table 1, we control the packet arrival interval to the receiver buffer. Also, a codec can be automatically switched between Opus and iSAC. The WebRTC project includes many codecs including but not limited to Opus, iSAC, G.711, G.722. In our experiment, iSAC and Opus were only deployed as they represent currently the mostly deployed codecs in WebRTC. Regarding the packet arrival to the receiver, the following values, 10, 15, and 20 ms, were used in our experiments. However, in real world scenarios, these values would not be constant as each packet delay may vary. Therefore, we plan to run a further study involving real world measurements. It is worth noting here that iSAC uses speech frames of length of 30ms and the WebRTC Opus implementation deploys frames of length 20 ms by default. The values 10 and 15 ms were chosen as they represent 50% of the default packetisation for Opus and iSAC respectively, whereas 20 ms was chosen to see the impact for Opus under ideal conditions.

Test Conditions Table 1

Configuration	Values
Arrival interval	10 | 15 | 20 (ms)
Codec	Opus | iSAC

2.2. Speech Quality Assessment

In order to assess how time scale modification impacts on listening quality experienced by the end user of WebRTC, we used both an objective method based on perceptual modelling and expert subjective listening. The objective method employed is POLQA version 1.1 (Perceptual Objective Listening Quality Assessment, ITU-T Recommendation P.863) [17]. We then compared and contrasted the POLQA MOS scores with an expert listener.

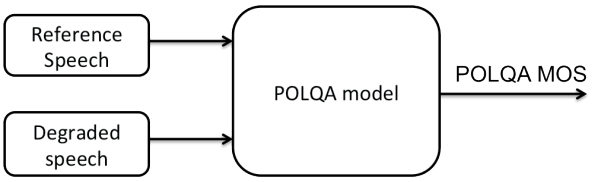


Fig. 2 Objective Listening Quality Tests

As the incoming packets are played out at the receiver, the output is recorded into a file for quality assessment. The reference speech file and recorded one (degraded speech file) are together passed into executable that represent POLQA prediction model to generate MOS scores for the corresponding speech sample as depicted in Fig. 2.

2.3. Speech Samples

We used 25 certified speech samples in the experiments. The samples are taken from SWB conformance databases included in the ITU Recommendation P.863 [17]. Speech files consist of a pair of utterances from 14 male and 11 female speakers with a pause in between. Durations of the speech files are between 8 and 10 seconds and stored in 16 bits, 48000 Hz linear PCM.

3. Results

In this section, we present results and analysis of the black box experiments carried out as well as observations discovered related to the voice aspects of WebRTC. Results can also be found on a website [18] with further details including interactive charts and waveforms.

3.1. General Analysis

We use the term acceleration rate to refer to the duration difference between original speech and degraded speech. We calculate the acceleration rate according to:

$$Ar = \left(1 + \frac{Td}{To}\right) * 100$$

where *Ar* is the acceleration rate, *Td* is duration of degraded speech file and *To* is the duration of original speech file.

Fig. 3 and Fig. 4 separately show the acceleration rates for two of the samples used in experiments. Most of the acceleration rates are close to the results in Fig. 3, lying between 15% and 25% for a packet arrival of 10ms and 15ms. However, there are some samples, which produced extremely low acceleration rates, as shown in Fig. 4, i.e. between 3% and 5% for the same packet arrival configurations.

One key finding is that the Opus codec consistently accelerates (scales) more than iSAC for packet arrival intervals tested of 10 ms or 15 ms as can be seen in Fig. 3. With a packet arrival interval of 20 ms, we expected not to see any acceleration as the default Opus packets in WebRTC are 20 ms. Interestingly, results show that Opus applies acceleration, which shortens the playout duration by about 2% to 3%, for all of the 25 samples used in the experiments. Figure 4 shows an example for one of the samples. It can be clearly seen from Fig. 4 that the acceleration rate is 2% at 20 ms arrival rate.

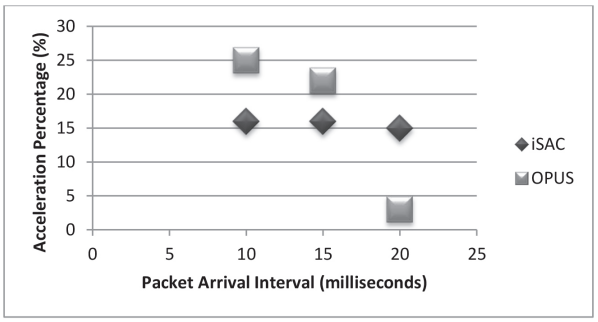


Fig. 3 Acceleration Percentage vs Packet arrival interval - file: CH_m1_s4_file_18.48k.pcm

An investigation of the time scale modification code of NetEQ algorithm in WebRTC showed us that it is codec agnostic. Therefore, it is an interesting outcome that Opus accelerating % is higher than iSAC for 10 and 15 ms arrival interval. This is especially interesting as iSAC has a larger packet size of 30 ms and thus 10/15 ms arrival interval represents a more extreme network burst scenario. Further investigation is needed to explain this behaviour in more detail.

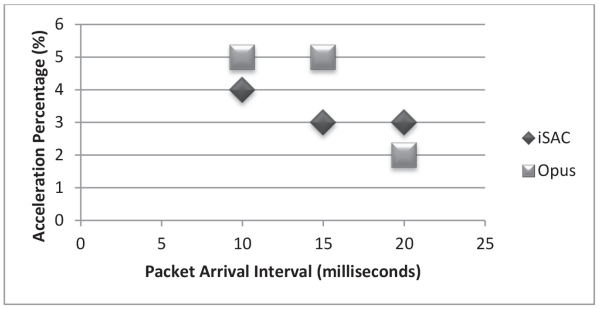


Fig. 4 Acceleration Percentage vs Packet arrival interval - file: CH_m2_s4_file_25.48k.pcm

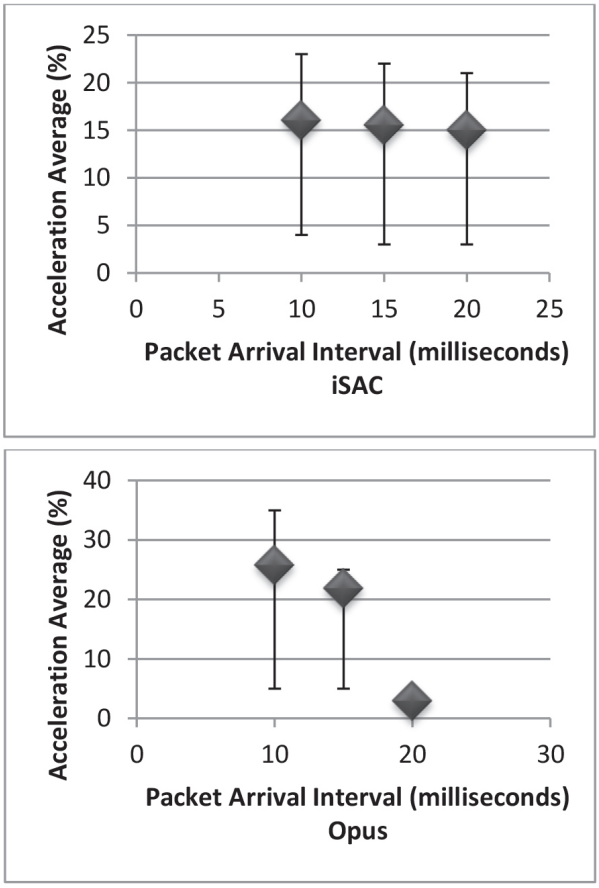


Fig. 5 Average Acceleration for each Packet Arrival

Figure 5 presents the average of the accelerations obtained for all the involved files applied by each codec per packet arrival interval. It is clearly depicted here that Opus applies more acceleration than iSAC, except unsurprisingly, for 20ms packet arrival. It is also quite interesting that whilst for Opus, there is a trend of decreased acceleration as interval increases from 10 to 15, for iSAC there is little variation in acceleration % across all 3 arrival settings.

3.2. Listening Quality Assessment and Time Scaling

We carefully listened to the each degraded speech file in order to see how perceptible are the degradations introduced by the time scaling algorithm of WebRTC. We found that acceleration higher than 15% is evident and perceptible. Moreover, we found that it is only marginally impacting the intelligibility and listening quality. On the other hand, a lower acceleration, i.e. less than 15%, is hardly perceptible. Overall, we think that if the samples undergo a formal subjective listening test, most, if not all, of the samples would be rated as 4 MOS, or close to it.

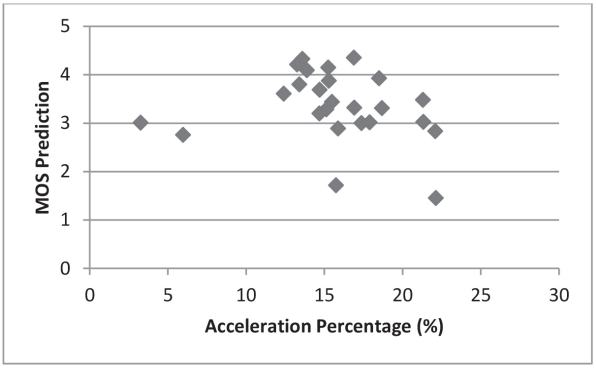


Fig. 6 POLQA MOS vs Acceleration rate - iSAC

We performed an objective quality assessment on the produced speech files and obtained POLQA predictions for all the files involved in this study. We obtained wide-ranging results and they mostly contradict the results from expert listening and thus the results we would expect from formal subjective listening tests according to [16].

Figure 6 shows the POLQA results for iSAC and Fig. 7 presents the results for Opus for a packet arrival interval of 15ms. We can clearly see from both figures that POLQA does not correctly predict quality for the samples degraded by higher acceleration. It is again interesting to note that despite fixed arrival rates set to half the packetisation rate for iSAC (15 ms vs. 30 ms) and 3/4 for Opus (15 ms vs. 20 ms) for all the samples, the acceleration % ranged from 5 to more than 20 depending on sample, with a significant cluster around 12-22 for iSAC and 15-25 for Opus.

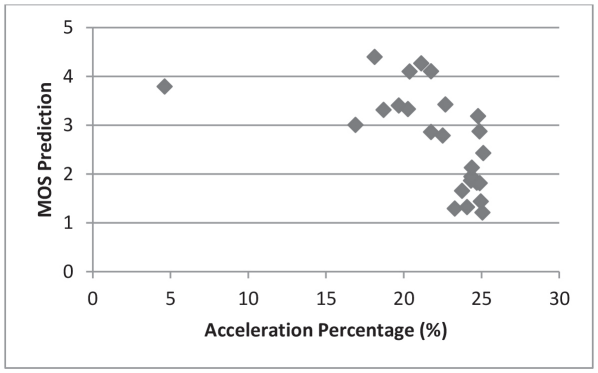


Fig. 7 POLQA MOS vs Acceleration rate - Opus

4. Conclusion and Future Work

In this paper, we have investigated through black-box testing, time scale modification deployed in the adaptive jitter buffer algorithm of WebRTC with a focus on two aspects. The first is the extent of acceleration/scaling for two codecs, Opus and iSAC, under different packet arrival intervals and secondly, the consequent impact on listening quality experienced by the end user. We examined the MOS scores predicted by POLQA and also executed expert listening tests. Two questions, as outlined in Section 1.1, are addressed in this study.

Regarding the first question, we observed that the default WebRTC employs a time scale modification algorithm to provide an adaptive jitter buffer mechanism. Its maximum buffer size is set to 50 packets and if it receives a new packet while it is full, it is designed to drop all of the 50 packets. The extent of scaling/acceleration was seen to differ greatly depending on codec and speech signal. We saw a higher acceleration ratio for Opus even though the simulated network bursts were more extreme for iSAC with a packetisation interval of 30ms. Also, the acceleration varied hugely across the samples even when network burst ratio and codec were kept constant. The reasons behind this behaviour require further research.

Addressing the second question, our results show that POLQA (version 1.1) produces very varied results at the level of acceleration used in this research. We did not observe any correlation in the results achieved. Related research shows that POLQA can deal with scaling represented by a simple resampling of the speech signal as long as it is less than 3%. Our research applies extreme packet bursts and results in much more significant time scaling, up to around 25%. On the other hand, our informal subjective tests performed by the authors, did not reveal such degradations in the perceived quality. Therefore we conclude that the MOS-LOQ scores predicted by POLQA model do not reflect subjective results. Moreover, it is worth noting here that inability of POLQA to provide reliable results at the acceleration levels

used in this research is not caused at all by the modified buffer size deployed in this study.

Finally, whilst extreme, it is theoretically possible to see situations where there is a very high arrival rate of packets into the jitter buffer under certain network conditions for a period of time. The approach taken in this study, which is to increase the buffer size and sending packets faster than packet spacing at the origin, is an attempt to simulate those conditions. This simulation is a first attempt and we plan to emulate a broader range of real world network conditions in future work.

Acknowledgement

This work has been partially supported by the ICT COST Action IC1304 - Autonomous Control for a Reliable Internet of Services (ACROSS), November 14, 2013 – November 13, 2017, funded by European Union.

References

[1] BERGKVIST, A., BURNETT, D., JENNINGS, C., NARAYANAN, A.: *WebRTC 1.0: Real-time Communication Between Browsers. W3C Editor's Draft, W3C*. Retrieved from W3C: http://www.w3.org/TR/webrtc/, 2012.

[2] IETF. (n.d.). *Real-Time Communication in WEB-browsers*. Retrieved from IETF: https://tools.ietf.org/wg/rtcweb/

[3] CINAR, Y.: *An Objective Black-box Evaluation of Voice Quality within the WebRTC Project in Presence of Network Jitter,* 2013.

[4] WebRTC. (n.d.). *WebRTC*. Retrieved from www.webrtc.org

[5] POCTA, P., MELVIN, H., HINES, A.: An Analysis of the Impact of Playout Delay Adjustments introduced by VoIP Jitter Buffers on Listening Speech Quality. *Acta Acustica united with Acustica , 101* (3), 616-631, 2015.

[6] MOON, S. B., KUROSE, J., TOWSLEY, D.: Packet Audio Playout Delay Adjustment: Performance Bounds and Algorithms. *Multimedia systems* , 2/1, 17-28, 1998.

[7] LIU, F., KIM, J., KUO, C-C. J.: *Adaptive Delay Concealment for Internet Voice Applications with Packet-based Time-scale Modification. IEEE ICASSP'2001.* IEEE, 2001.

[8] LIANG, Y. J., FARBER, N., GIROD, B.: *Adaptive Playout Scheduling Using Time-scale Modification in Packet Voice Communications.* Acoustics, Speech, and Signal Processing, 2001. Proc. of (ICASSP'01), 2001 IEEE International Conference on. 3, pp. 1445-1448, IEEE.

[9] LIU, F., KIM, J., KUO, C-C. J.: *Quality Enhancement of Packet Audio with Time-scale Modification.* ITCom 2002: The Convergence of Information Technologies and Communications, pp. 163-173. International Society for Optics and Photonics, 2002.

[10] SCHMIDMER, C.: *POLQA Characterization for Time Scaling Conditions,* ITU-T, 2011.

[11] MELVIN, H.: *The Use of Synchronised Time in Voice over IP (VoIP) Applications*. PhD Thesis, University College Dublin, October 2004.

[12] International Telecommunications Union: *ITU-T Rec. P.800: Methods for subjective determination of transmission quality.* Geneva, 1996.

[13] International Telecommunications Union: *ITU-T Rec. P.862: Perceptual evaluation of speech qaulity (PESQ).* Geneva, 2001.

[14] International Telecommunication Union: *ITU-T Rec. G.107: The E-Model: a computational model for use in transmission planning.* Geneva, 2009.

[15] TechCrunch. (n.d.): *TechCrunch*. Retrieved from The WebRTC Race Begins Today: http://techcrunch.com/2015/02/28/1123773/

[16] HINES, A., SKOGLUND, J., KOKARAM, A. C., HARTE, N.: ViSQOL: An Objective Speech Quality Model. *EURASIP J. on Audio, Speech, and Music Processing,* 5/17, 2015. [17] International Telecommunications Union: *ITU-T Rec. 863: Perceptual objective listening quality assessment.* Geneva, 2011.

[18] *WebRTC Quality of Experience*. Retrieved from webrtcquality.cloudapp.net.

Jozef Polacky - Peter Pocta - Roman Jarina *

AN IMPACT OF NARROWBAND SPEECH CODEC MISMATCH ON A PERFORMANCE OF GMM-UBM SPEAKER RECOGNITION OVER TELECOMMUNICATION CHANNEL

The automatic identification of person's identity from their voice is a part of modern telecommunication services. In order to execute the identification task, speech signal has to be transmitted to a remote server. So a performance of the recognition/identification system can be influenced by various distortions that occur when transmitting speech signal through a communication channel. This paper studies an effect of telecommunication channel, particularly commonly used narrowband (NB) speech codecs in current telecommunication networks, on a performance of automatic speaker recognition in the context of a channel/codec mismatch between enrollment and test utterances. An influence of speech coding on speaker identification is assessed by using the reference GMM-UBM method. The results show that the partially mismatched scenario offers better results than the fully matched scenario when speaker recognition is done on speech utterances degraded by the different NB codecs. Moreover, deploying EVS and G.711 codecs in a training process of the recognition system provides the best success rate in the fully mismatched scenario. It should be noted here that the both EVS and G.711codecs offer the best speech quality among the codecs deployed in this study. This finding also fully corresponds with the finding presented by Janicki & Staroszczyk in [1] focusing on other speech codecs.

Keywords: *Speaker identification, GMM-UBM, MFCC features, TIMIT, Speech codecs, Narrowband voice transmission.*

1. Introduction

Over the past decades, Automatic Speaker Recognition (ASR) has become a very popular area of research in pattern recognition and machine learning. Scientists from around the world have been constantly working on improving speaker recognition systems and have also been looking for more effective procedures, which increase the actual recognition rate. ASR is a general term for both speaker identification and speaker verification tasks. A principle of a speaker identification and verification is displayed in Figs. 1 and 2 respectively.

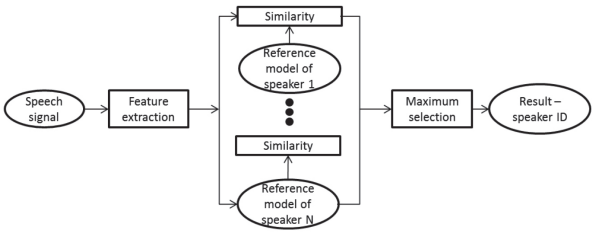


Fig. 1 Speaker identification

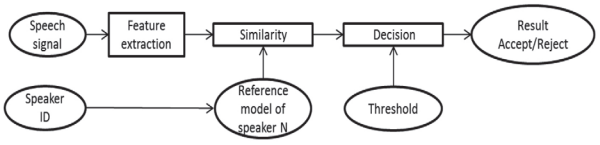


Fig. 2 Speaker verification

ASR technique can be used to verify speaker´s identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information and remote access to computers. Most of the above listed applications require a transmission of the user's voice to the remote server that executes an identity validation. Gaussian Mixture Model -Universal Background Model (GMM-UBM) technique represents the currently most popular technique for this task. Its performance has been tested under different conditions in [2 - 4]. ASR research is currently also focused on reducing the within-speaker variations which are caused by a channel mismatch. The channel mismatch occurs when the utterances for enrollment and the utterances for testing are transmitted through channels with different characteristics.

* **Jozef Polacky, Peter Pocta, Roman Jarina**
Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia
E-mail: jozef.polacky@fel.uniza.sk

A speech compression plays a significant role in mobile communication, Voice Over Internet Protocol (VoIP), voicemail and gaming communication. In all of the above mentioned cases, lossy speech codecs are deployed. A purpose of speech codec is to compress a speech signal by reducing a number of bits needed for its transmission while maintaining the intelligibility of speech once decoded. The distortions introduced by speech codecs may have a significant impact on a performance of speaker recognition system. An importance of this problem even increases when it comes to the channel mismatch as there is a huge difference between an impact of different codecs deployed in current telecommunication networks on a performance of speaker recognition system. Moreover, it is worth noting here that codec degradations are currently considered as one of the most prominent degradations encountered in current telecommunication networks. Therefore, an analysis of codec-induced degradations in the context of speaker recognition and development of the ASR techniques that are robust against this type of degradations, are of great interest to researchers around the world.

An effect of codec mismatch on a performance of a speaker recognition system has been investigated with different classifiers such as Hidden Markov Models (HMM) [5], GMM-UBM [6], Support Vector Machine (SVM) systems [1] and i-vector techniques [7]. The experiments published in [5] have revealed a significant degradation of performance under mismatched conditions: Pulse-code modulation (PCM) data–trained models vs. Code-excited linera prediction (CELP) coded test data. Finally, two techniques for improving the performance in these situations were examined, namely the maximum a posteriori (MAP) adaptation strategy and the Affine transform strategy. Significant improvements in the performance over mismatched conditions for both cases were achieved. In [6], the authors have demonstrated that a performance of speaker verification system based on GMM-UBM decreases, for all the codecs involved in the study, as the degree of mismatch between training and testing conditions increases. Both the fully matched and mismatched conditions have been investigated in [1] deploying SVM. In the mismatched conditions, Speex codec was shown to perform best for creating robust speaker models. The authors in [7] have focused on the benefits provided by an extended bandwidth used for voice communication in the context of codec mismatch and bandwidth mismatch, using i-vector approach for speaker recognition. Their results show that the performance of ASR on wideband data is significantly better than that employing narrowband (NB) signals for both matched and codec-mismatched conditions.

In this paper, we also focus on the influence of codec mismatch between an enrollment and test utterances. In comparison to the previous studies [1, 5, 6 and 7], we have expanded a set of codecs used for an investigation in terms of their setups, and types of signal degradations introduced by the codecs. The codecs under the tests represent NB speech codecs commonly implemented in current telecommunication networks. We also included a brand new 3GPP EVS codec standardized recently [8]. As a back-end we have applied GMM-UBM classifier as it can provide high recognition rate and is very easy-to-implement [9 and 10]. It overcomes also more sophisticated i-vector approach if only constrained amount of training data is available [11].

The rest of the paper is organized as follows: In Section 2, GMM-UBM approach for speaker's enrollment in speaker identification system is presented. Section 3 describes experiment and experimental results. Finally, Section 4 concludes the paper and suggests a future work.

2. GMM-UBM identification approach

GMMs used in combination with MAP adaptation represent the main technology of most of the state-of-the-art text-independent speaker recognition systems [9 and 12]. GMM based speaker-specific models are derived from Universal Background Model - i.e. a generic speaker-independent GMM statistical model, which has been trained on a great amount of multi-speaker speech data.

2.1 Gaussian Mixture Model

Gaussian Mixture Model (GMM) [13] is a stochastic model, which can be considered as a reference method for speaker recognition. The Gaussian mixture probability density function of model λ consists of a sum of K weighted component densities, given by the following equation:

$$p(x \mid \lambda) = \sum_{k=1}^{K} P_k N(x \mid \mu_k, \Sigma_k) \tag{1}$$

where K is the number of Gaussian components, P_k is the prior probability (mixture weight) of the k-th Gaussian component, and

$$N(x \mid \mu_k, \Sigma_k) = \\ = (2\pi)^{-\frac{d}{2}} \mid \Sigma_k \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\} \tag{2}$$

is the d-variate Gaussian density function with mean vector μ_k and covariance matrix $\sum_k$. The prior probabilities $P_k \geq 0$ are constrained as $\sum_{k=1}^{K} P_k = 1$.

For numerical and computational reasons, the covariance matrices of the GMM are usually diagonal. Training a GMM consists of estimating the parameters $\lambda = \{P_k, \mu_k, \Sigma_k\}_{k=1}^{K}$ from a training sample $X = \{\vec{x}_1, ... \vec{x}_T\}$. The basic approach is to maximize likelihood of X with respect to model λ defined as:

$$p(X \mid \lambda) = \prod_{t=1}^{T} p(\vec{x}_t \mid \lambda) \quad (3)$$

The goal is to obtain Maximum-likelihood (ML) parameter estimation. The process is an iterative calculation called the Expectation-Maximization (EM) algorithm [14]. Note that K-means [15] can be used as an initialization method for EM algorithm.

In the identification process, a set of test utterances and its model is compared with each model of the training database. From each comparison between test and training model is obtained a likelihood and the model with the highest score corresponds to the unknown speaker.

Let us assume a group of speakers S_p=*1,2,...S* represented by GMM´s $\lambda_1, \lambda_2, ...\lambda_S$. Unknown speaker model is identified to each model:

$$\hat{S} = \arg\max_{1 \le k \le S} \sum_{t=1}^{T} \log(p(\vec{x}_t \mid \lambda_k)) \quad (4)$$

2.2 Universal Background Model

Universal Background Model (UBM) is an improvement in the field of speaker recognition using GMM. It is typically characterized as a single Gaussian Mixture Model trained with a large set of speakers using the EM algorithm.

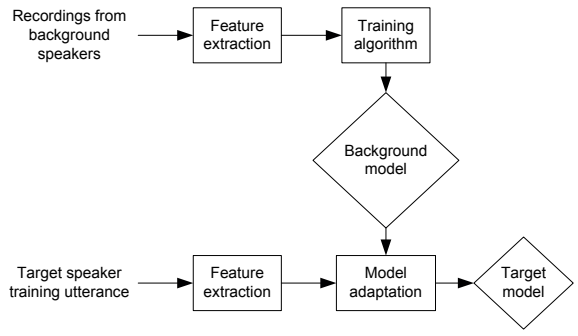


Fig. 3 UBM adaptation based speaker enrollment

As shown in Fig. 3 UBM is used as an initial model for training speaker-specific GMM during speaker enrollment. This process prevents from need for estimating the parameters of the speaker model from scratch. There are multiple ways how to adapt the UBM. It is possible to adapt one or more of its parameters as well as all parameters. Adaptation of the parameters is usually done using MAP technique. The background model (UBM) must be built from utterances with common characteristics in the meaning of type and quality of speech, uttered by great number of speakers. For example, an identification/verification system that uses only telephone channel and female speakers must be trained using only telephone speech spoken by female speakers. For a system where the gender composition is an unknown parameter, the model will be trained using both male and female utterances.

The following average log-likelihood formula gives the final score in recognition process [16]:

$$LLR(X, \lambda_{target}, \lambda_{UBM}) = \\ = \frac{1}{T} \sum_{t=1}^{T} \{\log(p(\vec{x}_t \mid \lambda_{target})) - \log(p(\vec{x}_t \mid \lambda_{UBM}))\} \quad (5)$$

Where $X = \{\vec{x}_1, ...\vec{x}_T\}$ corresponds to the set of observation or test feature vectors. The higher the score, the more likely the test features belong to the speaker-model with which they are compared.

3. Experiment

The experiment consists of three different scenarios with aim to examine an effect of codec-based speech distortion on a performance of the speaker identification. In the first scenario, speaker enrollment is performed on clean speech while testing is done on degraded speech affected by a codec (named as *partially mismatched* case hereinafter). In the *fully matched* case, both training (speaker enrollment) and testing are carried out on speech data coded by the same codec. In the third scenario, training and testing are carried out also on coded speech but coded by different type of codecs (assigned as *fully mismatched* case hereinafter). In all the scenarios, speaker models were obtained by MAP based adaptation of the UBM model, which was composed of 512 Gaussians. For UBM training, EM algorithm was used, a relevance factor was set to 10, and K-means algorithm with 100 iteration was applied. Note that all of the GMM parameters (i.e. weights, means and covariance matrices) were modified during UBM-GMM.

In order to achieve a high performance of speaker recognition system, the system has to be based on powerful statistical models whose parameters have to be derived by using an adequate amount of training data. Therefore, we have decided to use a widely known and acoustically and phonetically rich TIMIT database [17] containing recordings of phonetically-balanced English speech of 630 speakers of eight major dialects regions of the United States (each reading ten phonetically rich sentences resulting in 6300 sentences). Approximately 70% of the speakers are male and rests are female. It is worth noting here that this database, in spite of its original design intention (designed for a speech recognition), is currently also widely used for a speaker recognition [7 and 18]. As a sampling rate of the recordings is 16 kHz, we have downsampled all of them to 8 kHz via an anti-aliasing low-pass FIR filter with no further processing, with the aim to stick to a sampling rate of NB speech communication. Furthermore, the speech samples were coded by the following codecs at the specified bit rates to

Summary of results for individual conditions Table 1

		G.711	G.729	AMR_5.9	AMR_7.95	AMR_12.2	EVS_5.9	EVS_8	EVS_13.2
partially mismatched	success rate[%]	76.00	63.00	60.11	62.95	68.53	67.53	71.37	74.47
fully matched	success rate[%]	78.79	49.00	39.84	44.32	36.58	69.53	67.32	71.74
partially mismatched	variance[%]	6.16	10.75	27.91	19.12	12.68	17.73	10.32	8.27
fully matched	variance[%]	7.93	12.60	13.80	13.23	20.51	14.43	25.07	11.30

introduce codec specific degradations induced when speech is transmitted over telecommunication network:

- G.711 speech codec [19] (a typical PCM (Pulse-Code Modulation) speech codec) operating at 64 kbps
- G.729 speech codec [20] (a very popular parametric codec dominantly deployed in fixed networks) operating at 8kbps
- AMR-NB [21] speech codec (typically deployed in 3G mobile networks) operating at 5.9kbps, 7.95kbps and 12.2kbps
- EVS speech codec [8] (a brand new 3GPP codec recently standardized by 3GPP and designed to be deployed in 4G (LTE) networks) operating at 5.9 kbps, 8 kbps and 13.2 kbps

Remark that codec degradations are currently considered as one of most prominent degradations encountered in current telecommunication networks. Moreover, the codecs selected for the experiment represent the ones commonly used in current telecommunication networks and also cover all range of degradations currently introduced by NB codecs. The selected bit rates cover the most popular ones.

All speakers in total (630) were used for the UBM training:10 clean (uncoded) recordings and 8 coded speech recordings per speaker were deployed. Note that each of these eight coded recordings was coded by different codec or the same codec operating at different bit rate, as listed above.

For the speaker enrollment (i.e. speaker specific GMM adaptation) phase as well asthe testing phase, 190 speakers from the first three dialect regions of the "training part" of the TIMIT database were used in each scenario. A set of utterances (folder of 10 audio recordings) of each speaker was divided into two non-overlapping parts. One half of the utterances of each speaker were utilized during the enrollment and the rest for the testing and vice versa.

Thus, overall 1900 runs (190 speakers x 10 recordings) were conducted for each test condition (type of codec or its bit rate) for each scenario. For each session, a recognition performance was evaluated by calculating a success rate in percentage. The obtained values were averaged and a mean value and variance were calculated.

As a front-end, speech analysis was performed frame-by-frame with 16 ms frame duration and 50% overlap, 12 Mel Frequency Cepstral Coefficients (MFCCs) (excl. 0-th coefficient) were extracted from each speech frame. The MFCCs are the most common speech features used in both speech and speaker recognition [22].

3.1 Experimental results

Figure 4 shows a success rate of the recognition process for the first two scenarios, namely fully matched and partially mismatched scenarios. It can be seen from the graphs that the recognition rate for the partially mismatched conditions is higher in the most of the cases compared to the fully matched conditions, except for G.711 codec, and EVS codec operating at 5.9 kbps. But the difference between the fully matched and partially mismatched case with G.711 and EVS codecs is rather small, less than 3%. The worst performance of the speaker identification system was achieved for AMR codec operating at 5.9 and 7.95 kbps. On the other hand, the best performance was obtained for G.711 codec closely followed by EVS codec operating at 13.2 kbps. A maximum difference between the success rate achieved for partially mismatched and fully matched scenario was attained for AMR codec (all bit rates). As can be also clearly seen from Fig. 4, the success rate grows with increasing bit rate and results are more statistically balanced (see Fig. 5 for statistical variances) for the AMR and EVS codec in the partially mismatched scenario. It is worth noting here that EVS codec has achieved quite good success rate in this experiment despite the fact that this codec represents lossy parametric codecs. This can be considered as a very promising result as this codec is going to be widely deployed in a voice communication over LTE (a successor of 3G mobile communication system). Therefore, a huge amount of the voice communication is going to be coded by this codec in the near future as mobile networks generate a dominant portion of voice communication nowadays. Table 1 summarizes the results for both experimental scenarios and all the test conditions. For comparison, in the case when both training and testing were carried out only on clean uncoded speech, recognition rate reached 96% as resulting from our previous experiments [23].

It is not a trivial task to select a codec for training offering a good performance over wide range of NB codecs currently deployed in telecommunication networks, as a performance of the recognition system is a codec-specific (see for instance results presented above). Moreover, an identification application mostly does not have access to a communication protocol and thus does not have information about the codec used for the voice transmission.

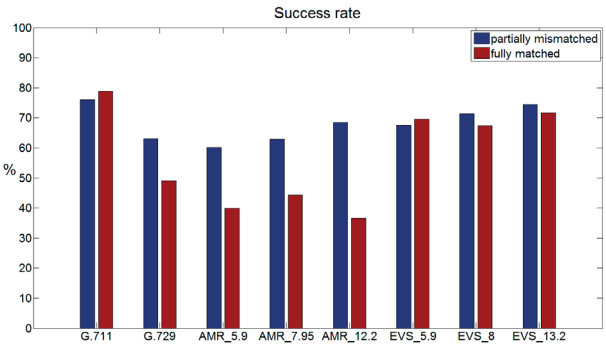


Fig. 4 Average success rate for the first two scenarios

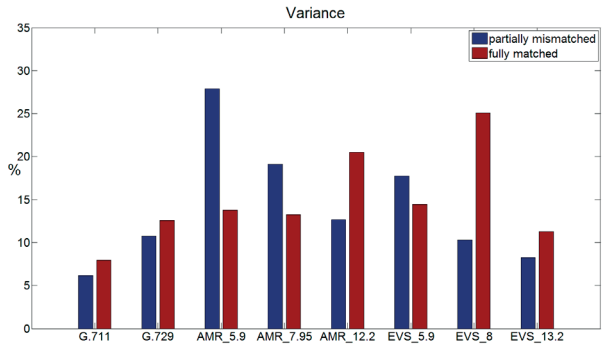


Fig. 5 Variance of success rate for the first two scenarios

In order to build the robust system covering all the prospective coding/degradation situations, a good performance in codec mismatch situations is also very important for ASR applications. Such codec mismatch situations are represented by the third scenario used in this study, defined above as the *fully mismatched* scenario, in which training and test materials come from the different codecs. As a fully factorial design would be very demanding in this case, we have decided to use only one bit rate of those offered by AMR and EVS codec for this part of the experiment. We have chosen a bit rate of 5.9 kbps because we believe that this bit rate covers most of the degradations introduced by the particular codecs. Due to clarity reasons, the results are only presented in a tabulated form, see Table 2. The best results were obtained for G.711 and EVS codec deployed for speaker specific model training. It is of a great surprise that the success rate obtained for G.711 and EVS is very similar. The corresponding difference is rather small, less than 1%. Note that EVS codec, as a lossy parametric codec, offers significantly better results than G.729 and AMR codecs. So, we can conclude that EVS and G.711 codecs offering the best speech quality from the codecs involved in this study (based on our expert listening), have achieved the highest success rate. This is in line with the finding presented in [1] involving different codecs.

Speaker recognition accuracy [%] for systems trained and tested with different codecs — Table 2

enrollment/test	G.711	G.729	AMR_5.9	EVS_5.9
G.711	-	61.00	63.37	67.79
G.729	43.26	-	44.16	39.58
AMR_5.9	33.89	38.16	-	33.53
EVS_5.9	67.16	64.21	61.63	-

4. Conclusions and future work

In this paper we analyze an impact of digital communication channel, particularly commonly used narrowband (NB) speech codecs in current telecommunication networks, on a performance of automatic speaker recognition in the context of codec mismatch between enrollment and test utterances. We have constructed the speaker identification system (Fig. 1) using the UBM-GMM model for the three different scenarios, namely fully matched, partially mismatched and fully mismatched. Surprising finding is that it is better to use the partially mismatched conditions (the system trained on clean uncoded data) than the fully matched conditions/scenario (the system trained and tested on the coded data using the same coding scheme). In the case of fully mismatched scenario (the system trained and tested on the coded data but using diverse coding schemes), the best recognition rate is achieved if G.711 or EVS codecs are deployed during the speaker enrollment (speaker specific GMM training). Note that both codecs offer the best speech quality from all the codecs deployed in this study. This finding fully correlates with the finding presented in [1] focusing on different speech codecs.

