T.C.
MARMARA UNIVERSITY
INSTITUTE FOR GRADUATE STUDIES IN
PURE AND APPLIED SCIENCES

COMPARISON OF DIGITAL AUDIO WATERMARKING TECHNIQUES FOR THE SECURITY OF VOIP COMMUNICATIONS

Füsun Çıtak ER

THESIS
FOR THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SUPERVISOR
Prof. Dr. Ensar GÜL

İSTANBUL 2011

T.C.
MARMARA UNIVERSITY
INSTITUTE FOR GRADUATE STUDIES IN
PURE AND APPLIED SCIENCES

COMPARISON OF DIGITAL AUDIO WATERMARKING TECHNIQUES FOR THE SECURITY OF VOIP COMMUNICATIONS

Füsun Çıtak ER
(141524120089003)

THESIS
FOR THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SUPERVISOR
Prof. Dr. Ensar GÜL

İSTANBUL 2011

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my thesis adviser, Prof. Dr. Ensar Gül for his invaluable guidance and help during the preparation of this dissertation.

Also, special thanks to Prof. Dr. Haluk Topçuoğlu and Prof. Dr. Murat Doğruel.

This thesis is dedicated to my family.

01, 2011 **Füsun Çıtak ER**

TABLE OF CONTENTS

ABSTRACT

COMPARISON OF DIGITAL AUDIO WATERMARKING TECHNIQUES FOR THE SECURITY OF VOIP COMMUNICATIONS

In this thesis, it is presented that digital audio watermarking techniques can be utilized to improve security of Real-Time Communications using Session Initiation Protocol (SIP), such as Voice over IP (VoIP), for source origin authentication.

Such watermark-enabled VoIP mechanism utilizes audio watermarking techniques as a covert channel between calling parties to send source origin indicator information during conversation phase of VoIP. In order to exchange source origin indicator information between calling parties, available steganographic techniques for SIP that are used for creating covert channels during signaling phase of VoIP call are evaluated. The effects of audio watermarking were measured using the Signal-to-Noise Ratio, watermark extract durations and the effects of a-law encoding during transportation phase of VoIP. Moreover, various audio watermarking algorithms were implemented to demonstrate applicability of defined security solution in terms of certain parameters, like: robustness, evaluation times, complexity and capacity.

The Experimental results showed that some watermarking algorithms are applicable in VoIP while some are not suitable for source origin authentication.

01, 2011 **Füsun Çıtak ER**

ÖZET

VOIP HABERLEŞME SİSTEMLERİNİN GÜVENLİĞİ İÇİN DİJİTAL SES DAMGALAMA YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Bu tezde, digital ses damgalama teknikleri, oturum başlatma protokolü olarak SIP kullanan gerçek zamanlı iletişim sistemlerinde, kaynak kökenli kimlik doğrulama yöntemi olarak kullanilabileceği sunulmuştur. Önerilen yöntem, iletişim kuran taraflar arasında kaynak köken bilgisi iletimini sağlamak için digital ses damgalama tekniklerini kullanmaktadır. İletişim kuran taraflar, karşılıklı olarak kaynak köken bilgilerini SIP(Oturum Başlatma Protokolü) nün steganographic tekniklerini kullanarak birbirlerine iletirler.

Bu çalışmada, digital ses damgalama teknikleri, önerilen yöntemin uygulanabilirliğini göstermek amacı ile gerçek zamanlı çoklu ortam iletişim sistemleri üzerinde uygulandı, gerçek zamanlı çoklu ortam iletişim sistemleri sistemleri sağlamlık, güvenlik, şeffaflık, karmaşıklık, kapasite, doğrulama ve geri alınabilme gibi ölçüt parametreleri açısından değerlendirildi.

Deney sonuçları bazı kısıtlar ile birlikte önerilen yönetimin gerçek zamanlı çoklu ortam iletişim sistemlerinde uygulanabilirliğini göstermektedir.

01, 2011 **Füsun Çıtak ER**

LIST OF SYMBOLS

t **:** time

N(w) **:** Length of embedded information using digital audio watermarking.

A **:** Amplitude of digital audio signal

LIST OF ABBREVIATIONS

CM : Continuous Media
DCT : Discrete Cosine Transform
DFT : Discrete Fourier Transform
DM : Discrete Media
DSSS : Direct Sequence Spread Spectrum
FFT : Fast Fourier Transform
FHSS : Frequency Hopping Spread Spectrum
FTP : File Transfer Protocol
GSM : Global System for Mobile Communications
HTTP : Hypertext Transfer Protocol
LSB : Least Significant Bit
RSVP : Resource Reservation Protocol
RT : Real Time
RTI : Real Time Intolerant
RTCP : Real-Time Transport Control Protocol
RTPS : Real-Time Transport Protocol
RTP : Secure Real-Time Transport Protocol
NRT : Non Real Time
SDP : Session Description Protocol
SIP : Session Initiation Protocol
SMTP : Simple Mail Transfer Protocol
SNR : Signal-to-Noise Ratio
VoIP : Voice over IP
THSS : Time Hopping Spread Spectrum
WM : Watermarking
3GPP : Third-Generation Partnership Project

LIST OF FIGURES

PAGE NO

LIST OF TABLES

PAGE NO

CHAPTER I

INTRODUCTION and AIM

I.1. INTRODUCTION

VoIP is a real-time technology that allows voice conversations through Internet. Fourth generation (4G) cell phone networks will be pure IP and SIP, so that, the 3GPP have chosen SIP as the protocol underlying many of the important interfaces between elements in a 4G network. Securing VoIP is an important topic due to this fact. Up to present some security techniques have been proposed for VoIP communications with the necessity of the trade off between providing security and the low latency for real time service. Nowadays, Digital Audio Watermarking is used for several purposes in VoIP. Mazurczyk et al [1] utilize digital audio watermarking techniques for FEC(Forward Error Correction). In [2], digital audio watermarking used an alternative data integrity measurement method against SRTP.

In this thesis, digital audio watermarking used for security purposes for source origin authentication, such a mechanism is implemented combining SIP level key exchange as described in [3] and embedding source origin indicator in conversation phase using digital audio watermarking as described in [1]. Digital audio watermarking is utilized for this purpose due to its bandwidth consuming is insignificant and redundant to packet loss. Up to the present Digital Audio Watermarking techniques have been proposed to secure VoIP conversations. In this thesis, it is aimed to evaluate applicability of several audio watermarking techniques as VoIP security mechanisms.

I.2. AIM

In this thesis, we define a security mechanism for real-time VoIP services based on SIP as a signaling to authenticate the source origin using digital audio watermarking. Our primary concern is to examine various digital audio watermarking algorithms that are suitable for real-time VoIP Systems in terms of speech quality and delay time. Exploring SIP steganographic techniques in order to create covert channels in SIP messages during signaling phase of VoIP is our secondary concern. Those covert channels are used to exchange source origin indicators among calling parties. Such a VoIP system is called as "WM-Enabled VoIP System" and explained in Section I.2.1.

