Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 7 (2012) pp. 147-155
© MIR Labs, www.mirlabs.net/jias/index.html

Digital Audio Watermarking Techniques for the security of VOIP Communications

Fusun C. Er[1] and Ensar Gul[2]

[1] PhER Technology, Kartal,
Istanbul, Turkey
fusun.er@pher.com.tr

[2] Dept. of Computer Engineering, Marmara University,
Goztepe Kampusu, Istanbul,Turkey
ensar.gul@marmara.edu.tr

***Abstract*: Digital Audio Watermarking techniques can be used as a lightweight security mechanism for source origin authentication in voice over IP (VOIP) systems. During the transportation phase of VOIP audio-marked voice packets are used as a source origin indicator. In this paper several audio watermarking algorithms were evaluated to demonstrate the applicability of the source authentication solution in terms of certain parameters such as: robustness, evaluation time, complexity and capacity. The effects of the audio watermarking were also measured using the Signal-to-Noise ratio and watermark extraction times. The evaluated VOIP codecs are: G711 a-law and u-law, GSM, G723.1 and ILBC.**

Keywords: VoIP Authentication, Audio Watermarking, Session Initiation protocol, voice codecs**.**

I. Introduction

VoIP is a real-time communication technology that enables voice conversations via the Internet. Fourth generation (4G) cell phone networks will be pure IP and SIP[1]; thus the 3GPP have chosen SIP as the protocol underlying many of the important interfaces between elements in a 4G network. Due to this factor securing VOIP has become an important topic. Some security techniques have been proposed for VOIP communications; however those that have been suggested to date include a trade-off between security and low latency for real time service. Recently, digital audio watermarking has been used for several purposes in VoIP. Mazurczyk et al [2] utilized digital audio watermarking techniques for FEC (Forward Error Correction). In [3], digital audio watermarking was used as an alternative data integrity measurement method against SRTP [4]. In [5] genetic algorithms were used to embed to higher LSB layers to increase the robustness against the attacks which attempts to reveal the hidden message. A methodology is proposed to design secure VOIP infrastructure in [6]. This methodology uses security policies to define security requirements. Security risks such as denial of service attacks and man in the middle attacks in VOIP systems are discussed and some solutions are proposed in [7]. Covert channels can also be used as a security threats in VOIP systems as described in [8]. Content-fragile watermarking and invertible watermarking approach were introduced for digital audio content authentication in [9]. Benchmarking for content based audio-watermarking are discussed in [10].

Digital audio watermarking used in source origin authentication, such a mechanism was implemented by combining SIP level key exchange as described in [11] and embedding the source origin indicator in transportation phase using digital audio watermarking as described in [2]. This paper is an updated and expanded version of the paper which was presented at IAS conference [12].

The next section presents VOIP systems and SIP. The implemented digital audio watermarking techniques are described in section 3; in addition capacities and complexities of algorithms are also mentioned in this section. In Section 4, experimental results are provided. Implemented digital audio watermarking algorithms are compared in terms of evaluation times, capacity against a-law encoding, SNR in dB and robustness. Finally, Section 5 concludes the paper.

II. Background

VoIP is a real-time technology that allows voice conversations via the Internet. VoIP communication is mainly structured by two phases: Signaling Phase and Conversation Phase.

In signaling phase, calling parties are authenticated and authorized to create, modify and terminate VoIP sessions. The 3GPP have chosen SIP as the signaling protocol in 4G IP Networks. SIP is an application-layer control protocol that works with both IPv4 and IPv6.

After establishing a connection between calling parties, the conversation phase started. The most frequently used transport protocol in the conversation phase is Real Time Protocol RTP [13], which provides end-to-end network transport functions suitable for applications that transmit real-time audio. RTP defines a profile for video or audio applications associated with payload formats.

VoIP is a real-time service that is needed to provide some QoS (Quality of Service) parameters, such as dropped packets, delay, jitter, latency, out of order delivery error.

Due to the importance of QoS parameters satisfaction during VoIP communication, many security mechanisms implemented in traditional data networks just are not applicable to VOIP in their current form. Most security mechanisms create high latency. Because of the time-critical nature of VOIP and its low tolerance for latency and packet loss, we are facing with trade-off between providing security and the low latency for VoIP.

Security concern can be classified regarding its compromise on confidentiality, integrity, or availability of the VoIP system. Some security mechanisms are dealing with confidentiality for media data. Confidentiality refers to the need to keep information secure and private and cannot be accessed by unauthorized parties [14]. Confidentiality threats generally expose the content of the conversation between two parties, but could also include exposure of call data (telephone numbers dialed, call durations). Threatening the ability to trust the identity of the caller, the message, the identity of the recipient named as integrity threads. Some security mechanisms deal with integrity of content, which produce solution to protect content from alteration by unauthorized users. Availability means stay up-and-running services for use when needed. The proportion of the whole time of a system is in functioning condition gives availability. Availability threats corrupt the ability to make or receive call.

RTP can be used to deliver audio/video data in IP networks. The confidentiality of RTP provided by RTPS (secure RTP) at the application level. The confidentiality of RTP is provided by IPsec at the IP level. Audio watermarking covert channel that is created in RTP audio stream can also be used as a lightweight security solution for confidentiality of audio content.

A. Digital Audio Watermarking

Digital watermarking is an imperceptible, robust and secure communication of data that is related to the host signal. Embedded watermark information follows the watermarked multimedia; it is expected to endure unintentional modifications and intentional removal attempts. The principal design is based on embedded watermark reliably as detected by a watermark detector [15], [16] and [17].