Since GMM training is sensitive to amount of the training data we plan to extend the experiments by using other speech databases (note TIMIT contents less than 20 seconds of speech per speaker), and also apply different types of speech features. In addition to that we would like to improve a performance of the proposed system. Moreover, we also plan to extend a codec set used in this experiment towards wideband speech codecs in order to reflect a current movement in a voice communication over telecommunication networks.

References

[1] JANICKI, A., STAROSZCZYK, T.: *Speaker Recognition from Coded Speech Using Support Vector Machines,* TSD 2011, LNAI 6836, pp. 291-298, 2011.

[2] BHATTACHARJEE, U., SARMAH, K.: GMM-UBM Based Speaker Verification in Multilingual Environments, *IJCSI Intern. J. of Computer Science Issues*, vol. 9, No. 6, November 2012.

[3] ASBAI, N., AMROUCHE, A., DEBYECHE, M.: *Performances Evaluation of GMM-UBM and GMM-SVM for Speaker Recognition in Realistic World*, ICONIP 2011, Part II, LNCS 7063, pp. 284-291, 2011.

[4] PILLAY, S. G., ARIYAEEINIA, A., PAWLEWSKI, M., SIVAKUMARAN, P. **Speaker Verification under Mismatched Data Conditions,** *IET Signal Processing*, vol. 3, No. 4, 2009, pp. 236-246.

[5] FAKHR, W., ABDELSALAM, A., HAMDY, N.: *Enhancement of Mismatched Conditions in Speaker Recognition for Multimedia Applications,* ICASSP, vol. 1, pp. 377-80, 2004.

[6] QUATIERI, T. F., SINGER, E., DUNN, R. B., REYNOLDS, D. A., CAMPBELL, J. P.: Speaker and Language Recognition Using Speech Codec Parameters, *Eurospeech*, vol. 2, pp. 787-790, 1999.

[7] GALLARDO, L. F., WAGNER, M., MOLLER, S.: *I-vector Speaker Verification for Speech Degraded by Narrowband and Wideband Channels,* ITG-Fachbericht 252: Speech Communication, Erlangen, September 2014.

[8] 3GPP: EVS Codec Detailed Algorithmic Description, Third Generation Partnership Project, 3GPP TS 26.445, 2014.

[9] REYNOLDS, D. A., QUATIERI, T. F., DUNN, R. B.: Speaker Verification Using Adapted Gaussian Mixture Models, *Digital Signal Processing,* vol. 10, No. 1-3, 2000, pp. 19-41.

[10] BECKER, T., JESSEN, M., GRIGORAS, C.: *Forensic Speaker Verification Using Formant Features and Gaussian Mixture Models*. Interspeech 2008, pp. 1505-1508.

[11] SORDO MARTINEZ, P. L., FAUVE, B., LARCHER, A., MASON, J. S.: *Speaker Verification Performance with Constrained Durations.* Intern. Workshop on Biometrics and Forensics (IWBF), 2014, pp. 1-6.

[12] TOGNERI, R., PULLELLA, D.: An Overview of Speaker Identification: Accuracy and Robustness Issues. *Circuits and Systems Magazine*, IEEE, 11(2), 2011, 23-61.

[13] REYNOLDS, D. A., ROSE, R.: *Robust Text-independent Speaker Identification Using Gaussian Mixture Speakers Models,* IEEE Trans. On Speech and Audio Processing 3, 1995, pp. 72-83.

[14] BISHOP, C.: *Pattern Recognition and Machine Learning, Springer Science+Business Media*, LLC : New York, 2006.

[15] LINDE, Y., BUZO, A., GRAY, R.: *An Algorithm for Vector Quantizer Design.* IEEE Transactions on Communications 28, 1980, pp. 84-95.

[16] KINNUNEM, T., LI, H. *An Overview of Text-Independent Speaker Recognition: From Features to Supervectors*, Speech Communication, 2009

[17] GAROFOLO, J., LAMEL, J. et al.: *DARPA, TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM*. National Institute of Standards and Technology, 1990.

[18] RAJESHWARA, R. R., PRASAD, A., KEDARI RAO, CH.: Robust Features for Automatic Text-Independent Speaker Recognition Using Gaussian Mixture Model, *Intern. J. of Soft Computing and Engineering (IJSCE)*, vol. 1, No. 5, November 2011.

[19] ITU: *Pulse Code Modulation (PCM) of Voice Frequencies*, Intern. Telecommunication Union : Geneva, ITU-T Rec. G.711, 1988.

[20] ITU: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-exited linear prediction (CS-ACELP), Intern. Telecommunication Union : Geneva, ITU-T Rec. G.729, 2007.

[21] 3GPP: Mandatory Speech CODEC Speech Processing Functions; AMR speech Codec; General description, Third Generation Partnership Project, 3GPP TS 26.071, 2012.

[22] REYNOLDS, D. A.: *An Overview of Automatic Speaker Recognition Technology,* IEEE, 2002.

[23] POLACKY, J., GUOTH, I.: *Comparative Evaluation of GMM and GMM/UBM Speaker Identification Systems,* Proc. of intern. conference TRANSCOM 2015, University of Zilina, June 2015.

Peter Sykora - Patrik Kamencay - Robert Hudec - Miroslav Benco *

A NEW ALGORITHM FOR KEY FRAME EXTRACTION BASED ON DEPTH MAP USING KINECT

In this paper, a new algorithm for key frame extraction based on depth map for hand gesture recognition is presented. The all input sequences are captured by Microsoft Kinect camera system. These methods extract three key frames from captured depth video sequence. These key frames describe dynamic gesture. The proposed extraction method is composed of two parts. The first part, labelled as space segmentation extracts the region of hand from background. The second part labelled as time segmentation splits captured sequence into three parts and marks one frame per part as the key frame. A new gesture database for evaluation of proposed method was created. The proposed method to human evaluators was compared. The experimental results show that the proposed system obtained accuracy about 90%.

Keywords: *Gesture recognition, 3D video, Microsoft Kinect, depth map, image segmentation.*

1. Introduction

The hand gesture recognition, as a part of non-verbal communication methods of humans, can be useful in human-machine interaction. In many situations, the voice command cannot be executed, e.g. a noisy environment, a person cannot use voice, etc. Methods described in this paper aim for the pre-processing part of the gesture recognition. There are many steps for gesture recognition calculations. First step is the recording and pre-processing. These methods take raw 3D video from Microsoft Kinect and extract only frames that are relevant for dynamic gesture description. The rest of the paper is organised as follows: Section 2 gives a brief overview of the state-of-the-art in gesture recognition problem. The proposed methods are described in section 3. Finally, experimental results, implementation issues and conclusions are discussed in Section 4, or in Section 5.

2. Related Work

A lot of present gesture recognition systems use some hardware peripheries to estimate the gesture variables, e.g. data-gloves [1 - 2]. Other approaches use markers to track the position of various parts of hand, mostly the fingers [3 - 5]. Even the systems working with video data use some gloves or markers to track the position [6]. There are some difficulties in segmentation of hand. For example, the skin of two humans can vary in dominant colour [7]. There are some other problems related to the hand description following feature extractions and their classification. Some methods relay on physiology of human hand [8]. Likewise, the gesture can be represented by motion of hand [9]. The aim of our research is to improve the methods for gesture feature extraction.

3. Proposed System

In this section the proposed system for key frame extraction is presented. The flowchart of the proposed system is illustrated in Fig. 1. The space segmentation (blue section) is the first part of system. The hand region is extracted from video sequence. Subsequently, hand sequences are processed by time segmentation part of the system (green section). This part splits the sequence to three sections. Next, the extraction of a defined number of key-frames is provided. It is assumed that these frames represent the entire sequence. These frames are used in another part of the overall hand recognition system.

* **Peter Sykora, Patrik Kamencay, Robert Hudec, Miroslav Benco**
Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia
E-mail: sykora@fel.uniza.sk

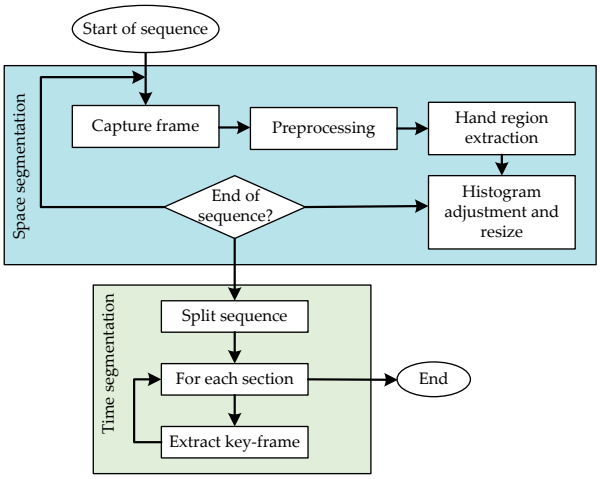


Fig. 1 Flowchart of the segmentation algorithm

The scene is captured by the camera as a first step. It is a 640 by 480 resolution depth map picture with depth of 16 bits. Pixels with the zero value cover locations of the scene when Kinect camera fails to calculate the distance. As zero values represent errors, high values represent a region with high distance from camera. Thus the hand region falls between these two extremes. The values of nonzero pixels are divided from the maximum of 16 bit (65535). Now the image appears as the disparity map. The low values of pixels represent background and errors of depth map. Next, the histogram is stretched to highlight details (Fig. 2).

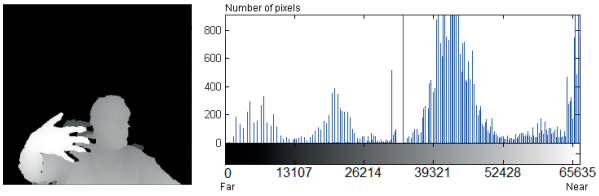


Fig. 2 Disparity map of captured scene and its histogram

3.1 Space Segmentation

The finding of the nearest pixel to the camera is the first step of the space segmentation. The value of pixel is defined as the distance of represented object from the camera. The algorithm finds pixel with the highest value. This value is stored as maxim. From this maxim the border is calculated as border=maxim-3500. The value 3500 is set subjectively and it represents the depth of segmented object. For the human hand a depth of all pixels with a value lower than this border is set to zero. The image now contains only the hand and black background. After this, pixel values of the hand are in the close range of high values. Stretching the histogram of this image is the next step. The details of third dimension are more visible after this. Cropping the excessive areas is the next step. The resizing of the image is the last part. To avoid the hand to be stretched, additional black pixels are added. Resulted image can be seen in Fig. 3.

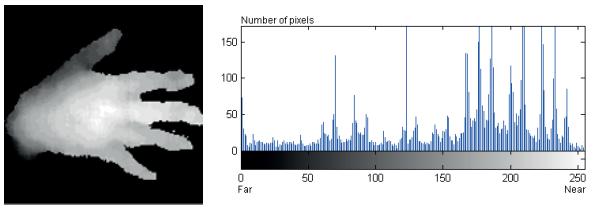


Fig. 3 Resulted picture of one frame and its histogram

3.2 Time Segmentation Methods

When the whole duration of dynamic gesture is captured and segmented, the key frame extraction part is started. These algorithms pick up a defined number of images from the sequence. Frames are picked up by their statistical position in the sequence (described below). Firstly, the methods split the sequence to more parts based on the significant change of shape. For this system, all methods divide the sequence to three parts (begin, centre and end as seen in Fig. 4).

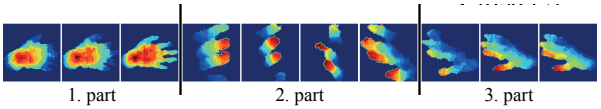


Fig. 4 Depth map sequence splitting to parts

A method, labelled as D_A, calculates Euclidean distance of neighbour frames. This is a distance of two vectors (frames) in high-dimensional space. If sequence has 14 frames, there are 13 neighbour distances. Let $\boldsymbol{d}$ be a vector of distances, **S** be a sequence of frames of length N and *mean* be the function that calculates mean number of a vector. Thus $\boldsymbol{d}_i$ is calculated as seen in following formula:

$$d_i = mean(S_{i-1} - S_i), i \in \{0, 1, ..., N-1\}. \quad (1)$$

A method, labelled M_A, calculates division image of two neighbouring frames. Next it calculates the mean pixel value of this image. Let $\boldsymbol{m}$ be a vector of distances and L_2 be the function that calculates Euclidean distance of two vectors. Thus mi is calculated as seen in Equation 2.

$$m_i = L_2(S_i, S_{i+1}), i \in \{0, 1..., N-1\}. \quad (2)$$

When visualised as curve, the highest peak determinates the first division of sequence. It can be seen as top green circle in Fig. 5. After remembering its position, this value is

set to zero. New highest peak determinates second division (purple circle). The highest peaks in first or last frame can be problematic. There are no frames before first frame. To resolve this, algorithm ignores the first and last values from curve (blue circles). Also, after zeroing first peak, neighbour values are set to zero too (red circles). At least one frame in the cut is ensured by this.

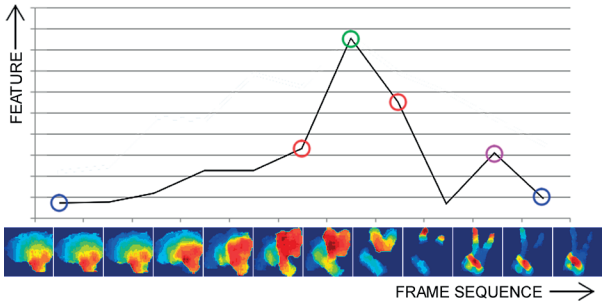


Fig. 5 Visualisation of significant points in the distance vector

Secondly, the key frames are extracted from the cuts. Two methods are used again. First method is labelled D_B. It calculates Euclidean distance for all combinations of frames in the sub-sequence. Frame with the lowest values to all other frames is picked up. It is the best representation for this sub-sequence. Let **Q** be the section of sequence with length *M*. The resulted key-frame **Fk** is picked as the frame with the smallest value in overall distance vector ***a*** (Equation 3). The vector ***a*** is calculated as sum of columns from distance matrix **A** (Equation 4). The position in matrix **A** represents the Euclidean distance of two frames of sequence. E.g. A at 1, 2 positions represent distance between frames 1 and 2:

$$\mathrm{F}_k = \mathrm{S}_i, i \rightarrow a_i = \min(a), \tag{3}$$

$$a_i = \sum_{k=0}^{M-1} \mathrm{A}_{ki}, i,k \in \{0,1,...,M-1\}, \tag{4}$$

$$\mathrm{A}_{ij} = \begin{cases} L_2(\mathrm{Q}_i, \mathrm{Q}_j), i \neq j \\ 0, \quad otherwise \end{cases}; i,k \in \{0,1,...,M-1\}. \tag{5}$$

The second method is labelled M_B. It calculates the mean pixel value from all sub-sequence pixels. This creates the so called mean picture. The mean picture is composed of mean pixels. The frame from sequence is taken. Pixel from this frame is compared with pixel from the mean picture. The equal counter is increased if they are the same. Each frame has this counter. The highest counter points to the frame that has the most pixels as the mean picture. This method ignores the zero valued pixels as they represent the background. Key-frame **Fk** as frame with maximal value in counter vector ***b*** is picked (Equation 6). The value in vector ***b*** on position *i* is incremented if pixel value of frame **P** is the same as in **MED** picture (Equation 7). As the calculation goes pixel by pixel, the vector **P** stores pixel values at given position from whole section **Q** (Equation 8). The **MED** picture is calculated by median value of pixels of section **Q** at given position:

$$\mathbf{Fk} = \mathrm{S}_i, i \rightarrow b_i = \max(\boldsymbol{b}), \tag{6}$$

$$b_i = \begin{cases} b_i = b_i + 1, \mathbf{MED} = \mathrm{P}_i \\ b_i \qquad otherwise \end{cases}; i \in \{0,1,...,M-1\}, \tag{7}$$

$$\mathrm{P}_i = \mathrm{Q}_{xyi} \quad \begin{matrix} x \in \{0,1,...,width\} \\ y \in \{0,1,...,height\} \end{matrix}, \tag{8}$$

$$\mathrm{MED}_{xy} = mean(\mathrm{Q}_{xyi}). \tag{9}$$

From these methods four combinations are created. These algorithms are labelled M_A-M_B, M_A-D_B, D_A-M_B and D_A-D_B.

4. Evaluation of Proposed Methods

The description of experiments and the results of proposed methods are presented in this section. The algorithms are implemented in C++ programming language with OpenCV library in Microsoft Visual Studio. The KinectSDK only for communication with Microsoft Kinect camera system is used. The experiment was run on the Microsoft Windows 7 Professional 64-bit computer with two AMD Opteron 6134 processors and 16 GB RAM memory.

4.1 Database of Depth Video Sequences

To compare these methods the database of depth video sequences is created. The existing gesture databases contain static gestures, or the dynamics is in the motion of hand rather than in the shape change [10 - 12]. Ten dynamic gestures were presented by 10 actors for this database. From them 3 were females and 7 were males. Total 100 depth video sequences are in dataset. The Microsoft Kinect camera was used for capturing. The sequences don't have the same length but they are processed by the first stage of segmentation. They have the same resolution of 150 by 150 pixels. Captured sequences of dynamic hand gestures are shown in Fig. 6. For this example, all have the same length. For example, the first gesture (first row in Fig. 6) is the hand waving from down to up. The second gesture (second row in Fig. 6) is in reverse. More gestures have the same shape change but in different direction. Gestures of such pairs are in rows 1-2, 3-4 and 9-10.

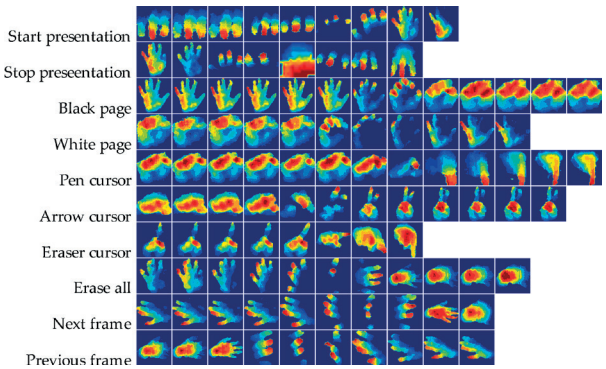


Fig. 6 Short example sequences of all gestures

4.2 Experimental Results

The goal of the experiment is to segment (extract key frames) all sequences with 4 presented methods. In the result there are 3 images that represent the whole sequence. They contain the information about dynamic shape change of hand. To evaluate these results, 10 respondents were appointed. Each respondent is supposed to divide the sequence to three parts. Next, they pick up the most representative picture of part. If they can't distinguish between multiple images, several similar images can be pointed. This gives 10 reference results per sequence. If system picks up the same picture as some of human evaluators, this result is considered to be correct. It basically means that this system acts as human.

Overall results for the four methods are in Table 1. The best results are represented by bold font. It is clear that method D_A-D_B gives the best results. Figure 7 shows results of all methods by the order of frames. For the first extracted frame all methods give almost the same results as humans. High accuracy is given on the last frames too. The results for the middle frame are worse than for the borders, because the dynamics of shape change is higher. The D_A-D_B method achieves overall accuracy of 90 %.

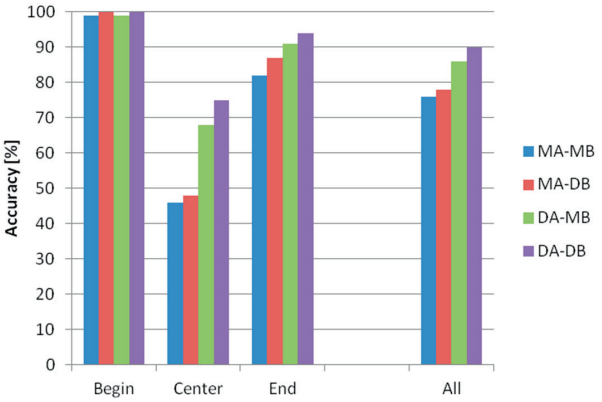


Fig. 7 Graph representing accuracy of time segmentation methods

Next Table 2 shows the time needed for calculation of each method based on the length of the sequence. For testing the length of sequence it is in the range from 8 frames to 30 frames. Results are in milliseconds. Again the method D_A-D_B provides the best conclusion as it is the fastest. For example, method D_A-D_B needs 2.3 ms to calculate three key-frames from sequence long 22 frames. On the other hand, the method M_A-M_B needs 10.3 ms to do the same.

In Fig. 8 these values are shown as graphs. Each method appears to have a trend. The processing time is in direct correlation with length of the sequence. From all these results the method D_A-D_B gives the best accuracy and the fastest calculation.

Accuracy of methods by actor — Table 1

	Actor	0	1	2	3	4	5	6	7	8	9
Accuracy	M_A-M_B	0.77	0.63	0.77	0.87	0.70	0.83	0.73	0.80	0.83	0.63
	M_A-D_B	0.77	0.73	0.83	0.87	0.70	0.87	0.70	0.87	0.83	0.67
	D_A-M_B	0.77	**0.97**	**0.90**	0.97	0.90	0.83	0.80	0.90	**0.90**	0.67
	D_A-D_B	**0.87**	**0.97**	**0.90**	**1.00**	**0.93**	**0.90**	**0.83**	**0.93**	0.90	0.73

Calculation time for each method by the length of the sequence — Table 2

Number of frames		8	12	16	20	24	28	30
Calculation time [ms]	D_A-D_B	**1.400**	**1.516**	**1.759**	**2.386**	**2.198**	**2.612**	**2.992**
	D_A-M_B	7.267	7.509	8.392	8.189	8.636	9.079	8.869
	M_A-D_B	1.572	2.448	2.822	2.890	3.793	3.969	4.299
	M_A-M_B	7.937	8.633	9.770	9.863	10.176	10.676	9.889

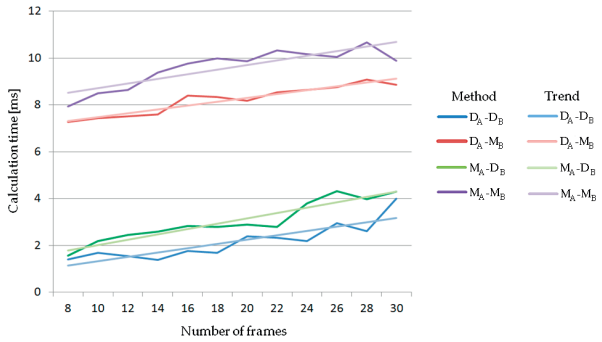


Fig. 8 Graph representing time needed for calculation of each method

5. Conclusion

In this paper a novel segmentation method for hand gesture recognition was described. This method takes raw depth video from Microsoft Kinect camera. It extracts the key-frames for further processing. These frames contain information about the whole dynamic gesture. This system is divided into two parts, the space segmentation and time segmentation. Firstly, the space segmentation extracts the region of human hand from depth video sequence. Next, it resizes the image to have resolution 150 by 150 pixels and stretches the histogram to highlight details. Secondly, the time segmentation extracts the key-frames. From the tests it is clear that system D_A-D_B extracts the frames with accuracy of 90%. As this pre-processing method provides three pictures it is up to the next stage to process them and recognize the whole dynamic gesture.

To the future, we plan to include this algorithm as the first step in dynamic gesture recognition system. The features should be extracted [13] from these three frames and combined for the classification [14].

Acknowledgments

This contribution is the result of the project's implementation at the Centre of Excellence for Systems and Services of Intelligent Transport, ITMS 26220120028. It was supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] YOSHIMURA, Y., OZAWA, R.: *A Supervisory Control System for a Multi-fingered Robotic Hand using Datagloves and a Haptic Device.* Intern. Conference Intelligent Robots and Systems (IROS), 2012, p. 5414-5419

[2] CARBONARO, N., MURA, G. D., LORUSSI, F., PARADISO, R., DE ROSSI, D., TOGNETTI, A.: Exploiting Wearable Goniometer Technology for Motion Sensing Gloves. *IEEE J. of Biomedical and Health Informatics*, 2014, vol. 18, No. 6, p. 1788-1795

[3] MINNEN, D., ZAFRULLA, Z.: *Towards Robust Cross-user Hand Tracking and Shape Recognition.* IEEE Intern. Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 1235-1241

[4] CHANG, L. Y., POLLARD, N. S., MITCHELL, T. M., XING, E. P.: *Feature Selection for Grasp Recognition from Optical Markers.* Intern. Conference on Intelligent Robots and Systems, 2007, pp. 2944-2950

[5] TANGSUKSANT, W., ADHAN, S., PINTAVIROOJ, C.: *American Sign Language Recognition by using 3D Geometric Invariant Feature and ANN Classification.* 7th Intern. Conference Biomedical Engineering (BMEiCON), 2014, pp. 1-5

[6] MEYER, J., KUDERER, M., MULLER, J., BURGARD, W.: *Online Marker Labeling for Fully Automatic Skeleton Tracking in Optical Motion Capture.* IEEE Intern. Conference on Robotics and Automation (ICRA), 2014, pp. 5652-5657,

[7] LUKAC, P., HUDEC, R., BENCO, M., KAMENCAY, P., DUBCOVA, Z., ZACHARIASOVA, M.: *Simple Comparison of Image Segmentation Algorithms Based on Evaluation Criterion.* 21st Intern. Conference Radioelektronika (RADIOELEKTRONIKA), 2011, pp. 1-4

[8] QURAISHI, M. I., DHAL, K. G., CHOUDHURY, J. P., GHOSH, P., SAI, P., DE, M.: *A Novel Human Hand Finger Gesture Recognition using Machine Learning.* 2nd IEEE Intern. Conference on Parallel Distributed and Grid Computing (PDGC), 2012 , pp. 882-887

[9] JUAN, W., GUIFANG, Q., JUN, Z., YING, Z., GUANGMING, S.: Hand Motion-Based Remote Control Interface with Vibrotactile Feedback for Home Robots. *Int. J. Adv Robot Syst*, 2013, vol. 10:270. doi: 10.5772/56617

[10] ZHOU, R., JUNSONG, Y., WENYU, L., ZHENGYOU, Z.: Minimum Near-Convex Shape Decomposition and its Application in Hand Gesture Recognition. *Intern. J. of Computer Vision* (*IJCV*), 2011

[11] KIM, T. K., WONG, S. F., CIPOLLA, R.: *Tensor Canonical Correlation Analysis for Action Classification.* Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007

[12] LIU, L., SHAO, L.: *Learning Discriminative Representations from RGB-D Video Data.* Proc. of Intern. Joint Conference on Artificial Intelligence (IJCAI), Beijing, 2013

[13] CHENYANG, Z., XIAODONG, Y., YINGLI, T.: Histogram of 3D Facets: *A characteristic Descriptor for Hand Gesture Recognition.* 10th IEEE Intern. Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013, pp. 1-8

[14] DARDAS, N. H., GEORGANAS, N. D. Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques. *IEEE Transactions on Instrumentation and Measurement*, 2011, vol. 60, No. 11, p. 3592-3607.

Michaela Solanska - Miroslav Markovic - Milan Dado *

RESOURCES RESERVATION IN THE SEGMENT OF OPTICAL NETWORK WITH OBS

The growing demands on transmission quality, flexibility and granularity can be achieved by powerful reservation protocols which are an essential part of networks with burst switching. In today's high-speed optical networks there exist a lot of reservation protocols. In this article we present a design of a new reservation protocol called Search&Compare. This reservation protocol is designed according to the well-known Segment-based Robust Fast Optical Reservation Protocol. Search&Compare is parallel-segment based and uses parallel link reservation. Our goal is to describe the behavior of reservation protocol in the transmission with setting the priority for incoming bursts with the QoS utilization.

Keywords: *Wavelength-division multiplexing, optical burst switching, reservation protocols, services.*

1. Introduction

At present there is increasing demand for transmission bandwidth for Internet access from users. Providers must design their core networks to satisfy user's claims in the future. To make this possible they must use the multiplexing methods. The different multiplexing techniques are used to meet the increasing demands for transmission capacity and its quality. Currently the most common method of multiplexing to create high-speed full-optical networks is wavelength-division multiplexing (WDM). This technique allows more efficient data transmission via multiple wavelengths transmitting in single optical fiber. The advantage of WDM is the ability to transmit the data with different transmission speed, modulation and format in every single wavelength [1 - 3].

2. Optical switching

The networks nodes are very important in high-speed full-optical networks. Mainly two ways of incoming data switching are used in the nodes: Optical Packet Switching (OPS) and Optical Burst Switching (OBS).

2.1. Optical Packet Switching

At present the packet switching is preferred by many telecommunication providers. This is due to its large deployment in the Internet, Wide Area Networks (WAN), Metropolitan Area Networks (MAN) and Local Area Networks (LAN). Next Generation Networks task will be to unite each type of networks which will work on IP protocol. Therefore, the optical networks working with packet switching are considered for the future. However, this type of network contains various problems, mainly optical memory, because a packet needs to be saved in memory before its routing, until a routing decision is made. It is obvious that minimal bulk of optical memories has to be equal to maximum bulk of packet. However the constant length of packets is considered in many proposals of OPS nodes. For on time adjustment of optical devices during switching, it will be necessary to insert the guard intervals, which will be at the beginning and at the end of the packet information part. OPS nodes typically demand the low level of aggregation at transmission speed over 10 $Gb.s^{-1}$. The packet switching time is much shorter than packet transmission time. For the packet length, which is in μs, the switching time in the range of ns will be required [4 - 6].

2.2. Optical Burst Switching

Optical burst switching is the type of optical switching which combines the best elements of optical packet switching and wavelength routing, while it avoids its insufficiency (see Fig. 1). It provides the mechanism for big amount of bursts data transportation via transparent optical switching network and could serve as bridging technology between existing wavelength routing optical networks and future networks, using optical packet switching [7 - 11].

* **Michaela Solanska, Miroslav Markovic, Milan Dado**
Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia
E-mail: michaela.solanska@fel.uniza.sk

In OBS networks the incoming data are collected into bursts. This is performed by collecting multiple data into the one bigger unit called burst. The burst can have a constant or variable length. Each burst has generated the header, as a necessity to correct the setting of the switching structure in every node, through which the burst will pass. Just before the burst is transported, the control header is sent to inform all the nodes along the road about the burst arrival. The control header is sent through the independent channel which is on different or same wavelength as transferred burst. In case that a control header and burst are transmitted on the same wavelength, we call it in-band signaling. Another case is out-of-band signaling. In each node, the control header is converted by OEO conversion, because of the unavailability of the all optical control and processing circuits. On the basis of the information contained in the control header the setting of switching node structure is made, which is processed in the control circuit. The information contains the time delay between the control header and burst, size of the burst, required output port, incoming wavelength. Then the new header for transmitting burst is generated. This header is again sent from the node ahead of the data burst. This repeats until the data burst arrives to the destination node. The purpose of the destination node is to segment the data burst [7 - 11].

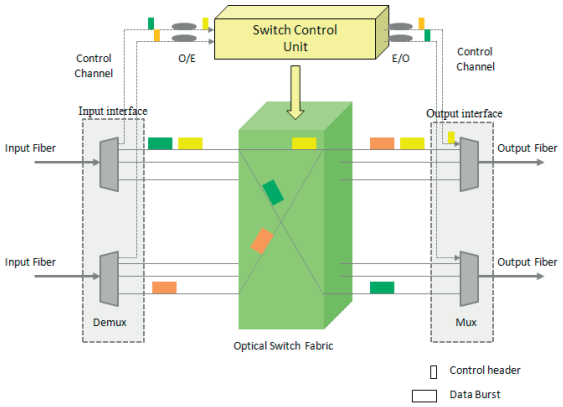


Fig. 1 The architecture of OBS node

2.3. Optical Switches

Figure 1 shows the architecture of OBS node. The general structure of OBS node consists of five main blocks, both input and output interface, control unit with optical-electrical-optical conversion (O/E/O), optical switch (see Table 1) and wavelength convertors. Optical switches could be either electronic or optical. In electronic switches it is required to convert incoming data burst and header into the electrical domain, followed by processing and switching to required output port. Only the control header is converted in optical switches. The data burst switching is performed without conversion. Optical switches should meet the following requirements [9 and 10]:

- High switching speed: switching time must be shorter than the one required for burst transmission, i.e. ones to tens μs,
- Dimension of switching structure,
- Low insertion losses or sufficient amplification,
- Low crosstalk between channels,
- Simplicity and easy integrability.

3. Reservation Protocols

Reservation protocols are important part of OBS networks. The main reason of reservation protocols in OBS nodes is to reserve their resources for some time period, which is needed for burst transmission. Recently, there are some variations of reservation protocols designed, which are trying to use the resources of OBS nodes most efficiently with the lowest blocking possibility. The best known reservation protocols are: Segment-based Robust Fast Optical Reservation Protocol (S-RFORP), Robust Fast Optical Reservation Protocol (RFORP), Resource Reservation Protocol-Traffic Engineering (RSVP-TE) and Intermediate-node Initiated Reservation (IIR) [4 and 10].

The suggested reservation protocol Search&Compare (S&C) is based on S-RFORP reservation protocol, because it has the best features of the mentioned reservation protocols. It uses S-RFORP to minimize the probability of blocking the parallel among-segment discovery and reservation, and simultaneously stock inter-segment discovery and reservation. Reservation protocol S&C is an improved version of S-RFORP protocol. Unlike the S-RFORP, it uses parallel inter-segment discovery and reservation through which is possible to achieve less time of inter-segment discovery and reservation [4 and 10].

The reservation protocol S&C uses out-of-band signaling and its advantage is the possibility to setup the priority for each incoming data burst into the segment due to which it utilizes the QoS. If OBS nodes in networks do not have enough resources for transmitting the required burst, it is necessary to prioritize the bursts based on the recognition of the burst priority for transmission.

3.1. Reservation of network resources in the segment

Reservation depending on the type of transmitted services is executed through segments in which the data of different types are transmitted. A segment contains an input edge node which, in this case, simply gathers incoming data (packets) that lead to the same destination node. In the input edge node a wavelength is reserved for each burst which can be the same through the whole burst transmission in OBS network (see Fig. 2).

Switching Technology

Table 1

Technology	Properties				
	Dimension	**Switching time**	**Insertion loss [dB]**	**Crosstalk [dB]**	**PDL loss [dB]**
Optomechanical	16 x 16	< 10 ms	3	< -55	0.2
2D MEMS	32 x 32	< 10 ms	1.7-6.9	< -60	0.11-0.16
3D MEMS	350 x 350 160 x 160	< 10 ms < 10 ms	6 ± 1 < 2	< -60 < -55	0.4 0.5
Thermooptical **Silica** **Polymer**	 8 x 8 16 x 16	 < 10 ms < 10 ms	 8 6	 < -35 < -30	 0.5 0.4
Liquid - Crystal	2 x 2	< 10 ms	1.5	< -35	0.1
Acoustooptical	1 x N	< 3 µs	6	< -35	
Electrooptical **$LiNbO_3$** **InP**	 8 x 8 1 x 2	 < 10 ns < 10 ns	 9	 < -35 < -25	 0.5
SOA		1 ns	~ 0	< -50	< 1

In the case of unavailability of the spectral channel with distinct wavelength, on which the burst is transmitted, the node will choose another appropriate wavelength. In the output node the burst is disassembled to data (packets) which are further processed [4 and 7].