In this thesis, several digital audio watermarking techniques are implemented and compared in terms of robustness, evaluation times, complexity and capacity to demonstrate applicability and feasibility of "WM-Enabled VoIP System".

I.2.1 WM-Enabled VoIP System

VoIP is a real-time technology that allows voice conversations through Internet. VoIP communications are composed of three sections: Signaling Phase, Conversation Phase and Ending Communication.

In signaling phase, calling parties are authenticated and authorized to create, modify and terminate VoIP sessions using SIP protocol.

Signaling Phase of WM-Enabled VoIP System is modified to carry ID-key of the caller to the callee using SIP/SDP Protocols Steganography, the caller embeds its ID-key into covert channels of INVITE(with SDP) message. the callee extracts ID-key of the caller from INVITE(with SDP) message and stores during calling session.

Conversation Phase of WM-Enabled VoIP System is modified to sent ID-key of the caller periodically to the callee using Audio Watermarking Techniques. The callee extracts ID-key of the caller embedded into RTP audio stream.

If extracted ID-key is different than gained ID-key in Signaling Phase then Conversation is Ended, which is named as Ending Communication Phase.

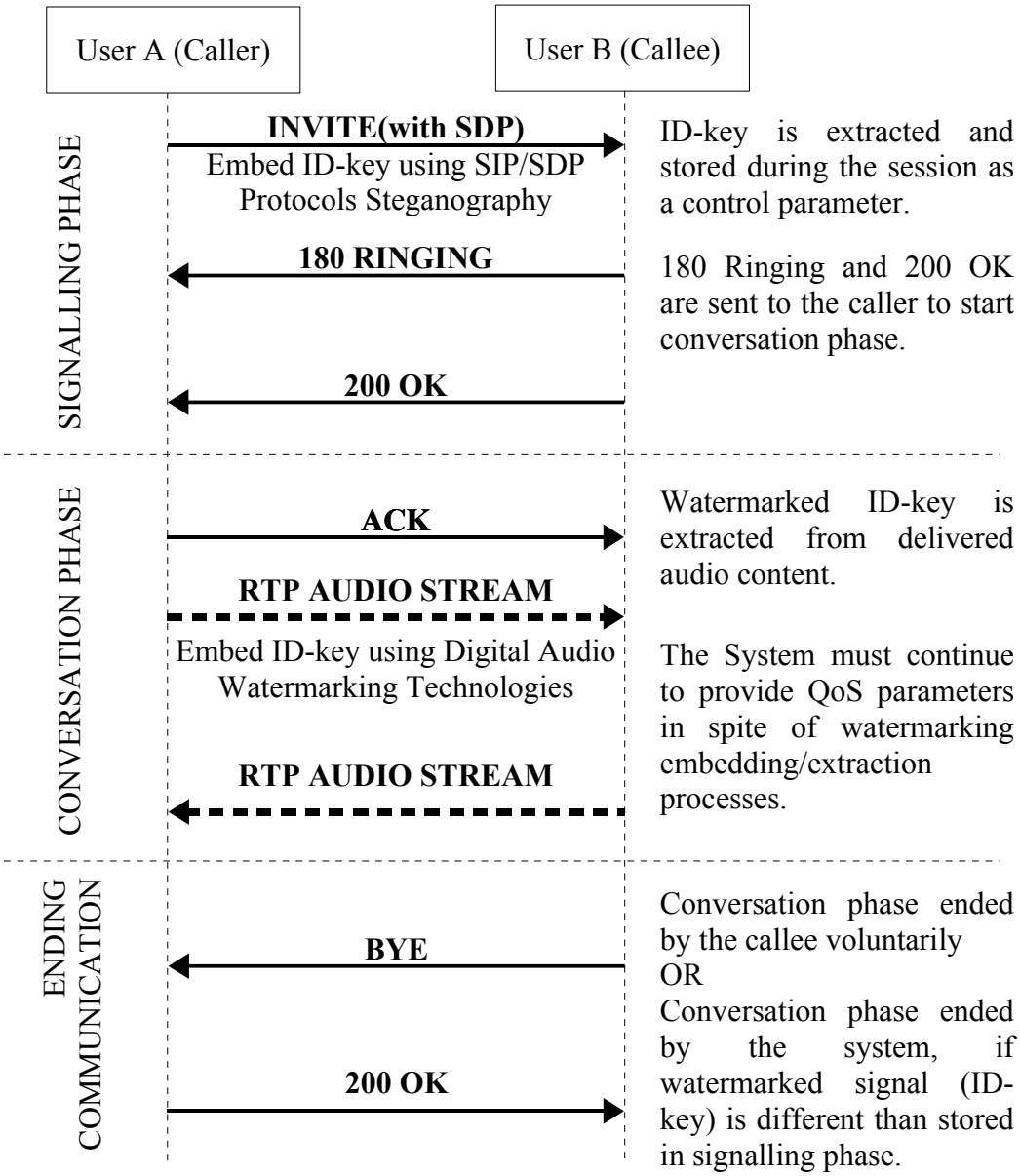


Figure I.1: General Structure of WM-Enabled VoIP System

I.3. OUTLINE OF THESIS

Chapter II presents the background information about Session Initiation Protocol and digital audio watermarking techniques. In Chapter III, the implementation of WM-Enabled VoIP System is expressed in detail. Experimental results are provided and interpreted in Chapter IV. Finally, discussions and future work in Chapter V.

CHAPTER II

GENERAL BACKGROUND

II.1. MULTIMEDIA NETWORKS AND MEDIA TYPES

The term 'multimedia' refers to diverse classes of media employed to represent information. Multimedia traffic refers to the transmission of data representing diverse media over communication networks. The media classified into three groups: text, visuals, and sound. Textual material may include the traditional unformatted plain text, numerous control characters, mathematical expressions, phonetic transcription of speech, music scores and other symbolic representations such as hypertext. The visual material may include line drawings, maps, gray-scale or colored images and photographs, as well as animation, simulation, virtual reality objects, video and teleconferencing. The sound material may include telephone/broadcast-quality speech to represent voice, wide band audio for music reproduction, and recordings of sounds such as electrocardiograms or other biomedical signals.[4]

From a networking perspective, all media types can be classified as either Real-Time (RT) or Non Real-Time (NRT). RT media types require either hard or soft bounds on the end-to-end packet delay/jitter, while NRT media types, do not have any strict delay constraints, but may have rigid constraints on error. Applications that require an error-free delivery of NRT media, typically use TCP for transport.[4]

The RT media types are further classified as Discrete media (DM) or Continuous media (CM), depending on whether the data is transmitted in discrete quantum as a file or message, or continuously as a stream of messages with inter-message dependency. The RT continuous type of media can further be classified as delay tolerant or delay intolerant. Examples of RT, continuous, and delay-intolerant media are audio and video streams used in audio or video

conferencing systems, and remote desktop applications. Streaming audio/video media, used in applications like Internet web cast, are examples of delay-tolerant media types.[4].