Basically, digital watermark technologies can be divided into blind and non-blind detection techniques, both are strongly related to the decoding process. If the detection of the digital watermark can be done without the original data, such techniques are called blind. On the other hand, non-blind techniques use the original source to extract the watermark data.

B. The Different Types of Digital Audio Watermarking Which can be used to Secure a VoIP Systems

An audio watermarking covert channel which is created in an RTP audio stream can be used as a lightweight security solution to protect confidentiality of audio content. However, there is a trade-off here between the audio quality and security of the VoIP call. Audio Watermarking Techniques that require less executions time, which have higher capacities, and which are more robust and less destroyable are applicable for VoIP communication. Otherwise, VoIP communication would not be able to satisfy the following QoS parameters delay, availability, confidentiality and integrity [14].

The callee receives only watermarked audio content in VoIP conversations. Thus, non-blind techniques are not applicable in such a system.

III. Watermarking Algorithms

In this study, several digital audio watermarking techniques were implemented and compared in terms of robustness, evaluation time, complexities and capacities in order to demonstrate the applicability and feasibility of the digital audio watermarking technique in VoIP systems.

Our primary concern was to examine various digital watermarking algorithms that make the system feasible in terms of speech quality and delay time on several types of VoIP systems. Watermarking (data hiding) algorithms for still images, text and audio are explored in [18].

A. Audio Encoding Techniques

Pulse-Code Modulation (PCM) is the simplest method for converting analog audio signals to digital representation with fixed precision. Most VoIP client devices capture analog audio content in PCM 8bit 8000 Hz mono format. In this paper, PCM 16bit 16000 Hz mono was also used in experiments.

In PCM, the amplitude of the analogue signal is sampled at regular time intervals. The bandwidth of the system, divided into the quantization levels, increases uniformly; this is known as linear quantization. However, linear quantization is not suitable for the Human Auditory System, since it uses the natural logarithmic process for quantization.

Two international non-linear companding standards are a-law and u-law. The u-law algorithm is the accepted standard of digital telecommunication systems of the U.S. and Japan, while the a-law algorithm is the European accepted standard. The main difference is that the u-law algorithm provides a slightly larger dynamic range than the a-law at the cost of worse proportional distortion for small signals.

G.723.1 is an ITU-T recommendation mostly used in Voice over IP (VoIP) applications due to its low bandwidth requirement.

B. Audio Watermarking Techniques

1) Least Significant Bit (LSB)

The LSB method is one of the earliest techniques proposed for audio watermarking. In LSB, the least significant bits of the audio signal are used to store watermark information bits.

The main advantage of the LSB method is a very high watermark channel capacity, e.g. the capacity of LSB is 8kbps for 8 kHz sampling rate.

The second advantage of the LSB is a low computational complexity of the algorithm, so that this algorithm has a very small algorithmic delay. This makes the LSB convenient for real-time application.

In fact, in practice the LSB is one of the simplest algorithms, as it is applied to selected subset of all available host audio samples by the watermark embedders; here this subset is determined by a secret key. The watermark extractors simply extract the watermark by reading the value of the selected bits from the watermarked audio.

The main disadvantage is considerably low robustness; on the other hand, the watermark would survive digital to analogue and analog to digital conversion.

2) DC-Level Shift (DCSHIFT)

The DCSHIFT has been proposed by Uludag et al [19] and involves shifting the DC level for the input audio signal to a negative and positive level according to the binary watermark sequence. Watermark information data is embedded in the lower frequency components of the audio signal. The lower frequency components of the audio signal are below the perceptual threshold of the human auditory system [20]].

The audio signal is divided into several frames that have equally fixed-sizes. In order to compute DC component of the frame. the Discrete Fourier Transform (DFT) is computed for each frame, x[n]. Frame means and frame powers are calculated for each frame.

$$FramePower = (1/N)(x[n])^2 \quad (1)$$

The first element of the frame vector obtained through DFT is modified to represent the watermark bit as follows: If the bit to be embedded is a zero, then the DC level of the corresponding frame is shifted to a negative level with the value:

$$Level0 = -DCBiasMultiplier * FramePower \quad (2)$$

If the bit to embed is a one, DC level the corresponding frame is shifted to a positive level with the value:

$$Level1 = +DCBiasMultiplier * FramePower \quad (2)$$

Finally, the Inverse Discrete Fourier Transform (IDFT) is computed to get a modified frame for each original frame.
For the decoding process, the audio signal is divided into several fixed-sized frames with the frame size being equal to that used during encoding. Frame means are calculated as in the embedding process. The signs of the frame mean give the extracted binary watermark sequence.

Capacity of this method is calculated with the frame size, where each frame holds one binary watermark data.

3) Frequency-Hopping Spread Spectrum (FHSS)

The Frequency Hopping Spread Spectrum (FHSS) is a spread spectrum modulation technique. Recently, FHSS has been used by the military to secure radio signals.

FHSS relies on the imperfections of the human auditory system (HAS) in that it is insensitive to small spectral magnitude changes in the frequency domain [21].

In the embedding phase, the audio signal is divided into several fixed-sized frames. For each frame, the DCT transform is computed so that the watermark is embedded to only a selected set of DCT coefficients determined by PN sequences.

In order to extract the watermark, the watermarked signal is divided into fixed-sized frames in which the frame size is equal to that used during encoding. The DCT of each block is computed where the sign of the correlation between the DCT is the coefficients of the selected components of each block and PN sequence.