To verify the proposed solution, we made the numerical computer model in VPIphotonics environment (VPI). As shown in Fig. 3, to develop one segment, the connection between optical switches had to be proposed and the verification bits of single switches had to be set. The proposed model contains five optical transmitters with RZ modulation, multiplexor, demultiplexor, six optical switches and delay segments.

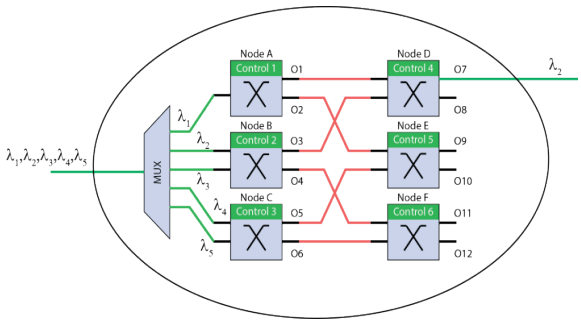


Fig. 2 Transmission depending on the type of transmitted services within one segment

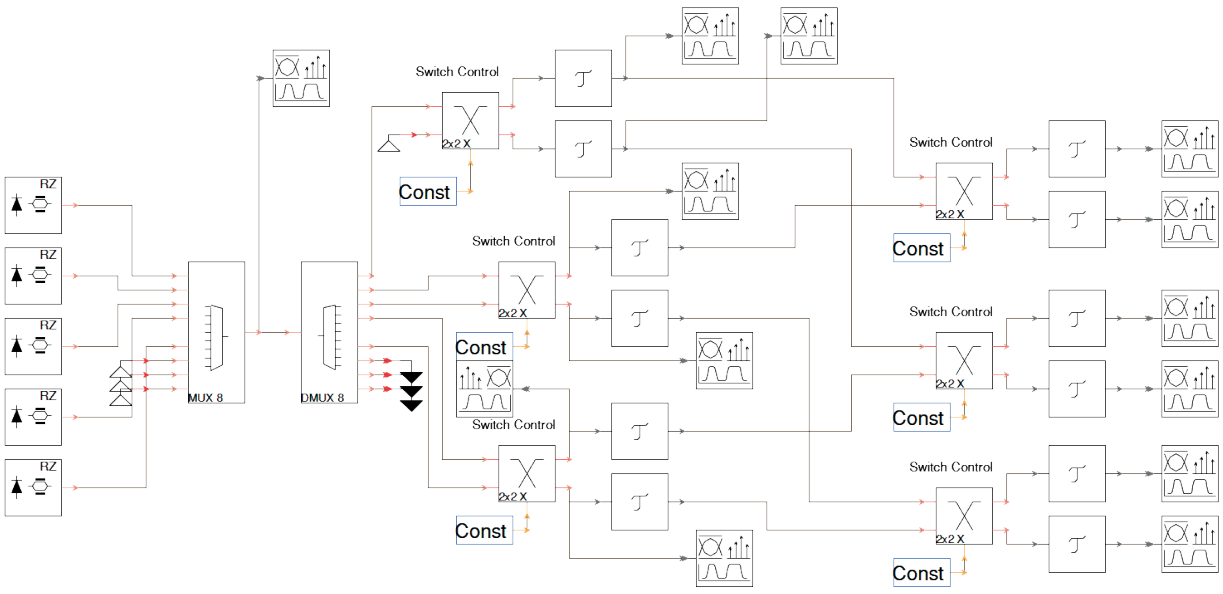


Fig. 3 Scheme of five-channel system in simulation software VPI

The control bits setting and output wavelengths for outputs O1 to O12 Table 2

Variant 1							
		O1	**O2**	**O3**	**O4**	**O5**	**O6**
Control 1	0	λ_1	-	λ_2	λ_3	λ_4	λ_5
Control 2	0	λ_1	-	λ_2	λ_3	λ_4	λ_5
Control 3	0	λ_1	-	λ_2	λ_3	λ_4	λ_5
		O7	**O8**	**O9**	**O10**	**O11**	**O12**
Control 4	0	λ_1	λ_2	-	λ_4	λ_3	λ_5
Control 5	0	λ_1	λ_2	-	λ_4	λ_3	λ_5
Control 6	0	λ_1	λ_2	-	λ_4	λ_3	λ_5
Variant 2							
		O1	**O2**	**O3**	**O4**	**O5**	**O6**
Control 1	1	-	λ_1	λ_2	λ_3	λ_5	λ_4
Control 2	0	-	λ_1	λ_2	λ_3	λ_5	λ_4
Control 3	1	-	λ_1	λ_2	λ_3	λ_5	λ_4
		O7	O8	O9	O10	O11	O12
Control 4	1	λ_2	-	λ_1	λ_5	λ_4	λ_3
Control 5	0	λ_2	-	λ_1	λ_5	λ_4	λ_3
Control 6	1	λ_2	-	λ_1	λ_5	λ_4	λ_3

To control optical switches, the verification bit was chosen, which is shown in Fig. 3 as Const. This constant can be changed to 1 or 0. Since the ideal optical switch was chosen, it was necessary to add the delay pad (segment) to each switch, which simulates the necessary time to switch the OXC. To illustrate the output signals, an optical signal analyzer was chosen.

By the control bits it is possible to set the switching structure of the switch. In case the control bit is set to 0, the input ports will not switch, i. e. input 1 will be transferred to the output port, input port 2 will be transferred to output port 2. If the control bit is set to λ, input port 1 will be switched to output port 2 and input port 2 will be transferred to output port 1. Ports O1–O6 represent the output ports from the first three optical switches, ports O7–O12 represent the output ports of output optical switches. For video transmission the l_2 wavelength was chosen, which was necessary to transmit from the input edge node to output node of the given segment.

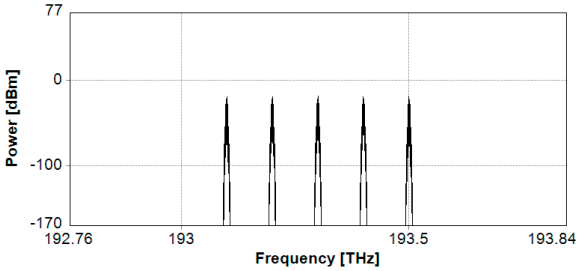


Fig. 4 Optical spectrum of a five-channel system

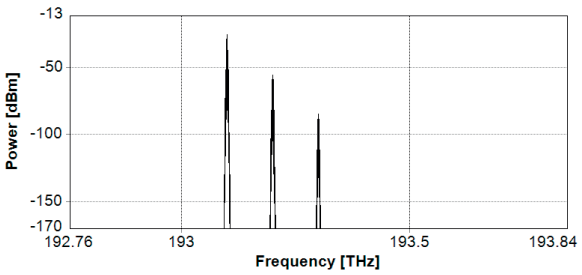


Fig. 5 Optical spectrum on output port 07

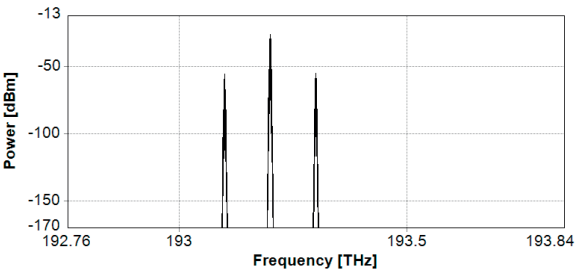


Fig. 6 Optical spectrum on output port 08

Figure 4 shows the optical spectrum of a five-channel system with wavelengths λ_1-λ_5.

For combination of the control bits 1 0 1 1 0 1 the required wavelength λ_2 will get on output port O7 (as shown in Fig. 5 and in Table 2). If we need the required wavelength λ_2 on a different

output port, e.g. on O8 (see the Fig. 6) as the combination of control bits we will choose 0 0 0 0 0 0.

4. Conclusion

In this document we focused on the development of the numerical computer model of reservation of the resources in one segment. We developed the simulation model in VPIphotonics environment. In simulation model the input parameters were wavelengths of optical transmitters and insertion losses of multiplexor and demultiplexor. Since the optical switch was ideal, it was necessary to add the delay pad to each switch, which simulates the necessary time to switch the OXC. The switching time was set according to Table 1. The output of the simulation was described in Chapter 3. Our next goal will be the development of the numerical computer model of the optical network which will consist of multiple segments.

Acknowledgements

This work is supported by the Slovak Research and Development Agency under the project APVV-0025-12 ("Mitigation of stochastic effects in high-bit rate all-optical networks") and the European Regional Development Fund and the Slovak state budget for the project "Research Centre of University of Zilina", ITMS 26220220183.

References

[1] SIVALINGAM, K. M., SUBRAMANIAM, S.: *Optical WDM Networks: Principles and Practice,* Massachusetts : Kluwer Academic, 2000.

[2] SICHANI, A. V., MOUFTAH, H. T. A.: *Novel Progressive Reservation Protocol for WDM All-optical Networks.* IEEE Intern. Conference on Communications, ICC 2003, 1463-1467, 2003.

[3] BENEDIKOVIC, D., LITVIK, J., KUBA, M., DADO, M., DUBOVAN, J.: Influence of Nonlinear Effects in WDM System with Non-equidistant Channel Spacing using Different Types of High-order PSK and QAM Modulation Formats. *Optical Modelling and Design II,* SPIE, vol. 8429, 2012.

[4] MARKOVIC, M.: *Reservation Protocols in High-speed All-optical Networks with Burst Switching,* Dissertation thesis, 2010, University of Zilina.

[5] POBORIL, R., LATAL, J., KOUDELKA, P., VITASEK, J., SISKA, P., SKAPA, J., VASINEK, V.: A Concept of a Hybrid WDM/TDM Topology using the Fabry-Perot Laser in the Optiwave Simulation Environment, *Advances in Electrical and Electronic Engineering*, vol. 9, 2011.

[6] MEI, Z., ZHAO, D., LIN, J. A.: *Survey of Optical Networks and its Visual Management and Control,* 3rd Intern. Conference on Advanced Computer Control, ICACC 2011, 473- 477, 2011.

[7] XU, H., YU, O., YIN, L., LIAO, M. *Segment-based Robust Fast Optical Reservation Protocol,* High-Speed Networks Workshop, 36-40, 2007.

[8] SALEH, A. M. M, SIMMONS, J. M.: All-optical Networking-evolution, Benefits, Challenges, and Future Vision, *IEEE J. & Magazines*, vol. 100, No. 5, 2012.

[9] O'MAHONY, M. J., POLITI, CH., KLONIDIS, D., NEJABITI, R., SIMEONIDOU, D.: Future Optical Networks, *J. of Lightwave Technology,* vol. 24, No. 12, 2006.

[10] MARKOVIC, M., DUBOVAN, J., DADO, M.: Search-&-Compare (S&C) - Reservation Protocol in High-Speed Optical Networks, *Electronics and Electrical Engineering,* vol. 114, No. 8, 2011.

[11] PAPADIMITRIOU, G., PAPAZOGLOU, CH, POMPORTSIS, A.: *Optical Switching,* Hoboken : Wiley - Interscience, 2007.

Miroslav Voznak - Ivo Zbranek - Miralem Mehic - Dan Komosny - Homero Toral-Cruz - Jerry Chun-Wei Lin *

COVERT CHANNEL IN RTP PAYLOAD USING A POINTER IN SIP HEADER

The paper addresses the issue of hiding data in the network flow. The authors discuss a new proposal of the steganographic approach in IP telephony: transmitting texts using the pointer in the SIP header by means of a data stream within the RTP protocol. This method is based on tagging the binary sequences in payload of the RTP packets, with the individual binary sequences representing individual steganogram characters. Subsequently, the position of the binary sequences is recorded in the SIP header in the Via field and the branch parameter. The proposed way of hidden data transmission cannot be detected by existing anomaly detectors; and does not represent an approach to statistical detection of covert channels. In fact, it is a new contribution to covert communication in ordinary VoIP traffic.

Keywords: *Steganography, VoIP, RTP, payload, SIP, pointer.*

1. Introduction

In the context of the constant rapid development of the data network and related services, the security of transmitted data is increasingly important. One of the most progressive services offered by these networks is IP telephony, using protocols applied across the Internet network. Consequently, real threats such as sensitive personal data, internal or corporate data leaks or server overloads causing the unavailability of the service, etc. arise. One of the possible ways to transmit sensitive or classified data is to use a covert communication channel. The creation of covert communication channels is the subject addressed by the scientific branch referred to as steganography the results of which can also be applied in IP telephony [1]. The most significant results of the research in this area were published by Mazurczyk and Szczypiorski and yield methods which enable creating a covert communication channel [2 and 3]. All data communication in digital networks is realised by means of a data transmission described in the binary scale. This means a vast amount of zeros and ones transmitted one after another. Given the large data volumes, the probability that a sequence, which would allow us to interpret the subsequent information in the required form, is found, is high. Such sequences can also be found within voice communication that uses the IP protocol. The highest volume of data to be exploited is the data stream itself, serving to transmit voice data.

Individual characters which we need to transmit are transformed to binary sequences according to the ASCII table and the transmitted RTP packets within session are locally stored by the sending User Agent. Subsequently, binary sequences representing individual characters are found out in local stored RTP packets and these position are recorded in absolute/relative values. The ASCII table is again applied to transform characters to HEX format and the obtained chain is used in the parameter "branch" as a randomly generated chain which is transferred within ordinary nearest transaction. The receiving User Agent can easily store incoming RTP packet and realized reverse process. Robustness can be enhanced by using shared key for the transform table combining steganography and cryptography.

2. Related Work

The idea of hiding data within the network flow combines two elements: utilisation of unused packet fields and information encoding in traffic behaviour. The first element

* [1]Miroslav Voznak, [1]Ivo Zbranek, [1]Miralem Mehic, [2]Dan Komosny, [3]Homero Toral-Cruz, [4]Jerry Chun-Wei Lin
[1]VSB-Technical University of Ostrava, Czech Republic
[2]Brno University of Technology, Czech Republic
[3]University of Quintana Roo, Mexico
[4]School of Computer Science and Technology, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China
E-mail: miroslav.voznak@vsb.cz

is a well-known technique that emerged from old Xmas packets. These packets, with every single option set for the used protocol, are included in a well-known nmap network scanning tool, and they were named Xmas packets because they resemble bright bulbs on a Christmas tree. These packets can be easily detected by intrusion-detection systems (IDS), or more advanced firewalls with an anomaly detection feature [4 and 5]. The second element, i.e. encoding information in the traffic behaviour, was first presented in [6] and this idea was further modified in [2] by Mazurczyk and presented as the LACK (Lost Audio Packets Steganography) solution for VoIP communication.

Covert channels were first observed and defined in the mid-80s as a result of the rapid development of communication networks. Lampson from Cambridge classified communication channels into three categories: Storage, Legitimate, and Covert. He also gave the first definition of covert channels stating that covert channels are the channels which are used for information transmission even though they are neither designed nor intended to transfer information at all [7]. In the last ten years, a large number of covert channels were introduced, and a further development of new techniques is expected in the upcoming years. All these techniques can significantly affect the level of security and reputation that certain communication solutions offer. Viewed from the client's point of view, it is reasonable to doubt the safety and quality of a particular communication solution which has weak points in the system that enable unnoticed leaks of confidential data. Because of that, covert channels are under a close supervision of governments and security companies that seek to prevent these leaks. One of the first papers on this subject [6] presents an approach to statistical detection of covert channel embedded in network packet delays. This simple technique implies the existence of clear differences between the packet delay and it is based on the probability of the existence of a covert channel, which is calculated as follows (1):

$$P_{CovChan} = 1 - \frac{C_{\mu}}{C_{\max}} \quad (1)$$

where C_{μ} is the packet count at the mean and C_{max} is the maximum packet-count of the histogram formed from the number of packets received with a given time, the experiment is explained in [6]. According to Mazurczyk and Szczypiorski [2 and 3], there are several types of steganographic techniques in VoIP networks but all of them can be classified into the following three categories:

- Packet modification steganography;
- PDUs time relations;
- Technique that requires hardware modification of device.

The first one represents a technique which takes advantage of unused fields in the protocols, mostly in IP (Internet Protocol), UDP (User Datagram Protocol), TCP (Transmission Control Protocol) or even RTP and RTCP (RTP Control Protocol) packets [2]. This method is susceptible to detection by IDS. Instead of using a separate RTCP flow, the authors proposed embedding the control information into the actual RTP flow. Unused bits in the IP/UDP/RTP headers signal the type of parameters, whereas the parameter values are embedded as a watermark in the voice data [8 and 9]. The second method is based on the deliberate data delay, since the VoIP content is very sensitive to delay and jitter variations. This method is related to the LACK technique mentioned above that is based on deliberate delays of the VoIP packets. The receiver will only consider the packets that are delivered on time and discard any packets that are delayed. Using this technique, instead of discarding the delayed packets, the receiver will read them for the purpose of a steganographic analysis. This method is quite hard to identify in the network, and equally hard to implement. There are a few variations of this technique such as affecting the order of packets, modifying the inter-packet delay or introducing the intentional losses [6]. The third method consists in using modified hardware for the purpose of a steganographic analysis. This technique is referred to as HICCUPS (Hidden Communication System for Corrupted Networks) for the VoWLAN (Voice over Wireless LAN) specific environment. However, the steganographic method is quite difficult to implement and detect, since it takes advantage of the imperfections of the transmission medium environment [10]. In addition, the authors of this paper have published the results related to the IP telephony security analysis and monitoring [11 and 12]. Their recently published papers discuss the computation of the available steganographic bandwidth in a VoIP covert channel and proved detecting irregularities in SIP flows caused by the injection of SIP headers or by the increased amount of SIP messages [13 and 14].

This article is organised as follows: the main idea is explained in the third section, the experiment and the results are presented and discussed in the fourth section and the conclusion is provided in the fifth section.

3. Main Idea Explanation

The RTP protocol has been designed to ensure media data transmission. In our case voice data was digitalised using codec G.711 and packets are generated every 20ms.. So within each individual packet, RTP transmits 160B of data, enabling to describe voice data, the so called payload. A five-minute phone call has potential of 2400000 sequences that can represent the character required by us. To convert data, internationally

standardised (ASCII) or self-designed converting tables can be applied. In the latter case, the self-designed tables need to be shared with the party receiving the steganogram. In case we intend to describe the entire 7-bit ASCII table, we need to establish 128 unique sequences represented in the binary code. Each character from the ASCII table can be described using 7-bits, but the converting table assigns each character 8-bits, where the first most important bit equals binary 0. Where a further reduction of characters to be used is required, e.g. alphanumerical character incl. lower case characters, all we need is 62 unique sequences. To separate and precisely identify the position of a particular sequence, 3 to 4 specific characters should be allocated.

To identify the packet in which the necessary sequences are located, the *Sequencenumber* value provided by the RTP protocol can be used. This value is unique and allows for uniquely determining the position of data. Individual sequences that correspond to the required character can be looked up in the payload of the packet concerned and we can record information about their position. Where the packet does not contain a particular sequence, we need to use a different packet. Since we need to provide the counterparty with the information which packet is the relevant one and the position of the sequence within the payload, we use the header of the SIP protocol, placing the information about the packet and position into the payload of the given packet. Once the information contained within the SIP header has been transmitted, the counterparty, once it has been provided with the converting table, can start to search for particular sequence positions representing individual characters. Once the binary sequences have been converted using the converting table, a steganogram can be drawn, sent by the first party or the sender. This method consists of several individual phases as is depicted in Fig. 1.

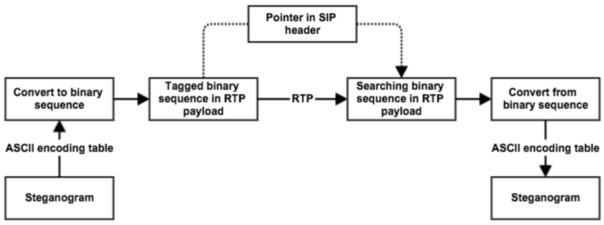
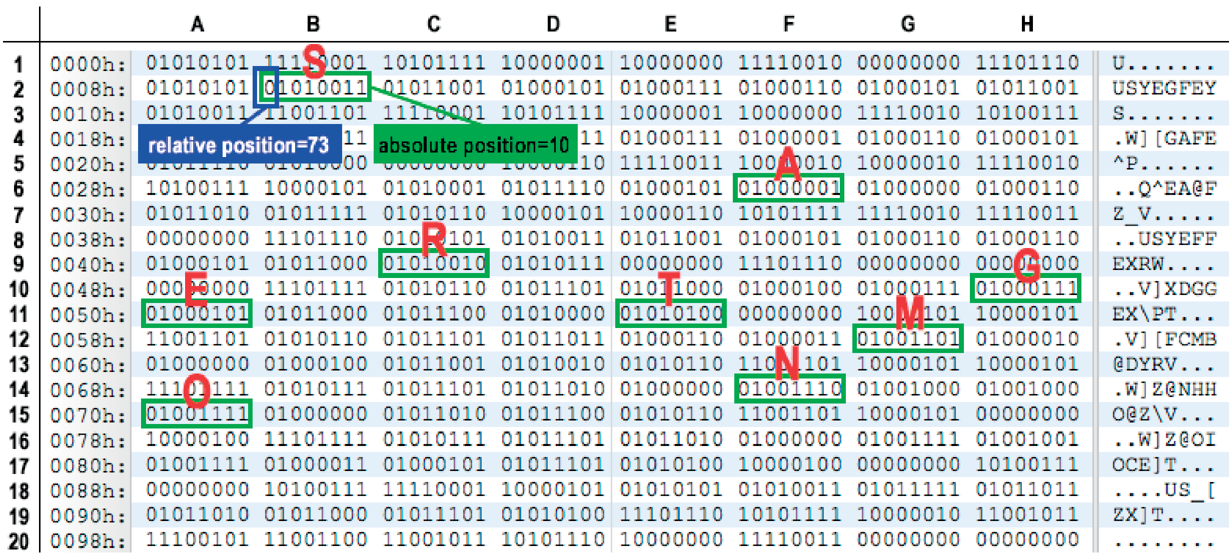


Fig. 1 The phases of the proposed steganographic approach

First, individual characters (ASCII) are converted from a steganogram into corresponding binary values (BIN) using the converting table. Then, these binary sequences are tagged within the payload of the RTP packet. Individual binary sequences stand for individual characters within the steganogram. The position of particular binary sequences together with the packet sequence number is recorded in the pointer located in the SIP header. In order to enhance the confidentiality, the pointer in the SIP header is re-converted from the decimal into the hexadecimal format. Using the information contained within the pointer, the counterparty is able to look up the binary sequences in the payload of the RTP packet. Lastly, the binary (BIN) sequences are transmitted back into alphanumeric characters ASCII) which form the steganogram.

3.1 Steganogram Conversion to Binary Sequence

To convert a steganogram into the binary format, both internationally standardised and self-developed converting



		A	B	C	D	E	F	G	H	
1	0000h:	01010101	[illegible]	10101111	10000001	10000000	11110010	00000000	11101110	U.......
2	0008h:	01010101	01010011	01011001	01000101	01000111	01000110	01000101	01011001	USYEGFEY
3	0010h:	01010011	11001101	[illegible]	10101111	10000001	10000000	11110010	10100111	S.......
4	0018h:	[illegible]	[illegible]	[illegible]	[illegible]	01000111	01000001	01000110	01000101	.W][GAFE
5	0020h:	[illegible]	[illegible]	[illegible]	[illegible]	11110011	[illegible]	10000010	11110010	^P......
6	0028h:	10100111	10000101	01010001	01011110	01000101	01000001	01000000	01000110	..Q^EA@F
7	0030h:	01011010	01011111	01010110	10000101	10000110	10101111	11110010	11110011	Z_V.....
8	0038h:	00000000	11101110	[illegible]	01010011	01011001	01000101	01000110	01000110	..USYEFF
9	0040h:	01000101	01011000	01010010	01010111	00000000	11101110	00000000	[illegible]	EXRW....
10	0048h:	[illegible]	11101111	01010110	01011101	[illegible]	01000100	01000111	01000111	..V]XDGG
11	0050h:	01000101	01011000	01011100	01010000	01010100	00000000	[illegible]	10000101	EX\PT...
12	0058h:	11001101	01010110	01011101	01011011	01000110	01000011	01001101	01000010	.V][FCMB
13	0060h:	01000000	01000100	01011001	01010010	01010110	[illegible]	10000101	10000101	@DYRV...
14	0068h:	[illegible]	01010111	01011101	01011010	01000000	01001110	01001000	01001000	.W]Z@NHH
15	0070h:	01001111	01000000	01011010	01011100	01010110	11001101	10000101	00000000	O@Z\V...
16	0078h:	10000100	11101111	01010111	01011101	01011010	01000000	01001111	01001001	..W]Z@OI
17	0080h:	01001111	01000011	01000101	01011101	01010100	10000100	00000000	10100111	OCE]T...
18	0088h:	00000000	10100111	11110001	10000101	01010101	01010011	01011111	01011011	....US_[
19	0090h:	01011010	01011000	01011101	01010100	11101110	10101111	10000010	11001011	ZX]T....
20	0098h:	11100101	11001100	11001011	10101110	10000000	11110011	00000000	00000000	....

Fig. 2 The payload of an RTP packet (sequence number 39796)

tables can be applied. In the latter case, the tables need to be provided to the counterparty. Using such a table, we obtain binary sequences that correspond to individual alphanumeric characters. In order to encode basic English alphanumeric characters, including lower case ones, a total of 62 unique sequences is required.

To separate and unambiguously identify the position of the sequences in the payload of the RTP packet, it is necessary to allocate 4 specific characters, where the relative positions of binary sequences are recorded, or 3 specific characters, where absolute positions of binary sequences are recorded. Using 7-bit ASCII tables, we obtain 128 unique sequences, an amount sufficient to transmit basic alphanumeric characters. When converting the characters from the steganogram, a character of the steganogram is looked up in the ASCII column in the converting table ASCII [15]. Its corresponding binary sequence value can be found in the Binary (BIN) column. This procedure is repeated until we obtain binary sequences for all characters of the steganogram. Subsequently, the binary sequences are tagged in the payload of the RTP packet, as is depicted in Fig. 2 in the real RTP packet for a selected word STEGANOGRAM.

3.1 Tagged Binary Sequence in RTP Payload

To tag the required binary sequence in the payload of the RTP packet, the packet sniffer should be applied (tcpdump is enough). It enables to display the payload of an RTP packet in the binary format. Even better results can be obtained by applying the network protocol analyzer which allows for filtering individual protocols and separating the payload of individual packets from each other. Based on the sequence, we subsequently look up the number of the RTP packet in which the required binary sequences are located. A particular sequence is tagged as follows: a binary sequence of a particular character obtained by means of the converting table is looked up in the payload of a particular RTP packet and its position is tagged. Where the required binary sequences failed to be identified in the packet, we proceed to search through further packets which contain the required binary sequence. To determine the position of a binary sequence we can apply both relative and absolute values. Afterwards, the sequence numbers and the position of individual binary sequences are recorded into the pointer in the SIP header.

3.1.1 The relative position of a sequence

The relative position of a sequence is defined by the first position of the binary sequence representing the first character of the steganogram. This position (the first position in the binary sequence) is the starting point for the determination of the distance between the first character of the octet and subsequent binary sequences. To determine the individual position, the following equation is applied (2):

$$P = B - F \tag{2}$$

where P is the final position of the given character, B is the number of the starting position of the sequence representing the given character and F is the number of the position of the sequence of the given character. We repeat this procedure until we have found all the necessary binary sequences, representing all the characters within the steganogram. Where the binary sequence of the first character is located at the end of the payload of the RTP packet and the subsequent characters are positioned before the first binary sequence, the relative distance values can be negative.

3.1.2 The absolute position of a sequence

When applying the absolute position, there is no need to determine the distance of the positions between individual binary sequences. All we need to establish is the value of the position of the given octet. This approach simplifies and speeds up the tagging of a particular binary sequence, as is depicted in Fig. 3. It also reduces the number of the specific characters required since it is not necessary to use the minus sign. The positions of individual binary sequences may gain 160 different values in a particular payload of the RTP packet.

	A	B	C	D	E	F	G	H
1	1	2	3	4	5	6	7	8
2	9	10 [S]	11	12	13	14	15	16
3	17	18	19	20	21	22	23	24
4	25	26	27	28	29	30	31	32
5	33	34	35	36	37	38	39	40
6	41	42	43	44	45	46 [A]	47	48
7	49	50	51	52	53	54	55	56
8	57	58	59	60	61	62	63	64
9	65	66	67 [R]	68	69	70	71	72
10	73	74	75	76	77	78	79	80 [G]
11	81 [E]	82	83	84	85 [T]	86	87	88
12	89	90	91	92	93	94	95 [M]	96
13	97	98	99	100	101	102	103	104
14	105	106	107	108	109	110 [N]	111	112
15	113 [O]	114	115	116	117	118	119	120
16	121	122	123	124	125	126	127	128
17	129	130	131	132	133	134	135	136
18	137	138	139	140	141	142	143	144
19	145	146	147	148	149	150	151	152
20	153	154	155	156	157	158	159	160

Fig. 3 The absolute position of a sequence – the payload of an RTP packet (sequence number 39796)

3.2 The pointer in a SIP header

In order to obtain information about the position of the binary sequences within the payload of the RTP packet,

we need to share such information with the counterparty. The information necessary to unambiguously identify a binary sequence representing a particular character of the steganogram is the *Sequencenumber* or the *relative* or *absolute* position of individual binary sequences within the RTP packet concerned. To share information, we use the *branch* parameter in the *Via* field. Subsequently, a random number of indicators can be shared with the counterparty using any SIP method.

3.2.1 Branch parameter

This parameter serves to unambiguously identify SIP transactions. Its value needs to be unique in time and space for all and any requests sent by the User Agent. The structure of the *branch* chain needs to contain alphanumeric characters. A typical length of the *branch* chain generated using the IP telephony application referred to as softphones differs, see Table 1.

Implementation of Branch Lengths String in Softphones Table 1

Softphone	String length (without magic cookie)
Media5-fone/4.1.3.3034 iOS/8.3	17
SessionTalk Version 5.11 iOS/8.3	28
X-Lite 4.8.0 75950-02930038 OSX	28
YATE/5.0.0 Linux	9

The maximal length of the *Via* field may return values from 0-65535 B. The typical length, however, is much lower. For instance in the SNORTSIP pre-processor, it is pre-set to 1024 B. The Branch parameter contains a so-called *magiccookie*, a chain which is the prefix for every transaction. This chain consists of the following characters: *z9hG4bK* and is constant. We use this parameter to transmit information about the sequence position in selected packets, where the requested information is found after prefix *z9hG4bK*.

3.2.2 Encoding pointers in a SIP header

When applying the *branch* parameter, it is recommended to convert the indicator values so that they correspond better with the standard values typical for this parameter. This approach enhances the chances that the covert communication channel will not be detected. A converting table is used to encode the pointer. Individual data from the indicator is encoded into the hexadecimal format. 7-bit ASCII table provides 128 values. 122 out of these 128 values are used to encode information contained directly in the pointer. Another 3 or 4 values are used for specific characters depending on whether relative or absolute binary sequence positions are applied (Table 2). Every specific character is encoded independently. Hexadecimal values can be recorded using lower case (a, b, c, d, e, f), upper case (A, B, C, D, E, F) characters or a combination there of which further inhibits the possibility to uncover the covert communication channel.

Special Characters for Encoding SIP Header Table 2

ASCII Character	HEX	Function
{	7b	Beginning set of pointers in RTP packet
| |	7c	Separation of individual pointers in RTP packet
}	7d	Ending set of pointers in RTP packet
~	7e	Character minus

When encoding, we go through the pointer (the branch chain), taking three characters at a time. Where the numerical value is higher than 122, we take away one character from the right and we encode only the two-character value. This approach is applied to the entire pointer with the exception of specific characters that are encoded independently.

Encoding the pointer in a SIP header -the relative position. To encode the pointer using the relative positions to tag a binary sequence, four specific characters need to be applied (see Table 2.). The character tagging the start of the pointer set referring to the binary sequences relating to a particular packet (Sequence number), the character for tagging the end of the pointer set referring to the binary sequences relating to a particular packet (Sequence number), the character for splitting individual pointers (binary sequence positions) within the RTP packet concerned and the character representing the minus sign.

Encoding the pointer in a SIP header – the absolute position. To encode the pointer using the absolute positions to tag a binary sequence, three specific characters need to be applied, since the specific character representing the minus sign is omitted. Individual octets representing binary sequences are tagged using values ranging from 1 to 160, by which their position in the payload of the RTP packet is determined.

3.2.3 Transfer of pointers during a call

The Re-INVITE (INVITE) method of the SIP protocol can be used to transmit information during the call or with any next transaction in the SIP dialogue (SIP method BYE). By changing the parameter of an established connection using the Re-INVITE method, a SIP message with a modified header in the *branch* field can be transmitted repeatedly

and it is a new transaction in the established SIP dialogue. The *branch* field of the INVITE method allows for verifying whether the counterparty is ready to receive the steganogram; or the field can remain empty (magiccookie). With the Re-INVITE method, the *branch* field already contains an encoded chain which represents the positions of binary sequences corresponding to the steganogram. Any number of *Re-INVITE* methods can be created during a call as necessary. The Re-INVITE method serves to modify the parameter of an established connection. Fields From, To, Call-ID are usually equally set as in the original INVITE message, but the remaining parameters can be modified.

3.2.4 Searching a binary sequence in the RTP payload

Searching for individual binary sequences that represent particular characters of the steganogram is done based on the information from the pointer in the SIP header in which such information is encoded. Decoding information is done by means of the conversion of hexadecimal values to decimal values using the converting table in which the HEX value (Hexadecimal value consists of two characters) has a corresponding value in DEC. Thus we obtain non-coded information in the pointer, where one can see individual binary sequence positions in the payload of the RTP packet identified through a sequence number.

3.2.5 Convert a steganogram from a binary sequence

Once the pointer in the SIP header has been decoded, the sequence number of the given packet and the positions of individual binary sequences in this packet are displayed. Binary sequences are converted into ASCII again by means of the converting table in which each binary section has a corresponding ASCII representation. Once all binary sequences have been converted, we obtain a steganogram sent by the counterparty.

4. Experimental Verification and Discussion

To verify a feasibility of the steganographic approach, the codec G.711 A-law is used with packetisation period of 20 ms. Values of a single sample may range from 0 to 255, or from 00000000 to 11111111 in the binary format. These values describe the amplitude of the given sample. For this reason, it would be great to obtain a phonetically diverse analogous voice signal, thus increasing the likelihood that many different binary sequences would occur. As a part of the experiment, a standard five-minute audio call was simulated. Asterisk software, to which the two stations were connected, was used as a SIP softswitch. The SIP softswitch itself ran on the Linux (Ubuntu 14.04) operation system. The first station hosted Windows 7 OS and softphone application Yate 5.4.2. The second station hosted OS X 10.10 with softphone application X-Lite 4.8.0.

Binary Representation and Sequence Position of Steganogram with Text Steganogram - Everything Needed For Transmission Steganogram Table 3

Char Sequence number	Binary sequence	Relative position	Absolute position
S 39796	01010011	73	10
T 39796	01010100	600	85
E 39796	01000101	568	81
G 39796	01000111	560	80
A 39796	01000001	288	46
N 39796	01001110	800	110
O 39796	01001111	824	113
G 39796	01000111	560	80
R 39796	01010010	456	67
A 39796	01000001	288	46
M 39796	01001101	680	95

Bob (Originator) sends a steganogram with text *STEGANOGRAM* to Alice (Recipient). Bob converts the characters in the steganogram by means of the converting table [15] into the binary format as seen in the table (Table 3). A connection is established and lasts 5 minutes. Afterwards, Bob goes through the payload of individual RTP packets. Once he finds the binary sequences (depicted in Fig. 2) corresponding to the text of the steganogram, he notes down the sequence number of the packet concerned. All binary sequences for all steganogram characters were found in a single packet with sequence number 39796. Now, Bob can apply any of two approaches to record the position of individual sequences, i.e. the relative or the absolute approach. The non-coded pointer in the SIP header (the branch parameter with the absolute position of a sequence) is as follows:

branch=z9hG4bK39796[10|85|81|80|46|110|113|80|67|46|95]

We can check positions in Figs. 2 and 3, now the coded pointer (branch) in the SIP header is as follows:

branch=z9hG4bK274f067b0a7c557c517c507c2e7c6e7c717c507c437c2e7c5f7d

Table 3 shows that using the absolute positions to tag binary sequences is more efficient. Subsequently, these positions are recorded together with the sequence number of the packet concerned into the branch parameter using the specific characters (Table 2). Using the converting table, the pointer is then encoded into hexadecimal format. Once Alice receives the pointer in the SIP header, she must first decode the pointer from the hexadecimal format into the decimal format using the converting table [15]. Thus she obtains the original non-coded pointer which allows her by means of the sequence number 39796 to determine where the relevant binary sequences are positioned in the payload (Figure 2) of the RTP packet. Last, invidiual binary sequences (Figure 3) are converted back into the text format (ASCII) using the converting table [15].

