Paper

Authentication in VoIP Telephony with Use of the Echo Hiding Method

Jakub Rachoń, Zbigniew Piotrowski, and Piotr Gajewski

Military University of Technology, Warsaw, Poland

Abstract—The paper describes the method intended to authenticate identity of a VoIP subscriber with use of the data hiding technique that is specifically implemented by means of the echo hiding method. The scope includes presentation of experimental results related to transmission of information via a hidden channel with use of the SIP/SDP signalling protocol as well as results of subjective assessment on quality of a signal with an embedded watermark.

Keywords—authentication, BS 1116-1 test, digital watermarking, echo hiding.

1. Introduction

Nowadays, when the VoIP technology is being rapidly developed, the problem of subscriber authentication is appearing as the more and more important problem. The existing threats to safety of telephone connections [1], [2] have led to the need to seek for alternative methods robust against intentional attacks. Encryption of dedicated phone calls is only a partial solution of the problem as most of PSTN calls that are currently made are incapable to cope with confidentiality of connections. There appeared an attempt to resolve the problem with use of the electronic appliance called personal trusted terminal (PTT) [3] that bases on an objective (numerical) verification of radio subscribers and that is also applicable to VoIP, GSM and PSTN networks. The concept associated with objective verification of telephone subscribers on telecom lines with use of the information hiding technique can be also applied to hidden authentication of subscribers for telephone and radio links [4]. The present study deals with the method of watermarking by means of the echo hiding technique. Alongside, results of studies on robustness of the watermark to variable conditions attributable to wide area networks (WAN) are presented. In addition, the emulation method for WANs link path conditions is also described, where the emulation is carried out with use of the VMware Player software and the Debian operating system that is derived from the Linux kernel. The mentioned software is free of charge and commonly available from Internet.

2. Echo Hiding Method

2.1. Embedder Design

The echo hiding method consists in filtering of the original signal where one of two filters with the finite impulse response is used. The two filters differ with the following parameters: response delay, rate of the response fading as well as number of delays that produce the echo (kernels). The specific filter is selected pursuant to the bit value of the transmitted watermark. The encryption method with use of the echo hiding technique is described with more details in [5] as well as in [6] and [7]. The applied algorithm implements the module of the detector that is capable to recognize sounding vowels and consonants, which makes it possible to incorporate the watermark to selected frames of the acoustic signal. For that purpose the average magnitude difference function (AMDF) is used. It is the method that for the first time was explained in [8]. Consequently, when information on structure of the applied detector of sounding consonants and vowels is unavailable it is infeasible to correctly detect the watermark on the receiver side.

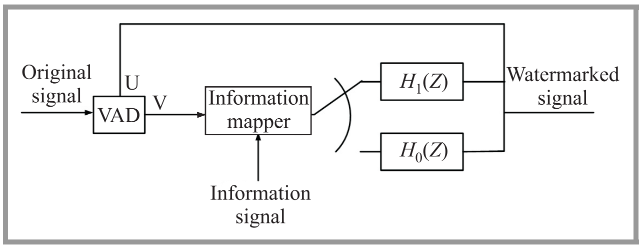

Fig. 1. Design of the watermark embedder.

In that way transmission of the watermark becomes hidden and adopts features of a confidential transmission as it is the case when the process of trials and errors is suitable to detect only selected fragments of the transmitted signal. Taking account for the fact that the information represented by the watermark is used only once for the specific phone connection session, it becomes much more difficult to fake identity of the subscriber. Design of the watermark embedder is shown in Fig. 1.

2.2. Design of the Watermark Embedder

Operation principle of the embedder consists in computation of the auto-cepstrum function intended to detect delay of the echo. It is the method that was described for the first time in [9]. The algorithms takes advantage of a voice activity detector (VAD) for sounding vowels and consonants as it is reasonable to find out only those fragments (frames) of the signal, where the watermark can be embedded. The same VAD is also used on the receiver side. The watermark extractor demands for more computation power than the embedder due to the reason that for each signal frame it must find out the auto-cepstrum function that is defined by the following equation:

$$f_d = \left\{ \mathrm{IFFT} \left[\log \left(\mathrm{FFT} \left(u_w(n) \right) \right)^2 \right] \right\}^2 , \tag{1}$$

where:
u_w – amplitude of subsequent samples within the frame,
FFT – fast Fourier transform,
IFFT – inversed fast Fourier transform