The entire network oriented classification has been shown in Figure II.1.

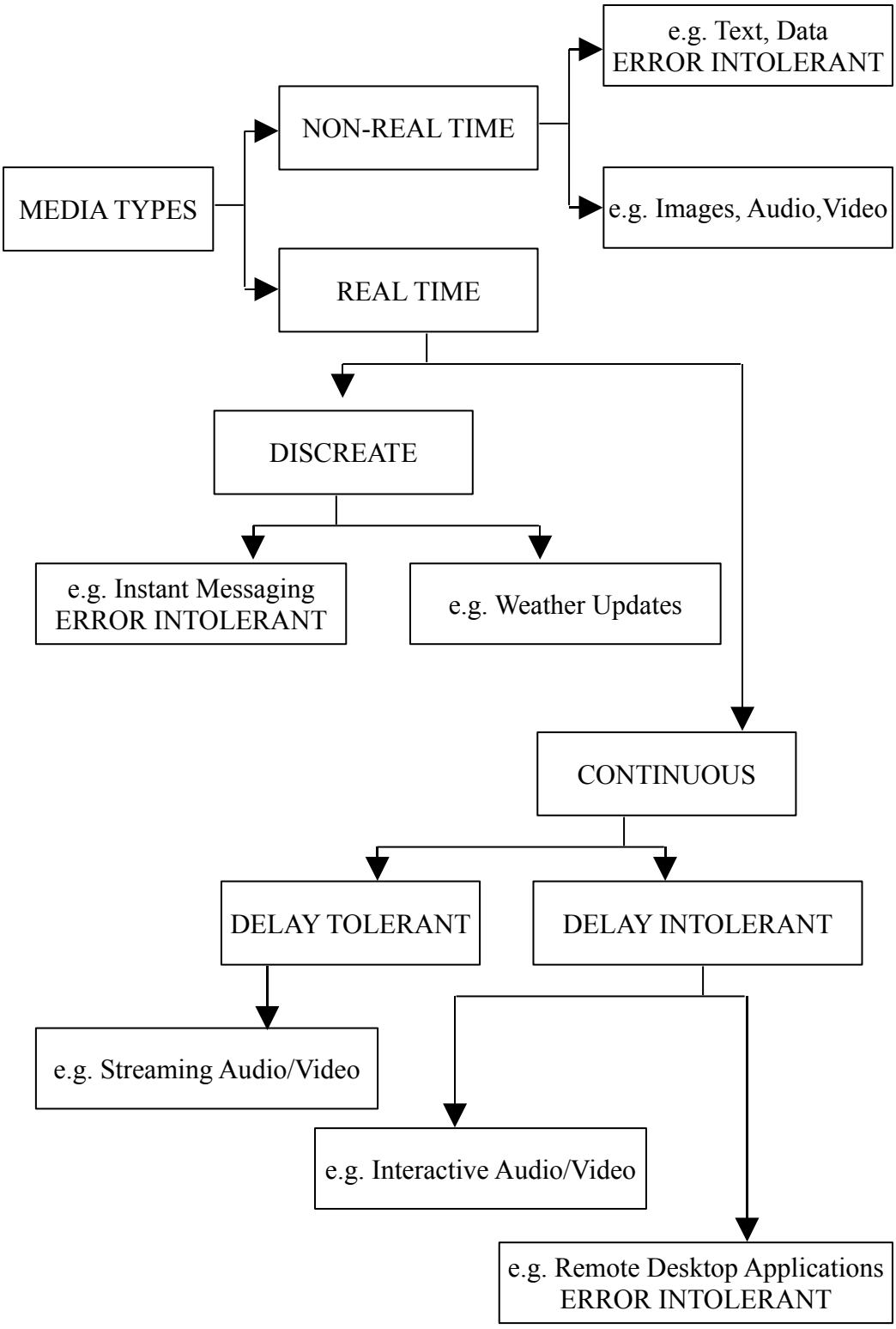


Figure II.1. Network-Oriented Classification of Media Types

Text and Audio are common media types. Text is the most popular of all the media types. It is distributed over the Internet in many forms including files or messages using different transfer protocols such as FTP, HTTP or SMTP. Audio media is sound/speech converted into digital form using sampling and quantization. Digitized audio media is transmitted as a stream of discrete packets over the network.[4]

The real-time requirements of audio strictly depend on the expected interactivity between the involved parties. Some applications like Internet-Telephony, which involves two-way communication, are highly interactive and require shorter response times. Applications that use this media type are called Real-Time Intolerant (RTI) applications. In most RTI applications the end-to-end delay must be limited to ~200 msec to get an acceptable performance. Other applications like Internet web cast, which involves one-way communication, have relatively low interactivity. Interactivity, in this case, is limited to commands that allow the user to change radio channels (say), which can tolerate higher response times. Such kind of media are termed as Real-Time Tolerant (RTT) applications. Streaming Audio is also used to refer to this media type.[4]

II.2. INTERNET TELEPHONY(VoIP)

VoIP is a real-time technology that allows voice conversations through Internet. A VoIP communication 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. H.323 [5] is also a signaling protocol but it is outdated.

After establishing connection between calling parties, conversation phase started. Mostly used transport protocol in conversation phase is Real Time Protocol RTP [6], which provides end-to-end network transport functions suitable for applications transmitting real-time audio. RTP defines a profile for video or audio applications those associated with payload formats. Some of the audio payload formats include: G.711, G.723, G.726, G.729, GSM, QCELP,

MP3, DTMF etc., and some of the video payload formats include: H.261, H.263, H.264, MPEG etc.

RTCP [7], SDP [8], and RSVP[9] are supplementary protocols that complete VoIP functionality.

II.2.1 Security Issues of VoIP

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 aren't applicable to VoIP in their current form. Most of security mechanisms raise high latency. Because of the time-critical nature of VoIP, and its low tolerance for latency and packet loss, we are facing with the 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.

1) Confidentiality: Some security mechanisms dealing with to provide confidentiality for media data. Confidentiality refers to the need to keep information secure and private and cannot be accessed by unauthorized parties[11]. 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).

RTP is the default standard for audio/video transport in IP networks which is a framework for audio or video data delivery, and it has some profiles for particular uses. The confidentiality of RTP is provided by RTPS (secure RTP) at the application level. The confidentiality of RTP is provided by IPSec at the IP level.