FHSS is little influenced by noises, reflections, other radio stations or other environmental factors thus making FHSS a very robust technology [22].

4) Direct-Sequence Spread Spectrum (DSSS)

The Direct Sequence Spread Spectrum (DSSS) is the other main spread spectrum modulation technique.

The DSSS is an algorithm that is evaluated by effectively multiplying the watermark signal and a pseudo-noise (PN) digital signal. PN is a pseudo random sequence of 1 and −1 values that have a flat frequency response over the frequency range, i.e. white noise. As a consequence, the spectrum of the watermark signal is spread over the available band.

The extraction process depends on the sign of the correlation between the block samples and the PN sequence for each block.

The main advantage of DSSS is its resistance to intended or unintended jamming. The capacity of DSSS is much greater than that of FHSS.

C. Audio Quality Evaluation Techniques

The signal-to-noise ratio is a measure to quantify how much a signal has been corrupted by noise. SNR compares the level of a desired signal to the level of background noise. The higher the ratio is the less the obtrusive the background noise is. SNR is often expressed using the logarithmic decibel scale in speech and audio sciences to quantify audio signal quality, is known as SNR in dB.

The SNR in dB is defined as

$$SNR_{dB} = 20 * \left({}^{A_{signal}}\!\Big/_{A_{noise}} \right) \quad (3)$$

in which A is the root mean square amplitude

IV. Experiments and Results

A. Experiments

Six different experiments were created using different clips and VoIP codecs.

Table 1 summarizes the characteristics of those six experiments.

#	Experiment Name	Captured Audio Format	RTP Audio Encoding Format
#1	Experiment-1	PCM signed 16-bit 16000Hz	G.711 a-law 16000 Hz
#2	Experiment-2	PCM unsigned 8-bit 8000 Hz	G.711 a-law 8000 Hz
#3	Experiment-3	PCM unsigned 8-bit 8000 Hz	G.711 u-law 8000 Hz
#4	Experiment-4	PCM unsigned 8-bit 8000 Hz	GSM 6.10 8000 Hz
#5	Experiment-5	PCM unsigned 8-bit 8000 Hz	G.723.1 8000 Hz
#6	Experiment-6	PCM unsigned 16-bit 8000 Hz	ILBC 8000 Hz

Table 1. Experiments

In each experiment, all five sample clips were used to carry all four watermarks each separately embedded with all four watermarking algorithmic. A total of 480 experiments (6 VoIP services x 5 clips x 4 watermark x 4 watermarking algorithms) were carried out and the simulation results were gathered. These results were reproduced using a different computer than that was used previously [12], therefore absolute numbers are not the same. In this section, the simulation results were interpreted in terms of audio qualities after the watermark embedding processes, audio qualities after the audio encoding processes, embed/extract times of the watermarking algorithms and capacities of watermarking algorithms.

B. Sample Clips used in Experiment

The following English phrases suggested by ITU-T, recommendation P.800, were used in the experimental results:

- “I want a minute with the inspector” (~1.9 second)
- “Did he need any money?” (~ 1.14 second)
- “You will have to be very quiet.” (~ 1.87 second)
- “There was nothing to be seen.” (~ 1.48 second)
- “They worshiped wooden idols.” (~ 1.71 second)

C. Watermarks Transferred in Experiments

The following watermarks were embedded/extracted in each sample clip:

- WM-1 : [0, 1, 1, 0]
- WM-2 : [0, 1, 1, 0, 0, 1, 1, 0]
- WM-3 : [0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0]
- WM-4 : [0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0]

D. SNR in dB Comparison of Audio Qualities after the Watermark Embedding Process

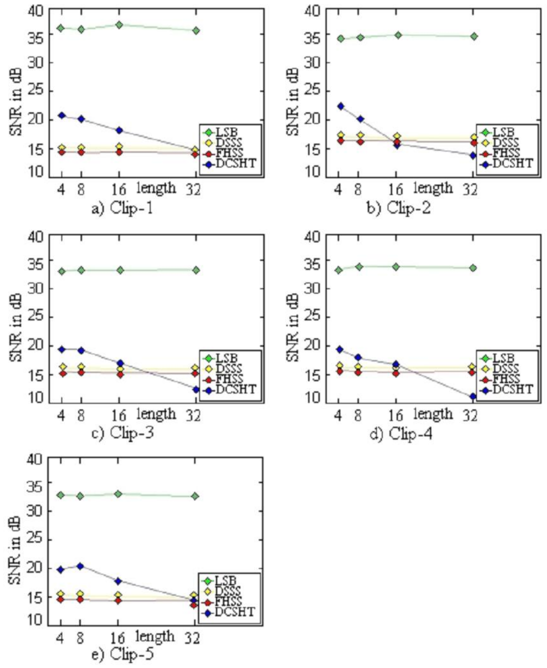
In experiment 1, the analog audio content was captured in PCM 16-bit 16000 Hz format. In experiment 2, 3, 4 and 5, the analog audio content was captured in PCM 8-bit 8000 Hz format. In experiment 6, the analog audio content was captured in PCM 16-bit 8000 Hz format.

SNR in dB values were measured after Watermark embedding processes in each experiment. Measurements were the same in experiment 2, 3, 4 and 5 as those experiments converted analog audio content into the same PCM format. Because of this, a comparison study is mentioned in terms of PCM formats for each audio watermarking algorithm.

