

Impact of IPsec on Speech Quality

Technical Report x/2009

Miroslav Vozňák – Filip Řezáč

CESNET z.s.p.o., Žitkova 4, Praha, Czech Republic

miroslav.voznak@vsb.cz, filip.rezac@vsb.cz

8.12.2009

Abstract

This technical report deals with an analysis of voice over secure communication links based on IPsec. The security increases an overhead, hence requires a change in a bandwidth allocation. We deal with issues such as its calculation and the impact of packet loss and delay on speech quality. Such basic information describing the transmission path is important to enable to estimate the overall speech quality. The achieved results should help in network design and optimizations, as network operators need to maintain certain levels of service quality.

1 Introduction

First of all, we would like to point out that our research has started last year and that we have already published a technical report [1] and a paper [2] concerning the impact of security on speech quality in TLS environment. We brought a general method of bandwidth calculation and found out significant overhead in the implementation of OpenVPN that requires a double bandwidth in case of G.729 or G.723.1 codecs.

Unfortunately, Datagram Transport Layer Security [3] has not been implemented yet in OpenVPN. We expect that future versions of OpenVPN will tend to use DTLS as the security layer. DTLS is very similar to TLS and adds minimal overhead to the previously non-DTLS capable application. In the past, it was not necessary to focus on UDP protocol. Nowadays, a number of application layer protocols designed to use UDP transport are used, RTP being one of them. This is the reason why OpenVPN appears to be noneffective.

The first part of the report is devoted to the general relations with respect to bandwidth requirements. The second and third part focus on SRTP and IPsec. The results achieved are summarized in tables and are compared with pure RTP. The following chapter describes our contribution to the computational model of speech quality. The last part of this report consists of a brief summary and an acknowledgement.

2 Bandwidth requirements

The relations which were presented in the technical report [1] take into account factors such as codec used, timing, payload size, number of concurrent calls and so on. The basic steps of speech processing on the transmission side are encoding and packetization. RTP packets are sent at given times and the difference between two consecutive packets depends of the timing variable. The total bandwidth BW_M can be determined as:

$$BW_M = M \cdot C_R \cdot \left(1 + \frac{H_{RTP} + \sum_{j=1}^3 H_j}{P_S} \right) \quad (1)$$

Variable M represents the number of concurrent calls, P_s [b] payload size and C_R [bps] codec rate. Besides H_{RTP} including a packet header at the application layer there is the sum of lower located headers of the OSI model where H_1 [b] is media access layer header, H_2 [b] Internet layer header and H_3 [b] is transport layer header. Figure 1 presents bandwidth as a function of timing and number of concurrent calls for G.729 codec. If we want to calculate the required bandwidth, we first need to determine contribution at particular layers.