2) Integrity: Threatening the ability to trust the identity of the caller, the message, the identity of the recipient named as integrity threats. Some security mechanisms deals with integrity of content, which produce solution to protect content from alteration by unauthorized users.

3) Availability: Availability means stay up-and-running services for use when needed. The proportion of the whole time of a system is in a functioning

condition gives availability. Availability threats corrupt the ability to make or receive call.

II.3. SESSION INITIATION PROTOCOL(SIP)

SIP is one of the most popular application-layer (TCP/IP model) signaling protocols for IP Telephony that can establish, modify, and terminate multimedia sessions, such as VoIP calls. It is text-based and simple. SIP specification defines only six main methods: REGISTER for registering contact information, INVITE, ACK, and CANCEL for setting up sessions, BYE for terminating sessions and OPTIONS for querying servers about their capabilities. SIP uses network elements called proxy or redirect servers to help route requests to the user's current location, authenticate and authorize users for services, implement provider’s call-routing policies, and provide features to users.[10]

II.4. DIGITAL WATERMARKING TECHNOLOGY

Digital watermarking is an imperceptible, robust and secure communication of data related to the host signal. Embedded watermark information follows the watermarked multimedia and expected that it endures unintentional modifications and intentional removal attempts. The principal design based on embedded watermark reliably detected by a watermark detector[3].

Digital watermarking technology has numerous application areas. Intellectual property protection is currently the main driving force behind research in this area. Watermarking used as a proof of ownership in digital audio content. Some applications are designed to use watermarking technology for authentication and tampering detection. Another application area for digital watermarking is broadcasting. Identification information is coded using digital watermarking technology[16].

Basically, digital watermark technologies can be divided into blind and non-blind detection techniques, which 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. In this perspective, In WM-Enabled VoIP System, the callee receive only watermarked audio content. So that, non-blind techniques can not be applicable in such a system.

Basic schema of embedding and extraction processes of digital watermarking technology is shown in Figure II.2.

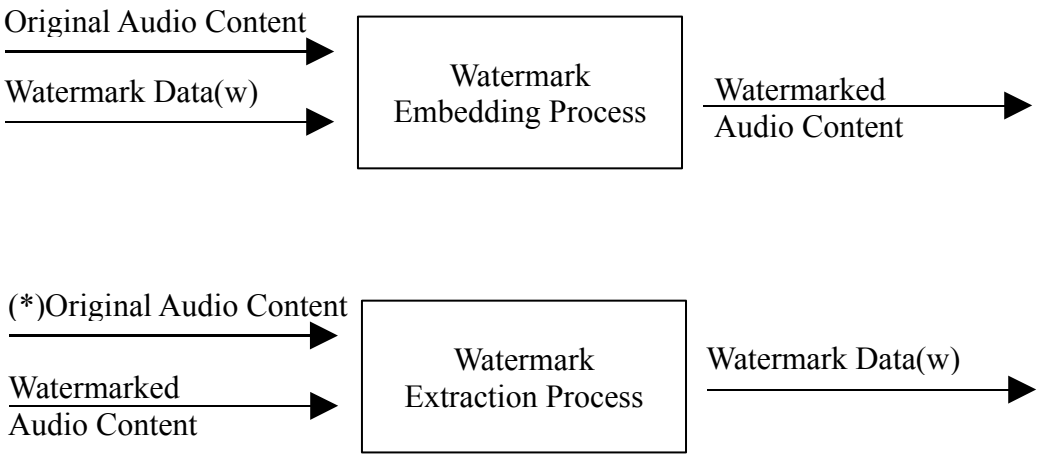


Figure II.2. Watermarking Embedding and Extraction Processes

CHAPTER III

STUDY

In this thesis, we define a security mechanism for real-time VoIP services based on SIP as a signaling to authenticate source origin using digital audio watermarking. Our primary concern is to examine various digital watermarking algorithms that make system feasible in terms of speech quality and delay time. Exploring SIP steganographic techniques in order to create covert channels in SIP messages during signaling phase of VoIP is our secondary concern. Those covert channels are used to exchange source origin indicators among calling parties.

In this thesis, several digital audio watermarking techniques are implemented and compared in terms of robustness, evaluation times, complexity and capacity to demonstrate applicability and feasibility of overall security system described in section I.2.1. This system is called as "WM-Enabled VoIP System" in this thesis.

III.1. WM-ENABLED VoIP SYSTEM

Defined security solution for real-time VoIP Systems using digital audio watermarking is defined in Section I.2.1, which is also called in this thesis as WM-Enabled VoIP System. In this section, how this system covers security issues of VoIP will be discussed.

Security concerns of VoIP are mainly classified into three categories: confidentiality, integrity and availability, see section II.2.1. WM-Enabled VoIP System deals with confidentiality and integrity.

Audio watermarking covert channel in RTP audio stream can be used as a lightweight security solution for confidentiality of audio content. However, it makes trade off between the quality and security of the VoIP call.

Steganographic techniques that extract available covert channels in SIP protocol are used as supplementary security mechanism described in this paper to provide integrity of content. The identity of the recipient is shared over covert channels in SIP protocol.

Figure III.1 shows detailed schema of signaling phase of WM-Enabled VoIP system. ID-key of the caller is embedded into created INVITE message usign SIP/SDP steganographic techniques. INVITE message with hidden ID-key is carried to the callee over SIP protocol. The callee extracts ID-key of the caller from received INVITE message. This is the handshake stage of the system, in which the callee will authenticate the caller with this ID-key during conversation phase of VoIP.