Experimental results for the PCM 16-bit 16000 Hz format and PCM 8-bit 8000 Hz format have the same characteristics. After the watermark embedding process before the audio encoding process, LSB offers better SNR in dB performance than the other methods. SNR in dB performances of LSB, DSSS and FHSS were not affected by the length of embedded watermark. However, the DC-SHIFT SNR in dB performance dramatically decreased with the water mark length.

The main difference between PCM 16000 and PCM 8000 is that PCM 16000 offers higher audio quality than the PCM 8000 format.

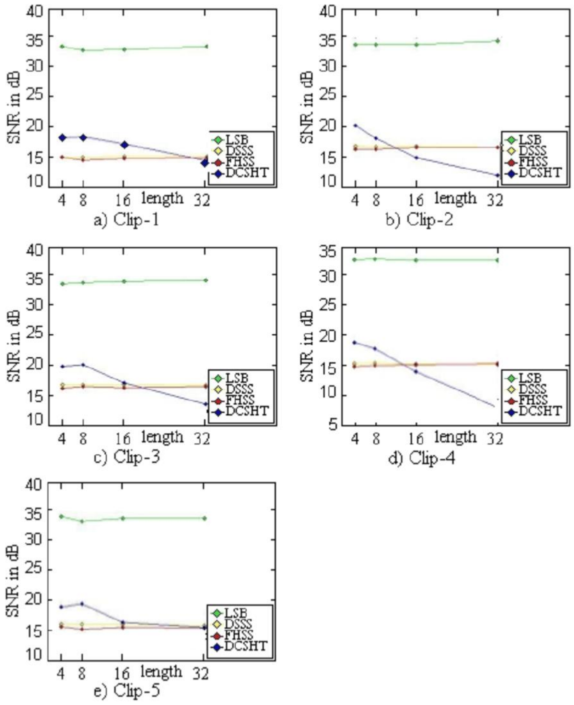
Fig. 1 shows a comparison between the PCM 16bit 16000 Hz audio qualities after the watermarking embedding process for all watermarking algorithms; each clip belongs to experiment-1.



a) Clip-1
b) Clip-2
c) Clip-3
d) Clip-4
e) Clip-5

Figure 1. SNR in dB Values for Each Clips in PCM 16-bit 16000 Hz mono Format Before Encoding

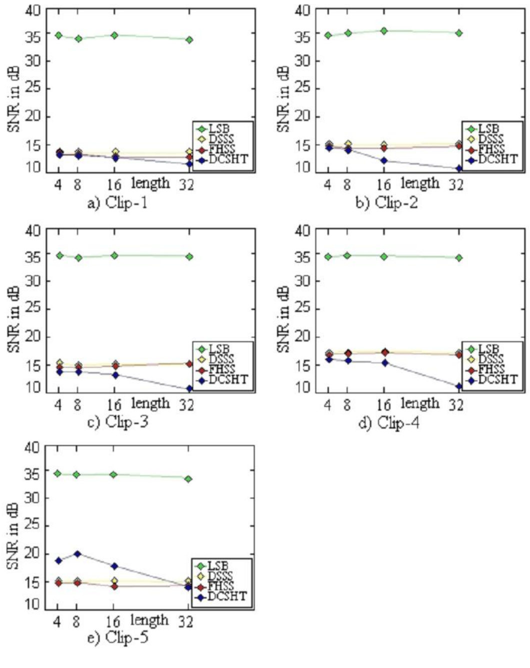
Fig. 2 shows the comparison of PCM 8bit 8000 Hz audio qualities after the watermarking embedding process for all watermarking algorithms; only for each clip belongs to experiment-2, experiment-3, experiment-4 and experiment-5.



a) Clip-1 b) Clip-2 c) Clip-3 d) Clip-4 e) Clip-5

Figure 2. SNR in dB Values for Each Clip in PCM 8-bit 8000 Hz mono Format before Encoding.

Fig. 3 shows the comparison of PCM 16bit 8000 audio qualities after the watermarking embedding process for all watermarking algorithms; only for each clip belongs to experiment-6.



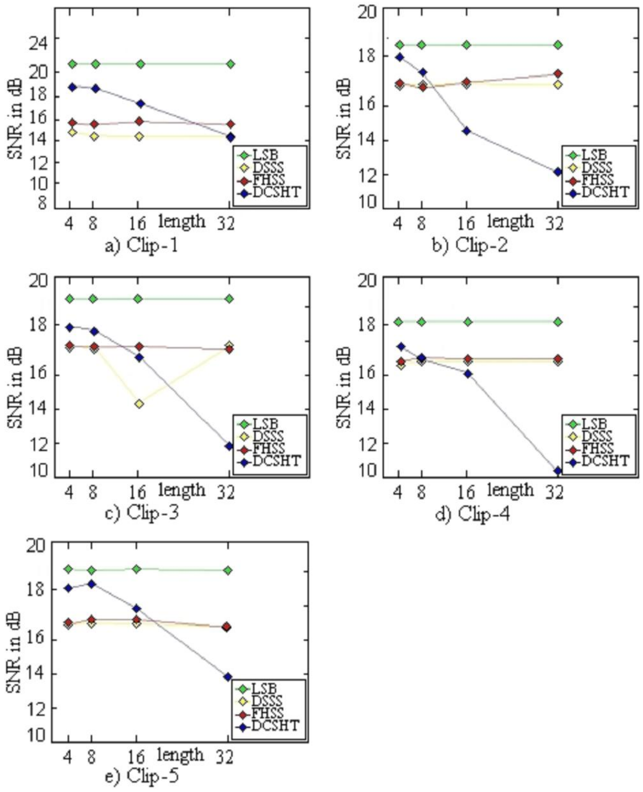
a) Clip-1 b) Clip-2 c) Clip-3 d) Clip-4 e) Clip-5