The total amount of data transmitted in the covert channel reflects the codec used to encode the voice, the frequency of packet generation, the size of the payload of the RTP packet and the length of the call. G.711 codec with A-law logarithmic compression is used in our experiment and the amount theoretical transmission capacity of the covert channel can be determined using the following equation (3):

$$CS = \frac{PC}{PF} \cdot t \quad [\mathrm{B}] \tag{3}$$

where CS is the total amount of data transmitted in the covert channel [B], PF is the frequency of packet generation [s], PC is the size of the payload within a single packet [B] and t is the length of the call [s]. Entering the values in the equation returns 2,400,000 B in case of 5 minutes long call, it is the maximum number of sequences which can appear in RTP packets. In reality, we only have a few opportunities to send a steganogram within one SIP session due to the existence of only several transactions. Although it is possible to apply the re-INVITE method to insert a new transaction in SIP dialogue, the anomaly detectors would be able to detect these transactions too often. On the other hand, known techniques cannotcannot detect such steganograms in ordinary transactions and nowadays no methods to detect existence of such a covert channel exist. Of course, we can take into account the possibility to enhance security based on the shared secret and the total amount of data transmitted in the covert channel can be enhanced by optimising the converting table.

Next issue to be addressed is the length of the branch value. A typical exact value for the length of the branch parameter has not been defined. For each softphone, the implementation of the branch parameter is different. In the softphone group tested, the length of the branch parameter ranged from 9 to 28 characters (without magiccookie). If we draw on information included in definition RFC 3261, each implementation of a SIP protocol by means of UDP can process messages of up to 65535 B in size. The basic length of the branch parameter within the SIP pre-processor for SNORT is set to 1024 B, but can be adjusted within the range from 0 to 65535 B. OpenSIPS implementation refers to the length of 32 B. If we want to prevent the application of the method described in this paper by limiting the length of the parameter to a particular value, e.g. the value corresponding to the typical length of this parameter, it could, under certain circumstances, restrict the VoIP communication as the implementation specification defined in RFC 3261 had not been adhered to. Where the value of the length of the branch parameter is preset to a typical value due to security reasons, the pointer in the SIP header could be adjusted to such limitations by shortening the length which is considered typical. Where the pointer in the SIP header has the typical length of the branch parameter, this steganographic method becomes virtually untraceable.

Another issue is a limitation due to the placement of B2BUA (Back-to-back User Agent) into the communication path. The described steganographic method allows establishing a covert communication between two end users. B2BUA inserted between two communicating parties terminates the connection from the one UA and establishes a completely new connection targeted towards the second UA. By establishing a new connection, new information is fed into the SIP header. As a result, information about the indicators contained within the branch parameter in the Via field is lost. Similar limitations also apply where a SBC (Session Border Controller) element is inserted into the communication path. Nevertheless, such limitations are not restrictive, since a new communication path containing two SIP proxy, which are interconnected by means of SIP trunk, could be created; or a communication could be set up directly between the communicating UAs. The existence of B2BUA or SBC elements in the communication path can be easily detected and communication can be set up directly between both parties.

5. Conclusion

This paper discusses the steganographic method enabling to cover communications within the data stream of the RTP protocol. This method consists in tagging binary sequences in the RTP stream and placing tags into SIP header. These sequences present characters encoded using a converting table. The position of individual binary sequences within the

packet in which it is located is reflected in the branch value, in the Via field in the SIP header. In order to enhance the protection against uncovering the covert communication channel, the pointer in the SIP header has been converted into the hexadecimal format. The transmission of the pointer series in the SIP header can be performed by means of the Re-INVITE method, nevertheless it is recommended to insert only ordinary transactions in the SIP dialogue and not to insert more Re-INVITE modifications artificially since these are prone to be detected as an anomaly. Certain limitations of this method occur when the communication path is interrupted by an element which impairs the original value of the branch parameter of the Via field such as B2BUA or SBC. The benefit of the steganographic method is that it does not modify the transmitted data, thus no degradation or alteration of the transmitted data occurs. In addition, this approach prevents the use of the steganoanalysis. Where the length of the branch parameter corresponds to the commonly used typical values, this method becomes virtually untraceable.

Acknowledgement

The research leading to these results received funding from the grant of SGS reg. no. SP2015/82 conducted at VSB-Technical University of Ostrava, Czech Republic and was supported by the National Sustainability Program under grant LO1401. This research was also partially supported by the Moravian-Silesian Region within the project VSB-Technical University of Ostrava activities with China. The research was performed using the infrastructure of the SIX Centre.

References

[1] KLIMO, M., KOVACIKOVA, T., SEGEC, P.: Selected Issues of IP Telephony. *Communications - Scientific Letters of the University Of Zilina*, 6 (4), 2004, 63-70.

[2] MAZURCZYK, W., SZCZYPIORSKI, K.: Steganography of VOIP Streams, *Lecture Notes in Computer Science*, 5332 LNCS (PART 2), 2008, 1001-1018.

[3] MAZURCZYK, W., SZAGA, P., SZCZYPIORSKI, K.: Using Transcoding for Hidden Communication in IP Telephony. *Multimedia Tools and Applications*, 70 (3), 2014, 2139-2165.

[4] NEVLUD, P., BURES, M., KAPICAK, L., ZDRALEK, J.: Anomaly-Based Network Intrusion Detection Methods. *Advances in Electrical and Electronic Engineering*, 11 (6), 2013, 468-474.

[5] VOZNAK, M., SAFARIK, J., REZAC, F.: Threat Prevention and Intrusion Detection in VOIP Infrastructures. *International J. of Mathematics and Computers in Simulation*, 7 (1), 2013, 69-76.

[6] BERK, V., GIANI, A., CYBENKO, G.: *Detection of Covert Channel Encoding in Network Packet Delays*. Dartmouth College : Hanover, Technical Report TR536, 2005.

[7] LAMPSON, B.W.: Note on the Confinement Problem, Communications of the ACM, 16 (10), 1973, 613-615.

[8] JANICKI, A., MAZURCZYK, W., SZCZYPIORSKI, K.: Steganalysis of Transcoding Steganography. Annales des Telecommunications/Annals of Telecommunications, 69 (7-8), 2014, 449-460.

[9] MAZURCZYK, W., KOTULSKI, Z.: New VOIP Traffic Security Scheme with Digital Watermarking. *Lecture Notes in Computer Science*, 4166 LNCS, 2006, 170-181.

[10] SZCZYPIORSKI, K. HICCUPS: *Hidden Communication System for Corrupted Networks*. Proc. of Intern. multi-conference on Advanced Computer Systems, October 2004, 31-40.

[11] SAFARIK, J., VOZNAK, M., REZAC, F., MACURA, L.: IP Telephony Server Emulation for Monitoring and Analysis of Malicious Activity in VOIP Network. *Communications - Scientific Letters of the University of Zilina*, 15 (2A), 2013, 191-196.

[12] REZAC, F., VOZNAK, M., TOMALA, K., ROZHON, J., VYCHODIL, J.: Security Analysis System to Detect Threats on a Sip VOIP Infrastructure Elements. *Advances in Electrical and Electronic Engineering*, 9 (5), 2011, 225-232.

[13] MEHIC, M., SLACHTA, J., VOZNAK, M.: *Hiding Data in SIP Session*, Proc. of 38th Intern. Conference on Telecommunications and Signal Processing (TSP), 2015, 1-5, doi: 10.1109/TSP.2015.7296445.

[14] MEHIC, M., MIKULEC, M., VOZNAK, M., KAPICAK, L.: Creating Covert Channel using SIP. *Communications in Computer and Information Science*, 429, 2014, 182-192.

[15] ASCII CONVERSION CHART, online available url https://designthatsit.files.wordpress.com/2013/12/ascii20conversion20chart.gif.

Jiri Slachta - Miroslav Voznak - Dan Komosny - Homero Toral-Cruz - Peppino Fazio *

AUTOMATICALLY PROVISIONED EMBEDDED COMMUNICATION SYSTEM BASED ON OPENWRT PLATFORM

The article deals with a design of a system that provides tools for creation of automatically provisioned embedded communication system and its components. As the key feature of the BEESIP platform (Bright Efficient Embedded Solution for IP Telephony) a unique building and provisioning system of the network devices has been developed allowing the administrators to fully control the firmware and configuration of the devices even in the remote and inaccessible locations. The process of custom firmware building and device provisioning eases the mass deployment of the BEESIP based hardware to cover the needs of small to medium business in the vast range of services.

Keywords: *BEESIP, SIP, IP Telephony, RTP, Asterisk, Kamailio, Snort, OpenWrt.*

1. Introduction

In order to reduce costs for maintaining communication systems, both physical devices and services running on them, many companies make a decision on moving the service infrastructure to cloud service providers. Nevertheless, during the movement of infrastructure to third-party several questions might arise. Services in cloud infrastructure were designed to be used for everyone, but they usually lack the broad configurability. In addition to these issues, the security questions have to be resolved, since the infrastructure is not under the control. During the years of development, a platform to address the mentioned issues has been created [1]. Among desired characteristics belongs an easy integration of such device into almost any computer network. In mid of 2011, a new project (BESIP) was established under strong support of the CESNET association (Association of Czech universities and Academy of Science), aiming at a robust and secure VoIP telephony infrastructure with additional key components that make this solution easily adaptable and configurable even without the deep knowledge of the technologies used by the components. It also aims to be a scalable solution with the unified configuration in mind [1]. The given name BESIP has had to be changed to BEESIP (The Bright Efficient Embedded Solution for IP Telephony) in 2014 because our BESIP trademark registration was rejected by the Czech Industrial Property Office due to the same existing trademark in field of public transport. In last two years, next important features have been implemented, and the automatic provisioning belongs to them. Soon after being ported Asterisk to OpenWrt Linux distribution within BEESIP project, we became responsible for maintenance of Telephony repository in OpenWrt, it includes any packages and patches which are connected with telephony.

2. Related work

As mentioned in the introduction, we discuss the implementation of a SIP communication server solution which would be an alternative to several current implementations. The main advantage of our solution is the ability to easily and quickly set up a full-featured PBX on almost any hardware. We can presume that almost all implementations are based on open-source Asterisk PBX, web-interface for Asterisk and with a GNU/Linux distribution on the base layer. At present, there are several projects that offer multipurpose IP telephony solutions for embedded devices and for household or enterprise platforms [2 and 3]. The initial project of a GNU/Linux distribution which offers an easy set-up of IP telephony in a few steps is the Asterisk@Home project. The first version of this project was released on 29 April 2005.

* [1]Jiri Slachta, [1]Miroslav Voznak, [2]Dan Komosny, [3]Homero Toral-Cruz, [4]Peppino Fazio
[1]VSB-Technical University of Ostrava, Czech Republic
[2]Brno University of Technology, Czech Republic
[3]University of Quintana Roo, Col. del Bosque, Mexico
[4]University of Calabria, Arcavacata di Rende, Italy
E-mail: jiri.slachta@vsb.cz

This project integrated a web interface for Asterisk, Flash Operators Panel to control and monitor PBX in real-time and also offered a full FAX support within one bootable image for almost any x86 PC. On 3rd May 2006 the development of this project was discontinued and was replaced by its successor Trixbox. However, the development of Trixbox does not seem to continue any more. Two existing projects - AsteriskNOW and Elastix – now offer an alternative to Trixbox. The former, AsteriskNOW appears to be similar to Trixbox – a packed GNU/Linux distribution with Asterisk with a FreePBX web interface on top of it. The latter, Elastix, is a bit more modular. Compared to any other project, it offers a slightly more modular hierarchy to facilitate the applicability to a multiple service server [4 and 5]. The increasing popularity of embedded devices, such as Raspberry Pi, is the reason why the Micro Elastix distribution was born. However, all of those projects are either prepared for x86 machines only or for specific hardware. Micro Elastix only supports three platforms, namely PICO-SAM9G45, MCUZONE and Raspberry Pi [6].

None of the projects includes a security module that would offer a complete IPS and IDS system to prevent attacks against the SIP Registrar server. Also, there is no module that would monitor the quality of voice calls transmitted through an integrated PBX [2 and 7]. Thanks to the portability of the OpenWrt distribution we prepare a BEESIP bootable image for almost any device.

3. Platform architecture

One of the biggest challenges during BEESIP development was to create or modify any existing Linux distribution to serve our expectations. We needed to create an environment that would be fully customizable to any purpose and also to be easily maintainable through the time the BEESIP would be developed. The choice of Linux distribution, we wanted to modify, fell on OpenWrt Linux distribution. The reason, why we chose that system, was the approach for building firmware, the toolchain, cross-compiler and all applications are downloaded, patched and built by scratch. It means that OpenWrt does not contain any source code, it does only have its build system with templates, patches and Makefiles with procedures how to build a system and its packages for targeted device. This approach allows us to create custom procedures for build system and packages that can be modified at any stage. A simplified view on BEESIP architecture is depicted in Fig. 1 which describes how the architecture is designed.

The first block, the build system, is a wrapper on the top of the OpenWrt build system. It is designed for easy creation of firmware images within the single text file which describes what should be built for specific architecture and device we are

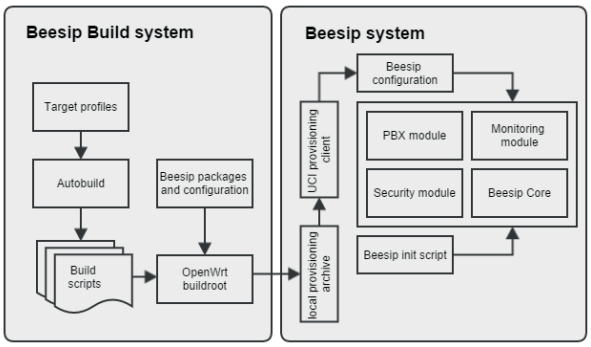


Fig. 1 Architecture of BeeSIP system

targeting on. The secondary part of BEESIP is the OpenWrt Linux distribution which uses packages that provide desired functionality. In this case to provide modules from packages that serve as PBX, monitoring system, security system, management system and the core connecting modules among themselves. In almost each home and small office there has been distributed and deployed embedded networking equipment for routing the Internet connection, providing multiple media services and securing the network behind the device. Despite the fact that there has been some focus on the security of network itself, those devices have received small attention to prevent the attackers to abuse the open vulnerabilities on such equipment. The absence of computational capacity on such devices is also another fact that has to be resolved to prevent denial of service. The solution of security in BEESIP is based on SNORT application with cooperation of SNORTSam and iptables [8 and 9]. In addition to preventing multimedia systems being unusable during DoS attacks the system has to protect itself. For limiting and blocking the attacks over VoIP traffic the ratelimit and pike modules from Kamailio package are used. The architecture of BEESIP system focuses mainly on providing multimedia services, such as IP telephony. Administrators of this service have to ensure if the content is delivered reliably, securely and the voice traffic should also follow given quality parameters as well due to an impact of background traffic on speech quality [10 and 11]. In the beginning of the project the first version of monitoring system was proposed [12]. For monitoring purposes the measurements of IP telephony traffic are achieved directly on the device. This solution exploited a tshark package and our java application interpreting the results from tshark. This one-time measurement gives information about particular speech quality. However, this solution providing one-time measurements was not robust enough. To address the anomalies in the network infrastructure the successor of the previous application has been made. Nowadays, the monitoring module works as an agent in the system which provides continuous monitoring evaluated immediately on the monitoring server and our work in this field received a best paper award at 22nd International

Science Conference on Computer Networks [13]. The rest of the monitoring functionality is handled by the Zabbix agent. The described modules deliver heterogeneous services which have to work in conjunction with each other. The last part of the system are core services that provide abstraction layer on the top of the modules. It is represented by a shell library (providing functions for scripts) with executable files to make the system working. As a configuration provider we developed UCI provisioning client that prepares the configuration for the system.

4. Build system

Before the concept of BEESIP system is described, it is necessary to introduce the build system which reduces the building procedure into one script call. As said above, BEESIP is based on GNU/Linux distribution OpenWrt which is built on top of the OpenWrt Buildroot. Buildroot is a set of Makefiles and files that allows to compile cross-compilation toolchain and to generate by that toolchain resulting cross-compiled applications into a root filesystem image to be used in the targeting device. Cross-compilation toolchain is compiled by host compilation system which is provided by any GNU/Linux distribution. In the beginning of BEESIP development we met issues that were holding us back. We could not test all changes immediately, we had to recompile all codes and generate images nearly always when we ported new application, modified post installation scripts or when cross-compilation toolchain has changed. Also, the system behaves differently during testing if it is new root filesystem image, or modified root filesystem that has been run more than once. At least those issues led us to create an easy interface that will ease the creation, automation and functional testing for system images. The BEESIP build system is a set of scripts, Makefiles and definition files that make an easy interface to OpenWrt Buildroot. We can consider the main Makefile to be as a core of the BEESIP build system. It performs all atomic operations with OpenWrt Buildroot, works with source code management systems (to update/revert/any operation with local copies of OpenWrt source codes), patches OpenWrt Buildroot and executes images as virtual machines. Those commands might be used by any user or by autobuild scripts, which will be described after. On the top of the core Makefile is autobuild.sh script. This script calls all atomic operations within more complex parameterized operations whose variables are defined in specific target files. Those target files are user defined and on the basis of those files are configuration files for OpenWrt Buildroot created. Once we have configuration files the system images could be created by calling autobuild.sh script with command build and parameter containing the name of the target file. Such techniques can be used for any purpose of automated building system images for any device or platform supported by OpenWrt. Those could be firmware images for campus access points, specialized network probes, virtualized multimedia servers or any other devices.

5. Core module

The role of the Core module is to provide a glue among all services that served by all BEESIP modules. The most important part of the Core module is the BEESIP shell library that provides functions for all utilities and scripts used by BEESIP system. Functionality of a Core module complements utilities for configuration management and for simplified configuration of system image. With all those utilities comes along also default configuration which prepares all module services into fully functional state with all BEESIP modules running and operational. Also, the role of this module is to switch any existing OpenWrt environment to BEESIP environment while the device is booted the first time or the BEESIP environment is used and ran the first time.

5.1. BEESIP telephony environment

Since IP telephony in the Czech national research and education network is highly developed and we interconnected via VoIP nearly all PBXs' of Czech universities 15 years ago [12], we keep information about academic network infrastructure (more than 50 VoIP Gateways and PBXs behind them). The project BEESIP should draw benefits from its nature. Each participant of this network stores the data about their VoIP gateways in the IPTelix system, which is the database for the VoIP gateways connected to the Czech academic research and education network. The main focus of the system is to maintain and to monitor prefixes to the gateways. The data are provided from the database in the JSON format, our scripts in BEESIP automatically prepare configuration files for internal Asterisk and set up the outgoing traffic to establish trunks against these gateways. To identify the VoIP traffic the data from BEESIP UCI configuration file are obtained and subsequently the IP telephony prefixes and the numbering plan is set up. Phone provisioning tool, which is connected to internal Asterisk in the system, creates phone provisioning data according to the type of the phone connected to it.

5.2. Provisioning client

The impetus for development of provisioning tool arose during the period when firmware images created by BEESIP build system were deployed to computers, routers and wireless access points. Those machines were not configured for target networks, which were supposed to be deployed on. Because the target configuration does not depend on a person who builds the system, but on the network administrator, then configuration should lie outside of a BEESIP firmware image. The creation of such tool brings a question how the target device should fetch and apply its configuration. In the build system, we can pass static information about our provisioning server which provides configuration (during build time). We can also change this information in firmware image. This information can be used for protocols which translate one kind of information to another. As an example we can use DNS protocol and its TXT records. The target configuration could be stored on a server designated within an URI in a variable from TXT record which is obtained from static URL provided by build system. This solution is replicable for any protocol which allows distribution of that kind of information (LLDP, DHCP or any other else). An example how to resolve UCI provisioning URI:

```
host -t txt provdomain \
 provdomain descriptive text \ "provuri=http://12.34.56.78/uciprov/"
```

If a device knows where to obtain configuration from, then the device can construct all provisioning URI addresses for each device state it needs. This approach is necessary when system administrator needs to differentiate configuration for devices which start up the first time. UCI provisioning client currently handles not only configuration files that are handled by UCI system (Unified Configuration Interface) for centralized configuration, but also the tar archives that consist of compressed overlay. If a device knows where to obtain configuration from, then it can obtain configuration data from ordinary transport protocols designated in provisioning URI. The benefits that BEESIP draws from OpenWrt builds upon the UCI configuration system which is based on plain text configuration files with firmly defined structure. This configuration is obtained using software for file retrieval from network resources, e.g. wget, and immediately imported into UCI. From the introductory part of motivation for the techniques it is clear why provisioning is a needed component for configuration deployment on higher number of such devices. During the development of any application or any system the developers need to simplify the process of deployment of applications and its configuration, thus the UCI provisioning tool was developed, known under abbreviation uciprov. The architecture of the uciprov tool stands on the two separate parts. The first part of the uciprov tool is located in the build system of OpenWrt. The package itself supports the selection of used protocols for discovery of provisioning URI, and also offers user to add specific variables during the build time. Within the package we can also work with macros, thus all variables do not have to be static at all. This can be used when the domain has to be resolved, but the URL structure is known. There are several macros that mostly identify the hostname, domain, MAC address, IP address, release number and many others and finally, we are able to make system upgrades or automatic system reconfiguration without administrator intervention on the targeting device. An example of used macros in the build system for the uciprov tool can be seen below where the URI to image for the system upgrade is distributed.

```
string "Static URI for sysupgrade"
   depends on UCIPROV_USE_STATIC
   default "{base_uri}/image{fd}.bin"
```

The second side of UCI provisioning tool is an installable package to OpenWrt system. Despite the fact that the tool was designed for the distribution of UCI configuration among all desired devices, it can call any function that we hook to any stage of uciprov tool. The modules for this tool must hook their functions with the following specific uciprov stage call „uciprov_hook_add uciprov_stage custom-function-name“ in the script preamble. Thus we are able to upgrade the BEESIP system, distribute SSH keys or configuration files that do not comply with UCI syntax. Since we are working only with variables, we are able to move the application logic into scripts. This led to the implementation of URI resolver scripts (DNSSEC, DNS, HTTPS...) and also to scripts for handling the constructed URIs (system upgrade, public key distribution).

The server side of UCI provisioning is currently solved by providing static file structure with files which consist of an export provided by UCI system. The flow diagram is depicted in Fig. 2 and the description, how UCI provisioning works, is following:

1. Waiting for the system to be ready to be provisioned.
2. Stage 1 (preinit):
   a. During the first stage the URIs are obtained.
   b. Uciprov macro functions are set up from UCI configuration file. The same applies to every variable.
   c. Subsequently the uciprov_geturi hook is called. This stage calls every URI resolver script.
   d. Call user hooked scripts.
3. Stage 2 (obtain configuration from URI):
   a. URI addresses are validated.
   b. Obtain configuration or files from user modules.
   c. Call user hooked scripts (preapply).

d. If obtaining configuration failed, retry stage 2.
4. Stage 3 - apply received configuration:
5. Call user hooked scripts (postapply, reboot).

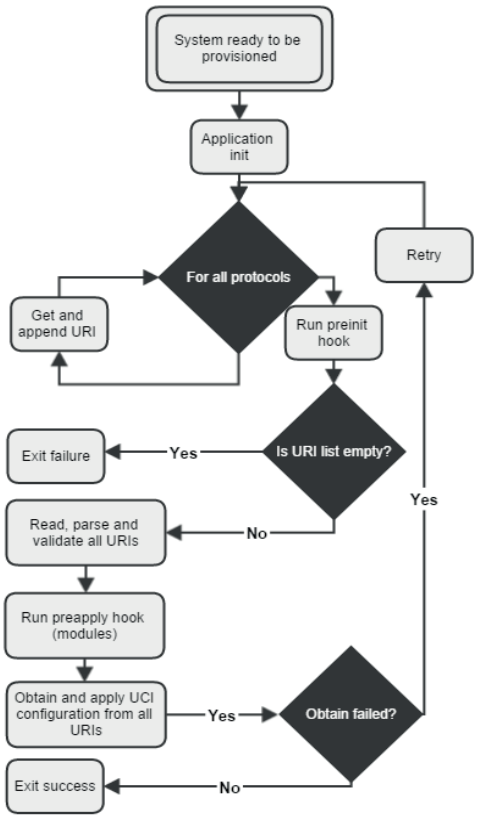


Fig. 2 Flow diagram of UCI provisioning client

6. PBX module

The PBX module is a key part of the BEESIP project. It operates as SIP proxy or SIP B2BUA, depending on configuration, and ensures a call routing. Asterisk is used for call manipulation and for the PBX services. Kamailio is used as a complementary part of PBX module for the proxying SIP requests, the traffic normalization and for the security. There are always two factors when developing VoIP solution. The first one is high availability and reliability, the second one is an issue of advanced functions. Many developers try to find a compromise, we have implemented both, and our BEESIP is able to adapt to the user requirements [12]. More complex system can handle many PBX functions such as a call recording or an interactive voice response but due to the bigger complexity it is more susceptible to fault. On the opposite side, pure SIP proxy is easier software, which can perform call routing, more fault tolerant, but it is more difficult to use the advanced PBX functions.

7. Security module

The security module is next element of BEESIP and all the time, it was considered to make the developed system as secure as possible [8]. Next to this, the entire system has to be fault-tolerant, monitored and protected from attacks. If a security incident is detected, BEESIP immediately solves the situation and notifies this event in a detailed report to administrator. The attacks are recognized and processed by SNORT rules, the source IP address is automatically sent into the firewall by SNORTSam and the intruder's IP is blocked. This is very flexible, reliable and efficient solution. Dropping attack based on IP directly in the Linux kernel is much more efficient than to check messages on the application level. Only first messages are going to SNORT filter. When SNORT identifies a suspicious traffic, next messages from the same IP are blocked. If more soft faults appear from some IP, it is blocked at the IPTABLES level; this approach can effectively block incorrectly configured clients and servers. For example, if a client sends REGISTER with proper credentials, it is not obviously security attack but the client attempts to register again and again, with every registration requires computing sources at SIP REGISTRAR server, see Fig. 3. Such attempts can be denoted and blocked for a time interval, the line IPS (Intrusion Protection System) represents the CPU load in case of active security module in BEESIP. The dependencies clearly prove the ability of security module to mitigate the performed attacks.

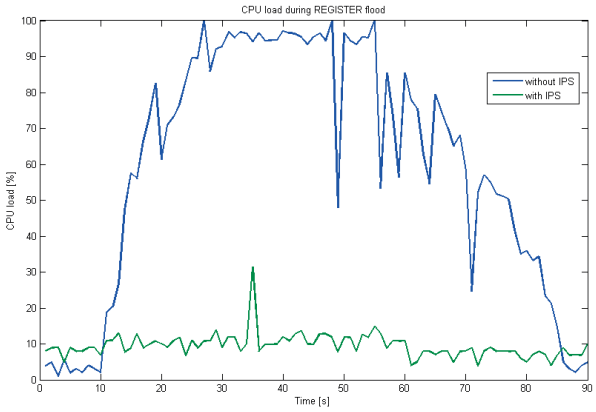


Fig. 3 Attack effectiveness based on REGISTER flood

Administrators can use Zabbix agent inside BEESIP to gather all information directly into their monitoring system. Partially, BEESIP is resistant to some kind of DoS attacks. It depends on hardware used. If the hardware is strong enough to detect some security incidents on application level, the source IP is immediately dropped. Low-performance hardware cannot handle such detection on application level. In such a case, it is more suitable to stop DoS attack before it reaches

BEESIP. Therefore, the SNORT running on a dedicated machine provides more flexible and robust solution than the SNORT as an integral part of VoIP system [8].

8. Monitoring module

Due to the nature of the BEESIP architecture, which is focused mainly on the embedded and low-profile devices, BEESIP can be deployed as the network monitoring probe as well, as is depicted in Fig. 4. For these purposes, libraries allowing continuous monitoring of speech quality have been developed and incorporated into the BEESIP system. The monitoring module takes the advantage of the Asterisk PBX, which is the part of the BEESIP's PBX module, and is designed to work as the network probe [13].

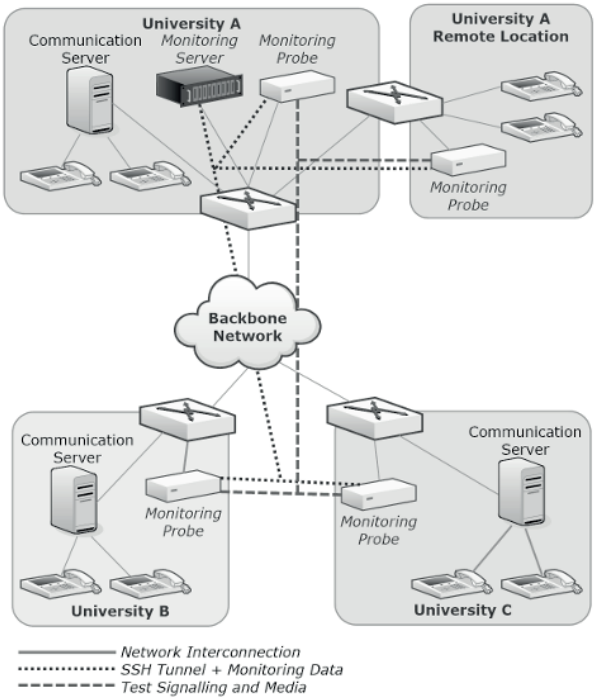


Fig. 4 The monitoring system architecture

For the BEESIP's monitoring module to work properly there are two necessary components. The first one is the BEESIP itself in the role of speech monitoring probe, the second one is the data collection and presentation server. While the former is fully implemented in the BEESIP, the latter requires only a mainstream platform capable of running HTTP server with Zabbix monitoring server and can therefore be run on any commonly used server platform. The server role is to collect the recorded phone calls and calculate the Mean Opinion Score for the sound files, the client part looks into the database of the partner probes, which is obtained from the Zabbix server, and then establishes a call to all these partner probes in regular intervals. The call media consist of sound samples that conform to the ITU-T P.862 recommendation, the media payload is identical for the both directions of the call and is recorded to the wav file. The resulting wav file is then sent to the server for analysis and presentation of the results. Due to the relatively simple implementation of the monitoring probe a highly scalable solution for speech quality monitoring can be deployed using the cost-effective hardware. The BEESIP ensures the automatic addition of new probes to the monitoring system, but it requires the access to the local DHCP (Dynamic Host Control Protocol) service, where the administrator needs to enter the records the probe requires to work properly. First of all, it is the probe's IP address and hostname, which are used to communicate with the rest of the network. Then it is the server's IP address or its resolvable domain name to allow the probe to communicate with the server and the last item (apart from standard ones, e.g. default gateway, etc.) is the path from which the probe can download its private key. The last item has to be kept secret and the communication must be secured using HTTPS protocol and restricted only for intended hosts. With all this information in place the probe then can register itself to the server without the user interaction, which is useful in large deployment scenarios. The BEESIP provides the secure communication based on public keys infrastructure, which is also used to provide access control, so that only authorized probes can join the system and fetch the sensitive data about the other probes. The transfer of network information from the server to the probes uses a module, which is built upon the Zabbix monitoring server [13]. By using the custom made tools and standard Zabbix communication application interface the direct link from the probes to network information database of the Zabbix server is created. The probe then reads the complete list of the probes, which is then used for testing. A standard Zabbix web interface is used together with visualization tool that allows to render a map of probes and the tested interconnections, so that the results of measurements are easily accessible and understandable. Together with the information about the partner probes, server sends the parameters of the test as well. The most important parameter carries the information about the time period between individual test rounds. All the test calls initiated by one probe are performed at once, which in case of precise time synchronization across the network will result in all calls to be performed at once. Due to relatively low bandwidth of a single call it is very difficult to reach measurable network load in the intended environment of backbone networks. When the call (or test round) is finished, the recorded wav files are then transferred to the server for analysis. The test rounds are then performed periodically and their frequency depends on the time interval set by the administrator of the network. This way

the compromise between network load and responsiveness of the system can be defined thus making the system usable in a vast number of different network environments.

9. Use cases

In this paper, main ideas of the BEESIP architecture are described, the BEESIP includes own provisioning and build tool that is able to build this OpenWrt-based system for almost any platform with desired characteristics. Up to this time there has been developed following use cases:

- Eduroam Access Points – Fully provisioned, adaptable and monitored Access Points (open-source and low-cost version supporting and spreading IP mobility and roaming within academic environment).
- Honeypots and network probes – fully provisioned honeypots and network probes that are able to evaluate network characteristics and collect network data.
- PBX and SBC (Session Border Controller) – Components for building an infrastructure that provides VoIP telephony.
- Speech quality monitoring probes – fully provisioned probes providing speech quality assessment with integration into the monitoring system Zabbix.

10. Conclusion

This paper describes the idea and the proposition of the BEESIP system, which is based on one of the popular Linux distributions for the embedded devices. The developed system is fully adoptable and each component is reusable on any other Linux distribution. This system introduces modules for several areas, such as security, PBX, provisioning or management. Some of them have been developed from scratch and the rest of the components have been fully adopted. New tools for speech quality assessment and provisioning have been especially designed for the deployment on embedded devices. The contribution of our work is in the overall design, implementation and maintenance of open-source communication system BEESIP. We are aware of the fact that our work represents applied research and experimental development and is highly practically oriented, nevertheless with many positive comments from research and industrial community (e.g. technology director at Cisco on linkedin). All source codes are released under GPL license and available as open-source software. Fully functional platform images are distributed and prepared mainly for x86 platform, which is also possible easily virtualized for testing or deployment purposes. Nowadays there are platform targets created for x86 platform and access points based on MIPS architecture (ar71xx platform). Binary images from the auto-build system can be downloaded from [14] and source codes can be cloned from GIT repository from the same page as well [14]. There are several example firmware images for several target devices, such as TP-Link access points or Raspberry PI computer. Configuration is available through web-browser, SSH client or to be provisioned using supported provisioning protocols.

The approach used in the build system can help networkers to maintain increasing amount of devices that are under their administration. Moreover, we became responsible for maintenance of Telephony repository in OpenWrt and the trust given us by OpenWrt community is the real appreciation for our work in the BEESIP project. Now, we check and help to improve every patch and package which is submitted by developers in Telephony OpenWrt repository. We are closer to the needs of networkers and connected with OpenWrt community.

Acknowledgement

The research leading to these results received funding from the grant of SGS reg. no. SP2015/82 conducted at VSB-Technical University of Ostrava, Czech Republic and also was supported by the National Sustainability Program under grant LO1401. For the research, infrastructure of the SIX Center was used.

References

[1] VOZNAK, M., SLACHTA, J., MACURA, L., TOMALA, K.: Advanced Solution of SIP Communication Server with a New Approach to Management, *Telecommunication Systems*, vol. 59, No. 4, 2015, 541-549.

[2] SEGEC, P., KOVACIKOVA, T.: A Survey of Open Source Products for Building a SIP Communication Platform, *Advances in Multimedia*, no. 372591, 2011.