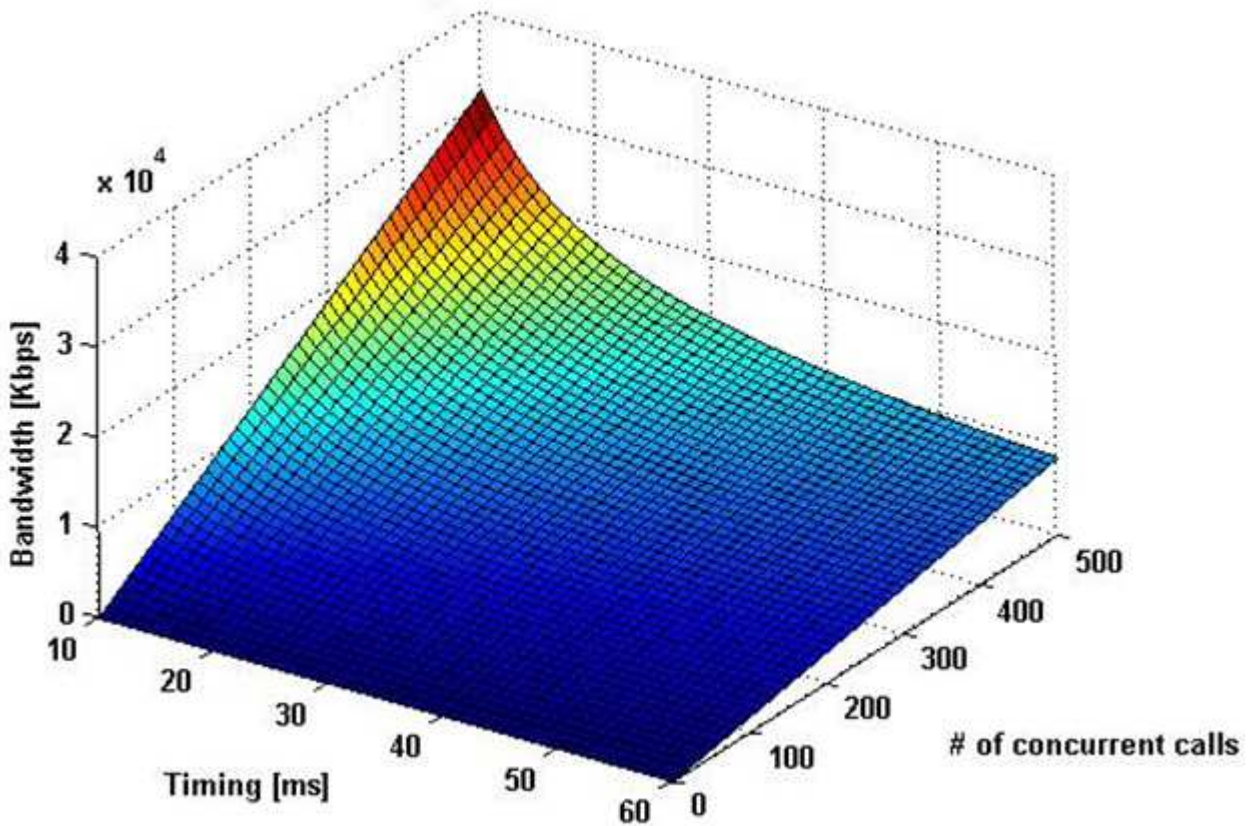


Figure 1: Bandwidth as function of timing and number of concurrent calls for G.729 codec

3 Secure RTP

SRTP is a transport protocol providing confidentiality and message authentication. Its specification is stated in recommendation RFC 3711 that provides a framework for the encryption and message authentication of RTP and RTCP streams [4]. SRTP defines a set of cryptographic transforms which are based on an additive stream cipher for encryption and a keyed-hash based function for message authentication. The format of an SRTP packet is described in RFC 3711 on page 5. When comparing it with the format of RTP we can claim that SRTP is a profile of RTP, the components of header are the same, only the payload is encrypted and the last item contains an authentication tag. This tag is not mandatory but is recommended, the authentication algorithm HMAC SHA-1 protects the integrity of the entire original RTP packet as is shown in figure 2.

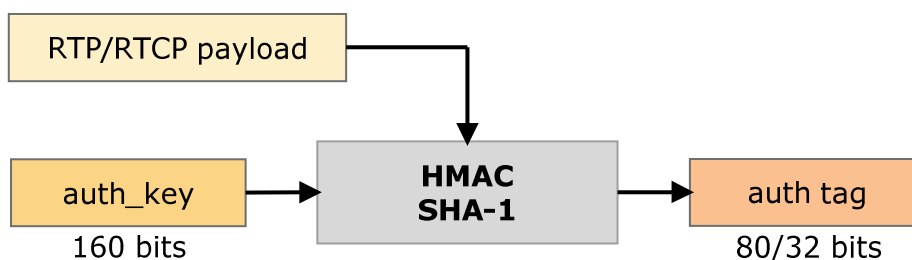
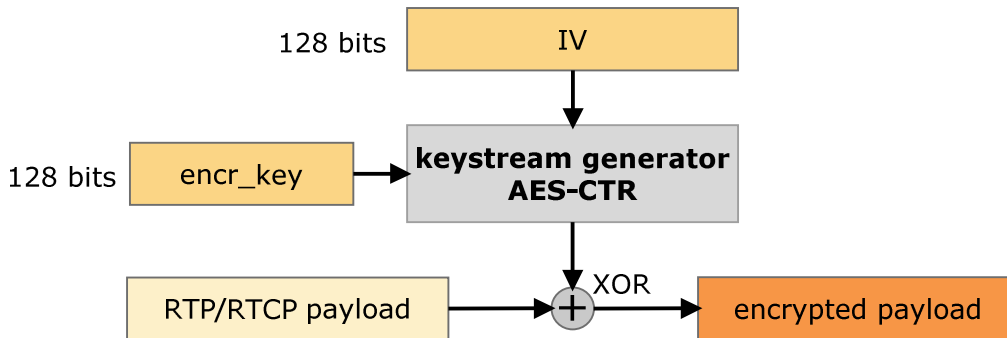