Figure 3. SNR in dB Values for Each Clip in PCM 16-bit 8000 Hz mono Format before Encoding

E. SNR in dB Comparison of Audio Qualities After Audio Encoding Process on Watermarked RTP Audio Content

In experiment 1, watermarked audio content was encoded with A-Law 16000 encoding during the transportation phase of VoIP. In experiment 2, 3, 4, 5 and 6 watermarked audio content were encoded with the following encoding algorithms, correspondingly, a-law 8000, u-law 8000, GSM 6.10, G.723.1 and ILBC.

Experiments show that LSB generally offers better SNR in dB performance than other methods; this was true in all experiments. The order of SNR in dB performances is not affected by the encoding process. However, SNR in dB performance of DC-SHIFT algorithm dramatically decreased according to the length of the embedded water mark data. Table 2 shows the decrease ratios of SNR in dB values in each experiment for each watermark algorithms. Table 2 shows audio contents which are encoded with ILBC encoding or audio contents which are encoded with G.723.1 encoding have a negative effect on perceived speech quality. According to Table 2, LSB has negative effect when u-law and GSM used.



a) Clip-1 b) Clip-2 c) Clip-3 d) Clip-4 e) Clip-5

Figure 4: SNR in dB Values for Each Clip in PCM 16bit 16000 Hz mono Format after G.711.1 a-law 16000Hz Encoding.

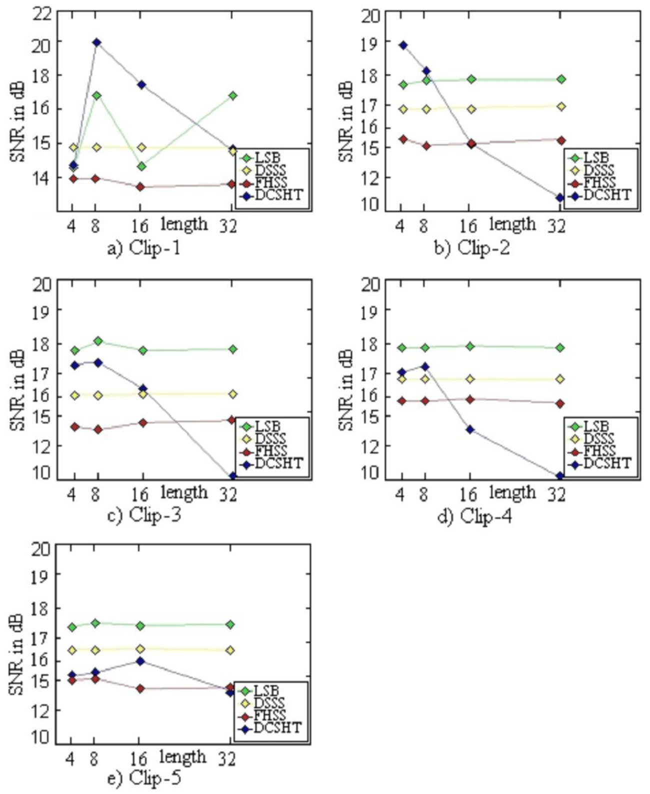


Figure 5: SNR in dB Values for Each Clip in PCM 8bit 8000 Hz mono Format after G.711.1 a-law 8000Hz Encoding.

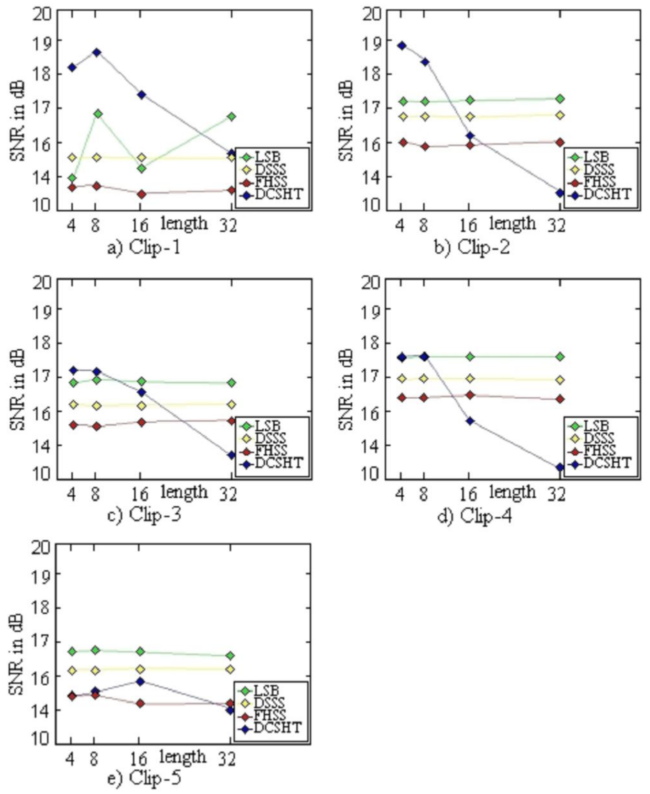


Figure 6: SNR in dB Values for Each Clip in PCM 8bit 8000 Hz mono Format after G.711.1 u-law 8000Hz Encoding.