2.3. The Real-Time Mode Process of Watermark Embedding and Extraction

Owing to the fact that the algorithm for watermark embedding is relatively uncomplicated, it is feasible to implement it in the real-time mode to system platforms with low computation capacities. It makes possible to use the algorithms in such portable devices as PDA, mobile phones or other appliances, where, e.g., the Java virtual machine is installed. Due to more demanding requirements of the extractor, the stream of received bits must split into two paths, where the first part is forwarded to the D/A converter and then delivered to the loud-speaker (handset), whereas the second part arrives to the watermark extractor. Therefore, it is possible to embed and extract watermarks with no compromise to the voice quality.

3. Measurement Test Bed

In order to measure robustness of the watermark to variable conditions typical for WANs links, two virtual machines created within the VMware Player [9] were used. VMware Player is the software application that makes it possible to assign a part of hardware computer resources to establish an isolated architecture that enables to run any operating system. Furthermore, the virtual machines are capable to communicate by means of the IP protocol. To emulate WAN environment one machine called router was used to launch the packets forwarding service (IP forwarding) and then the tool called traffic control was applied to manage traffic of outgoing packets (Fig. 2). The router forwards packets to the virtual telephone exchanger PBX Asterisk. The traffic control tool allows emulation of the following parameters:

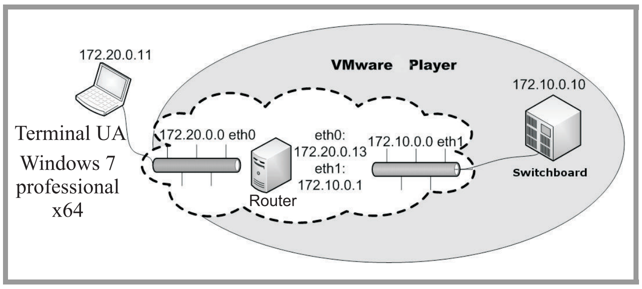

Fig. 2. WAN emulation.

- packet delay,
- packet loss,
- packet repeating,
- packet damage by random bit swapping,
- packet reordering.

The foregoing configuration of the test bed made it possible to carry out the following experiment. The stream of speech signal was redirected to the subscriber's voice mailbox, whereas the stream of RTP packets was controlled by the router, where traffic of outgoing packets was delayed in accordance with the Gaussian distribution with the constant mean value of 100 ms and increasing standard deviation across the *jitter* experiment. The acoustic signal transmitted with use of the RTP protocol was encoded by means of the G.711 μ-Law codec. The experiments were carried out in the following way:

1. The binary signature of the watermark was embedded by means of the echo hiding algorithm with use of the Matlab environment.
2. The WAVE file containing the examined soundtrack was reproduced as a sound source for the SIPCLI software. It is the software tool that is used to establish connections when the SIP protocol is applied.
3. The recorded voice message stored on the voice mailbox was transferred to the local disk.
4. The file with the recorded message was decoded by the watermark extractor within the Matlab environment.

The connection between the client and the machine, where the VMware Player environment was launched, was established with use of the Ethernet cable UTP cat. 5 with the length of 1.2 m. The connection was handled by network cards operating according to the IEEE 802.3u standard (100Base-TX Fast Ethernet).

4. Measurement Results

Figure 3 presents comparison between quality of the watermark extraction depending on the type of the finite impulse response (FIR) filter applied to embed the echo. The experiments were carried out within a closed loop for the male English speech. The signal was sampled with the frequency of 8 kHz and 16-bit resolution. The d0 parameter stands for the echo delay expressed as a number of samples for the bit with the low logic level (0), whereas the d1 parameter is meant for the echo delay expressed as a number of samples for the bit with the high logic level (1). The echo fading factor is 0.4 for the both cases. The values of d0 were selected on the basis of the experimental study described in [10]. The presented graphs serve as the proof that any increase in number of echo kernels is not enough to substantially improve the detection efficiency. Only the filter with the preceding or delayed kernel (also known as bilateral kernel) significantly improves extraction quality. On the other hand, the watermark quality estiamtion tests demonstrate that such a solution considerably deteriorates the original signal and leads to distortions that make the watermark easily hearable. To reach a compromise between satisfying results of the watermark extraction and the watermark inaudibility, the following parameters of the watermark embedder were selected for the male English speech:

- d0 = 15 samples,
- d1–d0 = 5 samples,
- number of echo kernels = 1,
- echo fading factor = 0,4.