Figure 2: Authentication tag in SRTP

The authentication tag is the only SRTP addition to the original RTP packet containing usually either 4 or 10

bytes. Authentication ensures that attackers can neither modify packets in the stream nor insert additional information. The authentication operation is performed after the encryption operation and protects the entire RTP packet. RTP and SRTP payload sizes match exactly. The default cipher is the AES in two modes of running. The first is a f8-mode (used in UMTS) and the latter is a counter mode AES-CTR commonly used in SRTP, depicted in figure 3.



$$IV = f(\text{salt_key}, \text{SSRC}, \text{packet index})$$

112 bits

Figure 3: Payload encryption in AES-CTR mode

Initialization vector IV is obtained from following expression 2:

$$IV = (k_s \cdot 2^{16}) \oplus (\text{SSRC} \cdot 2^{64}) \oplus (i \cdot 2^{16}) \quad (2)$$

SSRC identifies the synchronization source. The value is chosen randomly, with the intent that no two synchronization sources within the same RTP session will have the same SSRC. The parameter k_s represents a salt key and i is an index of SRTP packet on the sender side, its value is incremented with every SRTP packet sent. The encryption key encr_key and others, such as salt_key and auth_key , are achieved with a key derivation function. These respective security keys are securely derived from the master key. It means that a single master key provides keying material for confidentiality and integrity of SRTP stream. Now, we apply the relation (1) to the SRTP with 10 bytes authentication header and obtain the results depicted in table 1.

Table 1. Comparison of RTP and SRTP bandwidth requirements

Codec	Bit Rate [kbps]	Payload Size [bytes]	RTP bandwidth in Ethernet [kbps]	SRTP bandwidth in Ethernet [kbps]
G.711 PCM	64	160	90.4	94.4
G.729 CS-ACLEP	8	20	34.4	38.4
G.723.1 ACELP	5.3	20	22.9	25.6

A new protocol in the field of media security is ZRTP (Zimmermann RTP). It consists of security enhancement for SRTP [5]. ZRTP suggests a protocol for media path Diffie-Hellman exchange to agree on a session key and parameters for establishing SRTP sessions. The ZRTP protocol is media path keying because it is multiplexed on the same port as RTP and does not require any support in the signaling protocol. ZRTP does not assume a Public Key Infrastructure (PKI) or require the complexity of certificates in end devices. For the media session, ZRTP provides confidentiality, protection against man-in-the-middle

(MiTM).

4 IPsec

IPsec is a suite of protocols for securing IP communications by authenticating and encrypting each IP packet of a data stream. IPsec also includes protocols for establishing mutual authentication between agents at the beginning of the session and negotiation of cryptographic keys to be used during the session.

The baseline IPsec architecture and fundamental components of IPsec are defined in RFC2401. Among all IPsec protocols, there are two specific protocols to provide traffic security:

- Authentication Header (AH),
- Encapsulating Security Payload (ESP).