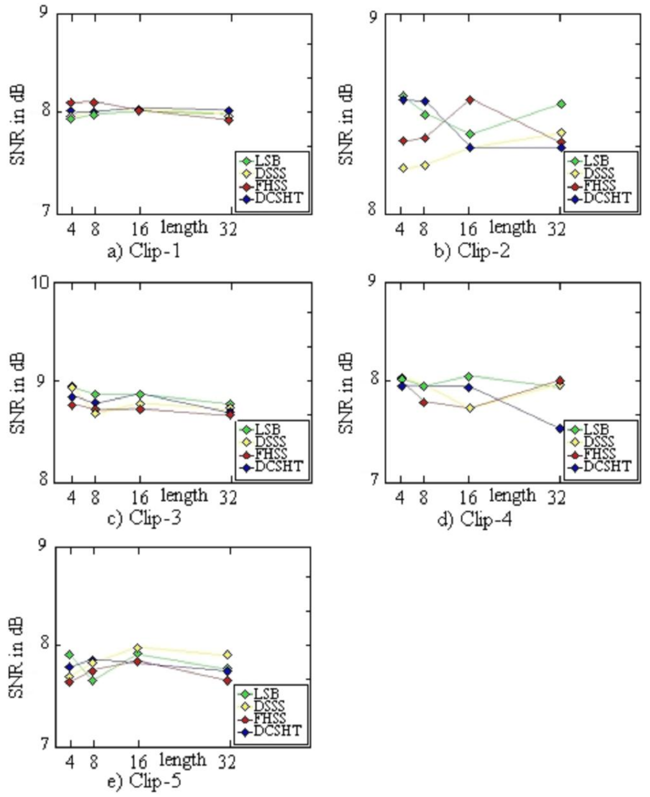


Figure 7: SNR in dB Values for Each Clip in PCM 8bit 8000 Hz mono Format after GSM 6.10 8000Hz Encoding.

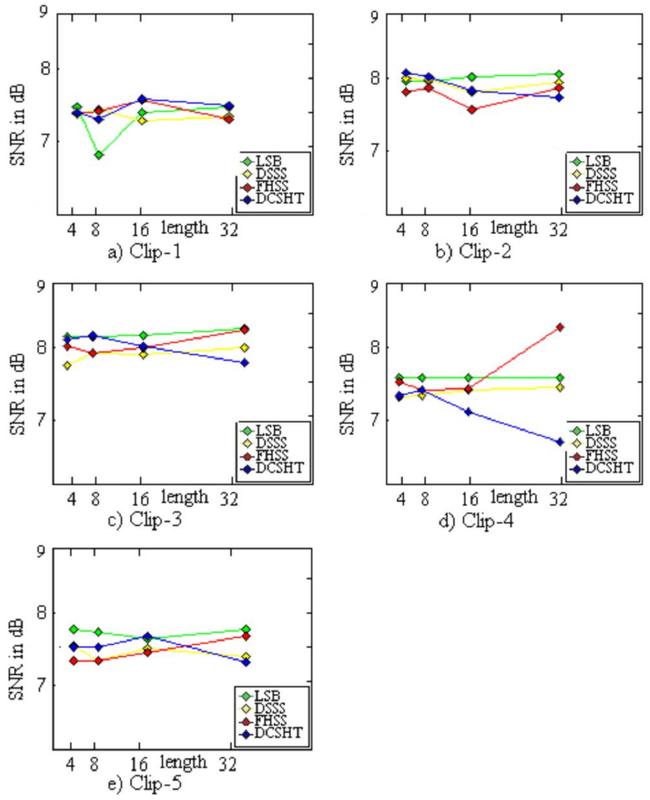


Figure 8: SNR in dB Values for Each Clip in PCM 8bit 8000 Hz mono Format after G.723.1 8000Hz Encoding.

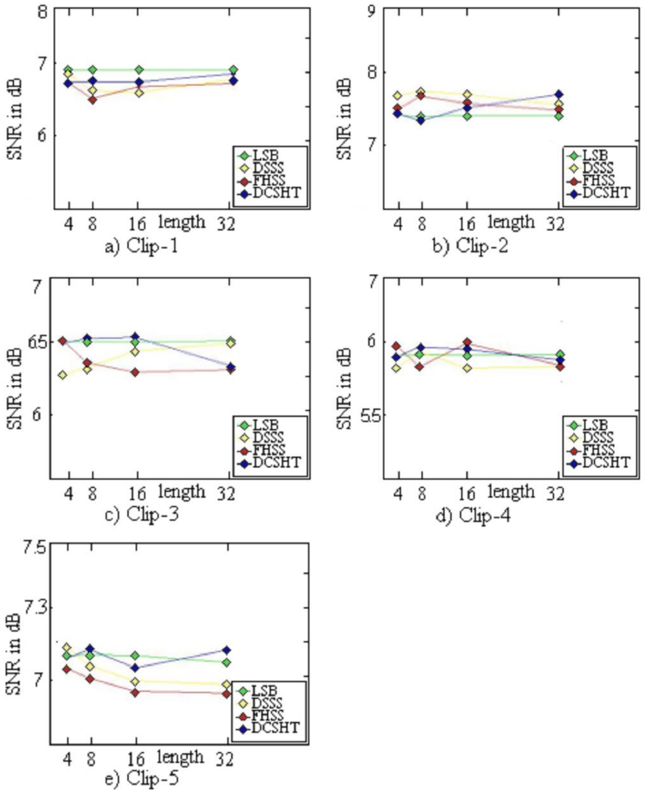


Figure 9: SNR in dB Values for Each Clip in PCM 16bit 8000 Hz mono Format after ILBC 8000Hz Encoding.