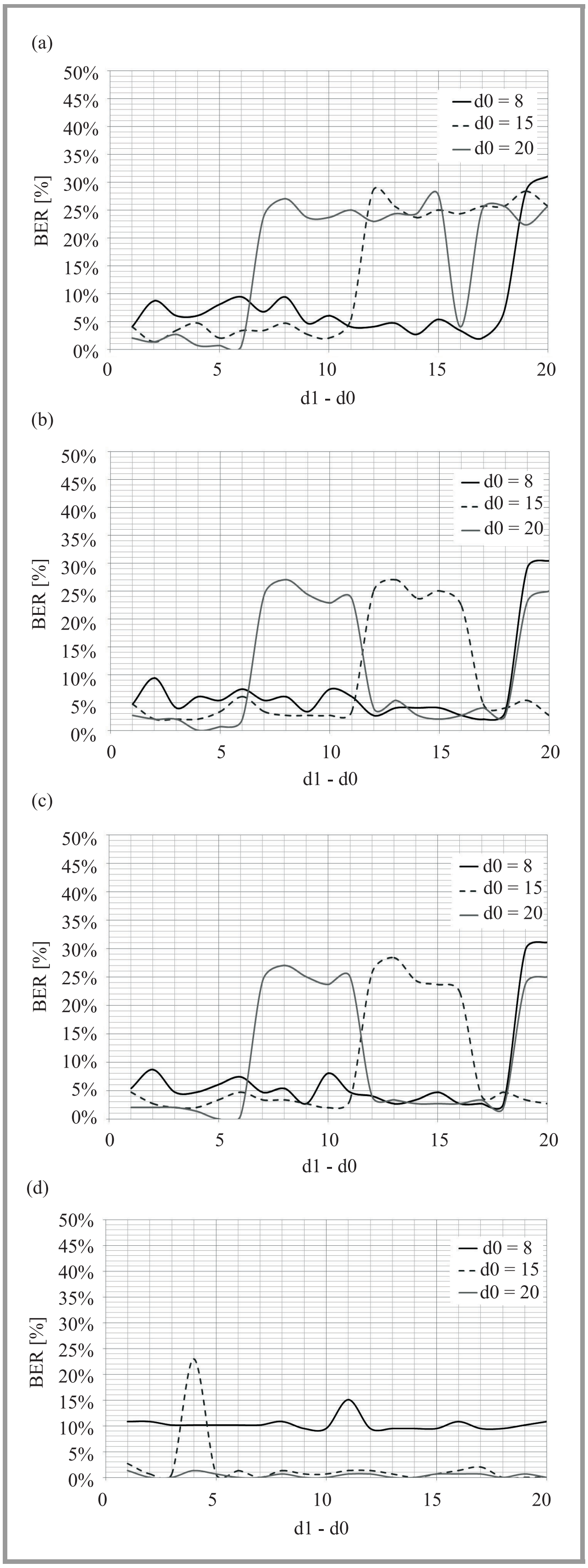

Fig. 3. Efficiency of detection as a function of echo delay and number of echo kernels: (a) = 1, (b) = 3, (c) = 5 for the English speech signal and for signature with 32 [b] payload (d) bilateral echo methode.

To guarantee correct detection of the signature secured by means of the error detecting and correcting code BCH, the elementary error rate must be below 10%. As one can see in Fig. 4 detection/extraction of the watermark is feasible when the standard deviation of packet delays ranges within the interval of 9 ms. It imposes the requirement to guarantee relative steady parameters of the line that was established for the needs of the voice connection to the SIP protocol. The average standard deviation for the packet delays is equivalent to the average value of the parameter that is referred to as *jitter*. For digital transmissions such as transmission with use of the RTP protocol the watermark is correctly extracted at the receiver side, even in case of conversion from the PCM format 16 bits per a sample to the G711 μ-Law 8 bits per a sample and the reverse conversion to the PCM format 16 bits/sample at the side of the PBX Asterisk switchboard. The approximate transmission watermark data payload is 7 bit/s. The following paragraph comprises statistical information on parameters of the RTP channel. The parameters have been determined with use of the RTP stream analysis software application incorporated