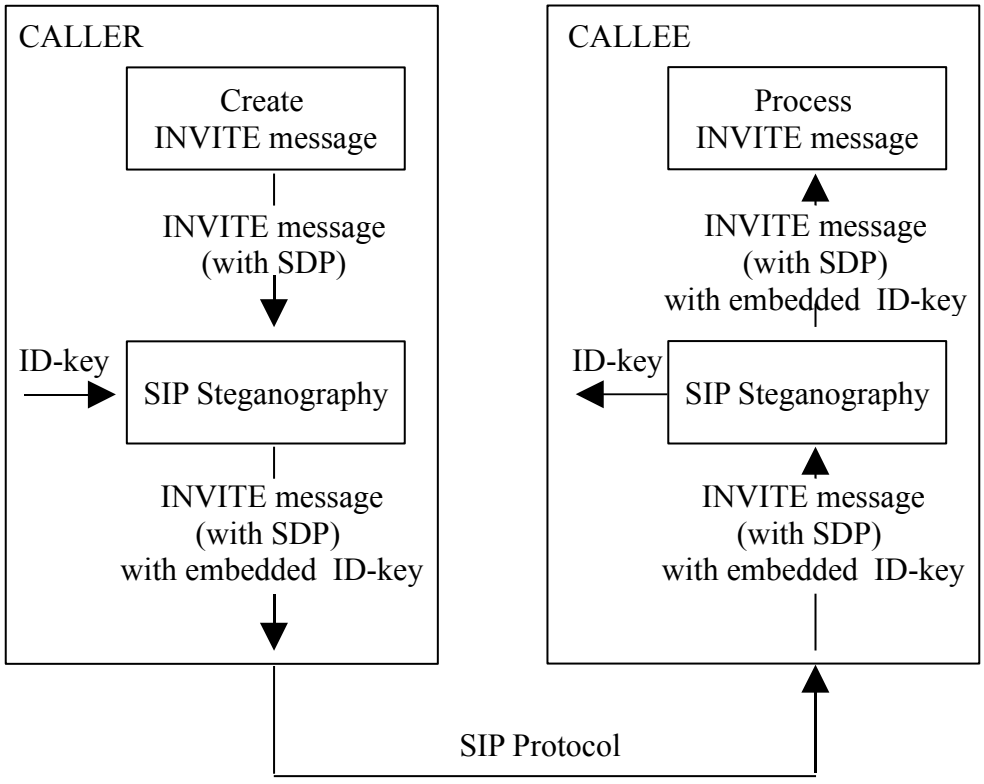


Figure III.1. Signaling-Phase Modifications

Figure III.2 shows detailed schema of conversation phase of WM-Enabled VoIP system. In conversation phase, ID-key of the caller is embedded into audio content using digital audio watermarking technologies. Encoded and watermarked audio content is transferred to the callee over RTP Protocol.

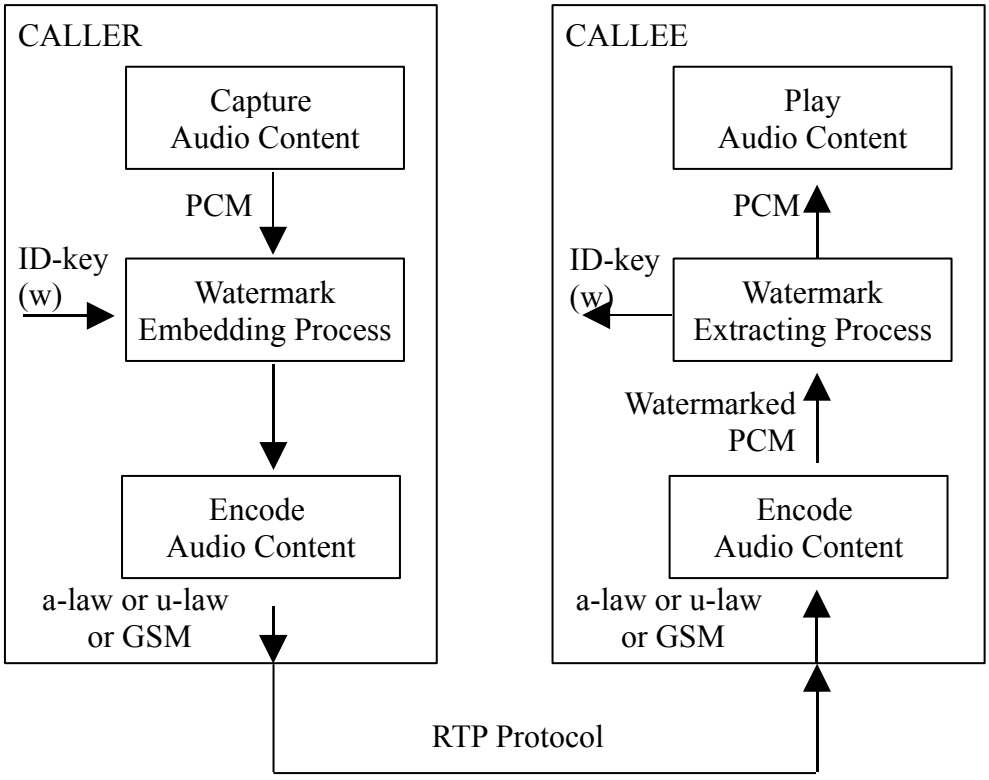


Figure III.2. Conversation-Phase Modifications

III.2. IMPLEMENTED STEGANOGRAPHIC TECHNIQUES FOR SIP

In WM-Enabled VoIP System, each caller has own ID-key for authentication. In Signaling Phase, the caller sends its ID-key to the callee. In this thesis, existing coverts channels in INVITE message are analyzed in order to show how much information may be transferred as ID-key. Covert channel capacity determines maximum length of an ID-key could be.

III.2.1. SIP/SDP Protocols Steganography

"Call-ID", "Contact" and "CSeq" fields are used as a low-bandwidth, one directional covert channel in this solution. "Call-ID"(which uniquely identifies a call), "Contact"(have no direct impact on the communication itself) and "CSeq"(initial sequence number that serves as a way to identify and order transactions) fields are generated randomly by the caller[3]

III.2.2. SIP Parameters, Tokens and Fields Steganography

Some SIP tags in INVITE message can be used as covert channel like “branch” tag and “t” tag those are forms transaction identifier and SIP dialog identifier correspondingly[3]

III.2.3. SDP Protocols Steganography

The following SDP fields are ignored by SIP : v (version), o (owner/creator), s (session name), t (time session is active) and k (potential encryption key if the secure communication is used)[3]

III.2.4. Case Insensitivity Steganography

According to SIP (RFC 3261)[7], SIP header fields are always case-insensitive, which creates covert channel. e.g. although “FROM” header field and “from” header field are the same, different means could be loaded correspondingly 1(one) and 0(zero)[3]

III.3. IMPLEMENTED AUDIO WATERMARKING TECHNIQUES

Audio watermarking techniques implemented in this thesis are:

- Least significant bit technique (LSB)
- DC-level shifting technique(DC-SHIFT)
- Direct sequence spread spectrum technique (DSSS)
- Frequency hopping spread spectrum technique (FHSS)

LSB is most popular and one of the simplest algorithm to implement. DC-level shifting is an algorithm that uses Discrete Fourier Transform. DSSS and FHSS are spread spectrum modulation techniques[17].

III.3.1. Least Significant Bit(LSB) Coding Technique

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[1].

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

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

III.3.2. DC-Level Shifting(DC-SHIFT) Technique

The DCSHIFT is proposed by Uludag et al [14] which involves shifting the DC level for the input audio signal to negative and positive level according to the binary watermark sequence. Watermark information data is embedded in lower frequency components of the audio signal. Lower frequency components of the audio signal are below the perceptual threshold of the human auditory system [11].