[3] ABID, F., IZEBOUDJEN, N., BAKIRI, M., TITRI, S., LOUIZ, F., LAZIB, D.: *Embedded Implementation of an IP-PBX/VOIP Gateway*, Proc. of 24th Intern. conference on microelectronics, No. 6471377, 2012.

[4] TITRI, N., LOUIZ, F., BAKIRI, M., ABID, F., LAZIB, D., REKAB, L.: *An Opencores/Open-Source Based Embedded System-On-Chip Platform for Voice Over Internet,* INTECH: VOIP Technologies, 2011, 145-172.

[5] PRASAD, J., KUMAR, B.: *Analysis of SIP and Realization of Advanced IP-PBX Features*, Proc. 3rd Intern. conference on electronics computer technology, vol. 6, no. 5942085, 2011, 218-222.
[6] SENTHIL, S. K., DHIVYALEKSHMI, B. S., PREETHI, S., PERUMALRAJA, R.: PBX Implementation in Lan Using Asterisk Open Source Software, *Intern. J. of Applied Engineering Research*, vol. 10, no. 55, 2015, 66-69.
[7] ALAM, M., BOSE, S., RAHMAN, M., AL-MUMIN, M.: *Small Office Pbx Using Voice Over Internet Protocol (VOIP)*, Intern. conference Advanced communication technology, No. 4195481, 2007, 1618-1622.
[8] SAFARIK, J., VOZNAK, M., REZAC, F., MACURA, L.: IP Telephony Server Emulation for Monitoring and Analysis of Malicious Activity in VOIP Network, *Communications - Scientific Letters of the University of Zilina*, vol. 15, No. 2a, 2013, 191-196.
[9] CHI, R.: Intrusion Detection System Based on Snort, Lecture Notes in Electrical Engineering, *272 LNEE*, vol. 3, 2014, 657-664.
[10] POCTA, P., KORTIS, P., VACULIK, M.: Impact of Background Traffic on Speech Quality in VOWLAN, *Advances in Multimedia*, vol. 2007, 2007, No. 57423.
[11] MRVOVA, M., POCTA, P.: A Quality Estimation of Synthesized Speech Transmitted over IP Networks, *Communications - Scientific Letters of the University of Zilina*, vol. 16, No. 1, 2014, 121-126.
[12] VOZNAK, M., TOMALA, K., VYCHODIL, J., SLACHTA, J.: Advanced Concept of Voice Communication Server on Embedded Platform, *Przeglad Elektrotechniczny*, vol. 89, No. 2b, 2013, 228-233.
[13] REZAC, F., ROZHON, J., SLACHTA, J., VOZNAK, M.: Speech Quality Measurement in IP Telephony Networks by Using The Modular Probes, *Communications in Computer and Information Science*, vol. 522, 2015, 172-181.
[14] Project beesip, repository with source codes, available on url: https://homeproj.cesnet.cz/ projects/besip/wiki/ download.

Peter Kajan - Patrik Kamencay - Jan Hlubik - Robert Hudec - Miroslav Benco - Peter Sykora *

REAL-TIME FACIAL MOTION CAPTURE USING A WEBCAM

In this paper, we create a software which would be able to combine marker and marker-less face tracking for the purpose of the creation of the accurate and real time running 3D facial animation inside the Cinema 4D. The aim is develop a powerful piece of the software that can provide facial motion capture and share the motion coordinates on the server in real time. The software solution includes tracking software which is written in C++ and provides real time face tracking. Tracking data are sent to the server in real time. Main benefit is that this solution can provide tracking data on the server for any third software. At least, the smooth run of the above mentioned real time tracking software requires simple color webcam with VGA resolution, running on 30 frames per second and a multi-core processor including at least two threads.

Keywords: *Facial Motion Capture, C++, face tracking, 3D facial animation.*

1. Introduction

The motion capture is defined as the process of recording the motion of objects, people or animals. It is used in different areas including robotics, military, sports, medical applications and entertainment. Facial expressions of the human actors are recorded and processed by means of the computer vision algorithms, which provide us with required information about the motion. This information is used to animate the digital character within 2D or 3D computer animation. Face motion capture is focused only on the facial motion of an actor. For this purpose, in order to capture facial expressions, many different approaches are used. This capture can be done in two or three dimensions, depending on the number of the cameras, which are used. To get full three dimensional information about the play of the actor's face and rotation of the head professional solutions use multi camera rigs or laser marker systems. On the other hand, two dimensional solutions use only one simple camera which cooperates with suitable software and gets two relative dimensional coordinates of the facial elements, such as lips, eyes, eyebrows [1].

The aim of this paper is the practical verification of the theoretical knowledge in the field of the computer vision and image processing. With using only one simple webcam we experimentally verified the suitable software technology for the real-time face tracking and the real-time 3D facial animation [2]. The theoretical part describes the fundamentals of the digital image processing with focus on the facial motion capture. The practical part uses this knowledge in order to create solution for facial motion capture software. This software provides the face tracking in real time and uses the tracking information in order to provide the facial animation in the environment of the Cinema4D.

The rest of the paper is organized as follows: In Section 2, state of the art is briefly reviewed. The design and creation of the face tracker is discussed in Section 3. The conclusion and comparison of our face tracker with other commercial systems is described in Section 4.

2. State of the Art

Computer vision is a part of the artificial intelligence concerned with computational understanding and processing of the visual information provided by digital video cameras. The main goal of the computer vision is to duplicate the abilities of the human vision by means of the electronically perceived and understood digital image. This enables computers and robots to see and understand in nearly the same way as human beings do. Within the frame of the computer vision, such as object tracking, we are able to solve some specific tasks with individual algorithms [3].

2.1 Object Tracking

Object tracking can be defined as a problem of how to estimate the trajectory of the specific object in the image

* **Peter Kajan, Patrik Kamencay, Jan Hlubik, Robert Hudec, Miroslav Benco, Peter Sykora**
Department of Telecommunications, Faculty of Electrical Engineering, University of Zilina, Slovakia
E-mail: patrik.kamencay@fel.uniza.sk

plane, as it moves around the scene in time. Object tracking is a complex process which can be split into several categories. The main task of the object tracker is to generate over time trajectory of the object by locating its position in every frame of the video. There are two methods how the tracker can provide the detection and establish the correspondence between the objects across the frames. First of them establishes the correspondence separately. In every frame, the possible object regions are obtained by means of the object detection algorithm and then the tracker makes correspondence between the objects across the frames. The second method performs detection and correspondence at the same time. At the same time the object region and correspondence are estimated by iteratively updated object location and region information. All information is obtained from the previous frame [4].

2.2 Object Detection

Every tracking method requires a mechanism for the object detection. The common approach for the object detection is that we try to detect a particular object separately, in every single frame. To avoid detection errors the most advanced methods of the object detection evaluate the object position in a few consecutive frames. The most popular detection methods are listed below [4]:

- Point detectors - Point detector finds interesting points for the tracking in images. Interesting points for the tracking are usually some corners and areas with a significant texture. Commonly used interest point detectors are Harris interest point detector and SIFT detector.
- Background subtraction - The background subtraction is a method based on the assumption that the tracking objects move between the frames and that the background remains static. The tracking object is detected by this background subtraction.
- Segmentation - The segmentation algorithm splits the image into perceptually similar regions. There are two very important issues the segmentation algorithm has to deal with. First, it has to set up the right criteria for a good partition, and second, it has to achieve an efficient partitioning.

3. Design and Creation of the Face Tracker

The proposed face tracker is a real-time face tracking application which has been developed for the purposes of this paper. This face tracker uses two independent threads which work simultaneously, i.e. at the same time. One of them is the main thread, which manages the whole process and provides all the necessary tasks and resources which are requested for the second thread. The second tracking thread provides the tracking for each frame and after the processing of the tracking data, these data are sent to the server.

Very important part of the proposed face tracker is the tracking core. The core is split into two parts. The first part includes detection modules and the second part contains the data processing modules. This second part is called data processing part.

3.3 Detection Modules

The main task of the detection modules is to correctly detect the absolute position of the important face elements in the current frame. The flowchart of the detection modules is illustrated in Fig. 1.

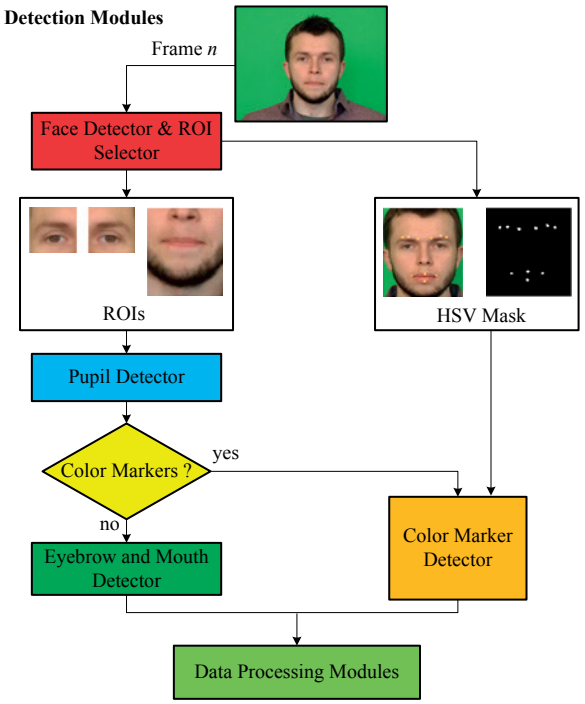


Fig. 1 Flowchart of the detection modules

There are two fundamental modes. First mode detects the face features without any color markers. This marker-less mode includes eyebrow and mouth detector. Second mode is used for more accurate tracking. This mode requires footage with color markers on the actor's face. As it is impossible to attach the markers on the eyes, it is the pupil detector which is always used for both modes.

3.3.1 Face Detector & ROI Selector

In order to detect the human face and other facial features, such as mouth or eyes, the Haar classifier cascades have to be trained. The OpenCV library contains xml files with classifiers which are trained for the human face and other facial features. In our case, we used the haarcascade_frontalface_alt.xml. This cascade is used only for the face detection with a very good accuracy – of more than 80%. The OpenCV function for the detection of the multi-scale object is used in the Face Tracker. The function returns in the vector of the rectangles. Each rectangle represents the detected face area of each face in the image. In our case, we assume only one face which is supposed to be in the image, so the function supposes to return in the vector which contains only one rectangle. In that case, more than one face will appear there, and the biggest one will be selected. The detected face is shown in Fig. 2.

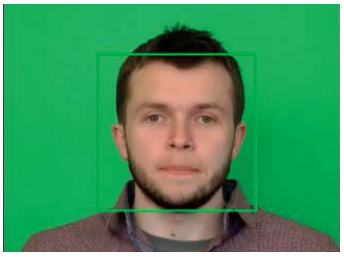


Fig. 2 Detected face ROI

Next step concerns the selection of the regions of interests (ROI) for eyes and mouth (see Fig. 3). The OpenCV Haar-Cascade also allows us to detect facial features, such as nose, eyes and mouth. Although this detection is quite accurate, a huge computational power is required. In order to keep the computational time as small as possible we have to take a different approach. According to the known proportions of an average human face, the ROIs are defined for each part.

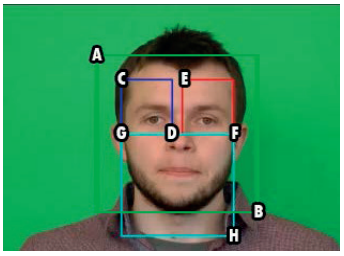


Fig. 3 ROI selection for each part of the face

Every ROI can be described by two 2D points. The width of the face rectangle is defined as follows:

$$W = Bx - Ax \tag{1}$$

The computation of other ROIs is defined below:

$$\begin{aligned} C \cdot x &= A \cdot x + (W * 0.15) \\ C \cdot y &= A \cdot y + (W * 0.15) \end{aligned} \tag{2}$$

$$\begin{aligned} F \cdot x &= A \cdot x + (W * 0.85) \\ F \cdot y &= A \cdot y + (W * 0.5) \end{aligned} \tag{3}$$

$$\begin{aligned} D \cdot x &= C \cdot x + \left(\frac{F \cdot x - C \cdot x}{2.2}\right) \\ D \cdot y &= F \cdot y \end{aligned} \tag{4}$$

$$\begin{aligned} E \cdot x &= F \cdot x + \left(\frac{F \cdot x - C \cdot x}{2.2}\right) \\ E \cdot y &= C \cdot y \end{aligned} \tag{5}$$

$$\begin{aligned} G \cdot x &= C \cdot x \\ G \cdot y &= D \cdot y \end{aligned} \tag{6}$$

$$\begin{aligned} H \cdot x &= F \cdot x \\ H \cdot y &= B \cdot y + (W * 1.15) \end{aligned} \tag{7}$$

where A, B, C, D, E, F, G, H are ROIs and W is width of face rectangle.

In order to achieve the best possible results, the constants in the above expressions are the results of many measurements and tests. Once the ROIs are defined, the facial features can be detected on the inside of those areas.

3.3.2 Pupil Detector

Correct and accurate pupil detection is crucial for the face tracking. Besides this, we can also detect the eye gaze and define more accurate ROIs for eyebrows and the relative vertical position of both pupils gives us very important information about the head rotation. There are many different approaches how to detect the eye pupil.

The pupil detector is based on the Cumulative Distribution Function (CDF) algorithm. This method assumes that the pupil is much dimmer than the cornea. The CDF algorithm is defined by:

$$CDF(r) = \sum_{w=0}^{r} p(w) \tag{8}$$

where $p(w)$ is probability to find a pixel with luminance which is equal to w.

Fig. 4 The pupil detection procedure

In the first step the ROI area is blurred and inverted. Secondly, the CDF function is applied. The next step is to restrict the ROI area only to the eye. The main reason for this restriction is that some people have very dark hair or eyebrows, which can be even darker than the pupil itself. The right down corner is blended. This could easily confuse the algorithm. After that the circle mask from the right down corner discloses the ROI area. In each step, the maximum value of the whole ROI is measured. Once the pupil is uncovered, the maximum value is reached. When the next value is smaller than the previous one the algorithm stops. This phase is illustrated in Fig. 4b. Next step is to measure the maximum value of the eye area. Then, a threshold of, for example, 95% of the whole range between min and max value is applied. The pupil candidate, which has best rate of all of those facilities, is selected as the pupil.

3.3.3 Eyebrow Detector

After detecting the pupil positions, the eyebrow ROI can be defined. The eyebrow ROI occurs inside the eye ROI with lower border above the detected eye. To detect the eyebrow shape reliably, the Sobel derivatives are used. The Sobel derivatives allow us to detect the edges in the gray scale image in horizontal or vertical direction. The eyebrow detector itself consists of two parts. The first part is the coarse detection of the eyebrow. The ROI is strongly blurred and the vertical Sobel is applied. The result is illustrated in Fig. 5b. The main reason for the strong blur filter is to eliminate the wrinkles and other features which can be considered to be the eyebrow (see Fig. 5g). After applying the adaptive threshold, the eyebrow´s thickest part is detected. Let us call this blob anchor. The anchor will be used later in the second part of the eyebrow detector. The second part provides the most accurate results in Fig. 5f. The Gaussian blur and then the Sobel filter are applied to the ROI image.

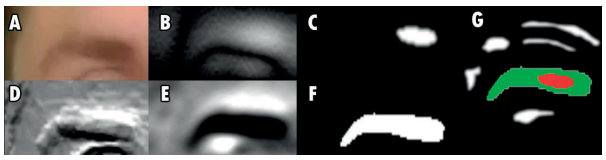


Fig. 5 Visual illustration of the eyebrow detection

To select the right candidate, the distance between the candidates and the anchor is measured. The candidate, which is positioned nearest the anchor, is the right one. Once the shape is clearly detected, the OpenCV function of find contours is applied. This function returns in the vector of points. Those contours points are illustrated in Fig. 6.

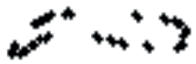


Fig. 6 The eyebrow contour points

At least a simple sorting algorithm is applied. For example, to get the right corner of the eyebrow, all points are upwardly sorted according to their horizontal position. The first three points will be selected and a 2D space centroid will be calculated. This centroid represents the right corner of the left eyebrow. As the centroid represents the average position of the three points, it will eliminate the disturbing noise.

3.3.4 Mouth Detector

The mouth detector consists of two parts. First part is color-based and provides the color transformation of the mouth ROI from the RGB color space to the gray scale I defined by transformation formula:

$$I_{(x,y)} = \frac{2G_{(x,y)} - 2R_{(x,y)} - 0.5B_{(x,y)}}{4} \quad (9)$$

where R, G, B are color components of a certain pixel in the mouth ROI and I is the illumination value of the result pixel. The transformation is based on the principle that the blue component has a reduced role within the lip and skin discrimination.

Fig. 7 Lips detection procedure

The mouth ROI after the color transformation is illustrated in Fig. 7b. The holes of the lips and nose are clearly separated from the background. When we add a binary threshold to the transformed image, we get a clean binary mask. Once we have the mask, the function of the find contours is applied. As the contours of the lips are very rugged, the approximation of an envelope is needed. The application of the convex hull function on the vector of the contours points returns to a smooth hull of the lip, Fig. 7c. This hull is also defined as the vector of points. These points are illustrated in Fig. 7d and are sorted and used to determine the corners of the lips.

3.3.5 Color Marker Detector

The above mentioned mouth and eyebrow detectors have some limitations. They are very sensitive to bad light conditions, like strong side illumination of the face. For example, if the light source has a red toned light spectrum,

this red light, which is reflected on the cheeks, can be wrongly detected as a part of the lips. The indistinctive eyebrow will not be probably detected correctly either. To avoid such mistakes and limitations, the color trackers which are glued to the actor's face are used. The trackers should have a unique color, which is clearly separated from the actor´s face and background. In the first step, the face ROI is transformed from RGB to the HSV color space. HSV (Hue, Saturation and Value) simply allows us to separate the trackers from the background. The lower and upper value of the threshold for each HSV channel is defined by the user. After that, the image is transformed into the binary mask, which is illustrated in Fig. 8.

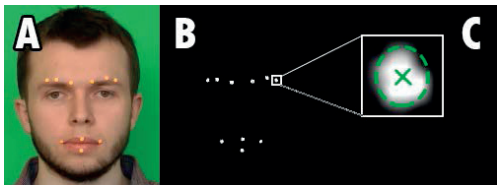


Fig. 8 The color point masking and detecting procedure

For every white blob, which represents one color tracker, the contour points are computed in the same way as in the previous cases. The contour points can be bounded by ellipse calling function fit Ellipse. The centre of the bounded ellipse is the centre of the pertaining tracker (see Fig. 8c). Detected points are sorted and put in the right order which represents the eyebrows and mouth.

3.3.6 Data Processing Modules

Processing modules transform the absolute positions of each tracking point to get some useful piece of information for the next handling. This part includes many tasks, such as alignment, relative positioning, stabilization, gaze detection, etc.

- Eye Gaze Detector - The eye gaze detector provides the eyes tracking. In our case, only the horizontal motion of the pupils is investigated.
- Trackers Smoother & Fixation Module - Besides the quantization noise, the position errors occur with the tracking process among the frames. These disturbing motions and errors should be minimized by smooth algorithm. This smooth algorithm is based on the principle of the memory buffer. In each frame, the values in the buffer shift one position down and the current value is added to the top. The last value is removed.
- Frame Drawer Module - The drawer module is required to give the user proper feedback and information about the tracking accuracy.

4. Conclusions

In this paper, the software which would be able to combine marker and marker-less face tracking for the purpose of the creation of the accurate and real time running 3D facial animation inside the Cinema 4D was created. Within the field of computer vision, the facial motion capture is a complex problematic which can afford a lot of challenges. The most difficult part is to learn the computer to be able to reliably recognize the facial features and to estimate their motions. The computer vision algorithms are very sophisticated and require a lot of computing power. In order to create a real-time face tracking software, many difficult tasks had to be solved, including multithreading, effectiveness and accuracy. However, the C++ programming language with OpenCV libraries and Qt Creator allow us to create a very stable and powerful application with minimal computing power demands. Nowadays, the movie and advertising businesses increasingly work with computer generated imagery (CGI). Furthermore, the game industry also uses the computer vision and motion capture systems in order to provide more realistic and enjoyable games.

For facial capture, a few commercial products have already been available. The proposed Face Tracker meets almost all requirements for the face tracking software. It does not need any special additional hardware and it has a very few computational requirements. The proposed system smoothly runs on double cored Intel Core i5 with 2 GB RAM.

Acknowledgements

This contribution/publication is the result of the project implementation at the Centre of excellence for systems and services of intelligent transport, ITMS 26220120028 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] GANAPATHI, V., PLAGEMANN, C., KOLLER, D., THRUN, S.: *Real Time Motion Capture using a Single Time-of-flight Camera,* IEEE Intern. Conference on Computer Vision and Pattern Recognition (CVPR), pp. 755-762, ISBN 978-1-4244-6984-0, 2010.

[2] AGARWAL, A., TRIGGS, B.: *Recovering 3D Human Pose from Monocular Images,* Image Processing, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, No. 1, pp. 44-58, ISSN 0162-8828, 2006, doi: 10.1109/TPAMI.2006.21.

[3] SPIRIK, J., ZATYIK, J.: *Image extrapolation using sparse methods*, Communications - Scientific Letters of the University of Zilina, vol. 15, No. 2A, pp.174-179, ISSN 1335-4205, 2013.

[4] MARKOVIC, I., CHAUMETTE, F., PETROVIC, I.: *Moving Object Detection, Tracking and Following using an Omnidirectional Camera on a Mobile Robot,* IEEE Intern. Conference on Robotics and Automation (ICRA), pp. 5630-5635, 2014, doi: 10.1109/ICRA.2014.6907687.

Jaromir Hrad - Jiri Vodrazka *

MODELLING OF DATA COMMUNICATION OVER MEDIUM-VOLTAGE POWER DISTRIBUTION LINES

The paper is dedicated to different topologies of medium-voltage power distribution lines with respect to their properties influencing the transmission of communication signals. We provide very basic introduction to the problems of signal transmission over these lines, discuss their parameters and examine the typical behaviour and characteristics. We conclude our considerations with some specific recommendations concerning the most advantageous configurations for the purposes of Power Line Communication (PLC). The data transmission capacity is calculated for different topologies and network conditions.

Keywords: *PLC; power line communication; Smart Grid; medium-voltage lines.*

1. Introduction

The concept of Smart Grids in power distribution systems [1] naturally implies the need for transmission of communication signals over medium-voltage lines, and therefore it is necessary to solve some specific problems related to this environment, sometimes substantially different from the issues known in conventional telecommunication lines and cables [2]. In this paper we introduce and analyse some of these problems, with special emphasis on typical topologies and their particular consequences.

2. State of the Art

A) Basic model of the line

Actual resources relevant for the power line network modelling are [3] and [4]. If we want to determine the transmission function of a line composed of several sections with different parameters and containing also taps or other elements, we can express it from the product of matrices describing the individual cascaded sections. We should use the cascade parameters of a two-port network ABCD [5]. The resulting matrix of the cascade will be given by the product of the sectional matrices:

$$\mathrm{A} = \mathrm{A}_1 \cdot \mathrm{A}_2 \ldots\ldots \mathrm{A}_n = \begin{bmatrix} a(f) & b(f) \\ c(f) & d(f) \end{bmatrix} \tag{1}$$

where A_1 to A_n are matrices describing the cascaded elements and *a(f)* to *d(f)* are the resulting parameters of the entire cascade.

The matrix for section of a homogeneous line with length l_i, characteristic impedance Z_{ci} and propagation coefficient γ_i can be calculated using the following formula:

$$\mathrm{A}_{Li} = \begin{bmatrix} \left(\cosh\left(\gamma_i(f)\cdot l_i\right)\right) & \left(Z_{Ci}(f)\cdot\sinh\left(\gamma_i(f)\cdot l_i\right)\right) \\ \left(\dfrac{\sinh\left(\gamma_i(f)\cdot l_i\right)}{Z_{Ci}(f)}\right) & \left(\cosh\left(\gamma_i(f)\cdot l_i\right)\right) \end{bmatrix} \tag{2}$$

B) Modelling of medium-voltage lines

Let us consider the initial values for modelling of a 22kV line according to [6], which lists the following primary parameters: L = $1.9\cdot10^{-6}$ H/m, C = $8\cdot10^{-12}$ F/m, R(200kHz) = 0.03 Ω/m, and G(200kHz) = $1.5\cdot10^{-6}$ S/m. Some other important parameters of British Telecom model (BT9) [7] include: DC resistance r_0 = 1; frequency characteristics slope constant a_c = 0.0002; leakage for low frequencies g_0 = $7.5\cdot10^{-10}$; and growth coefficient towards higher frequencies g_e = 1. Then the characteristic impedance Z_c can be calculated as:

$$Z_C = \sqrt{\frac{L}{C}} = \sqrt{\frac{1.9}{8}10^6} = 487\Omega \tag{3}$$

and the propagation velocity *v* as:

$$v = \frac{Z_c}{L} = 256495 km/s \tag{4}$$

Local maximums of the attenuation characteristics can be estimated using the following formula:

* **Jaromir Hrad, Jiri Vodrazka**
Department of Telecommunication Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic
E-mail: hrad@fel.cvut.cz

$$f_{Amax} = (2k+1)\cdot\frac{v}{4l_{bt}},\ for\ k = 0;1;2;3... \quad (5)$$

where l_{bt} denotes the length of a bridged tap and v the signal propagation speed on the given line [8] and [9].

3. Medium-voltage Lines with Branches

Medium-voltage power distribution network has, in general, a tree topology with many branches (that we refer to as bridged taps, with respect to the common terminology used in the telecommunications area [5]). For the purposes of modelling of the medium-voltage transmission environment for PLC communication we will consider three basic topological configurations:

1. Tree topology,
2. Linear topology with very short taps,
3. Star topology.

These topologies will be compared in a situation with testing length 12.26 km between the central unit (placed in a HV/MV transformer station) and the most distant modem in a distribution transformer station (DTS). It is assumed that the taps are bridged with a combination of an equivalent impedance of the distribution transformer and a coupling element; so, we can assume that a PLC modem is connected in all DTS nodes.

A) Tree topology

The first considered topology (Fig. 1) is a tree with taps in nodes No. 2, 4, 6, 8, 9, 12, 14 and 15; their respective lengths are 152 m, 2630 m, 81 m, 1870 m, 54 m, 289 m, 42 m and 521 m, and the lengths of the individual sections are 385 m, 4491 m, 27 m, 1751 m, 3420 m, 1290 m and 923 m.

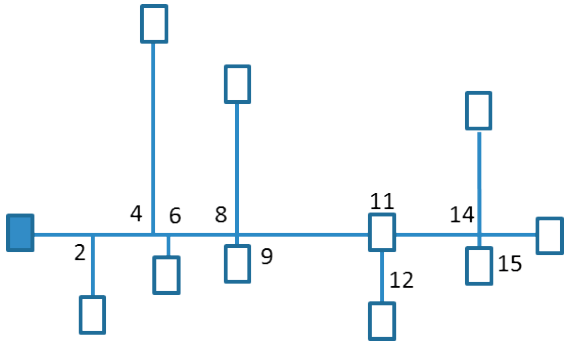


Fig. 1 Configuration #1

The graph in Fig. 2 shows modelled attenuation for the tree topology (blue curve) compared to a homogeneous line (green curve). In order to compare different situations, let us focus on an identical tree topology, but with extremely increased leakage (10 times higher than in the previous situation), which corresponds to very high values of humidity in rain and fog (see Fig. 3).

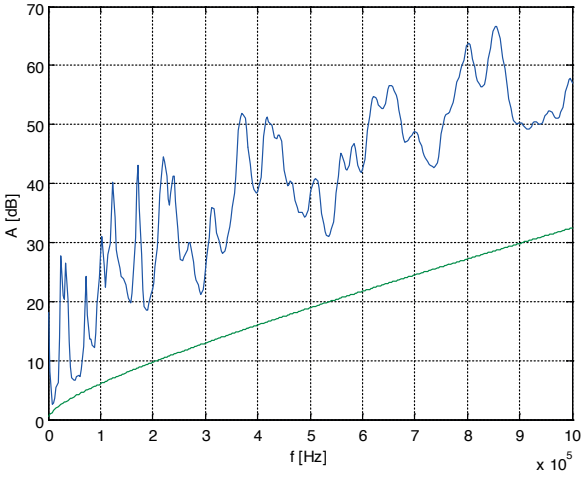


Fig. 2 Frequency dependence of attenuation

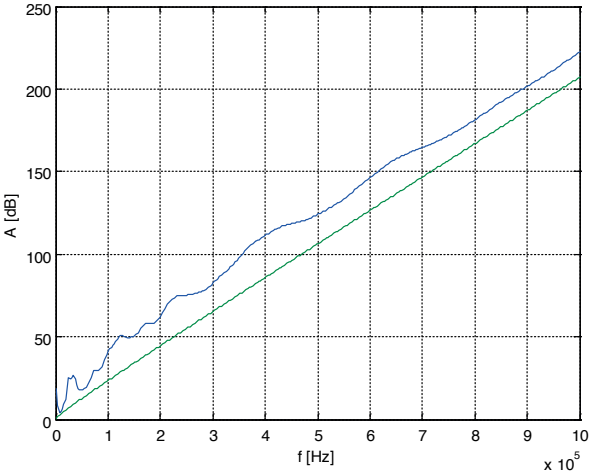


Fig. 3 Frequency dependence of attenuation (for extremely increased leakage)

It can be observed from the frequency dependence of attenuation that thanks to substantial increase of specific attenuation the reflections are attenuated as well (i.e. the maximums are flattened), but – especially in the higher frequency range – the values of attenuation are very high, and therefore the practical usability of such channel for transmission of communication signals is limited to the band up to 100 kHz at most.

C) Linear topology with very short taps

The next considered topology is linear, with short taps leading from the main line to the individual DTS (see Fig. 4).

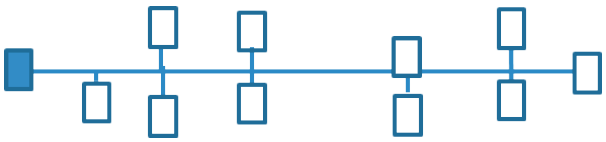


Fig. 4 Configuration #2 (linear topology with short taps)

The lengths of the respective short taps are 27 m, 81 m, 54 m, 42 m, 16 m, 18 m, 29 m and 98 m. The lengths of line sections remain the same as in the case of tree topology. The graph in Fig. 5 shows the modelled attenuation for linear topology with short taps (blue curve) compared to a homogeneous line (green curve).

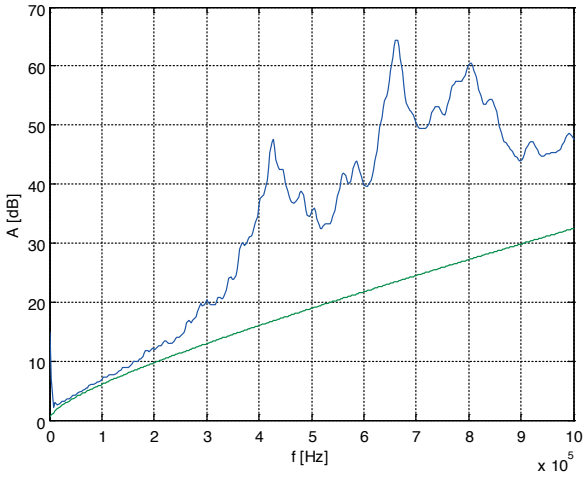


Fig. 5 Dependence of attenuation on frequency (linear topology with short taps)

From the graphs it is clear that up to approximately 300 kHz the taps do not disturb the transmission and the values of attenuation are almost comparable with homogeneous lines. Also the impulse response is much cleaner.

D) Star topology

The last considered case is a star topology – lines towards the individual DTS begin in the same node. Other transformers may be deployed on these lines, as we can see on the scheme in Fig. 6.

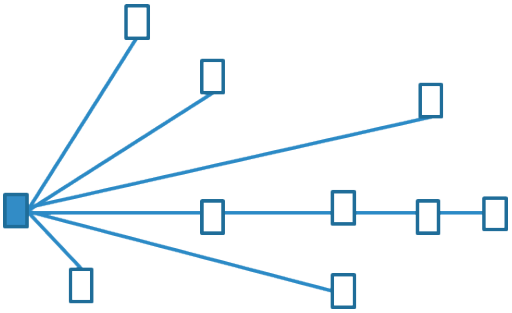


Fig. 6 Configuration #3 (star topology)

Besides the main examined line (12.26 km) we consider also five other "beams" of the following lengths: 3853 m; 1527 m; 4491 m; 6762 m; 2630 m. The graph in Fig. 7 shows the modelled attenuation for the star topology (blue curve) compared to a homogeneous line (green curve).

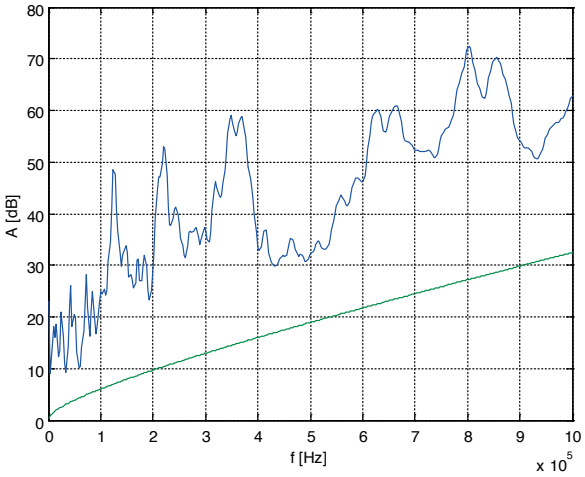


Fig. 7 Frequency dependence of attenuation (star topology)

The modelled dependencies show that their nature is totally different, which is caused by substantial influence of reflections. The most remarkable difference against the previous cases is substantially lower input impedance caused by connecting several lines in parallel. This fact can be used to decrease the internal impedance of modems in the central node of a network with star topology.

4. Data transmission performance

We have performed initial modelling, using the tree topology that appears to be the most critical one from the viewpoint of transmission parameters (exhibiting multiple reflections, the most rippled transfer function as well as impulse response). We considered OFDM scheme with 1kHz subchannel width and QAM modulation, noise margin NM=4 (6dB), permissible symbol error rate 10^{-7} (Shannon gap 9.5) and signal amplitude 0.1 V – RMS value per 1kHz channel (typically corresponding to 10 V_{pp} for transmission in the entire considered bandwidth). We did not consider any error correction method. The maximum number of allocated bits is 8, which corresponds to the maximum number of states for 256-QAM modulation.

Furthermore, we considered background noise typical for industrial environments, taken from low-voltage networks, and interference from LW radio transmitters – DCF77 (time code at 77.5 kHz) and LW radio broadcast (at 270 kHz, 153 kHz, 183 kHz, 234 kHz and 225 kHz).

Figure 8 shows power spectral density of the considered noise, and Fig. 9 illustrates frequency dependence of attenuation in the channel under optimum transmission conditions for line length (transmitter – receiver) 20 km.