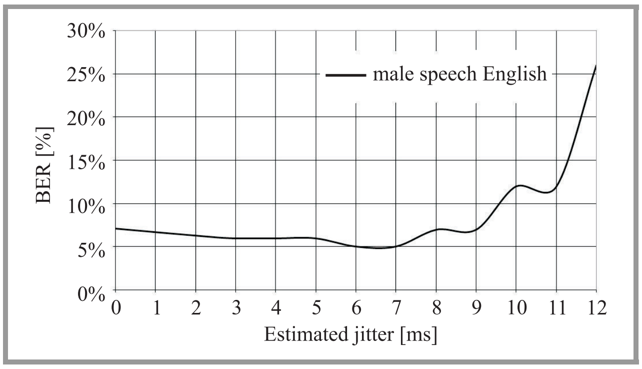

Fig. 4. Robustness of the watermark against jitter in the RTP channel, $fs = 8$ kHz

into the Wireshark package on the basis of data packets captured at the side of the PBX telephone switchboard.

```
Max delta = 66,16 ms at packet no. 1746
Max jitter = 15,23 ms.  Mean jitter = 10,19 ms.
Total RTP packets = 1150 (expected 1150)
Lost RTP packets = 0 (0,00%)
Sequence errors = 120
Duration 23 s
(-386 ms clock drift,
corresponding to 7868 Hz (-1,65%)
```

The above statistics have been found out for the connection with the emulated standard deviation of 9 ms. It turns out that for such circumstances the drift of phase angle amounting to –1,65% occurs. Therefore, the proposed method is also insensitive to the effect of phase angle drift, which is very common on telecom links.

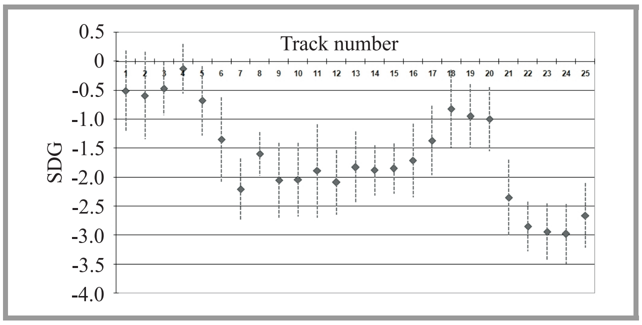

Fig. 5. Subjective fidelity assessment for the signal with the embedded watermark. The results are provided for the confidence interval of 95% and were determined with use of the ITU-R BS 1116-1 test. Details about sound tracks used in this test can be found in Table 1.

To subjectively assess fidelity of signals with embedded watermarks the test defined by the standard ITU R BS.1116-1 was carried out. The test was completed with participation of 10 listeners. Test results were subject to statistic computations with the confidence interval of 95% and then shown in Fig. 5. The SDG values above –1 mean that the watermark is not hearable. The obtained results enable to come out with the following conclusions:

- Authentication of SIP subscribers with use of watermarks is possible for networks where QoS is applied.
- Watermark signal is robust against variable delays of packets transmitted within the network.
- Increase both: echo signal numbers, as well as the echo scaling factor, makes the watermark better hearable on the background of the original signal.
- The subjective assessment how much embedded watermarks distort the original speech signal may vary with the language of conversation and gender of speakers.

The experimental results serve as the confirmation that authentication of subscribers who use VoIP telephony is possible when watermarks are embedded by means of the echo method and the original signal at the receiver side may remains unknown. It was also demonstrated that the proposed method is robust against conversion with use of the G.711 μ-Law codec. In addition, the experiments provided the proof that the method is robust against the phase angle drift (signal jitter) that commonly occurs in telecom channels. Therefore, the 'blind' extraction of watermarks is possible, i.e., the original signal at the receiver side is not necessary to correctly extract the binary signature that is represented by the embedded watermark.