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

$$\text{FramePower} = (1/N)\sum(x[n])^2 \quad (n=1..N) \qquad \textbf{Equation III.1}$$

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

$$\text{level0} = -\text{DCBiasMultiplier}*\text{FramePower} \qquad \textbf{Equation III.2}$$

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

level1 =+ DCBiasMultiplier*FramePower **Equation III.3**

Finally, The Inverse Discrete Fourier Transform (IDFT) is computed to get 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 embedding process. Signs of the frame means gives the extracted the binary watermark sequence.

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

III.3.3. Direct Sequence Spread Spectrum (DSSS) Technique

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

The DSSS is an algorithm 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 having a flat frequency response over the frequency range, e.g. white noise. As a consequence, the spectrum of the watermark signal is spread over the available band.

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

Main advantage of DSSS is its resistance to intended or unintended jamming. Capacity of DSSS is much higher than FHSS[15].

III.3.4. Frequency Hopping Spread Spectrum (FHSS) Technique

The Frequency Hopping Spread Spectrum (FHSS) is a spread spectrum modulation techniques. Nowadays, FHSS has a common usage area in Military to secure radio signals.

FHSS rely on the imperfections of the human auditory system (HAS) that is insensitive to small spectral magnitude changes in the frequency domain [2].

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 with the frame size being equal to that used during encoding. The DCT of each block is computed where the sign of the correlation between the DCT coefficients the selected components of each block and the PN sequence.

FHSS is having with little influence from noises, reflections, other radio stations or other environment factors that makes FHSS as a very robust technology [5]

III.4. IMPLEMENTED AUDIO ENCODING TECHNIQUES

III.4.1. PCM

Pulse-Code Modulation(PCM) is the simplest method for converting analog audio signals to its digital representation with fixed precision. PCM provides higher voice quality at a lower cost.

The amplitude of the analogue signal is sampled at regular time intervals. To carry a typical phone call over the PSTN, the analog voice signal is must be sampled at a rate of at least 8 kHz. 8 kHz sampling rate means the signal is sampled every 125 ms.

The bandwidth of the system divided into the quantization levels increase uniformly, which is called as linear quantization. Linear quantization over typical PSTN, per sample defined with at least 12 bits.

Since high amplitude signals have the same degree of resolution (same step size) as lower amplitude signals in linear quantization, quantization process resulted with unneeded quality for high amplitude signals. High amplitude signals need wider zones than lower amplitude signals. Linear quantization is not suitable for the Human Auditory System, since its natural logarithmic process for quantization.

Two international non-linear companding standards are a-law and u-law. u-law is the accepted standard of the digital telecommunication systems of the U.S. and Japan, while a-law is the European accepted standard.

III.4.2. U-LAW

U-law (μ-law or μu-law) companding is a form of logarithmic data compression for audio data, which represents 13-bits number as an 8-bits number. Compressed 8-bits composed of segment and quantization. Bits 6-4 of compressed code word called as segment that represents the logarithmic magnitude domain. While bits 3-0 of compressed code word stores the quantization. 8 segments can be defined in u-law companding algorithm. As if 8 segments are defined linearly, high amplitude signals would loose its most significant bits.

Biased Input Values														Compressed Code Word							
														Segment			Quantization				
Bit	12	11	10	9	8	7	6	5	4	3	2	1	0	Bit	6	5	4	3	2	1	0
	0	0	0	0	0	0	0	1	Q_3	Q_2	Q_1	Q_0	x		0	0	0	Q_3	Q_2	Q_1	Q_0
	0	0	0	0	0	0	1	Q_3	Q_2	Q_1	Q_0	x	x		0	0	1	Q_3	Q_2	Q_1	Q_0
	0	0	0	0	0	1	Q_3	Q_2	Q_1	Q_0	x	x	x		0	1	0	Q_3	Q_2	Q_1	Q_0
	0	0	0	0	1	Q_3	Q_2	Q_1	Q_0	x	x	x	x		0	1	1	Q_3	Q_2	Q_1	Q_0
	0	0	0	1	Q_3	Q_2	Q_1	Q_0	x	x	x	x	x		1	0	0	Q_3	Q_2	Q_1	Q_0
	0	0	1	Q_3	Q_2	Q_1	Q_0	x	x	x	x	x	x		1	0	1	Q_3	Q_2	Q_1	Q_0
	0	1	Q_3	Q_2	Q_1	Q_0	x	x	x	x	x	x	x		1	1	0	Q_3	Q_2	Q_1	Q_0
	1	Q_3	Q_2	Q_1	Q_0	x	x	x	x	x	x	x	x		1	1	1	Q_3	Q_2	Q_1	Q_0

Figure III.3. Sample Segmentation for U-Law Encoding

III.4.3. A-LAW

A-law is the CCITT recommended logarithmic(non-linear) companding algorithm that is slightly different than u-law. In the European, a-law is widely used in the digital telecommunication systems. A-law compresses 12-bits audio signal into 8-bits compressed code word.

As in Figure III.3, 5-bit length data quantized into one segment, on the other hand as in Figure III.4, 5-bit length data quantized into two segments. Results in, a-law algorithm has certain edges over the u-law algorithm. The u-law algorithm provides a slightly larger dynamic range than the a-law at the cost of worse proportional distortion for small signals.

Input Values												Compressed Code Word						
												Chord			Step			
bit:	11	10	9	8	7	6	5	4	3	2	1	0	bit:	6	5	4	3	2	1	0
	0	0	0	0	0	0	0	a	b	c	d	x		0	0	0	a	b	c	d
	0	0	0	0	0	0	1	a	b	c	d	x		0	0	1	a	b	c	d
	0	0	0	0	0	1	a	b	c	d	x	x		0	1	0	a	b	c	d
	0	0	0	0	1	a	b	c	d	x	x	x		0	1	1	a	b	c	d
	0	0	0	1	a	b	c	d	x	x	x	x		1	0	0	a	b	c	d
	0	0	1	a	b	c	d	x	x	x	x	x		1	0	1	a	b	c	d
	0	1	a	b	c	d	x	x	x	x	x	x		1	1	0	a	b	c	d
	1	a	b	c	d	x	x	x	x	x	x	x		1	1	1	a	b	c	d

Figure III.4. Sample Segmentation for A-Law Encoding

III.4.4. GSM 6.10

The GSM (Global System for Mobile communication) is the most popular standard for mobile telephony systems in the world, widely used in Europe.

The GSM uses two codecs according to the types of data channel they were allocated, called Half Rate(6.5 kbits/s) and Full Rate(13 kbits/s). Systems of those codecs systems base upon linear predictive coding(LPC), which is one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters. Audio signal is split up into frames, then LPC coefficients are found for each frame. Namely, spectral envelope information is transmitted over the GSM network.