AH provides connectionless integrity, data authentication and optional replay protection but, unlike ESP, it does not provide confidentiality, so AH is used to authenticate but not encrypt IP traffic. Consequently, it has a much simpler header than ESP. Authentication is performed by computing a cryptographic hash-based message authentication code over nearly all the fields of the IP packet, and stores this in a newly-added AH header and sends it to the other end.

ESP provides confidentiality, data integrity, and optional data origin authentication and anti-replay services. It provides these services by encrypting the original payload and encapsulating the packet between a header and a trailer.

IPsec may operate in two distinct ways, the transport and tunnel mode, depending upon whether the secure communication is between two endpoints directly connected or between two intermediate gateways to which the two endpoints are connected.

In transport mode, an IPsec header (AH or ESP) is inserted between the IP header and the upper layer protocol header. In this mode, the IP header is the same as that of the original IP packet except for the IP protocol field and the IP header checksum, which is recalculated. In this mode, the destination IP address in the IP header is not changed by the source IPsec endpoint. In tunnel mode, the original IP packet is encapsulated in another IP datagram and an IPsec header (AH or ESP) is inserted between the outer and inner headers.

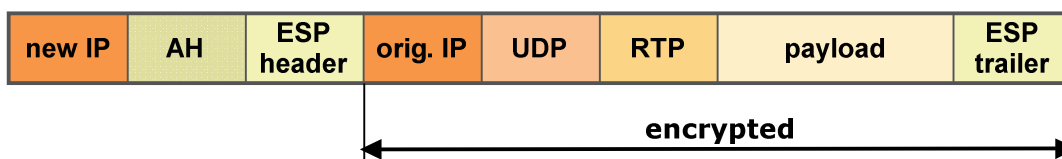


Figure 4. RTP packet in IPsec tunnel mode (ESP+AH)

For the calculation of bandwidth (1) to an IPsec scenario it is necessary to resize the payload size and to add the total contribution of the various headers of the distinct layers. An exact mathematical expression is possible but it is rather complex. We suggested an approximation based on empirical determination of constant for IP header H_{IPsec} which represents H_2 in sum $\sum_{j=1}^3 H_j$ applied to the basic expression (1). We have proved the validity of this approach in an experiment with OpenSwan.

Openswan is an Open Source implementation of IPsec for the Linux operating system. It is a continuation of former project FreeS/WAN. Empirically it was established that the calculation can be considerably simplified and still provide relevant results. The values below were achieved in measurements repeated three hundred times in IPsec environment using AES cipher (ESP tunnel mode included in AH, fig. 4). The values were designed for three codecs under the same conditions as in table 1:

- G.711, $H_{IPsec}=704$ b
- G.729, $H_{IPsec}=672$ b
- G.723.1, $H_{IPsec}=672$ b

We apply the values to relation (1), the results are depicted in table 2.

Table 2. Comparison of RTP and encapsulated RTP in IPsec

Codec	Bit Rate [kbps]	Payload Size [bytes]	RTP bandwidth in Ethernet [kbps]	RTP-IPsec bandwidth in Ethernet [kbps]
G.711 PCM	64	160	90.4	117.6
G.729 CS-ACLEP	8	20	34.4	60
G.723.1 ACELP	5.3	20	22.9	40

The results were verified in a testbed based on OpenSwan, on a total of 300 measurements. We determined the standard deviation δ , for G.711 $\delta=0.51$, for G.729 $\delta=0.59$, for G.723.1 $\delta=0.5$ and we can claim that the chosen approach provided the results with a high accuracy. The standard deviation is the root mean square deviation of its values from the mean.

$$\sigma = \sqrt{\frac{1}{N} \sum_i^N (x_i - \bar{x})^2} \quad (3)$$

where \bar{x} is the arithmetic mean of the values x_i and N is the number of elements.

5 R-factor calculation

Nowadays, the so-called E-Model is the preferred new approach. It has been described in ITU-T G.107 [10], updated in April 2009. It is mainly empirical by nature and was developed on the basis of large amounts of auditory test data (both Listening Only Tests and Conversation Tests).