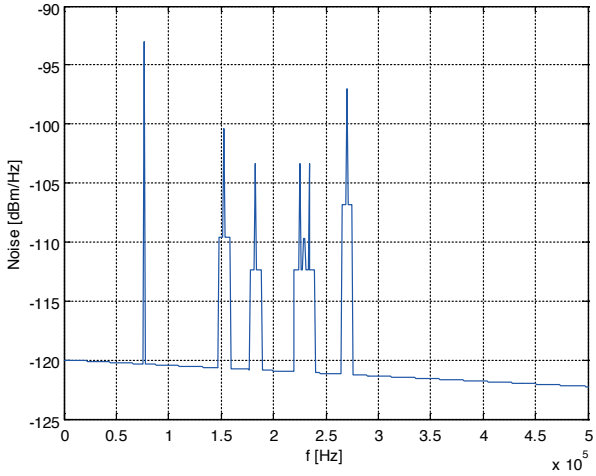


Fig. 8 Power spectral density of noise (star topology)

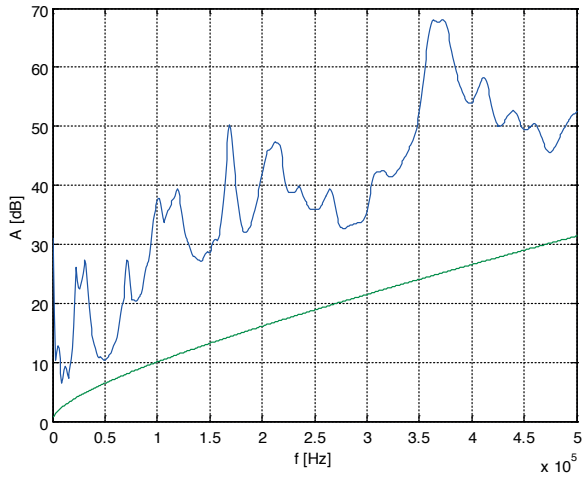


Fig. 9 Frequency dependence of attenuation for 20km length (configuration #1)

The graph in Fig. 10 shows theoretical maximum transmission speed – the blue curve represents line length dependence of the speed for QAM modulation and the green one for OFDM. The graph of bit allocation for OFDM (Fig. 11) shows the usability of different spectral bands (dependence on subchannel sequential number).

If the lower part of the band between 10 and 60 kHz is used for transmission, the resulting data rate is practically independent of the distance. The comparison above does not allow unambiguous decision whether it is more advantageous to use the single-carrier (QAM) or multi-carrier (OFDM) modulation. However, it is positive that acceptable data rates can be achieved even under worsened conditions when the frequencies in the lower part of the band are used.

The current trend in communication systems is the deployment of turbo codes [10] and [11], which leads to lower achievable data rate, but it also rapidly increases system robustness, i.e. noise immunity and acceptable attenuation of the transmission path under extreme conditions.

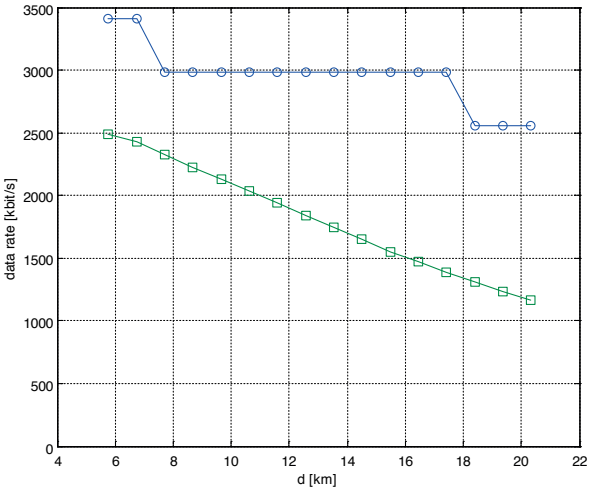


Fig. 10 Dependence of data rate on distance (configuration #1)

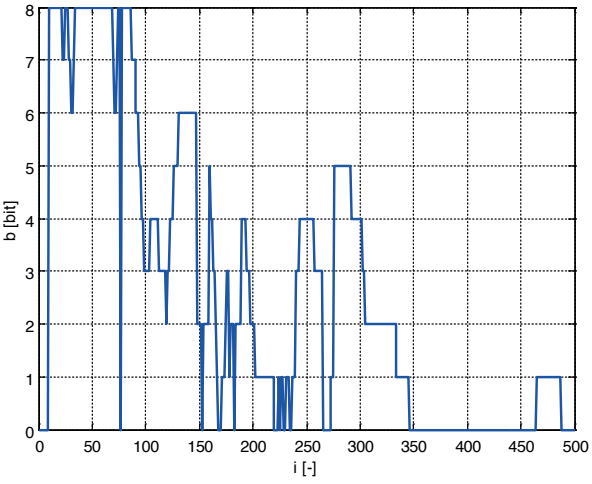


Fig. 11 Example of bit allocation for OFDM transmission format

5. Conclusion

The graphs displayed above clearly show that the resulting channel properties for the respective topologies differ substantially, being influenced mainly by reflections. The most remarkable difference of star topology against the other cases consists in

much lower input impedance, as several lines are connected in parallel; this approach can be used to decrease the internal impedance of a modem in the central node of a star network accordingly.

The most advantageous topology – from the viewpoint of reflections – is the linear one with short taps, where we can identify relatively wide band, suitable for PLC communication, the parameters of which nearly approach those of a homogeneous line.

Acknowledgement

This paper has originated thanks to grant support obtained from the Technology Agency of the Czech Republic within the project "Research and Development of New-Generation Communication Devices for Transmission Over High-Voltage Power Lines" (TA03011192).

References

[1] CEPA, L., KOCUR, Z., MULLER, Z.: Migration of the IT Technologies to the Smart Grids. *Elektronika ir Elektrotechnika,* 2012, vol. 7, No. 123, pp. 123-128. ISSN 1392-1215.

[2] NEVOSAD, M., VODRAZKA, J., LAFATA, P.: Parameters of Twisted Pairs Assessment for Frequency Band up to 1 GHz. *Advances in Electrical and Electronic Engineering.* 2012, vol. 10, No. 4, pp. 224-227. ISSN 1336-1376.

[3] SANZ, A.: *Evolution and Limit of Performance of OFDM-based Narrowband PLC. WSPLC15,* 9th Workshop on Power Line Communications, September 2015, Klagenfurt - Austria.

[4] PITTOLO, A., TONELLO, A.: *New Results in Channel Modelling. WSPLC15,* 9th Workshop on Power Line Communications, September 2015, Klagenfurt - Austria.

[5] VODRAZKA, J., JARES, P., SIMAK, B.: More Accurate Modelling of Digital Subscriber Lines with Inhomogeneity. *Communications - Scientific Letters of the University of Zilina*. 2013, vol. 15, No. 2A, pp. 129-133. ISSN 1335-4205.

[6] DABAK, A., KIM, I. H., VARADARAJAN, B., PANDE, T.: *Channel Modeling for MV/LV AMI Applications in the Frequency Range < 500kHz,* Workshop on Power Line Communications WSPLC, 2011.

[7] RAUSCHMAYERM, D. J.: ADSL/VDSL Principles: *A Practical and Precise Study of Asymmetric Digital Subscriber Lines and Very High Speed Digital Subscriber Lines.* Macmillan Technical Publishing, Indianapolis 1999.

[8] VODRAZKA, J., HRAD, J.: *Modeling of a Subscriber Line with Inhomogeneity,* 32nd Intern. Conference on Telecommunications and Signal Processing (TSP 2009), Dunakiliti, 2009, pp. 79-83, ISBN 978-963-06-7716-5.

[9] HRASNICA, H.: *Broadband Powerline Communications Networks.* John Wiley & Sons : Chichester 2004. ISBN 0-470-85741-2

[10] SEDY, J., SILHAVY, P., KRAJSA O., HROUZA, O.: Performance Analysis of Turbo Codes. *Communications - Scientific Letters of the University of Zilina,* 2013, vol. 15, No. 2A, pp. 167-173. ISSN 1335-4205.

[11] ZAVRTALEK, J., KEKRT, D., HRAD, J.: New Concept of PLC Modems: Multi-carrier System for Frequency Selective Slow-fading Channels Based on Layered SCCC Turbocodes. *Radioengineering,* 2015, vol. 24, No. 3, pp. 857-872. ISSN 1210-2512.

Peter Mucka *

SENSITIVITY OF INTERNATIONAL ROUGHNESS INDEX TO DISTRESSES OF CEMENT CONCRETE ROAD SURFACES

The paper examined the sensitivity of International Roughness Index (IRI) to the local discontinuities (various distresses, joints, joint and surface defects, other road features, etc.) of the cement concrete (CC) pavements. About 5 300 road records of total length 470 km from Long Term Pavement Performance (LTPP) program were processed. The raw profiles were separated into a random part and a distress part using the median filtering method. The median filter order was set to identify distresses of variable maximum width from 20 to 40 cm, and minimal height, 3 mm. About 26 000 distresses were separated and their dimensions were identified. The raw longitudinal road profiles were compared with the separated pure random parts. The mean relative increase in IRI caused by distresses was 3.3 % for maximum distress width 20 cm, 6.6 % (30 cm), and 10.7 % (40 cm).

Keywords: *Longitudinal road profile; road roughness; distress; crack; joint; slab; International Roughness Index; cement concrete (CC) pavement.*

1. Introduction

The cement concrete pavements present 0.6 % of the total length of road network in the Slovak Republic [1]. In some other countries the proportion of CC pavements is up to 20-30 % of total road network. The typical part of CC pavement is a joint between slabs. The joint defects are major pavement distress form of rigid pavement. Distresses of CC pavements negatively affect ride comfort and ride safety. The four basic types of joints are used in practice - contraction (control) joints, isolation joints, construction joints, and warping joints [2]. Miller and Bellinger [3] divided the distresses typical for pavements with jointed CC surfaces into four groups: (a) cracking (corner breaks, transverse cracking, etc.), (b) joints deficiencies (joint seal damage, spalling of joints, etc.), (c) surface defects (map cracking, polished aggregate), (d) miscellaneous distresses (faulting of transverse joints and cracks, patch/patch deterioration, water bleeding, etc.). Joint deteriorations such as spalling, breaking, cracking, chipping, or fraying of the slab edges usually occur within 50 mm of joints [4].

International Roughness Index (IRI) is used in Pavement Management System (PMS) in the Slovak Republic to characterize the longitudinal road unevenness [5 - 8].

The research in the field of road distresses influence on the road unevenness indicators is predominantly focused on the influence of the vertical faults, i.e., vertical shifts between adjacent slabs, on the IRI statistics [9 - 13]. Only several papers [14 - 16] examined the influence of joint width, joint depth, joints spacing, or the road data processing on the IRI. The published results were predominantly oriented to the real road sections and the change of IRI with time. Mucka [17] analysed the influence of the artificial random profiles with superimposed joints with controlled dimension on the twenty one road unevenness indicators. The published results did not allow distinguishing between the contribution of the random profile part and the distress part to the total value of IRI.

Denotation 'distress' in this study means local discontinuities of various shapes and origins such as joints, joint deficiencies, surface defects, various distress types of CC pavements or other road features that were separated from a raw profile by the median filter. The median filter was set to separate distresses of variable maximum width from 20 to 40 cm, and minimal height, 3 mm.

The question to ask is whether distresses of CC pavements are an important factor affecting IRI that will influence the longer-term rehabilitation decision making.

The main objectives of this study are as follows:

- Process the real CC road profiles by median filtering method and select a pure random part and a distress part from a raw profile;

* **Peter Mucka**
Institute of Materials and Machine Mechanics, Slovak Academy of Sciences, Bratislava, Slovakia
E-mail: ummsmuc@savba.sk

- Identify every single distress and its own parameters and evaluate the statistics of all the distress parameters;
- Compare the IRI values calculated for a pure random profile with those calculated for a raw profile with distresses;
- Quantify the influence of distresses on the IRI;
- Evaluate the influence of the maximum width of the separated distresses on the results.

2. International Roughness Index

The IRI is essentially a computer-based virtual response-type system based on the response of a mathematical quarter-car vehicle model to the road profile. The IRI is based on simulation of the roughness response of a car travelling at 80 km/h - it is the Reference Average Rectified Slope, which expresses a ratio of the accumulated suspension vertical motion of a vehicle, divided by the distance travelled during the test.

The IRI is a numeric that summarizes the roughness qualities impacting on vehicle response. The IRI is a dimensionless measure with units (mm/m) or (in/mi). The algorithm was proposed by Sayers et al. [18] and is also implemented in the prEN 13036-5 [19]. In Slovakia, the IRI is implemented in Technical specifications TP 04/2012 [5] and TP 05/2012 [20]. The road unevenness classification based on the IRI in Slovakia is shown in Table 1. PMS in Slovakia used 20 m interval for IRI estimation.

Limitations of the IRI to characterize the road unevenness were discussed in Mucka and Granlund [21] or Mucka [22]. Calculation of the IRI and further specifications were published in Kovac et al. [4] and Decky and Kovac [23].

3. Distress separation by median filtering method

Common practice in road profiles processing is using the moving average filter. The median filter removes large vertical distresses in a longitudinal profile without or with limited affecting the random nature of the surrounding parts of the road profile. This feature is a great advantage when compared with the standard moving average filtering.

Median filtering approach [24] was applied to longitudinal raw profile h_{RD} to select a random part of profile h_R and a distress part h_D ($h_D = h_{RD} - h_R$) of profile (Fig. 1). The distress dimensions – distress depth, d_D, distress width, w_D, and distance of successive distresses, l_D, are schematically depicted in Fig. 1. It follows from the definition of the median filter that for the complete filtering of the non-random component including m discrete data, the $n \geq 2m + 1$ window length is needed [24]. The maximum distress width was considered to be $w_{DMAX} \sim 20$ cm. The window length for median filter should be twice of the expected maximum distress width. Order of the median filter was selected to be as follows, $n = 16$ ($n\Delta l = 40$ cm, Δl is the sampling interval, $\Delta l = 2.5$ cm). Distresses higher than the limit absolute value $|d_{DLIM}| = 3$ mm were only detected. The d_{DUNC} quantity reflects the level of uncertainty of the distress residual random component.

The variable Δ_{IRI} quantifies the distress influence on the IRI and presents the difference between the IRI calculated for the raw profile with distresses (RUI_{RD}) and for the separated pure random part (IRI_R) of this profile as follows:

$$\Delta_{IRI}(\mathrm{mm/m}) = \mathrm{IRI}_{RD} - \mathrm{IRI}_R,$$
$$\Delta_{IRI}(\%) = 100 \times \frac{\mathrm{IRI}_{RD} - \mathrm{IRI}_R}{\mathrm{IRI}_R} \quad (1)$$

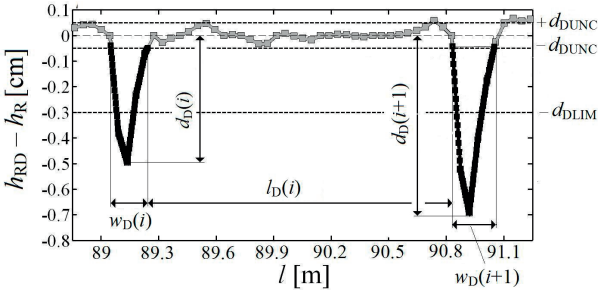


Fig. 1 Scheme of road distress characteristics obtained by median filtering method

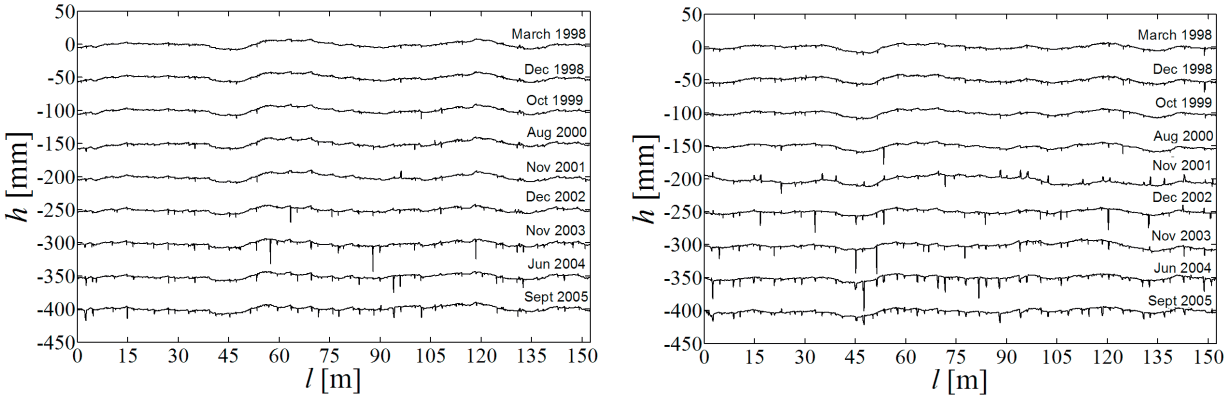
The positive sign of variable Δ_{IRI} indicates that the value of IRI calculated for a raw profile is higher than that obtained for a separated random part. All computations were provided in Matlab®.

The road unevenness classification based on the IRI in Slovakia [5] Table 1

Road class	Motorways and expressways	1st and 2nd class roads	3rd class roads and local highways
1 (very good)	< 1.90	< 1.90	< 3.30
2 (good)	1.91 – 3.30	1.91 – 3.30	3.31 – 5.00
3 (fair)	3.31 – 5.00	3.31 – 5.00	5.01 – 8.00
4 (poor)	5.01 – 8.00	5.01 – 10.00	8.01 – 14.00
5 (very poor)	> 8.00	> 10.00	> 14.00

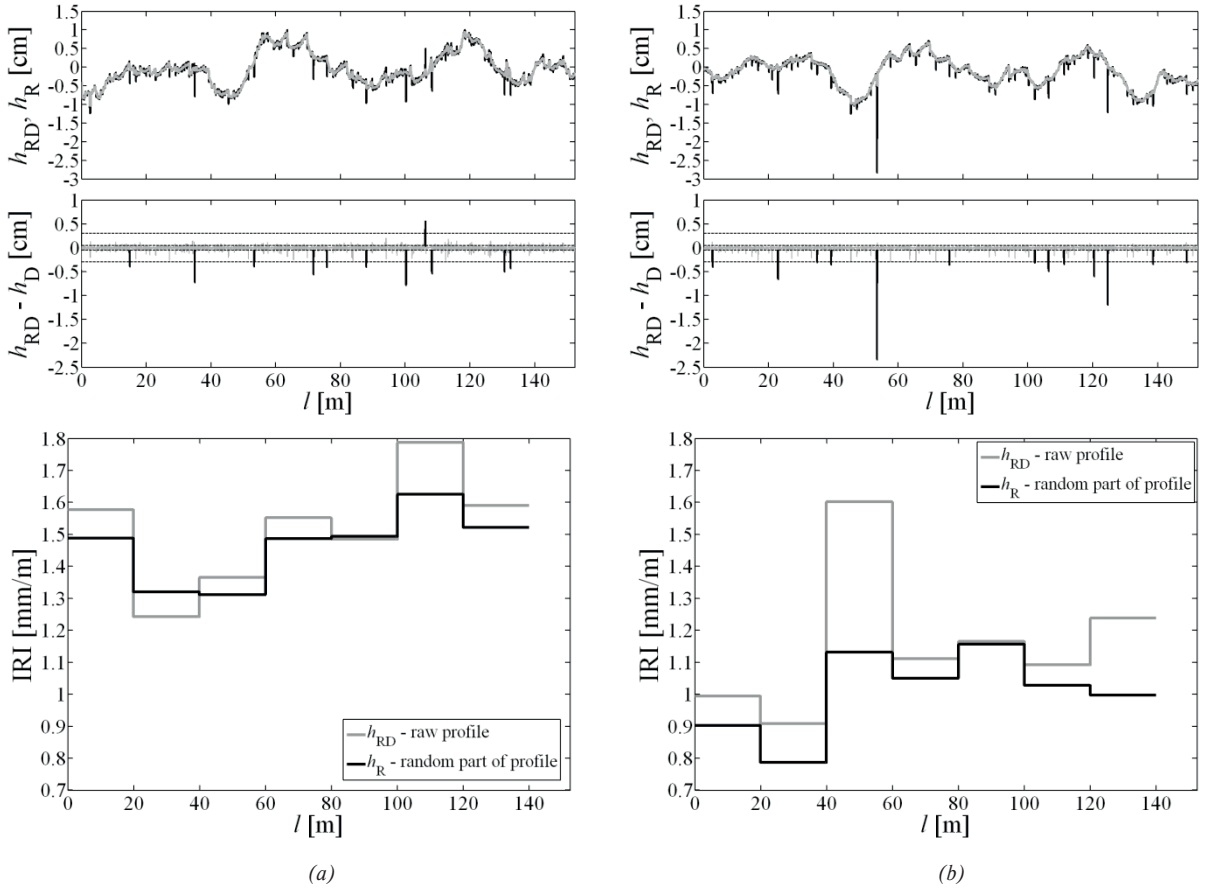
4. Results for road profile sample

Figure 2 shows nine measurements of the left and right track elevation of the road test section #180602 measured from March 1998 to September 2005. Section #180602 is a jointed plain concrete pavement (JPCP) with minimum preparation of original section. Partial depth patching other than at joint was provided on this section in April 2000 and June 2005 [25]. Two profilers provided the measurements - T-6600 profiler (K.J. Law) and MDR4086L3 profiler (International Cybernetics Corporation



(a) *(b)*

Fig. 2 Road elevation of the JPCP road test section #180602 measured from 1998 to 2005 (shifted by 50 mm): (a) left track, (b) right track



(a) *(b)*

Fig. 3 Detection of road distresses (black line) by median filtering method (n = 16) for JPCC road section #180602 measured in August 2000 and corresponding IRI values: (a) left track, (b) right track

Mean values of the identified road distress dimensions for JPCP section #180602 Table 2

Profiler	*Date*	Left track				Right track			
		N_D	mean($\|d_D\|$) (cm)	mean(w_D) (cm)	mean(l_D) (m)	N_D	mean($\|d_D\|$) (cm)	mean(w_D) (cm)	mean(l_D) (m)
K.J. Law	Mar 1998	5	0.40	10.8	27.8	20	0.43	12.5	7.6
K.J. Law	Dec 1998	7	0.37	10.9	20.5	41	0.47	11.1	3.6
K.J. Law	Oct 1999	10	0.45	11.5	14.3	15	0.47	11.9	10.3
K.J. Law	Aug 2000	11	0.52	12.9	11.6	13	0.62	13.0	12.1
K.J. Law	Nov 2001	7	0.50	10.8	19.2	27	0.52	16.6	5.6
ICC	Dec 2002	20	0.63	9.6	5.6	41	0.77	13.7	3.6
ICC	Nov 2003	37	0.83	12.2	4.0	47	0.75	14.1	3.1
ICC	Jun 2004	31	0.76	11.0	4.6	52	0.62	15.4	2.8
ICC	Sept 2005	35	0.64	12.4	4.1	48	0.45	14.7	3.1

IRI for JPCP section #180602 Table 3

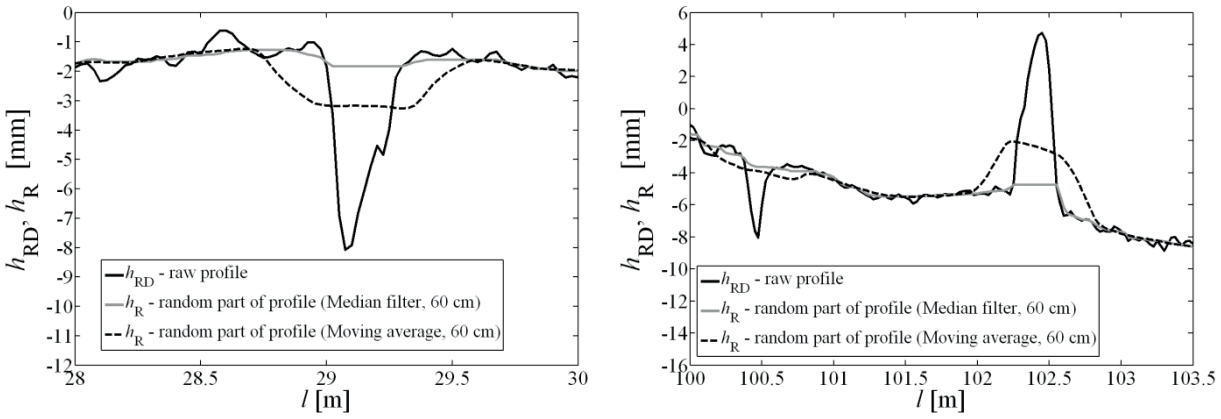
Profiler	Date	Left track					Right track				
		N_D	IRI_R (m/km)	IRI_{RD} (m/km)	Δ_{IRI} (m/km)	Δ_{IRI} (%)	N_D	IRI_R (m/km)	IRI_{RD} (m/km)	Δ_{IRI} (m/km)	Δ_{IRI} (%)
K.J. Law	Mar 1998	5	1.37	1.39	0.02	1.4	20	1.36	1.37	0.01	0.6
K.J. Law	Dec 1998	7	1.43	1.43	0	0.1	41	1.36	1.38	0.02	1.7
K.J. Law	Oct 1999	10	1.62	1.70	0.08	4.7	15	1.32	1.34	0.02	1.5
K.J. Law	Aug 2000	11	1.63	1.79	0.16	10	13	1.16	1.60	0.44	38.5
K.J. Law	Nov 2001	7	2.07	2.16	0.09	4.2	27	3.56	3.87	0.31	8.8
ICC	Dec 2002	20	1.67	1.76	0.09	5.8	41	1.61	2.63	1.02	62.7
ICC	Nov 2003	37	1.66	2.61	0.95	57.5	47	1.96	3.36	1.40	71.6
ICC	Jun 2004	31	1.58	2.15	0.57	36.4	52	4.70	6.27	1.57	33.2
ICC	Sept 2005	35	2.94	3.16	0.22	7.5	48	2.99	3.49	0.50	16.6
Mean (K.J. Law)					0.07	4.1				0.16	10.2
Mean (ICC)					0.46	26.8				1.12	46

(ICC)). The ICC profiler with the laser sensor measures a higher roughness because it includes spikes that the K.J. Law profiler with the infrared sensor did not. The smaller footprint of the ICC profiler caused a jump in number of identified distresses.

Figure 3 presents the detection of road distresses by median filtering method [24] in analysed test section #180602 measured in August 2000. The distress part of profile ($h_D = h_{RD} - h_R$) obtained by separation of the raw profile from the random part is shown in Fig. 3. The most common features presented in this profile are (a) spalling of transverse joints, (b) joint faulting, (c) transverse joint seal damage, (d) transverse crack, (e) asphalt concrete or CC patch, and (f) corner breaks [25]. The statistics of distresses in both tracks of profile #180602 is shown in Table 2. Value N_D presents the number of identified distresses.

Table 3 summarizes the IRI values for a pure random part of profile (IRI_R) and for a raw profile with distresses (IRI_{RD}) for nine measurements of JPCP road section #180602. The IRI values for the raw profiles were slightly higher in comparison to the pure random part of profile. The mean percentage change Δ_{IRI} [Eq. (1)] of IRI was 4.1 %/10.2 % (left track/right track) for K.J. Law profiler and 10.2 %/46 % for ICC profiler. Laser height sensors of ICC profiler have a circular footprint of about 1.5 mm. The smaller footprint of the ICC profiler caused a jump in number of identified distresses.

Figure 4 shows the separation of a random part and a distress part of the raw profile by the median filter of order $n = 24$ (length of window is $n\Delta l = 60$ cm, $\Delta l = 2.5$ cm) and by the moving average filter with the base length 60 cm. Figure 4 illustrates a limited ability of moving average filter to separate the distress part. The differences between the moving average filter and the median filter are function of raw profile nature, distance of the successive distresses, the base length of the filters, etc. In some cases the ability or disability to remove distresses from a raw profile may be similar for both types of filters (see Fig. 4b at $l = 100.5$ m).



(a) *(b)*

Fig. 4 Separation of the random part of the raw profile by the median filter (n = 24, window length, 60 cm) and by the moving average filter (base length = 60 cm): (a) Example # 1, (b) Example #2

The IRI algorithm contains the pre-processing of a raw profile with a 25-cm moving average [18]. The procedure for IRI calculation uses profile smoothing by the moving average filter to better represent the way in which tire of a vehicle envelops the ground. The pre-processing caused a lower sensitivity of IRI to the distresses. Further factor is the frequency response of the relative suspension velocity of the reference quarter car model intended for the IRI computation. This transfer function is most sensitive to the wavelengths ∼ 2 m with substantially lower gain corresponding to the wavelengths < 0.5 m. The comprehensive analysis of the road profile spectral content influence on the IRI response was provided in Mucka and Granlund [21].

5. Results for road profile database

Road profile data from Specific Pavement Study SPS-2 "Strategic Study of Structural Factors for Rigid Pavements" related to the CC surfaces were analysed. This study is included in LTPP Program governed by Federal Highway Administration.

Table 4 presents the basic statistics of analysed sections and distress dimension for SPS-2 database. About 22 300 road distresses were separated from 5 266 road records. A total length of analysed record with at least one distress was ∼ 470 km. Higher mean depth and width were detected in the right track.

Table 5 presents the influence of the median filter window length on the Δ_{IRI} for SPS-2 database. Basic statistics of Δ_{IRI} listed in Table 5 includes mean value, median, standard deviation (std), minimum (min), 95th percentile (P95) and maximum (max). The mean value of the difference Δ_{IRI} increased with window length as follows: Δ_{IRI} = 0.061 mm/m (w_{DMAX} = 20 cm), 0.118 mm/m (30 cm), and 0.180 mm/m (40 cm).

Figure 5 shows the probability density function (PDF) of the difference Δ_{IRI} as a function of the median filter window length. The increase of the window length increased the ability to separate wider distresses from the raw profiles.

Statistics of the analysed road sections and separated distresses in SPS-2 database Table 4

	Left	Right	Total
Profiles records analysed	2 633	2 633	5 266
Profiles records with detected distresses	1 480	1 603	3 083
Percentage of profiles with detected distresses (%)	56.2	60.9	58.6
Number of distresses	9 702	12 602	22 304
Total length of the analysed sections with at least one distress (km)	225.55	244.30	469.85
Number of distresses per km	43.01	51.58	47.47
Mean distress depth (cm)	0.45	0.52	0.51
Standard deviation of distress depth (cm)	0.48	0.46	0.47
Mean distress width (cm)	9.8	10.0	9.9
Standard deviation of distress width (cm)	2.5	2.4	2.5
Mean distance of successive distresses (m)	25.3	23.4	24.3
Standard deviation distance of successive distresses (m)	23.2	22.4	22.8

Statistics of difference Δ_{IRI} for SPS-2 database Table 5

Median filter order, *n*	Track	IRI_R (m/km)	IRI_{RD} (m/km)	Δ_{IRI} (m/km)					
		mean	mean	mean	median	std	min	P95	max
16 (w_{DMAX} = 20 cm)	Left	1.74	1.80	0.055	0.042	0.09	-0.16	0.14	1.71
	Right	1.84	1.91	0.067	0.045	0.16	-0.33	0.18	2.73
	Total	1.79	1.85	0.061	0.043	0.13	-0.33	0.15	2.73
24 (w_{DMAX} = 30 cm)	Left	1.69	1.80	0.107	0.090	0.13	-0.15	0.22	2.85
	Right	1.78	1.91	0.128	0.092	0.24	-0.22	0.27	5.26
	Total	1.73	1.85	0.118	0.091	0.20	-0.22	0.24	5.26
32 (w_{DMAX} = 40 cm)	Left	1.63	1.80	0.166	0.144	0.16	-0.12	0.34	3.37
	Right	1.71	1.91	0.194	0.149	0.32	-0.10	0.38	6.37
	Total	1.67	1.85	0.180	0.146	0.25	-0.12	0.36	6.37

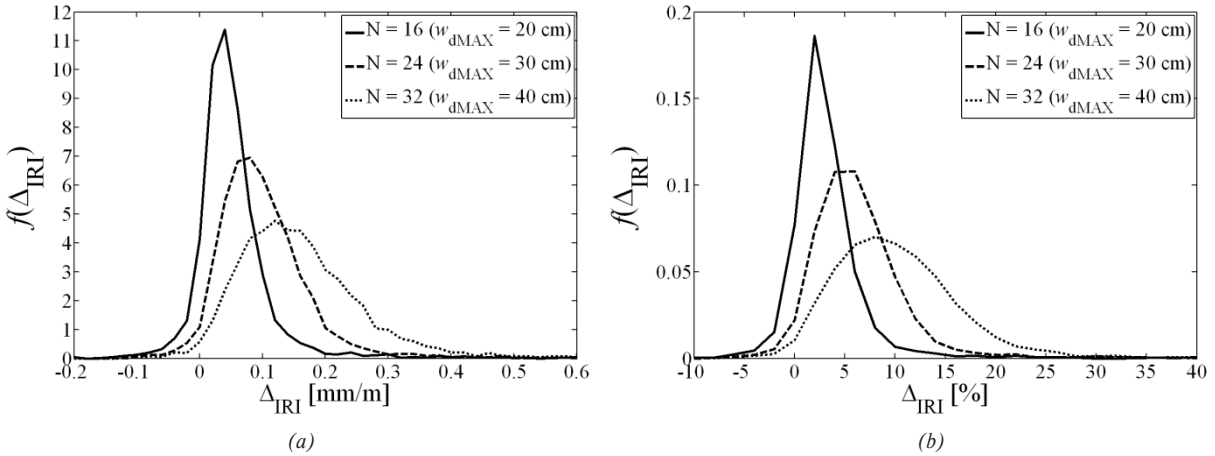


Fig 5 Probability density function of Δ_{IRI}: (a) Δ_{IRI} (mm/m), (b) Δ_{IRI} (%)

6. Conclusions

The presented study brings some advantages in comparison to the previously published papers [9 - 16]. From the results, following findings may be stated:

1. The median filtering approach is the productive tool to separate the random and non-random parts of a longitudinal profile. The median filtering method allows quantifying the sensitivity of IRI to the distresses presence in a longitudinal profile. The median filter affects only slightly the nature of a random part as well as a distress part of a raw profile in comparison to the moving average filter.
2. The influence of the road distresses on IRI is function of the selected bandwidth applied on the raw profile. The mean percentage increase in IRI caused by distresses was calculated as follows: 3.3 % (w_{DMAX} = 20 cm), 6.6 % (30 cm), and 10.7 % (40 cm). The mean increase in IRI in absolute values was Δ_{IRI} = 0.061 mm/m (w_{DMAX} = 20 cm), 0.118 mm/m (30 cm), and 0.180 mm/m (40 cm). Separated distresses of cement concrete pavements with maximum distress width about 40 cm have some influence on the IRI. The mean increase in IRI due to distresses of lower maximum width (20 cm and 30 cm) was relatively low.
3. Paper provides useful statistics of about 22 300 distresses and other road features dimensions.

Acknowledgements

This work has been undertaken within the grant No. 2/0058/13 of the VEGA Grant Agency of Slovak Academy of Sciences. The help of the LTPP InfoPave Support Team is highly appreciated.

References

[1] Review of road network data in the SR (in Slovak), Road databank : Slovak Road Administration, Bratislava, 2014, 77 p.

[2] MCGHEE, K. H.: *Design, Construction, and Maintenance of PCC Pavement Joints. NCHRP Synthesis of Highway Practice*, vol. 211, 1995, Transportation Research Board, Washington, DC.

[3] MILLER, J. S., BELLINGER, W. Y.: *Distress Identification Manual for the Long-Term Pavement Performance Program - 4th revised ed.,* Report No. FHWA-RD-03-031, Office of Infrastructure Research and Development, Federal Highway Administration, McLean, Virginia, 2003.

[4] KOVAC, M., REMISOVA, E., CELKO, J., DECKY, M., DURCANSKA, D.: *Diagnostic of Parameters of Roads Operational Capability (in Slovak).* Zilina: EDIS : University of Zilina, 2012, ISBN 978-80-554-0568-1.

[5] TP 04/2012. *Measurement and Evaluation of Road Roughness using Profilograph GE* (in Slovak). Ministry of Transport, Construction and Regional Development of the Slovak Republic, Bratislava, SR, 2012, 20 p.

[6] CELKO, J., DECKY, M., KOMACKA, J., KOVAC, M.: Pavement Diagnosis as Integrant of the Pavement System, *Communications - Scientific Letters of the University of Zilina*, 2008, vol. 10, No. 2, pp. 44-49, ISSN 1335-4205.