The GSM is very sensitive to errors, since the filter coefficients are transmitted directly. In other words, a very small error can distort audio signal, a small error might make the prediction filter unstable.

III.5. AUDIO QUALITY EVALUATION TECHNIQUES

III.5.1. HAS(Human Auditory System)

The human auditory system(HAS) is made up of the group of structures used in the process of hearing sound from the outer ear to the brain's auditory cortex.

The speech bandwidth for most adults is approximately 10 kHz. Typical hearing bandwidth for most adults is 15 kHz.

The HAS that is insensitive to small spectral magnitude changes in the frequency domain. Lower frequency components of the audio signal are below the perceptual threshold of the human auditory system.

In general, speech signals are composed of relatively fewer voiced phonemes than unvoiced phonemes. The HAS is a logarithmic process in which high amplitude sound(least likely to occur) does not require the same resolution as low amplitude sound(most likely to occur).

The human audio reflex time is about ~200 msec, which means that audio transmission delay must be limited to ~200 msec

III.5.2. SNR(Signal-to-Noise Ratio)

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 means the less obtrusive the background noise. SNR is often expressed using the logarithmic decibel scale in speech and audio sciences to quantify audio signal quality, called SNR in dB.

The SNR in dB is defined as $20*\log_{10}(A_{signal}/A_{noise})$, in which A is root mean square amplitude.

CHAPTER IV

RESULTS and DISCUSSIONS

IV.1. SIMULATION ENVIRONMENT

IV.1.1. Sample Clips Used in Experiments

Following English-language phrases suggested by ITU-T recommendation P.800 are used in experiments:

clip-1 : “I want a minute with the inspector”, ~1.9 second.

clip-2 : “Did he need any money?”, ~ 1.14 second

clip-3 : “You will have to be very quiet.”, ~ 1.87 second

clip-4 : “There was nothing to be seen.”, ~ 1.48 second

clip-5 : “They worshiped wooden idols”, ~ 1.71 second

For each clips, following two formats are used:

format-1 : PCM signed 16-bit, 16000Hz, 256kbps, mono

format-2 : PCM unsigned 8-bit, 8000Hz, 64kbps, mono

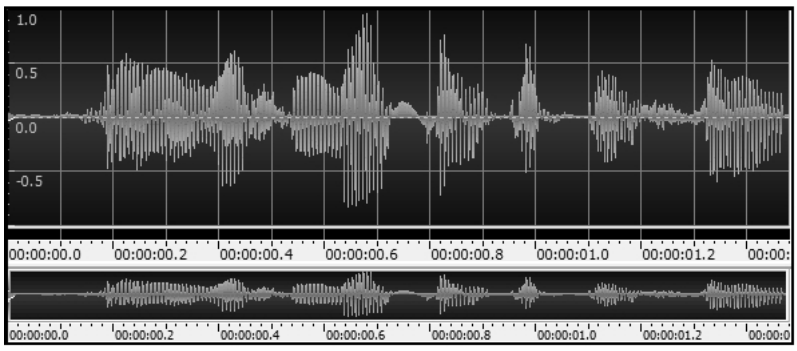


Figure IV.1. Clip-1: “I want a minute with the inspector”

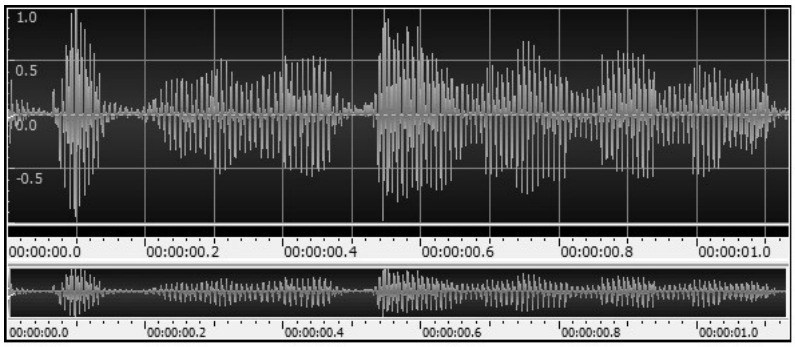


Figure IV.2. Clip-2: “Did he need any money?”

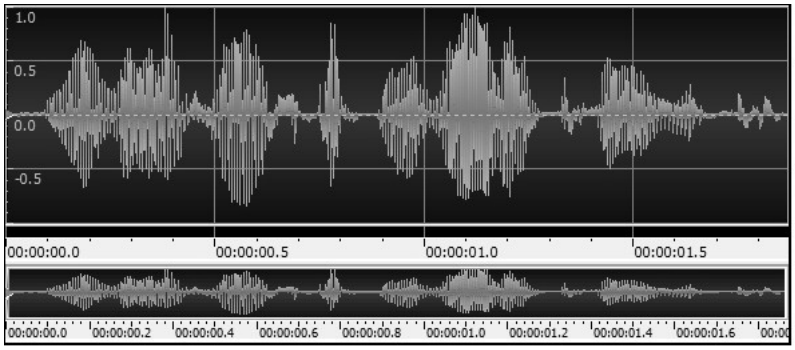


Figure IV.3. Clip-3: “You will have to be very quiet.”

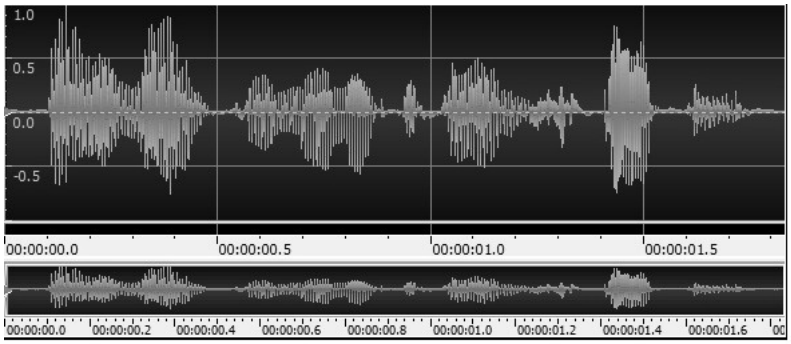


Figure IV.4. Clip-4: “There was nothing to be seen.”

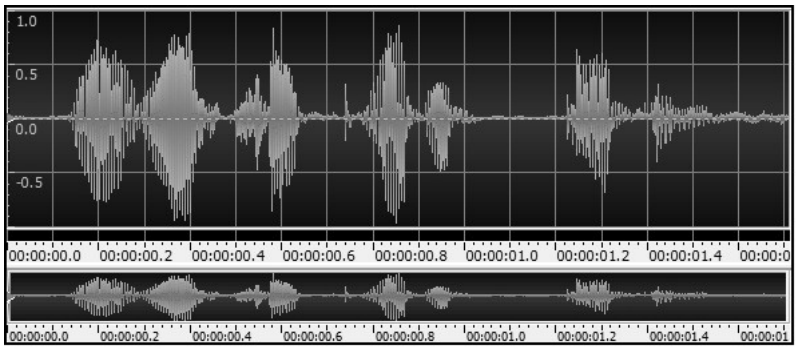


Figure IV.5. Clip-5: “They worshiped wooden idols.”