Experi ment #	LSB	DSSS	FHSS	DC SHIFT
#1	~%44.70	~%00.20	~%00.10	~%08.60
#2	~%49.90	~%01.20	~%06.80	~%00.10
#3	**~%52.20**	~%03.90	~%08.40	~%00.10
#4	**~%75.66**	~%49.20	~%47.90	~%47.10
#5	**~%77.13**	**~%77.55**	**~%52.10**	**~%51.86**
#6	**~%77.13**	**~%58.58**	**~%57.92**	**~%58.74**

Table 2. SNR in dB Decrease Ratios after RTP Encoding Processes

Higher decrease ratios are marked bold in Table 2.

F. Comparison of Embed/Extract Times of Watermarking Algorithms

Table 3 below, demonstrates embed and extract time durations of watermark algorithms for each experiments. The results are labeled bold, the total duration of which is above human tolerance of audio delay limit (~200ms) in RTI (real-time intolerant applications).

Experi ment #	LSB	DSSS	FHSS	DC SHIFT
#1	0.04/0.04	**0.39/0.16**	**0.14/0.12**	**0.12/0.09**
#2	0.03/0.02	0.08/0.02	0.09/0.04	0.05/0.06
#3	0.03/0.02	0.08/0.02	0.09/0.04	0.05/0.06
#4	0.03/0.04	0.08/0.02	0.09/0.02	0.05/0.13
#5	0.03/0.01	0.05/0.01	0.04/0.02	0.05/0.02
#6	0.01/0.01	0.02/ 0.002	0.015/ 0.005	0.01/ 0.002

Table 3. SNR in dB Decrease Ratios after RTP Encoding Processes

Higher embed/extract times are marked bold in Table 3.

As shown in Table 3, such VoIP systems as defined in experiment-1 cannot use watermarking techniques for security purposes due to of delay times. The LSB algorithm adds less extra time to communication than the other three algorithms used in all five experiment environments. DSSS and FHSS have extra delay times in all six experiment environments.

G. Comparison of Capacities of Watermarking Algorithms

Before encoding, all four watermarks were successfully extracted; however some were lost after encoding.

Experiment results are grouped according to WM algorithms as shown in Table 4.

As shown in Table 4, the simulated environment in experiment-4 and experiment-5 are useless for WM-Enabled VoIP communications. LSB algorithm is clip dependent, so it is also not of any use. ILBC encoding is also clip dependent, so it is also not of any use. More ever, according to Table 4, FHSS and DSSS were resistant to encoding in experiment-1, experiment-2 and experiment-3. FHSS and DSSS hold a greater capacity in a-law encoding environments (experiment-1 and experiment-2) than u-law encoding (experiment-3). Experiment-3 for DSSS algorithm provided a small degree of bit clip dependent results, thus making DSSS algorithm unreliable for u-law encoding.

Experiment #	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
#1	8	0	8	8	16
#2	4	4	0	4	0
#3	4	4	0	0	0
#4	0	0	0	0	0
#5	0	0	0	0	0
#6	0	4	0	8	4

Table 4-a. Capacities for LSB

Experiment #	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
#1	8	4	16	4	8
#2	4	4	8	8	8
#3	4	0	8	8	4
#4	0	0	0	0	0
#5	0	0	0	0	0
#6	0	0	0	0	0

Table 4-b. Capacities for DSSS

Experiment #	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
#1	8	8	8	8	8
#2	4	4	8	4	4
#3	4	4	8	4	4
#4	0	0	0	0	0
#5	0	0	0	0	0
#6	0	0	0	0	0

Table 4-c. Capacities for FHSS

Experiment #	Clip 1	Clip 2	Clip 3	Clip 4	Clip 5
#1	4	4	0	0	4
#2	0	0	0	0	0
#3	0	0	0	0	0
#4	0	0	0	0	0
#5	0	0	0	0	0
#6	0	0	0	0	0

Table 4-d. Capacities for DC-SHIFT

V. Conclusions

There are several methods proposed and implemented for the security of VOIP communication. One of the main issue to be addressed is the source origin authentication. Digital audio watermarking techniques can be used to identify the user who initiated the conversation.

In this paper, digital audio watermarking techniques were implemented to demonstrate the applicability of each system over VoIP for security purposes regarding of the source origin authentication. Evaluation times, capacity, complexity and robustness of the watermarking algorithms are important in real time systems therefore these were compared for different algorithms. Adding watermark can be considered as adding noise to the speech therefore, SNR for different algorithms were also compared. The experimental results demonstrated that LSB is the simplest algorithm to implement and offers better SNR (in dB) performance before encoding than others and gives better performance with a-law 16 bit 16000Hz encoding than others, unfortunately it is clip dependent. Although the evaluation times of DSSS and FHSS are high, these are more resistant to a-law encoding. FHSS has a greater capacity, despite a-law encoding. GSM and G.723.1 encoding correspondingly in experiment-4 and experiment-5 completely failed. ILBC encoding failed for all watermark algorithms except LSB but it is not usable due to its clip-dependency.