The E-model takes into consideration the impairments due to talker echo and transmission delay and can hence be applied to predict quality in conversation. It was compiled as a combination of different quality prediction models in the framework of the ETSI (European Telecommunications Standards Institute). The primary output of the E-model is the Rating Factor, R , also known as the R-factor. The R-factor is calculated based on adding assumed impairment factors in a transmission path, and is a result of an algorithm based on 21 parameters related to the terminal factor, environment factor, Network factor and so on. The computational tool was implemented in Java in accordance with the E-model and can be downloaded [6], including its source code, figure 5.

PARAMETR	ZKRATKA	VÝCHOZÍ	OMEZENÍ	HODNOTA	JEDNOTKA
Send loudness rating	SLR	(8)	(0 ÷ 18)	<input type="text" value="8.0"/>	[dB]
Receive loudness rating	RLR	(2)	(-5 ÷ 14)	<input type="text" value="2.0"/>	[dB]
Talker echo loudness rating	TELR	(65)	(5 ÷ 65)	<input type="text" value="65.0"/>	[dB]
Electric circuit noise	Nc	(-70)	(-80 ÷ -40)	<input type="text" value="-70.0"/>	[dBm0p]
Noise floor	Nfor	(-64)	-	<input type="text" value="-64.0"/>	[dBmp]
Room noise (send)	Ps	(35)	(35 ÷ 85)	<input type="text" value="35.0"/>	[dB(A)]
Room noise (receive)	Pr	(35)	(35 ÷ 85)	<input type="text" value="35.0"/>	[dB(A)]
Sidetone masking rating	STMR	(15)	(10 ÷ 20)	<input type="text" value="15.0"/>	[dB]
D-factor (send)	Ds	(3)	(-3 ÷ 3)	<input type="text" value="3.0"/>	
D-factor (receive)	Dr	(3)	(-3 ÷ 3)	<input type="text" value="3.0"/>	
Listener's sidetone rating	LSTR	STMR + Dr	-	<input type="text" value="18.0"/>	[dB]
Mean one-way delay	T	(0)	(0 ÷ 500)	<input type="text" value="0.0"/>	[ms]
Absolute delay	Ta	= T	-	<input type="text" value="0.0"/>	[ms]
Round-trip delay	Tr	= 2T	-	<input type="text" value="0.0"/>	[ms]
Weighted echo path loss	WEPL	(110)	(5 ÷ 110)	<input type="text" value="110.0"/>	[dB]
Quantizing distortion units	qdu	(1)	(1 ÷ 14)	<input type="text" value="1.0"/>	
Equipment impairment factor	le	(0)	(0 ÷ 40)	<input type="text" value="CS-ACELP - G.729 (8 kbit/s) - 10"/>	
Packet-loss robustness factor	Bpl	(1)	(1 ÷ 40)	<input type="text" value="1.0"/>	
Packet-loss probability	Ppl	(0)	(0 ÷ 20)	<input type="text" value="0.0"/>	[%]
Burst ratio	BurstR	(1)	(1 ÷ 8)	<input type="text" value="1.0"/>	
Advantage factor	A	(0)	(0 ÷ 20)	<input type="text" value="Pevný terminál - 0"/>	
<input type="button" value="Výpočet"/>					
Factor R	R			<input type="text" value="83,21"/>	÷
Mean opinion score	MOS-CQE			<input type="text" value="4,14"/>	÷

Figure 5. The computational tool in accordance with E-model

We can claim that the progress in the field of the automatic gain control and the echo cancellation reached high level. If we only considered the impact of IP network, the calculation could be significantly simplified. The modified E-model was designed with the knowledge of VoIP systems and includes terms for codec, delay and packet loss, depicted in figure 6.