[7] CELKO, J., KOVAC, M., DECKY, M.: Analysis of Selected Pavement Serviceability Parameters, *Communications - Scientific Letters of the University of Zilina*, 2011, vol. 13, No. 3, pp. 56-62. ISSN 1335-4205.

[8] MIKOLAJ, J., REMEK, L., PEPUCHA, L.: Overview of the Road Network Management System, *Communications - Scientific Letters of the University of Zilina*, 2014, vol. 16, No. 4, pp. 53-57. ISSN 1335-4205.

[9] KHAZANOVICH, L., DARTER, M. I., BARTLETT, R. J., MCPEAK, T.: *Common Characteristics of Good and Poorly Performing PCC Pavements.* Report No. FHWA-RD-97-131, Federal Highway Administration, U. S. Department of Transportation, Washington, DC, 1998.

[10] PERERA, R. W., BYRUM, C., KOHN, S. D.: *Investigation of Development of Pavement Roughness.* Report No. FHWA-RD-97-147, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1998.

[11] SELEZNEVA, O., JIANG, J., TAYABJI, S. D.: *Preliminary Evaluation and Analysis of LTPP Faulting Data.* Report No. FHWA-RD-00-076, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 2000.

[12] BYRUM, C. R., PERERA, R. W.: *The Effect of Faulting on IRI Values for Jointed Concrete Pavements.* Proc. of 8th Intern. Conference on Concrete Pavements: Innovations for Concrete Pavement: Technology Transfer for the Next Generation, Colorado Springs, CO, 14-18 August 2005, Colorado: International Society for Concrete Pavements, vol. 2, 2005, pp. 755-770.

[13] LIU, C., WANG, Z.: Influence of Joints on Ride Quality and Roughness Index. *Road Materials and Pavement Design*, 2008, vol. 9, No. 1, pp. 111-121, ISSN 1468-0629.

[14] MORIAN, D. A., GIBSON, S. D., EPPS, J. A.: *Concrete Pavement Maintenance Treatment Performance Review: SPS-4 5-year Data Analysis.* Report No. FHWA-RD-97-155, Federal Highway Administration, U.S. Department of Transportation, Washington, DC, 1998.

[15] BYRUM, C. R.: Analysis by High-speed Profile of Jointed Concrete Pavement Slab Curvature. *Transportation Research Record*, 1730, Transportation Research Board, Washington, DC, 2000.

[16] HALL, K. T., CROVETTI, J. A.: *LTPP Data Analysis: Relative Performance of Jointed Plain Concrete Pavement with Sealed and Unsealed Joints.* NCHRP Web Document 32, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC, 2000.

[17] MUCKA, P.: Sensitivity of Road Unevenness Indicators to Distresses of Composite Pavements, *Intern. J. of Pavement Research and Technology*, 2015, vol. 8, No. 2, pp. 72-84. ISSN 1997-1400.

[18] SAYERS, M. W., GILLESPIE, T. D., PATERSON, W. D. O.: Guidelines for Conducting and Calibrating Road Roughness Measurements. *Technical Paper No. 46*, The World Bank, Washington, DC, 1986.

[19] prEN 13036-5: 2015. *Road and Airfield Surface Characteristics - Test Methods. Part 5: Determination of longitudinal unevenness indices*. European Committee for Standardization (CEN), Brussels.

[20] TP 05/2012. *The Methodology for the use of HDM-4 in Terms of SROV* (in Slovak). Ministry of Transport, Construction and Regional Development of the Slovak Republic, Department for Road Traffic and Roads, Bratislava, SR, 2012, 76 p.

[21] MUCKA, P., GRANLUND, J.: Is the Road Quality Still Better? *J. of Transportation Engineering - ASCE*, 2012, vol. 138, No. 12, pp. 1520-1529. ISSN 0733-947X.

[22] MUCKA, P.: Correlation among Road Unevenness Indicators and Vehicle Vibration Response, *J. of Transportation Engineering - ASCE*, 2013, vol. 139, No. 8, pp. 771-786. ISSN 0733-947X.

[23] DECKY, M., KOVAC, M.: *The Longitudinal Road Evenness of the Road Network (in Slovak),* EDIS : University of Zilina, 2014, ISBN 978-80-554-0925-2.

[24] KROPAC, O., MUCKA, P.: Specification of Obstacles in the Longitudinal Road Profile by Median Filtering. *J. of Transportation Engineering - ASCE*, 2011, vol. 137, No. 3, pp. 214-226, ISSN 0733-947X.

[25] LTPP InfoPave [online]. 2014. Federal Highway Administration, US Department of Transportation, Available from: www.infopave.com [Accessed 20 February 2014].

Sarka Krocova - Miloslav Rezac *

INFRASTRUCTURE OPERATION RELIABILITY IN BUILT-UP AREAS

Reliability and safety of energy, drinking water deliveries and draining wastewater from built- up areas in cities and villages have many proportions and technical-operational conditions. Among basic requirements of the use of all types of residential or manufacturing objects in the 21st century is reliable delivery of drinking water, draining and cleaning of urban and industrial sewage. Disturbance or putting out of service of this type of infrastructure has always a domino effect. Gradually growing reliance of user base of all types of drinking water consumers leads especially in the section of the public or private infrastructure to restriction or complete stoppage of their operation. The following article defines in basic range the ways and means how to lower the danger that an emergency situation in operational systems of drinking water will be unmanageable and how to prevent natural or anthropogenic threats at the particular type of technical infrastructure.

Keywords: *Water, drinking water, water main, hydraulic surrounding, infrastructure of an area, risk, risk analysis.*

1. Introduction

Operational ability of infrastructure in cities and villages is today dependent from 100% on drinking water and energies supplies. Technical-operational problems appear even after a short term interruption of the supplies to media mentioned above. For example, a short term interruption of direct drinking water supplies to medical and accommodation services causes collapse in their working. Similar situation arises in the food processing plants and sanitary institutions of other infrastructure and in residential areas.

Should technical-operational facilities of water systems carry out their required function, they have to have sufficient hydraulic and operational preconditions. The input preconditions of hydraulic efficiency of water systems develop as soon as the waterworks are projected by the calculation or mathematical modelling of water mains and can be considered as sufficient. But real operational conditions change and in the terms of time get worse. The following article in its basic form implies under which conditions and in which situations this happens.

2. Operational and hydraulic preconditions for waterworks systems working

Water systems of crude and especially drinking water are significantly different compared to other types of technical infrastructure. They do not work only with media but with one of the most important kind of food, without which the human life, food processing and operating minimally all kinds of the city and village sanitary institutions are not possible.

2.1. Operational preconditions of waterworks systems function

Basic technical-operational and functional preconditions of waterworks systems of water mains for the public need in the Czech Republic must meet two preconditions given below, depending on the type of water used for production or subsequently treated drinking water in distribution plants.

Crude water meant for drinking water treatment

In EU drinking water can be made only from the water meeting conditions for the category of treatment A1, A2, A3 [1].

For the need of the technological facility of the water treatment plants the operator can set approximate treatment index (Iu) at selected indicators, when regarding higher quality unsteadiness of crude water during the year, it is not possible to fit the source clearly into one category [1]:

$$I_u = I_{u1}\frac{a}{100} + I_{u2}\frac{b}{100} + I_{u3}\frac{c}{100} + I_{u4}\frac{d}{100} \qquad (1)$$

* [1]Sarka Krocova, [2]Miloslav Rezac
[1]Faculty of Safety Engineering, VSB - Technical University of Ostrava, Ostrava - Vyskovice, Czech Republic
[2]Faculty of Civil Engineering, VSB - Technical University of Ostrava, Czech Republic
E-mail: miloslav.rezac@vsb.cz

Where:

a, b, c are frequency of presence of the indicator in category A1, A2, A3 in % and *d* is frequency of presence of the indicator in category higher than A3.

The treatment index (Iu) is a figure in the range of 1 to 3 matching categories A1 to A3. Forhigher figures that are for category A3 is Iu4. An increasing treatment index is related toworsening quality of the source. The treatment index corresponds with the difficulty index which is given by figure from 0 to 3.75. The following approximate treatment index (Iu) matches roughly the standard method of water treatment according to demands on the water treatment technology.

The quality of crude waters demanded by the regulation from the year 2001 [1] will be extremely difficult to keep in the new climate conditions of the 21st century. The average temperature, pollution of accumulated or running water and other negative effects will increase gradually, which will influence its treatment of drinking water.

Drinking water

Drinking water distributed by the waterworks systems of the water mains for the public use has very strict criteria which have to be kept during the whole process from its processing to the supply for the final consumer. The basic preconditions of the quality of drinking water and the qualities connected with drinking water are:

- Sanitation of drinking water while observing microbiological and physical-chemical index of water quality [1],
- water freshness in each profile of water distribution mains
- cerfications of all materials in contact with drinking water
- permanent monitoring of the drinking water quality in characteristic points of the distribution system.

The operational preconditions of the waterworks systems given above can be carried out in real operation only in coaction with hydraulic parameters by accumulation of drinking waters and specification of suitable dimension of water distribution mains.

2.2. Hydraulic precondition of function of distribution mains of drinking water

To guarantee the optimal working functions of all types water mains it is necessary to take into consideration surrounding in the pipe systems.

From the whole range of the matters dealing with the hydraulic of drinking water pipes and its output effect on the parameters relating consequently to the amount and quality of water it is inevitable to take into consideration the following terms and definitions.

Stationary flow is a flow the quantity of which does not change with time but only with location. The equation is:

$$Q = v \cdot S = \text{const.} \quad [m^3 \cdot s^{-1}] \tag{2}$$

Nonstationary flow is a flow the quantity of which changes with the location and time.

- stationary flow of a liquid (*Q=const.*)
- even $S_{1=}S_2$ = const., $v_1 = v_2$ = const.
- uneven delayed $S_1 < S_3$, $v_1 > v_3$
- accelerated $S_3 > S_4$, $v_3 < v_4$

A scheme of even and uneven liquid flows in pipes [2] is given in Fig. 1.

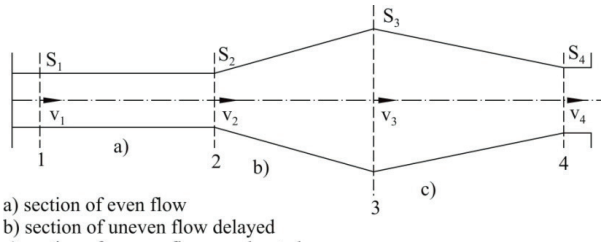


Fig. 1 Scheme of the liquid flow in pipes (adjusted by 2)

Water flow under the pressure

Resistence of the surrounding and the liquid is evident as:

- loss caused by friction Zt ,
- loss by locality Zm .

$$\text{Total loss } \sum Z = \sum Z_t + \sum Z_m \tag{3}$$

According to the type of designed pipes $Q = v \cdot S = \text{const.}$, the rate should move within the following limits:

$$d = \sqrt{\frac{4Q}{\pi v}} \tag{4}$$

$$v = \sqrt{\frac{2gH_0}{1 + \lambda \frac{1}{d}}} \tag{5}$$

d – searched average [m],
Q – flow rate [m^3/s],
v – proposal speed of water in the pipeline [m/s].

$$H_O = \frac{v_1^2}{2g} + \sum Z \tag{6}$$

According to the type of designed pipes, the rate should move within the following limits:

suction line	0.4 – 0.6 m.s^{-1},
discharge line	0.6 – 1.2 m.s^{-1},
gravitation line	1.0 – 1.5 m.s^{-1}.

The rate of water flow is always a key factor in the maintenance of drinking water quality, its freshness, and simultaneously the adherence to the stated values prevents the cavitation regime.

2.3 Methods of water supply system calculation

For the successful mathematical modelling of local water supply systems of towns and municipalities and simultaneously also inside water supply systems of industrial parks, the following conditions must be fulfilled [3 and 4]:

a) **Node condition:** The sum of inflows, outflows and quantities demanded at each node of the water supply system must be zero.

$$\sum_{i=1}^{m}(a_{ij}Q_i)+G_j=0 \qquad (7)$$

$\mathbf{Q_i}$ – rate of flow in section i [m^3.s^{-1}],
$\mathbf{a_{ij}}$ – expresses whether node j is the initial or final node of section i,
$\mathbf{G_j}$ – quantity demanded from node j [m^3.s^{-1}].

b) Loop condition: In each loop of the water supply system, pressures will be equalized. If we give the sign, which is identical with that of the direction of flow at the same basic orientation of the positive flow in the loop, to the head losses in the section, the sum of the head losses in the loop will be zero.

$$\sum_{i=1}^{m}(b_{ik}h_i)=0 \qquad (8)$$

$\mathbf{h_i}$ – head loss in section i [m],
$\mathbf{b_{ik}}$ – expresses whether the given section i is part of loop k.

c) Hydraulic condition: It describes a relation between head loss and flow rate in the section.

$$h_i=k_iQ_i^2 \qquad (9)$$

h_i – head loss in section i (m)

$$k_i=\frac{8}{g\pi^2}\lambda_i\frac{l_i}{d_i^5} \qquad (10)$$

l_i – friction factor,
d_i – internal pipe diameter in section i [m].

d) For the calculation of a looped system, which consists of m sections and n nodes, we shall get s loops:

$$s = m - n + 1 \qquad (11)$$

Source path condition: The loss in the source path equals to the difference between height levels, free discharges, etc. of end nodes of the source path. If the number of pressure-dependent nodes is y, the number of source paths is equal to y-1.

$$\sum_{i=1}^{m}(c_{iy}h_i)=H_y^P-H_y^K \qquad (12)$$

hi – head loss in section i [mm],
ciy – orientation of sections of the source path,

HyP,K– level of pressure head at nodes at the beginning and at the end of the source path.

Detailed knowledge of operating and hydraulic parameters of the water supply system being dealt with substantially increases the efficiency of safety planning and improves preventive preparation for coping with potential extraordinary events.

3. Safety risks of the functional changes of waterworks systems

It is not possible to prevent the cases of emergency in the water-supply engineering. It is possible to prepare on their progress and negative impact in advance from operational-safety analysis. The changes in functioning of the waterworks facilities are always caused by two causes with different potential of negative impact on the supplies of drinking water to consumers.

3.1. Natural influence

Natural influence is tightly connected with water ecosystems and also with the resources of surface and subterranean waters. Negative impact on the waterworks systems functioning can be caused by:

- the change of climatic conditions (long-term droughts, water temperature and floods),
- natural higher releasing of inorganic rocksubstances into water,
- earthquake and following breaks of the land relief including the disturbing of aquiferous levels,
- landslides disturbing linear constructions of the waterworks systems.

Natural operational-safety risks can be predicted by the means of risk analysis and reasonably the lowest negative impact can be prepared.

3.2. Anthropogenic contingencies

Anthropogenic contingencies have their occurrence in human activities. They show negatively on the water resources especially on distribution systems. They are mainly caused by the following characteristic features:

- breaking the regulations for activities and water management in water resourcesprotection zones, leakage of contamination substance from old ecological burdens into subterranean waters,
- breaking the regulations for economical use of drinking water by the operators of waterworks systems and subsequent secondary contamination of drinking water in accumulations or piping.
- leakage of dangerous substances from industrial activities or traffic accidents into surface waters,
- unintentional damage of waterworks linear construction during earthwork,
- international damage or terrorism.

Most of anthropogenic contingencies can be hardly predicted. In most cases they occur due to the lack of technological discipline and by breaking obligatory regulations.

4. Discussion

Considering that operability of waterworks systems including water resources and continuity in water supplying is everybody´s concern, it is important not only to run education but firstly to lead discussion among experts and look for ways how to make the present state better.

Professional discussion should be focused, for example, on the following fields:

- looking for new economically feasible ways improving the balance of water reserves in the climatic conditions of the Czech Republic.
- lowering the secondary risk of drinking water contamination during operation breakdown and finding methods how to prevent deliveries of contaminated water to consumers [5].
- considerable increasing of hydraulic efficiency of water mains.

Discussion on this and other issues in the field of water management can considerably advance and make more effective not only the waterworks but, in the wider context, also water management especially in the conditions of the forthcoming changes in the mild climatic zone.

5. Conclusion

This article in its basic range shows the ways and means for increasing the reliability of drinking water deliveries for built-up areas of the cities and villages to which the discussion and attention are focused with the aim to increase efficiency of operational waterworks systems. Regarding the forthcoming minimally regional shortage of water the present change of the approach to the water management can be considered absolutely necessary.

Acknowledgement

This contribution was prepared as a part of the grant project of the Ministry of the Interior of the Czech Republic under number VF20112015018, titled Safety of Citizens – Crisis Management (in Czech).

References:

[1] Decree of Ministry of Health No. 428/2001 Coll., implementing Act No. 274/2001 Coll., on Water Supply and *Sewerage* Systems for Public Use. Collection of Laws 2004, section 48, 2004.
[2] SLOVACKOVA, M., SULC, M.: *Technical Handbook for Water Managers - Secondary Building School (in Czech)*, Lipnik nad Becvou, 1998.
[3] INGEDULT, P., VYCITAL, J.: Mathematical Modelling of Water Supply Networks - Part 1 (in Czech), *Sovak*, vol. 8, No. 3, 1999.
[4] ODULA: *Program Agent Manual* (in Czech), DHI Hydroinform a. s., 2001.
[5] BERNATIK, A., SENOVSKY, P., SENOVSKY, M., REHAK, D.: Territorial Risk Analysis and Mapping. *Chemical Engineering Transactions*, 31, 79-84, 2013.

Roman Kozel - Milan Mikolas - Sarka Vilamova - Katerina Chuchrova - Marian Piecha *

PORTER'S ANALYSIS AS A STANDARDIZED PROCESS APPLICABLE IN INDUSTRIAL COMPANIES

Reliability and safety of energy, drinking water deliveries and draining wastewater from built- up areas in cities and villages have many proportions and technical-operational conditions. Among basic requirements of the use of all types of residential or manufacturing objects in the 21st century is reliable delivery of drinking water, draining and cleaning of urban and industrial sewage. Disturbance or putting out of service of this type of infrastructure has always a domino effect. Gradually growing reliance of user base of all types of drinking water consumers leads especially in the section of the public or private infrastructure to restriction or complete stoppage of their operation. The following article defines in basic range the ways and means how to lower the danger that an emergency situation in operational systems of drinking water will be unmanageable and how to prevent natural or anthropogenic threats at the particular type of technical infrastructure.

Keywords: *Water, drinking water, water main, hydraulic surrounding, infrastructure of an area, risk, risk analysis.*

1. Introduction

The business environment, especially in recent years, has been influenced by frequent and sudden changes as well as by strong competition. Early prediction of market opportunities, identifying threats and, of course, solving potential problems of strategic nature result in business success [1]. Therefore, the strategic analyses that can draw attention to potential opportunities and threats should be an integral part of the business activities of each economic entity. This is especially true for industrial companies, as demand for their products is by the definition of B2B market a derived demand. It is true that the ideal model of strategic management of the company does not exist; however, some similarities in taking certain linked steps can be traced as shown in Fig. 1, where the strategy life cycle is illustrated. It would be a mistake to try using universal strategies in all the companies.

Properly formed strategy must be tailor-made for each company, especially on B2B market, where there are fewer, but larger companies than on the consumer market. And an important starting point for formulating strategy is the result of strategic analysis. The aim is to identify and evaluate all relevant factors which may be presumed to influence the final choice of company's targets and strategy.

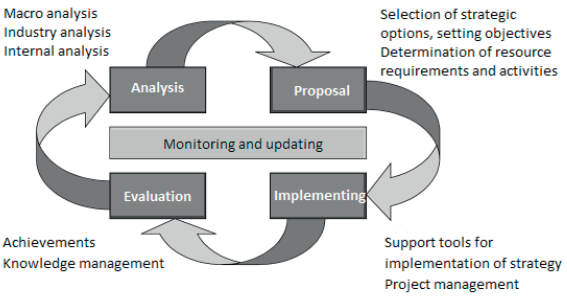


Fig. 1 Life cycle of a strategy [2]

The competition has a substantial influence on the achievement of specific objectives on the B2B market. An important method of the strategic analysis is the analysis of competitive forces. The most known is the **Porter five forces analysis** that identifies the degree of influence affecting the given industry or company. The goal of the model is to identify and understand the forces that take effect within the industry. Enterprise that wants to be successful should respond to these forces and deal with them and, if possible, change their influence to their advantage [3 and 4].

* [1]**Roman Kozel,** [2]**Milan Mikolas,** [1]**Sarka Vilamova,** [1]**Katerina Chuchrova,** [3]**Marian Piecha**
[1]Institute of Economics and Control Systems, Faculty of Mining and Geology, VSB - Technical University of Ostrava, Czech Republic
[2]Institute of Mining Engineering and Safety, Faculty of Mining and Geology, VSB - Technical University of Ostrava, Czech Republic
[3]Section of Support for Small and Medium-sized Enterprises, Ministry of Industry and Trade of the Czech Republic
E-mail: roman.kozel@vsb.cz

2. Analysis

There are no exact instructions for implementation of the Porter's analysis even though it is one of the most used analyses. This method is very often used for e.g. situational analysis or to set the competitive advantage, but its procedure and results are merely subjective, verbal interpretation of an expert applicable only to a problem which is currently being solved.

The aim of the research team is to create its standardized, visualized version bearing specific results suitable for presentation to the TOP management. The tendency is to create a user-friendly software environment so that the managers of the industrial companies, who are not usually involved in this matter, are able to create such analysis themselves.

Many authors pursued the topic of Porter's analysis in their published works. This analysis was, for instance, used in an article dealing with aspects of maritime safety, such as control of the sea, nuclear deterrence and defence against ballistic missiles, as well as naval consensus. [5] The analysis was also used when examining the competitive advantage of the Jordan Phosphate Mines Company (JPMC) based in Jordan and founded in 1949 [6] and in other articles following the force of the competition in the given industry [7].

Porter five forces analysis is also used when formulating the marketing strategy as it was, for example, used in the articles *Developing Marketing Strategy for Electronic Business by Using McCarthy's Four Marketing Mix Model and Porter's Five Competitive Forces* [8] and *Designing marketing strategy using the five competitive forces model by Michael E. Porter – case of small bakery in Croatia* [9].

The resulting data were visualized in this research and research data for the evaluation was collected on the basis of a questionnaire that was based on similar issues, which are defined in the model that is currently being constructed by the authors of this paper. Other articles deal primarily with innovation, modernization and improvement of Porter's ideas in relation to the new trends and facts in marketing and strategic areas [10 and 11]. The theoretical basis and application of Porter's analysis that were used in those articles will serve as groundwork for standardization of Porter's analysis with the use of suitable graphic outcomes.

3. Methods and Procedures

The aim of the research is to create a user-friendly software environment that offers processing of selected analyses, thus standardized inputs, built-in algorithms and methods of computation and subsequent standardized outputs in the form of tables and graphic outcomes. The system is therefore formed by the combination of Excel spreadsheets and VISUAL BASIC programming background to create the utility program.

Functionality of the standardized Porter's analysis in this software environment was tested by the means of competitive analysis implementation in this support system on real data of a company that is involved in the field of mining and processing of stones in the Moravian-Silesian Region. Due to the fact that data used in this article are real data and, therefore, are subject to secrecy, the mining company will be addressed as ABC and its suppliers / customers as companies Alpha, Beta, Gama, Delta and Omega.

As inputs for the system analysis the secondary data from internal company documents, data from previous analyses of the market and specific information and professional judgments of their managers were mostly used.

An important characteristic of the industry are competing forces influencing that sector. Analysis usually includes five Porter's forces, but it depends merely on the evaluator which of the forces affecting the company will be chosen. The menu of Porter's analysis guides the evaluator on how to proceed with individual analysis, see Fig. 2. For every force a main table, which includes a number of statements that express the level of dependence on the given group influencing the company, is defined.

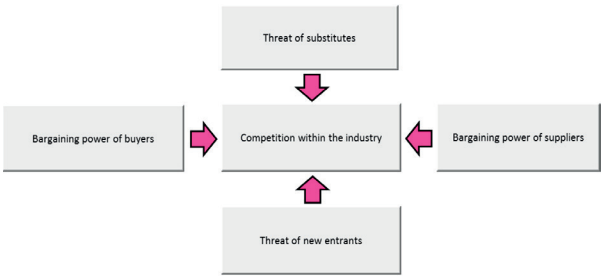


Fig. 2 Porter's analysis menu

The evaluators express the degree of compliance with individual claims on the scale of 0 - 10 (0 - absolutely disagree, 10 - agree completely), which also illustrates the strength of influence (0 - no influence, 10 - great influence). The values indicating the score of influence of individual factors are calculated by the use of an arithmetic average and the influence coefficient is derived for every analyzed group of factors:

- <0; 3.50> negotiating strength of the group is small
- <3.51; 7.00> negotiating strength of the group is average
- <7.01; 10.00> negotiating strength of the group is great.

When testing the functionality of the software support for standardization the Porter's analysis, all five forces (groups) were tested, and for each of the force at least 7 factors that were further evaluated had been set.

Analysis of the bargaining power of buyers – this partial analysis shows the influence of the buyers, customers. The question is what the power of our customers is. Strong customers may cause the loss of potential profit; they may negotiate many

CUSTOMERS OF THE COMPANY			
No.	Statement	Weight of statement	Customers
	Use pair comparison	☑	
1	We do not have enough customers	21%	7
2	Customers buy large volumes	21%	5
3	The industry has a large number of enterprises	21%	8
4	Our product is very standardized	9%	10
5	Customer purchases represent a large percentage of total sales in the industry	9%	2
6	Risk of integration of buyers	4%	1
7	The purchased product is an important input or necessary product.	7%	3
8	For the buyer it is economically more advantageous to purchase inputs from several suppliers.	7%	8
9		0%	
10		0%	
	Average rating	*100%*	6.18
	Total sales	CZK	

Fig. 3 Analysis of the customers

privileges and so on. Basic evaluating questions used when testing the system are stated in the table shown in Fig. 3.

Competition within the industry - Due to the limited number of customers on the B2B market, the intensity of this competitive force represents a noteworthy signal which predicates how much power the competing companies place into the efforts to gain a better market position.

When testing real data for this area the following statements were chosen: *Is there a large number of direct competitors in our industry? Is it a capital-intensive industry? Are there large differences in quality among the competitors? Are there large differences in price among the competitors? The competition offers better accompanying services. High fixed costs. Manufactured/offered product/service is poorly differentiated. High barriers to exit the industry. Slow growth of the industry / Slow demand growth.*

Threats of substitutes

Companies operating within one industry are in direct competition with companies in different industry if their products represent a good substitute. Competitive threats were during the testing learned by the use of the following factors: *The price of a substitute lower than the price of our product/service. The quality of the substitute is better than of our product/service. The substitute offers better accompanying services. Better shelf life of the product. Better handling of the substitute. Better recyclability of the substitute. Quicker delivery of the substitute. More favorable features of the substitute. The cost of the customer to switch to substitute is negligible* [12].

Analysis of the threat of new entrants

Low barriers to enter the industry could represent a significant threat of new entrants to the industry. That is why during the testing the following entry barriers were formulated: *Do you have any special technologies and know-how? There are too many legislative measures and frequent interference of the state. Your company has high demands regarding the human resource. The experience plays an important role within your industry. You are effective in achieving large volumes. You have dedicated customers. Your brand is well-known. The market penetration into your industry sector is investment demanding. You possess cost advantages. You possess distribution channels.*

Analysis of the bargaining power of suppliers

The power can be very low and the suppliers therefore have no impact on our business. Or their power can be so high that any interference from their side can have a great influence on our production. The evaluation theses could be, for instance, defined in the following way: *The supplier is hard to replace. There are not enough suppliers on the market. There is no substitute. The supplier is threating by possible integration. Supplier's product is an important entry for the business of the customer. The industry is not a primary aim of the supplier. What are the overall annual costs of the supplier?*

Figure 3 shows that pre-defined statements do not have to have the same influence for particular industry or company. Therefore, the possibility to set the weight of importance of individual evaluation criteria within the complex evaluation must exist in the way that the more importance the criterion has the more weight it has. In the basic design of the tested system are

	Statement	1	2	3	4	5	6	7	8	9	10	Frequency	Weight
		We do not have enough customers	Customers buy large volumes.	The industry has a large number of enterprises.	Our product is very standardized.	Customer purchases represent a large percentage of total sales in the industry.	Risk of integration of buyers.	The purchased product is an important input or necessary product.	For the buyer it is economically more advantageous to purchase inputs from several suppliers.	0.0	0.0		
1	We do not have enough customers.		1.2	1.3	1	1	1	1	1			6	21%
2	Customers buy large volumes.			2.3	2	2	2	2	2			6	21%
3	The industry has a large number of enterprises.				3	3	3	3	3			6	21%
4	Our product is very standardized.					4.5	4	4.7	4.8			2.5	9%
5	Customer purchases represent a large percentage of total sales in the industry.						5	5.7	5.8			2.5	9%
6	Risk of integration of buyers.							6.7	6.8			1	4%
7	The purchased product is an important input or necessary product.								7.8			2	7%
8	For the buyer it is economically more advantageous to purchase inputs from several suppliers.											2	7%
9	0.0											0	0%
10	0.0											0	0%
	Total											28	100%

Fig. 4 Table of pair comparison analysis of the customers

all statements of the same significance. When using the "use pair comparison" (see Fig. 3), it is possible to define the importance of individual statements and thereby determine their weight. The method of pair comparison was used to set the importance/weight during the testing phase. This method determines the preferential relationship between a pair of criteria. The purpose is to outline the number of preferences for each criterion in relation to all the other criteria of the given file (see diagram of Fig. 4).

In the upper right part of the table (upper triangular matrix) the evaluator examines whether the pair of criteria prefers the criteria stated in a line before the criterion set out in a column. If that is the case, the evaluator writes the number of criterion stated in the line into an appropriate field, if it be to the contrary, the evaluator writes the number stated in the column. If the evaluator perceives the criteria equally important, both criteria numbers must be written. Resulting weight of the criteria for each criterion stated in % is derived as a number of its preferences that is also equal to the sum of preferences in the line and column of that criterion. The result of each analysis is partly the strength of influence of individual groups on the company and partly the average evaluation of the group. After completing all the relevant groups influencing the course of events in the company, the final result of the influence can be seen on the spider diagram that describes the tested company with regard to the strength of the influencing powers. The number of angles depends on the number of selected influencing powers.

4. Results

The authors of this article filled in the programmed Porter's analysis together with the managers of company ABC. Due to the limited maximum number of pages of this article, the authors will only present the results of the implementation and outcomes of the suppliers analysis. At first it was necessary to appoint a supplier of company ABC. For the purpose of this analysis five different suppliers were selected. The preferences of individual statements were set by the means of pair comparison. The weight of the statement will consequently appear in the third column of the table in Fig. 5.

After completing the suppliers evaluation, the value of "Average evaluation" of a specific value of the average evaluation of individual suppliers, can be seen in the penultimate line. The last column states the average of the whole group. In this case the influence coefficient equals 7, which means that the negotiating power of this group is great.

In the last line "The total cost" the evaluator must state the annual costs associated with using the services of a particular supplier. As soon as all the required parameters were entered, it was possible to choose from the graphical tools menu and display the results of the analysis. As an example we can show, e.g., a chart evaluation of individual suppliers or a portfolio chart, the average rating of the supplier / costs of supplier. It depends purely on the requests, needs and choice of the evaluator.

In this article the chart showing the average evaluation of suppliers is shown in Fig. 6. This chart shows the rate of influence individual suppliers have on the company in question,

THE LARGEST SUPPLIERS													
No.	Statement	Weight of statement	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	Average
	Use pair comparison	☑	Alpha	Beta	Gamma	Delta	Omega						ø
1	Supplier is hard to replace.	33%	10	6	6	1	6						
2	There are only few suppliers.	27%	10	7	8	2	7						
3	There is no substitute.	13%	10	10	7	10	9						
4	Supplier is threating by possible integration.	10%	6	6	7	1	1						
5	Supplier's product is an important input for the business of a customer.	13%	10	10	10	10	9						
6	The industry is not an important customer of the supplier.	3%	10	8	10	10	3						
7		0%											
8		0%											
9		0%											
10		0%											
	Average rating	*100%*	9.60	7.40	7.43	3.97	*6.47*	0.00	0.00	0.00	0.00	0.00	7
	Total costs	CZK	1 170 000	260 000	2 400 000	10 680 000	6 000 000						4 102 000

Fig. 5 The analysis of suppliers

therefore, to what extent they are able to influence the business of the analyzed company. On top of that, each supplier has a different size of bubble on the 3D chart, showing the size of the supplier in terms of annual costs. The chart is designed so that those suppliers who occur in the upper third of the chart have the greatest influence on the analyzed company. The results obtained from company ABC imply that the greatest influence is shown by the supplier called Alpha that supplies the main raw material. The company has concluded a number of long-term contracts with this supplier until the year 2023 for excavation of spoils. Among other companies that have a great influence on the above mentioned company are also suppliers Beta and Gamma. The charts also clearly show that all the other analyzed suppliers are those with an average bargaining power. Overall, the smallest impact on the business of company ABC has supplier Delta, despite the fact that the annual costs spend on these suppliers are by far the greatest (see the biggest bubble).

Analyses of the remaining four forces of Porters analysis were carried out analogically as the partial competitive forces are programmed in a similar way as the bargaining power of suppliers. However, each of the forces has its own specifics. For instance, with the analysis of bargaining power of buyers, it is not necessary to name individual customers if the evaluator does not wish to do so, it is sufficient to evaluate them as a whole. The outcomes of the entire Porter's analysis in the tested system are a table and a chart "Average evaluation of competitive forces". It is, therefore, possible to create a model of competitive forces and determine the degree of rivalry between the individual parts, see Fig. 7.

The reference chart in the picture displays the result of system testing on real data, show the most significant risks and threats and at the same time opportunities for development of the company that can help to achieve either current or future goals. The biggest threat to company ABC represents the competition within the industry.

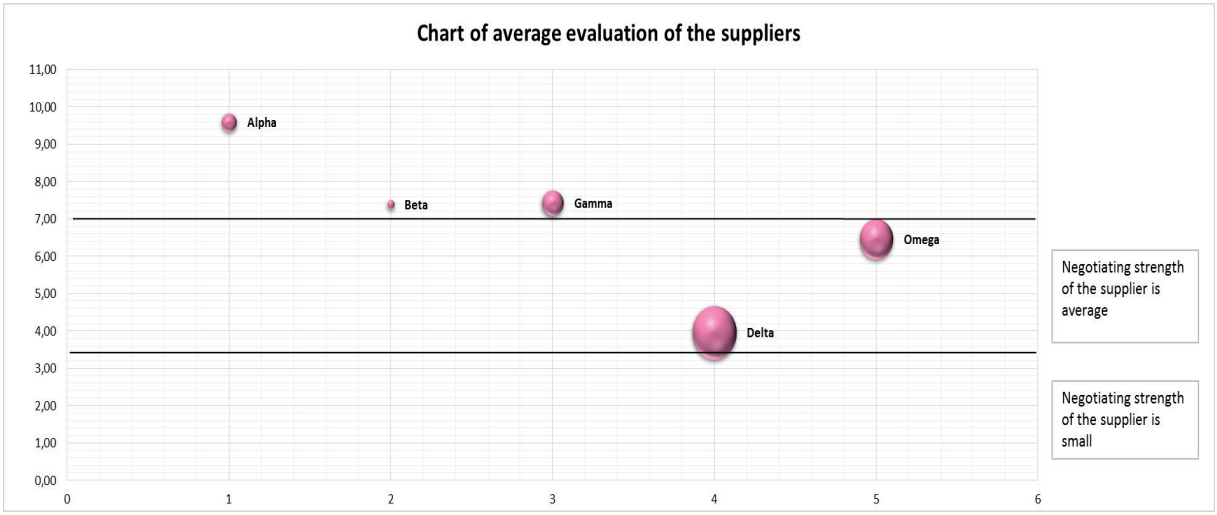


Fig. 6 Average evaluation of suppliers

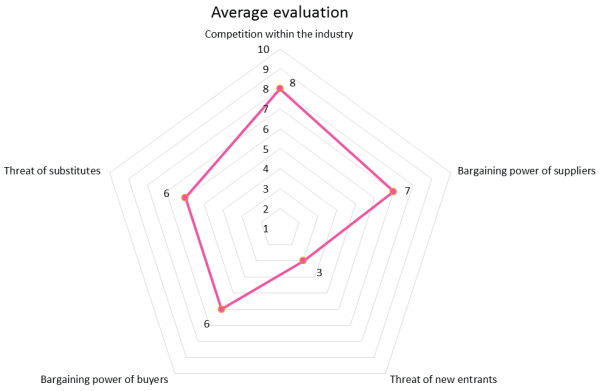


Fig. 7 Final evaluation

The company should, therefore, focus on their products and decide how to surpass its competitors and how to differentiate. Simultaneously it was found that other groups of competitive forces reach rather high or medium influence which alerts the managers to start focusing at these points as well.