In conclusion, FHSS and DSSS algorithms can be used in WM-Enabled VoIP mechanisms which may be utilized for source origin authentication. The source origins are represented here by 4-bit source origin indicators(ID-keys). However using the GSM codec 6.10 and G.723.1 and ILBC did not produce satisfactory results, and we do not advise using these codecs with watermarking.

The experiments showed that the embedded watermarks were not preserved in the lossy compression codecs. Using codecs which use compression are useful in low bandwidth communication systems. However increased bandwidth capacity of the internet by employing broadband networks and using quality of service to give high priority to voice communications decreases the importance of low bandwidth codecs. In-band tones which are used in call-center applications are not supported in these type of codecs as well. Therefore, it is likely that codecs which do not use compression will continue to be used in VOIP. As a result, watermarking can be used in the systems where uncompressed codecs used.

References

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks. “SIP: Session Initiation Protocol”, RFC 3261, IETF, June 2002.

[2] W. Mazurczyk, Z.Kotulski. “Adaptive VoIP with Audio Watermarking for Improved Call Quality and Security”, *Journal of Information Assurance and Security 2*, pp. 226-234, 2007.

[3] S. Yuan. S. A. Huss. “Audio Watermarking Algorithm for Real-time Speech Integrity and Authentication”, *MM&Sec'04*, Magdeburg, Germany, pp. 220-226, September 20–21, 2004.

[4] M. Baugher, McGrew. “The Secure Real-time Transport Protocol (SRTP)”, RFC 3711 IETF, March 2004 .

[5] M. Zamani, A. Manaf, R. Ahmed,F. Jaryani, F. Taherdoost,S. S. Chaiekar,H. R. Zeidanloo. “A Novel Approach for Genetic Audio Watermarking”, *Journal of Information Assurance and Security*, Vol. 5, pp. 102-111 , 2010.

[6] V. Casola, A. Mazzeo, N. Mazzocca, M. Rak. “Security Design and Evaluation in a VoIP Secure Infrastructure: A Policy Based Methodology”, *Journal of Information Assurance and Security*, Vol 3, pp. 205-216, 2006.

[7] A. Ghafarian, R. Draughorne, S. Hargraves, S. Grainger, S. High, C. Jackson. “Securing Voice over Internet Protocol”, *Journal of Information Assurance and Security*, Vol 2, pp. 200-204, 2007.

[8] T. Takahashi, W. Lee. “An assesment of VOIP Channel Threats”, *Third International Conference on Security and Privacy in Communication Networks*, Nice, France pp. 371-380, 2007.

[9] M. Steinebach, J. DittMann. “Watermarking Based Digital Audio Data Authentication”, *EURASIP Journal on Applied Signal Processing*, Vol 10, pp. 1001–1015, 2003.

[10] J. Dittmann1, M. Steinebach, A.Lang, S. Zmudizinski. “Advanced Audio Watermarking Benchmarking”, In *Proceedings of SPIE*, pp. 224-235, 2004.

[11] W. Mazurczyk, K. Szczypiorski. *Covert Channels in SIP for VoIP Signalling*, Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Telecommunications, Warsaw, Poland, 2008.

[12] F. C. Er and E. Gul. “Comparison of Digital Audio Watermarking Techniques for the security of VOIP Communications”, *Information Assurance and Security (IAS) Conference*, 2011.

[13] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications”, RFC 3550, IETF, July 2003.

[14] D. L. Evans, P. J. Bond , S. Phoha. *Security Considerations for Voice Over IP Systems*, NIST Special Publication 800-58. Gaithersburg, MD 20899-8930, January 2005.

[15] H. J. Kim. "Audio watermarking techniques," in *Pacific Rim Workshop on Digital Steganography*, Kyushu Institute of Technology, Kitakyushu, Japan, Jul. 3–4, 2003

[16] N. Cvejic, *Algorithms for Audio Watermarking and Steganography*, ISBN 951-42-7384-2, University of Oulu, Finland, 2004.

[17] N. Cvejic, T. Seppanen. *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks* , IGI Global, ISBN 1599045133, 2007.

[18] W. Bender, D. Gruhi. N. Morimoto, A. Lu. "Techniques for Data Hiding", *IBM Systems Journal*, Vol. 35, No 3&4, pp. 313-336, 1996.

[19] U. Uludag, L. Arslan. "Audio watermarking using DC-level shifting", *project report*, Bogazici University, 2001.

[20] T. C. Thanuja, R.Nagaraj. "Schemes for Evaluating Signal Processing Properties of Audio Watermarking", *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.7, July 2008.

[21] D. Kirovski, H. Malvar. "Robust Spread-Spectrum Audio Watermarking", *ICASSP*, IEEE 2001.

[22] S. M. Schewartz. "FHSS vs. DSSS", Seminar Notes, 2006.

Author Biographies

Fusun Citak Er was born in Turkey in 1980. She obtained her B.Sc. and M.Sc degree in Computer Science Engineering from the Marmara University in Istanbul, Turkey. She is co-founder of PhER Technology and directing the operation of software development operations of the organization. Her current research interests include Software Development Processes, VoIP Security and Watermarking.

Ensar Gul is currently a professor of Computer Engineering at the faculty of Engineering, Marmara University in Istanbul, Turkey. He received a BSc from Bogazici University, Turkey and MSc degree from Loughborough University of Technology, UK and PhD degree from Sussex University, UK. Besides his academic life, he also worked in telecommunication industry for 11 years. His research interests are parallel processing, real time software development and VoIP applications.