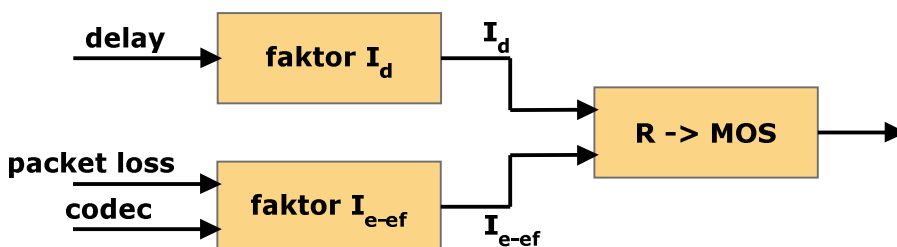


Figure 6. Simplified E-model

This makes it more easy to apply to VoIP systems than the other models. Following relations represent a modified E-model in the figure above. The impact factors I_s and A are suppressed in this simplified E-model, the variables and principles used are explained in another technical report [1] issued by CESNET last year.

$$R = R_0 - I_S - I_D - I_{E-EF} + A = 93.3553 - I_D - I_{E-EF} \quad (4)$$

Our new contribution concerning this field this year is the regress function describing the dependability of factor I_d on end to end delay. This regression can be used up to E2E delay 400 ms. We utilized the values achieved in measurements in AT&T laboratory [7].

$$I_d = \begin{cases} 0,0267 \cdot T & T < 175\text{ms} \\ 0,1194 \cdot T - 15,876 & 175\text{ms} \leq T \leq 400\text{ms} \end{cases} \quad (5)$$

Pearson's correlation coefficient between values achieved by measurements and the calculation in proposed regression $r=0.99$. A value of 1 implies that the equation describes the relationship perfectly.