The good news is that there is only a small threat of new entrants which is demonstrated by a very low influence coefficient. The cause may be high investment to enter the business in this field, announced regulations, etc.

5. Conclusion

The aim of the research team was to create standardized and visualized form of Porter's analysis, method that would be suitable for the determination of risk factors.

The article shows different input forms, procedures for the preparation and analysis of selected key graphical outputs. Given the scope of the article it was not possible to present all the specifics of this programmed strategic analysis. However, the functionality of the system together with the clarity of the standardized analysis outcomes that are needed for making strategic decisions in B2B companies was proven on the real data derived from the internal materials of company ABC. This model will, therefore, be gradually completed with further analyses and consequently tested on other significant industrial companies in other sectors so the programmed analyses are easy to complete, while offering clear and real results about the state of the analyzed company.

Acknowledgement

The work was supported by the specific university research of Ministry of Education, Youth and Sports of the Czech Republic No. SP2015/36 Creation of competitive analysis system for industrial companies.

References

[1] KOZEL, R., L., MYNAROVA, SVOBODOVA, H.: *Modern Methods and Techniques of Marketing Research (in Czech).* Praha : Grada, 2011. 304 s. ISBN 978-80-247-3527-6.

[2] SEDLACKOVA, H., BUCHTA. K.: *Strategic Analysis - 2nd ed. revised and supplemented (in Czech),* Praha : C. H. Beck, 2006, 121 p., ISBN 80-717-9367-1.

[3] JAKUBIKOVA, D.: *Strategic Marketing: Strategy and Trends*. Praha : Grada, 2008, 269 p., ISBN 978-80-247-2690-8.

[4] MALLYA, T.: *Principles of Strategic Management and Decision Making (in Czech)*. Praha : Grada, 2007, 246 p., ISBN 978-80-247-1911-5.

[5] YETKIN, U.: Revealing the Change in the Maritime Security Environment through Porter's Five Forces Analysis. *Defence Studies.* 2013, vol. 13, 4, pp., 458-484. ISSN 1470-2436.

[6] ALRAWASHDEH, R.: The Competitiveness of Jordan Phosphate Mines Company (JPMC) Using Porter Five Forces Analysis. *Intern. J. of Economics and Finance.* 2013, vol. 5, No. 1, pp. 191-200. ISSN 1916-971X.

[7] RICE, J. F.: Adaptation of Porter's Five Forces Model to Risk Management. *Defense AR J.,* 2010, vol. 17, No. 3, pp. 375-388. ISSN 1553-6408.

[8] AZADI, S., RAHIMZADEH, E.: Developing Marketing Strategy for Electronic Business by Using McCarthy's Four Marketing Mix Model and Porter's Five Competitive Forces. *EMAJ: Emerging Markets J.,* 2012, vol. 2, No. 2, pp., 46-58. ISSN 2159-242X.

[9] RENKO, N., SUSTIC, I. BUTIGAN, R.: Designing Marketing Strategy Using the Five Competitive Forces Model by M. E. Porter - Case of Small Bakery in Croatia. *Intern. J. of Management Cases.* 2011, vol. 13, No. 3, pp. 376-385. ISSN 1741-6264.

[10] GRUNDY, T.: Rethinking and Reinventing Michael Porter's Five Forces Model. *Strategic Change.* 2006, vol. 15, No. 5, pp. 213-229. Online ISSN: 1099-1697.

[11] HANZELKOVA, A. et al.: *Business Strategy: Step by Step (in Czech).* Praha : C. H. Beck, 2013, 159 p., ISBN 978-80-7400-455-1.

[12] MIKLOSIK, A.: Selected Aspects of Systemic Approach to Project Management. *Actual Problems of Economics: Scientific Economic J.,* 2014, No. 5, pp. 195-202. ISSN 1993-6788.

Ondrej Varta - Sarka Krocova *

THE LOCATION OF LPG FILLING STATIONS AND POTENTIAL RISKS OF INCIDENTS

The risk of unsuitable placement of LPG (i.e. Liquid Petroleum Gas) propane-butane filling stations in densely populated areas or in the immediate vicinity of busy roads may have, in the event of fire or explosion, incalculable consequences in terms of fatal or serious accidents and cause considerable economic damage to property. As a consequence, the positioning of any such LPG filling station should always be of crucial importance already in the filling station planning stage and its placement should be carefully assessed. The present paper aims at outlining a particular method of evaluating risks resulting from incorrect placement of an LPG filling station, including potential consequences. One such approach to assessment of installation safety of LPG filling stations is to implement the so called ALOHA screening method. This paper shall demonstrate practical implementation of the ALOHA screening, and that with regard to infrastructure, it is not always possible to find a suitable and safe location for LPG filling stations.

Keywords: *ALOHA, safe LPG filling station location, emergencies, safe distance.*

1. Introduction

The purpose of the following section is to describe, with the use of a model situation, the placement of an LPG filling station (Fig. 2) and to stress the importance of paying particular attention to safe placement of this installation for risks and incidents elimination with respect to the infrastructure, the occurrence of incidents and their adverse effects.

The word 'incident' in the context of the presently discussed situation conveys the extraordinary conditions and impact on the affected area, as well as on general public. The magnitude of the incident is the consequence of destructive effect and the nature of the affected area [1].

The consequences of the emergence of extraordinary conditions are numerous and they may be classified according to the following Fig. 1:

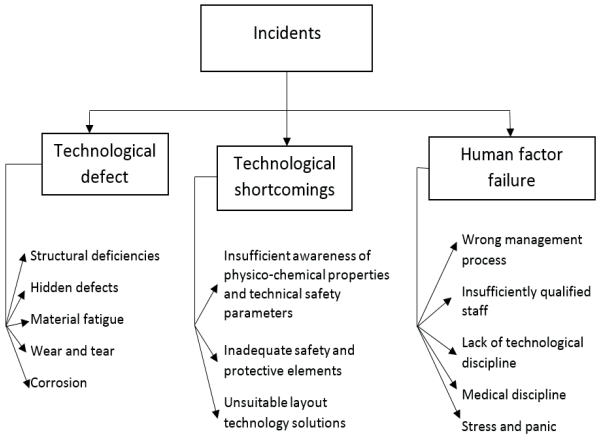


Fig. 1 The Emergence of Incidents [1]

For practical modeling of the ALOHA screening, a fictitious LPG filling station situated in a populated urban area close to a shopping center was chosen. The location of the LPG filling station was specified by a 50-metre distance from the shopping center and 10 meters from the road with normal city traffic. In terms of technical specification, a single-dispenser-unit

* [1]**Ondrej Varta,** [2]**Sarka Krocova**
[1]State Labour Inspection Office, Opava, Czech Republic
[2]VSB Technical University of Ostrava, Ostrava - Poruba, Czech Republic
E-mail: ondrej.varta@suip.cz

LPG filling station was considered, with a 4000-liter reservoir. The application of the ALOHA method will demonstrate the importance of resolving the matter of safe location for the LPG filling station prior to any construction works.

Fig. 2 An illustrative photograph of an LPG filling station (source: author – O. Varta)

2. ALOHA Screening Method and its Features

The ALOHA software application (Areal Locations Of Hazardous Atmospheres) is a dispersive model for Windows operating system that was developed by the United States Environment Protection Agency (US EPA) [2].

The software uses a series of Gauss distribution equations to evaluate the movement of pollutants released into the air.

The software works with inputs specified in the following fixed order:

- Site location data (place name, country, type of buildings);
- Information on the released agent (the software has a large database of chemical substances, including the necessary physical and chemical properties);
- Information on the state of atmosphere (the grade of air temperature stratification stability following the Pasquill's scale, wind strength and direction, air temperature, degree of cloud cover, etc.);
- Information on the source of the leakage; it is possible to enter 4 source types and their parameters (direct source, puddle, reservoir, pipeline).

The ALOHA software allows to have results in both textual and graphic form. It renders a cloud of an agent of specified concentration, dosage, and source capacity.

The limitations of the ALOHA software include:

- The software works with low wind velocities and stable atmospheric conditions;
- Small resolution of terrain topography;
- The software does not take into account changes in wind direction, impact of fire and chemical reactions, dispersion of solid particles and/or solutions;
- The agent leakage time is set to a period of 1 hour and the dispersion distance is restricted to 10 km.

The latest version of ALOHA software (5.4 of February 2006) added fire and explosion modeling functionality.

Apart from threats of toxicity, users may also evaluate fire hazards associated with the so called jet fires, puddle fires, vapor cloud explosions (VCE), boiling liquid expanding vapor explosions (BLEVE) and flashfires.

The ALOHA software is available free of charge on the US EPA official website [3].

3. Practical Steps for the Use of the ALOHA Method

The location selected for the LPG filling station was Opava, Moravian-Silesian Region, Czech Republic, random choice from the main streets available in this town. The practical calculation example involves toxic cloud dispersion and BLEVE explosion modeling (see Fig. 3 and 4).

4. Calculation pursuant Toxic Cloud Dispersion

SITE DATA:

Location:	TOWN CZECH REPUBLIC
Building Air Exchanges Per Hour:	0.59 (unsheltered single storied)
Time:	December 11, 2014 0931 hours ST (using computer's clock)

CHEMICAL DATA:

Chemical Name:	PROPANE
Molecular Weight:	44.10 g/mol
AEGL-1 (60 min):	5500 ppm
AEGL-2 (60 min):	17000 ppm
AEGL-3 (60 min):	33000 ppm
IDLH:	2100 ppm
LEL:	21000 ppm
UEL:	95000 ppm
Ambient Boiling Point:	-42.7° C
Vapor Pressure at Ambient Temperature:	greater than 1 atm
Ambient Saturation Concentration:	1,000,000 ppm or 100.0%

ATMOSPHERIC DATA: (MANUAL INPUT OF DATA)

Wind:	3 meters/second from 50° true at 3 meters
Ground Roughness:	urban or forest Cloud Cover: 5 tenths
Air Temperature:	20° C
Stability Class:	D
No Inversion Height	
Relative Humidity:	50%

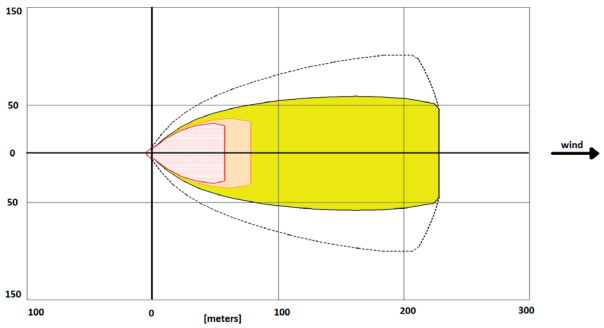
SOURCE STRENGTH:

Leak from short pipe or valve in horizontal cylindrical tank	
Flammable chemical escaping from tank (not burning)	
Tank Diameter:	1.8 meters
Tank Length:	1.57 meters
Tank Volume:	4.00 cubic meters
Tank contains liquid	
Internal Temperature:	20° C
Chemical Mass in Tank:	1.88 tons
Tank is 85% full	
Circular Opening Diameter:	5 centimeters
Opening is 0 meters from tank bottom	
Release Duration:	3 minutes
Max Average Sustained Release Rate:	755 kilograms/min (averaged over a minute or more)
Total Amount Released:	1,706 kilograms
Note:	The chemical escaped as a mixture of gas and aerosol (two phase flow).

THREAT ZONE:

Threat Modeled:	Flammable Area of Vapor Cloud
Model Run:	Heavy Gas

DISPERSION CLOUD RESULTS



Red: 58 meters – (21000 ppm = LEL)
Orange: 79 meters – (12600 ppm = 60% LEL = Flame Pockets)
Yellow: 229 meters – (2100 ppm = 60% LEL)

Fig. 3 Affected area in the vicinity of the LPG filling station [4]

5. BLEVE- Boiling Liquid Expanding Vapor Explosions

SITE DATA:

Location:	OPAVA, Tesinska ulice CZECH REPUBLIC
Building Air Exchanges Per Hour:	0.59 (unsheltered single storied)
Time:	December 11, 2014 0931 hours ST (using computer's clock)

CHEMICAL DATA:

Chemical Name:	PROPANE
Molecular Weight:	44.10 g/mol
AEGL-1 (60 min):	5500 ppm
AEGL-2 (60 min):	17000 ppm
AEGL-3 (60 min):	33000 ppm
IDLH:	2100 ppm
LEL:	21000 ppm
UEL:	95000 ppm
Ambient Boiling Point:	-42.7° C
Vapor Pressure at Ambient Temperature:	greater than 1 atm
Ambient Saturation Concentration:	1,000,000 ppm or 100.0%

ATMOSPHERIC DATA: (MANUAL INPUT OF DATA)

Wind:	3 meters/second from 50° true at 3 meters
Ground Roughness:	urban or forest
Cloud Cover:	5 tenths
Air Temperature:	20° C
Stability Class:	D
No Inversion Height	
Relative Humidity:	50%

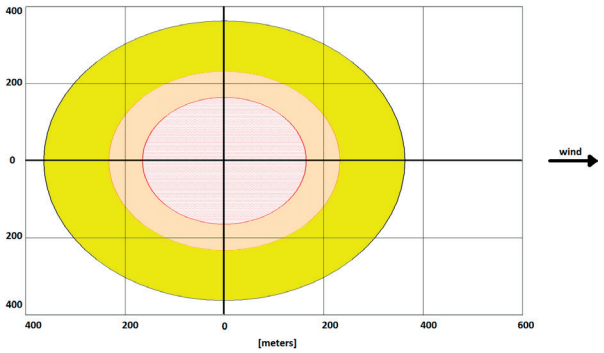
SOURCE STRENGTH:

BLEVE of flammable liquid in horizontal cylindrical tank	
Tank Diameter:	1.8 meters
Tank Length:	1.57 meters
Tank Volume:	4.00 cubic meters
Tank contains liquid	
Internal Storage Temperature:	20° C
Chemical Mass in Tank:	1.88 tons
Tank is 85% full	
Percentage of Tank Mass in Fireball:	100%
Fireball Diameter:	69 meters
Burn Duration:	6 seconds

THREAT ZONE:

Threat Modeled:	Thermal radiation from fireball

DISPERSION CLOUD RESULTS



Red: 165 meters – (10.0 kW/(sq m) = potentially lethal within 60 sec)
Orange: 233 meters – (5.0 kW/(sq m) = 2nd degree burns within 60 sec)
Yellow: 363 meters – (2.0 kW/(sq m) = pain within 60 sec)

Fig. 4 Affected area in the vicinity of the LPG filling station [5]

6. Conclusions

We shall now compare the ALOHA method calculation result with the distances entered to delimit the LPG filling station location; the original distances were:

a) 50 meters between the LPG filling station and the shopping center

b) 10 m meters from the road with normal city traffic and the calculation of the dispersion cloud with the use of ALOHA method:

Red: 58 meters – (21000 ppm = LEL)
Orange: 79 meters – (12600 ppm = 60% LEL = Flame Pockets)
Yellow: 229 meters – (2100 ppm = 60% LEL)

and the result for the BLEVE phenomenon:

Red: 165 meters – (10.0 kW/(sq m) = potentially lethal within 60 sec)
Orange: 233 meters – (5.0 kW/(sq m) = 2nd degree burns within 60 sec)
Yellow: 363 meters – (2.0 kW/(sq m) = pain within 60 sec)

From the above it is possible to come to a definite conclusion that the suggested location of the LPG filling station in Opava is completely **unsuitable** because of the proximity to a densely populated area and the shopping center.

The LPG filling station located as suggested above might be the source of an emergency incident following certain technological break-down causing injuries to a significant number of citizens, including subsequent likely explosion of BLEVE variety.

For the aforementioned reasons involving very high level of risk associated with emergency incidents, it is necessary to make use of feasibility study methods already in the course of assigning the relevant project documentation (e.g. the ALOHA screening method used in the above case study).

It is expected that the procedure set in this manner will lead to a decrease in the level of risks from stationary sources and emergency incidents.

References

[1] BARTLOVA, I., PESAK, M.: *Risk Analysis and Prevention of Industrial Emergencies II, Risk Analysis and Awareness of Industrial Emergency Situations,* EDICE SPBI SPEKTRUM 33, 2003. ISBN 80-86634-30-2.
[2] US EPA http://www.epa.gov/ceppo/cameo/aloha.htm
[3] BERNATIK, A.: *Prevention of Serious Emergency Incidents II, University textbook VSB, FBI, VSB Ostrava, 2006.*
[4] *Affected area in the vicinity of the LPG filling station, http://www.epa.gov/ceppo/cameo/ aloha.htm*
[5] *Affected area in the vicinity of the LPG filling station, http://www.epa.gov/ceppo/cameo/ aloha.htm.*

Pawel Szataniak - Magdalena Mazur - Robert Ulewicz - Frantisek Novy *

FRACTOGRAPHIC ANALYSIS OF HARDOX STEELS RESEARCH IN THE FIELD OF HIGH-CYCLE FATIGUE REGIME

The article presents preliminary fatigue tests results of two high-strength steels used in structural elements of trailers, wagons, farm devices, etc. Paper presents an analysis of fatigue test results of high-strength steels in the range from 104 to 107 numbers of cycles of applied load with a frequency of 40 Hz. Wöhler curve shows the results of the fatigue properties of steels Hardox 400 and Hardox 450, which are used in the construction of vehicles. Fatigue fractures were subjected to fractographic analysis in order to determine the mechanism of fatigue crack propagation.

Keywords: *Fatigue crack, Wohler curve, high-strength steels, high cycle fatigue tests.*

1. Introduction

Information about the fatigue strength of the material has become essential in solving the general problem of improving the reliability and durability of modern machines and constructions [1]. Demand for fatigue results of tested steel is reported by designers looking for new applications of high strength steel. Because manufacturers are interested in optimizing operational costs and increase of production, therefore, they pay attention to getting the greatest durability of produced objects. This is obtained by looking for new construction solutions of their products and application of appropriate materials (materials with appropriate properties). Material fatigue is a process of continuous accumulation of damages formed over a sufficiently long time. This process occurs due to variable mechanical stresses which cause nucleation and propagation of cracks, leading consequently to destruction of the material [2 - 4].

Extra high strength steels are used in structures such as truck chassis, cranes and excavators. In these applications, high strength steel is used in order to reduce weight while simultaneously increasing load capacity of the structure. According to the manufacturer information, Hardox steels are defined as "high-quality abrasion-resistant steels". They are characterized by high resistance to abrasive wear, the possibility of specialized machining tools, good weldability, excellent mechanical properties and resistance to impact loads. Hardox steels are produced in six types. Getting to know properties of the materials is meant to answer the question whether it is possible to substitute traditional steel with high strength steel, resistant to wear and easily cold-shaped fine-grained steel. The article presents preliminary fatigue tests results of two constructional steel grades used in structural elements of trailers, wagons, farm devices, etc. [5 and 6].

2. Fatigue tests

Fatigue characteristics (dependence $\sigma_a = f(N)$) of steel were determined at low testing frequency. The test was performed using the Rotoflex device which ensures loading specimens so that bending moment is constant over the entire working length of the specimen [7]. The correlations between stress amplitude per number of cycles until fracture (in S - N relation) and fatigue

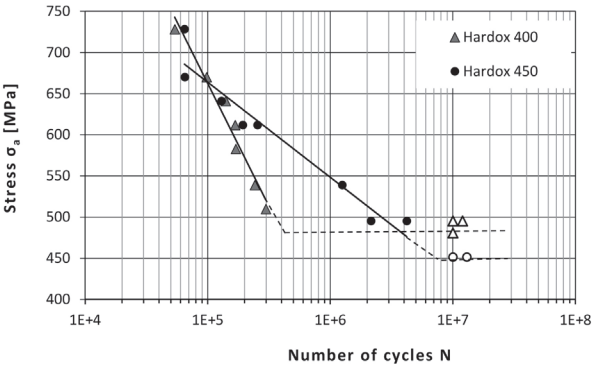


Fig. 1 Diagram of dependency between the stress amplitude and number of cycles of the tested steels

* [1]Pawel Szataniak, [2]Magdalena Mazur, [2]Robert Ulewicz, [3]Frantisek Novy
[1]WIELTON S. A., Wielun, Poland
[2]Institute of Production Engineering, Faculty of Management, Czestochowa University of Technology, Poland
[3]Department of Materials Engineering, Faculty of Mechanical Engineering, University of Zilina, Slovakia
E-mail: mazur.m@zim.pcz.pl

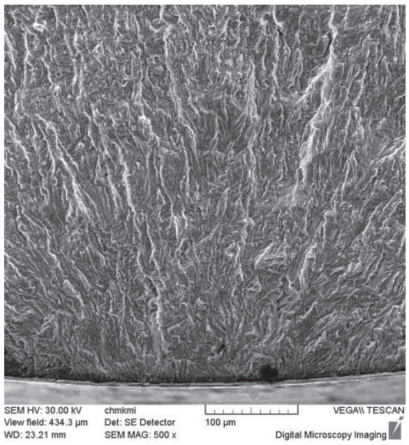


Fig. 2 Surface fatigue crack initiation for Hardox 400; σ_a = 330 MPa, N = 1.55·10^9 number of cycles

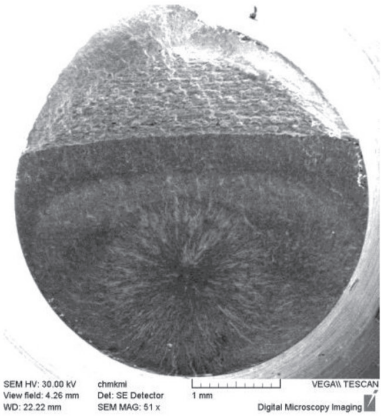


Fig. 3 Fatigue fracture Hardox 450 steel in the macro scale; subsurface fatigue crack initiation with structure "fish-eye"; σ_a = 455 MPa, N = 6.32·10^9 number of cycles

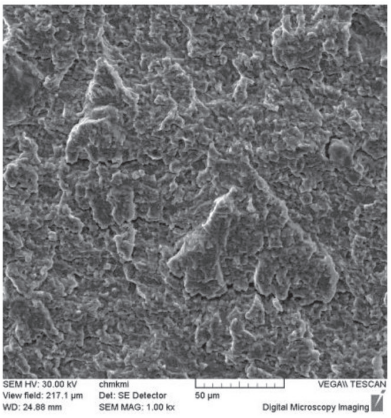


Fig. 4 Fatigue fracture surface, area of stable crack propagation for Hardox 450; σ_a = 524 MPa, N = 1.5·10^5 number of cycles

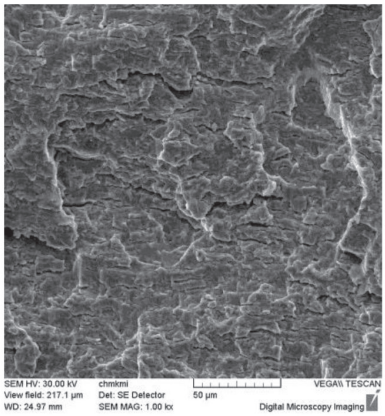


Fig. 5 Area of *unstable fatigue crack growth before rupture area with a clear prevalence of large secondary cracks and striations; Hardox 450 steel σ_a = 524 MPa, N = 1.5·10^5 number of cycles*

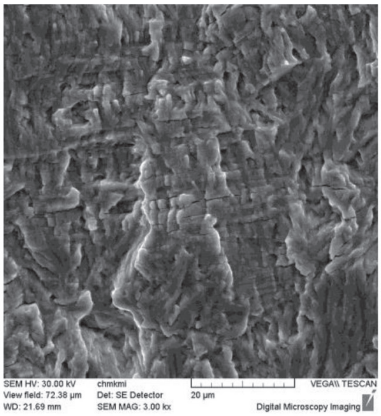


Fig. 6 Striations in the transcrystalline fatigue fracture of specimens Hardox 450 steel; σ_a = 600 MPa, N = 1.74·10^6 number of cycles

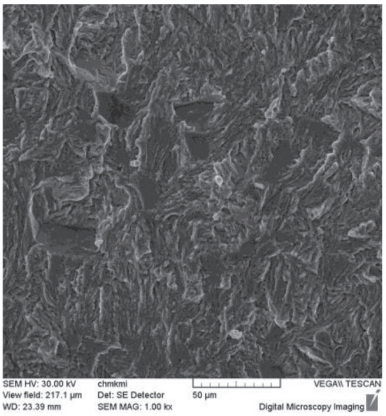


Fig. 7 Mixed character of trans- and inter crystalline fatigue fracture in fatigue failure of Hardox 450 steel; σ_a = 480 MPa, N = 7.28·10^8 number of cycles

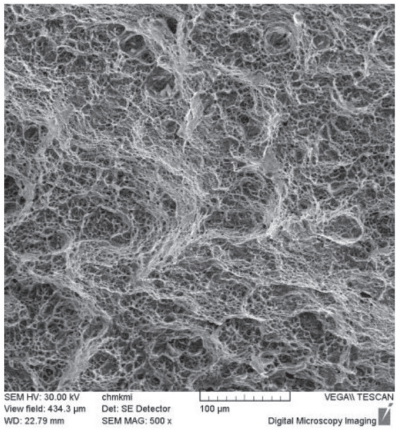


Fig. 8 Transcrystalline ductile fracture with dimple morphology in steel Hardox 400; σ_a = 330 MPa, N = 1.55·10^9 number of cycles

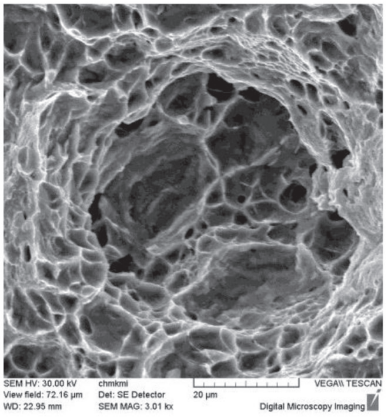


Fig. 9 Transcrystalline ductile fracture with dimple morphology in steel Hardox 400; σ_a = 330 MPa, N = 1.55·10^9 number of cycles

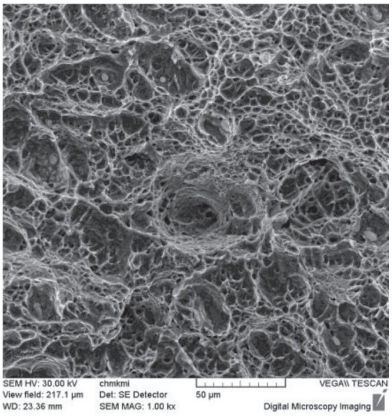


Fig. 10 Transcrystalline ductile fracture with dimple morphology in steel Hardox 450; σ_a = 480 MPa, N = 7.24·10^8 number of cycles

resistance are the main parameters for assessment of fatigue properties of structural steel shown in Fig. 1. The shape of the curve is characteristic for the Wohler curve.

The loading system was rotating bending equipment operating at the frequency of 40 Hz. The stress ratio of R = -1 was chosen. Both experiments were performed at the ambient temperature. During the test the working part of the specimens was cooled by means of fans [8]. The tests results formed a dependency curve between the amplitude of applied stress and number of cycles to the specimen crack, $\sigma_a = f(N)$ (Fig. 1). The tests results (for Hardox steels) formed a curve, which clearly shows that the stress amplitude σ_a decreases together with increase of cycles number N beyond conventional fatigue limit $N_c = 10^7$ cycles. The fatigue study showed that in the case of Hardox 400, fatigue limit was 490 MPa. In the case of Hardox 450 steel the fatigue limit was 460 MPa, at the $N = 10^7$ number of cycles. The results are close to each other, the difference is due to the higher strength and toughness properties of steel Hardox 450.

3. Fractographic analysis for the crack areas

In order to learn about the nature of fatigue process of the tested materials fractographic analysis for fatigue fractures was performed [9]. The tests were performed on a scanning electron microscope (SEM) Tescan II.

Test results for both materials Hardox 400 and 450 (Fig. 2) confirm the occurrence of only the surface fatigue crack initiation. Only two samples of steel Hardox 450 showed fatigue crack initiation point below the surface of the sample. In these cases, the fatigue fractures showed characteristic formations known in the literature by the term "fish eyes" (Fig. 3).

On the fatigue fracture surface of Hardox 450 specimens (Fig. 4) there are transcrystalline fatigue fractures of very fine particle morphology. In the area of unstable fatigue fracture just before rupture area there are many degrees of radial and secondary cracks (Fig. 5).

Transcrystalline fatigue fracture of tempering martensite was characterized for both analyzed Hardox steels. Analysis of both Hardox steels revealed transcrystalline fatigue failure of tempered martensite occurring locally in the form of intercrystalline facets present on the surface (Fig. 7). A transcrystalline propagation mechanism was observed in the area of stable fatigue crack propagation, reflected by presence of striations (Fig. 6).

Areas of rupture for two tested materials are characterized by ductile fatigue fracture with dimple morphology (Figs. 8 - 10).

4. Conclusion

The research demonstrates the possibility of using fine-grained Hardox 400 and 450 steels in building of semi-trailers. The use of wear-resistant materials extends the life and also reduces the weight of a trailer thanks to thinner sheets used for the construction of trailers. One of the key parameters is to ensure uniformity of properties throughout the cross-section of the sheet.

Results of this study confirm the classic shape of the Wöhler curve (function of dependencies $\sigma_a = f(N)$). The obtained graph clearly indicates the fatigue limit for all two tested materials. Fatigue tests for high cyclic fatigue, for tested high-strength steels showed a clear fatigue limit: in the case of Hardox 400, σ_c = 490 MPa and σ_c = 460 MPa in the case of steel Hardox 450.

The fractographic analysis showed that initiation of fatigue cracks was present on the surface of almost all specimens, which was also influenced by their loading. Only in the case of two specimens of Hardox 450, subsurface fatigue crack initiation was observed in the characteristic form of a "fish-eye" structure. However, it was impossible to determine the type of inclusions, which has an impact on the formation of the "fish-eye" structure. This was due to opening and closing of a fatigue crack in a very high number of cycles, which led to total fragmentation of particle inclusions.

In the case of Hardox steels, the share of intercrystalline fractures in the total fatigue fracture surface generally does not exceed 1%. Fatigue crack growth areas were characterized by striations that were oriented perpendicularly to the direction of fatigue crack propagation. The occurrence of fatigue striations is characteristic for the fatigue fracture of steels. Extending distances between individual striations can be observed on the fatigue fracture surface, running through the rupture area.

References

[1] VECHET, S., KOHOUT, J., BOKUVKA, O.: *Fatigue Properties of Cast Irons* (in Slovak), EDIS : University of Zilina, 2001.

[2] WYRZYKOWSKI J. W., PLESZAKOW E., SIENIAWSKI J.: *Deformation and Cracking of Metal* (in Polish), Wydawnictwo Naukowo-Techniczne, 1999.

[3] BURSAK, M., BOKUVKA, O.: Influence of Technological Factors on Fatigue Properties of Steel Sheets, *Communications - Scientific Letters of the University of Zilina*, vol. 8, No. 4, 2006, pp. 34-37.

[4] BURSAK, M., BOKUVKA, O.: Fatigue Properties of Steel with Increased Atmospheric Corrosion Resistance, *Communications - Scientific Letters of the University of Zilina*, vol. 11, No. 1, 2009, pp. 27-30.

[5] MAZUR, M., SZATANIAK, P., NOVY, F., ULEWICZ, R.: *Fatigue Properties of Fine-Grained Steel,* SEMDOK 2013, 18th Intern. seminar of PhD. Student's, Terchova, EDIS : University of Zilina, 2013.
[6] MAZUR, M., SZATANIAK, P., NOVY, F., ULEWICZ R.: *Fatigue Properties of Hardox 450 Steel,* SEMDOK 2014, 19th Intern. seminar of PhD. Student's, Terchova, EDIS : University of Zilina, 2014.
[7] TRSKO, L., MIKOVA, K., BOKUVKA, O.: *Fatigue Resistance of Shore Microalloyed Steels for Gas Transport,* Toyotarity, Heijunka, Yurii V. Makovetsky, Dnipropetrovsk, 2011.
[8] MAZUR, M., ULEWICZ, R., SZATANIAK, P.: *Fatigue Properties Selected Types of Steel Used in the Construction of Semi-trailers* (in Polish), Prace XL Szkoly Inzynierii Materialowej. Wydawnictwo Naukowe AKAPIT : Krakow, 2012.
[9] ZEMANDL, M.: *Fractography and its Practical Application in the Analysis of the Causes of Machine Part Fatigue* (in Czech), Letna skola unavy materialov, EDIS : University of Zilina, 2006.

COMMUNICATIONS – Scientific Letters of the University of Zilina
Writer's Guidelines

1. Submitted papers must be unpublished and must not be currently under review for any other publication.
2. Submitted manuscripts should not exceed 8 pages including figures and graphs (in Microsoft WORD – format A4, Times Roman size 12, page margins 2.5 cm).
3. Manuscripts written in good English must include abstract and keywords also written in English. The abstract should not exceed 10 lines.
4. Submission should be sent by e-mail – as an attachment – to one of the following addresses: komunikacie@uniza.sk or holesa@uniza.sk (or on CD to the following address: Zilinska univerzita, OVaV – Komunikacie, Univerzitna 1, SK – 010 26 Zilina, Slovakia).
5. Uncommon abbreviations must be defined the first time they are used in the text.
6. Figures, graphs and diagrams, if not processed in Microsoft WORD, must be sent in electronic form (as JPG, GIF, TIF, TTF or BMP files) or drawn in high contrast on white paper. Photographs for publication must be either contrastive or on a slide.
7. The numbered reference citation within text should be enclosed in square brackets. The reference list should appear at the end of the article (in compliance with ISO 690).
8. The numbered references (in square brackets), figures, tables and graphs must be also included in text – in numerical order.
9. The author's exact mailing address, full names, E-mail address, telephone or fax number, the name and address of the organization and workplace (also written in English) must be enclosed.
10. The editorial board will assess the submitted paper in its following session. If the manuscript is accepted for publication, it will be sent to peer review and language correction. After reviewing and incorporating the editor's comments, the final draft (before printing) will be sent to authors for final review and minor adjustments
11. Submission deadlines are: September 30, December 31, March 31 and June 30.

COMMUNICATIONS

SCIENTIFIC LETTERS OF THE UNIVERSITY OF ZILINA
VOLUME 18

Address of the editorial office:
Zilinská univerzita
Office for Science and Research
(OVaV)
Univerzitna 1
SK 010 26 Zilina
Slovakia

E-mail: komunikacie@uniza.sk

Each paper was reviewed by two reviewers.

Journal is excerpted in Compendex and Scopus.

It is published by the University of Zilina in
EDIS – Publishing Institution of Zilina University
Registered No: EV 3672/09
ISSN 1335-4205

Published quarterly

Single issues of the journal can be found on:
http://www.uniza.sk/komunikacie

ICO 00397 563
February 2016