Table 1
Description of sound tracks that were used to assess quality of the watermarking technique

Track number	Type of speech signal	Echo decay factor	Number of echo signals	BER [%]
1	French female	0.4	1	6.56
2	English female	0.4	1	5.71
3	English male	0.4	1	0.00
4	German male	0.4	1	5.56
5	English male	0.4	1	5.00
6	French female	0.8	1	1.64
7	English female	0.8	1	0.00
8	English male	0.8	1	0.00
9	German male	0.8	1	0.00
10	English male	0.8	1	1.67
11	French female	0.6	2	0.00
12	English female	0.6	2	0.00
13	English male	0.6	2	1.35
14	German male	0.6	2	2.78
15	English male	0.6	2	0.00
16	French female	0.4	3	4.92
17	English female	0.4	3	4.29
18	English male	0.4	3	1.35
19	German male	0.4	3	5.56
20	English male	0.4	3	6.67
21	French female	0.4	bilateral echo	3.33
22	English female	0.4	bilateral echo	20.29
23	English male	0.4	bilateral echo	1.35
24	German male	0.4	bilateral echo	1.39
25	English male	0.4	bilateral echo	0.00

5. Recommendations

The studies on implementation of watermark embedding with use of the echo hiding method serve as the evidence that the RTP channels set up for connections to both the H.323 and SIP protocols are suitable for transmission with watermarks embedded into voice signals.

References

[1] Z. Piotrowski and P. Gajewski, "Voice spoofing as an impersonation attack and the way of protection", *J. Inf. Assur. Secur.*, vol. 2, iss. 3, pp. 223–225, 2007.

[2] C. Roberts, "Voice over IP security", Center for Critical Infrastructure Protection, Wellington, New Zealand, March 2005.

[3] Z. Piotrowski, L. Zagoździński, P. Gajewski, and L. Nowosielski, "Handset with hidden authorization function", in *Proc. Eur. DSP Educ. Res. Symp. EDERS 2008*, Texas Instruments, pp. 201–205, 2008.

[4] Z. Piotrowski and P. Gajewski, "Novel method for watermarking system operating on the HF and VHF radio links", in *Computational Methods and Experimental Measurements XIII, CMEM XIII*, C. A. Brebbia and G. M. Carlomagnowit, Eds. Southampton, Boston: Wit Press, 2007, pp. 791–800.

[5] D. Gruhl, A. Lu, and W. Bender, "Echo hidding", Massachusetts Institute of Technology Media Laboratory, 1996, pp. 295–311.

[6] S. A. Chou and S. F. Hsieh, "An echo-hiding watermarking technique based on bilateral symmetric time spread kernel", in *Proc. IEEE ICASP 2006*, Toulouse, France, 2006.

[7] H. J. Kim "Audio watermarking techniques", in *Proc. Pacific Rim Workshop on Digital Steganography*, Kitakyushu, Japan, 2003.

[8] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg and H. J. Manley, "Average magnitude difference function pitch extractor", *IEEE Trans. Acoust., Speech and Sig. Proces.*, vol. 22, no. 5, pp. 353–362, 1974.

[9] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of time series for echoes: cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking", in *Proc. Symp. Time Series Anal.*, M. Rosenblatt, Ed., Chapter 15. New York: Wiley, 1963, pp. 209–243.

[10] Li Li, Ya-Qi Song, "Experimental research on parameter selection of echo hiding in voice", in *Proc. 8th Int. Conf. Machine Learn. Cybernet. ICMLC 2009* , Baoding, China, 2009, p. 2423–2426.

Jakub Rachoń received the M.Sc. in telecommunication systems from the Military University of Technology (MUT), Warsaw, in 2010. He spent one year at Ghent University in Belgium as an exchange student at faculty of engineering. His main area of interest are IT security, digital signal processing, mobile applications designing, intellectual property management and technology transfer processes.

e-mail: jakub.rachon@gmail.com
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Piotr Z. Gajewski received the M.Sc., and D.Sc. degrees from Military University of Technology (MUT) Warsaw, Poland in 1970, and 2001, respectively, both in telecommunication engineering. Since 1970 he has been working at Electronic Faculty of Military University of Technology (EF MUT) as a scientist and lecturer in communications systems (radios, cellular, microcellular), signal processing, adaptive techniques in communication and communications and information systems interoperability. He was an Associate Professor at Telecommunication System Institute of EF MUT from 1980 to 1990. From 1990 to 1993 he was Deputy Dean of EF MUT. Currently he is the Director of Telecommunication Institute of EF MUT. He is an author (co-author) of over 80 journal publications and conference papers as well as four monographs. He is a member of the IEEE Vehicular Technology and Communications Societies. He is also a founder member of the Polish Chapter of Armed Forces Communications and Electronics Association.

e-mail: Piotr.Gajewski@wat.edu.pl
Military University of Technology
Kaliskiego st 2
01-489 Warsaw, Poland

Zbigniew Piotrowski – for biography, see this issue, p. 16.