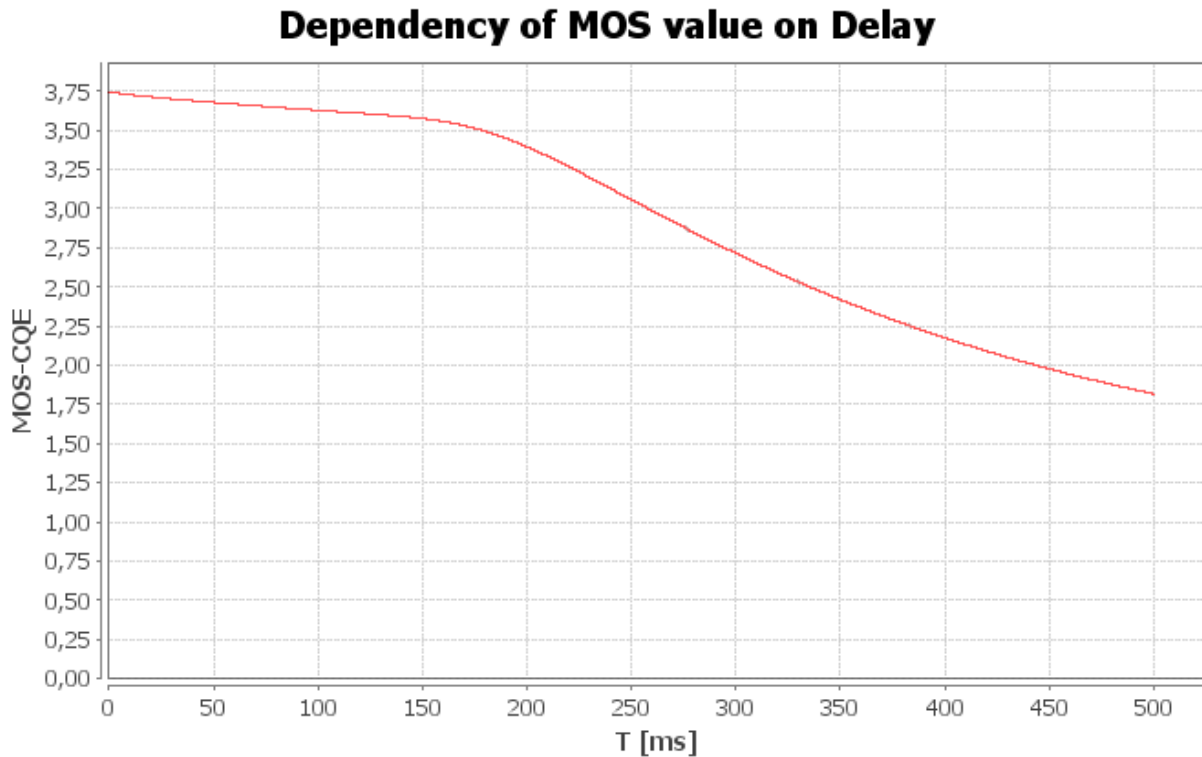


Fig. 7. Dependency of MOS on Delay, a validity for codec G.729

The simplified version of the E-model was also implemented in Java and can be downloaded [8] including its source code. The tool provides not only the result, i.e. the value of R-factor or MOS, but also a graph showing the dependency of the estimated speech quality on the delay or packet loss, figure 7.

6 Conclusion

This report is an extension of the previous work on the impact of security on the quality of VoIP calls in TLS environment [1]. The new contribution is the presented method of bandwidth calculation in the network using IPsec, the validity of general relation (1) that has also been proven for SRTP and finally the contribution to the computational model of speech quality, partly by the new regression for factor I_d and partly by the computational tool programmed in Java in accordance with recommendation ITU-T G.107 [9], [10], [11].

7 Acknowledgements

We would like to thank Saul Parada from University of Aveiro who spent one semester in the IP telephony lab in Ostrava in 2009 and carried out a lot of measurements for our research, and Martin Tomeš, a student

of a bachelor's degree at VŠB-Technical University of Ostrava.

This project has been supported by the "Optical Network of National Research and Its New Applications" (MŠM 6383917201) research intent. The report includes the results achieved by authors in the research activity Multimedia transmissions and collaborative environment, <http://www.ces.net/project/15/>.

8 References

- [1] VOZŇÁK, M. *Impact of security on speech quality*. CESNET: Technical Report 13/2008, December 2008
URL: <http://www.cesnet.cz/doc/techzpravy/2008/impact-of-network-security-on-speech-quality/>
- [2] VOZŇÁK, M. *Speech Bandwidth Requirements in IPsec and TLS Environment*. In Proceedings of the 13th WSEAS International Conference on Computers, p.217-220. Rodos, Greece, July 23-25.2009. ISBN 978-960-474-099-4
- [3] RESCOLA, E. *RFC 4347 : Datagram Transport Layer Security*. IETF, April 2006. URL: <http://www.rfc-editor.org/rfc/rfc4347.txt>
- [4] BAUGHER, M., and et. al. *RFC 3711: The Secure Real-time Transport Protocol (SRTP)*. IETF, March 2004. URL: <http://www.rfc-editor.org/rfc/rfc3711.txt>
- [5] ZIMMERMANN, P. and et.al. *ZRTP: Media Path Key Agreement for Secure RTP*. IETF: November 2009.
URL: <http://tools.ietf.org/html/draft-zimmermann-avt-zrtp-16>
- [6] TOMEŠ, M., VOZŇÁK, M. E-model in JAVA application. Ostrava, 2009. URL:
<http://homel.vsb.cz/~voz29/Emodel1.0.zip>
- [7] COLE, R., ROSENBLUTH. *Voice over IP performance monitoring*. ACM SIGCOMM Computer Communication, New York, 2001
- [8] TOMEŠ, M., VOZŇÁK, M. E-model in JAVA application. Ostrava, 2009. URL:
<http://homel.vsb.cz/~voz29/Emodel1.2.zip>
- [9] ITU-T Recommendation G.113. *Transmission impairments due to speech processing*, ITU: Geneva, 2007.
URL: <http://www.itu.int/ITU-T/>
- [10] ITU Recommendation G.107. *E-model, a computational model for use in transmission planning*. ITU, 2009. URL: <http://www.itu.int/ITU-T/>
- [11] HARDY, W. *VoIP service quality*. McGraw-Hill, 2003, New York, ISBN 0-07-141076-7.