IV.1.2. Transferred Watermark Data Used in Experiments

Following watermarks are embedded/extracted into/from clips in experiment, correspondingly having 4 bits, 8 bits, 16 bits and 32 bits:

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, 0, 1, 1, 0]

IV.1.3. Audio Compression Standards of Conversation Phase Used in Experiments.

Following audio compression standards are used in conversation phase of VoIP communications:

encoding-1 : G.711.1 A-LAW 16000 Hz, 128kbps, mono

encoding-2 : G.711.1 A-LAW 8000 Hz, 64kbps, mono

encoding-3 : G.711.1 μ-LAW 8000 Hz, 64kbps, mono

encoding-4 : GSM 6.10 8000 Hz, 13kbps, mono

IV.2. SIMULATION RESULTS

Four experiments are evaluated in order to compare watermark algorithms applied in VoIP in terms of audio quality, evaluation times, capacity and robustness. As explained in Section II.2, mostly used transport protocol in conversation phase is Real Time Protocol RTP, which provides end-to-end network transport functions suitable for applications transmitting real-time audio. Audio is encoded/decoded while transferring with RTP protocol. The encoding algorithms a-law, μ-law and GSM were used in the experiments.

Following Table IV.1. demonstrates attributes of evaluated experiments. The first column labeled as “Raw audio format before transfer” shows format of captured raw audio to be sent. The second column labeled as “RTP audio format while transfer” shows RTP encoding format. Finally, the last column labeled as “Raw audio format after transfer” shows delivery raw audio content format.

Table IV.1. Audio Formats Used in Conversation Phases of VoIP in Evaluated Experiments.

Experiments / Audio Formats	RAW AUDIO FORMAT BEFORE TRANSFER	RTP AUDIO FORMAT WHILE TRANSFER	RAW AUDIO FORMAT AFTER TRANSFER
Experiment – 1	PCM signed 16-bit	G.711.1.1 a-law 16000 Hz	PCM signed 16-bit
Experiment – 2	PCM unsigned 8-bit	G.711.1.1 a-law 8000 Hz	PCM unsigned 8-bit
Experiment – 3	PCM unsigned 8-bit	G.711.1.1 u-law 8000 Hz	PCM unsigned 8-bit
Experiment – 4	PCM unsigned 8-bit	GSM 6.10 8000 Hz	PCM unsigned 8-bit

In Experiment-1, all five audio clips(Clip-1, Clip-2, Clip-3, Clip-4, Clip-5) captured from sender with format format-1(PCM signed 16-bit, 16000Hz, 256kbps, mono) and watermarked with all four watermark data (wm-1, wm-2, wm-3, wm-4) using all four watermark algorithms (LSB, DC-SHIFT, FHSS, DSSS). Those audio clips were transferred with RTP in conversation phase of VoIP, encoded with encoding-1(G.711.1 A-LAW 16000 Hz, 128kbps, mono). Then, recipient decoded received audio clips to format-1(PCM signed 16-bit, 16000Hz, 256kbps, mono). Finally, watermark data extracted from those audio clips.

In Experiment-2, all five audio clips(Clip-1, Clip-2, Clip-3, Clip-4, Clip-5) captured from sender with format format-2(PCM unsigned 8-bit, 8000Hz, 64kbps, mono) and watermarked with all four watermark data (WM-1, WM-2, WM-3, WM-4) using all four watermark algorithms (LSB, DC-SHIFT, FHSS, DSSS). Those audio clips were transferred with RTP in conversation phase of VoIP, encoded with encoding-2(G.711.1 A-LAW 8000 Hz, 64kbps, mono). Then, recipient decoded received audio clips to format-2(PCM unsigned 8-bit, 8000Hz, 64kbps, mono). Finally, watermark data extracted from those audio clips.

In Experiment-3, all five audio clips(Clip-1, Clip-2, Clip-3, Clip-4, Clip-5) captured from sender with format format-2(PCM unsigned 8-bit, 8000Hz, 64kbps, mono) and watermarked with all four watermark data (WM-1, WM-2, WM-3, WM-4) using all four watermark algorithms (LSB, DC-SHIFT, FHSS, DSSS). Those audio clips were transferred with RTP in conversation phase of VoIP, encoded with encoding-3(G.711.1 μ-LAW 8000 Hz, 64kbps, mono). Then, recipient decoded received audio clips to format-2(PCM unsigned 8-bit,

8000Hz, 64kbps, mono). Finally, watermark data extracted from those audio clips.

In Experiment-4, all five audio clips(Clip-1, Clip-2, Clip-3, Clip-4, Clip-5) captured from sender with format format-2(PCM unsigned 8-bit, 8000Hz, 64kbps, mono) and watermarked with all four watermark data (WM-1, WM-2, WM-3, WM-4) using all four watermark algorithms (LSB, DC-SHIFT, FHSS, DSSS). Those audio clips were transferred with RTP in conversation phase of VoIP, encoded with encoding-4(GSM 6.10 8000 Hz, 13kbps, mono). Then, recipient decoded received audio clips to format-2(PCM unsigned 8-bit, 8000Hz, 64kbps, mono). Finally, watermark data extracted from those audio clips.

IV.2.1. Comparison of Audio Qualities After Watermark Embedding Process

In Experiment 1, analog audio content sampled and converted into digital representation in format-1, while format of raw audio content is format-2 in Experiments 2-3-4. Comparison of SNR in dB values after watermark embedding process using all four algorithms in Experiment 1 are demonstrated in Figure IV.6 Comparison of SNR in dB values after embedding process of all four algorithms in Experiment 2-3-4 are demonstrated in Figure IV.7.

As Shown in Figure IV.6, LSB offers better SNR in dB performance than others. SNR in dB performances of LSB, DSSS and FHSS are not affected by length of embedded watermark. However, DC-SHIFT SNR in dB performance dramatically decreased with WM length.

As Shown in Figure IV.7, LSB offers better SNR in dB performance than others also. SNR in dB performances of LSB, DSSS and FHSS are not affected by length of embedded watermark. However, DC-SHIFT SNR in dB performance dramatically decreased with WM length.

The main difference between Figure IV.6. and Figure IV.7 is that SNR in dB performance of DC-SHIFT with 4 bit Watermark data is much higher in format-1 than format-2. In overall sight, SNR in dB performances are slightly higher in format-1 then format-2.

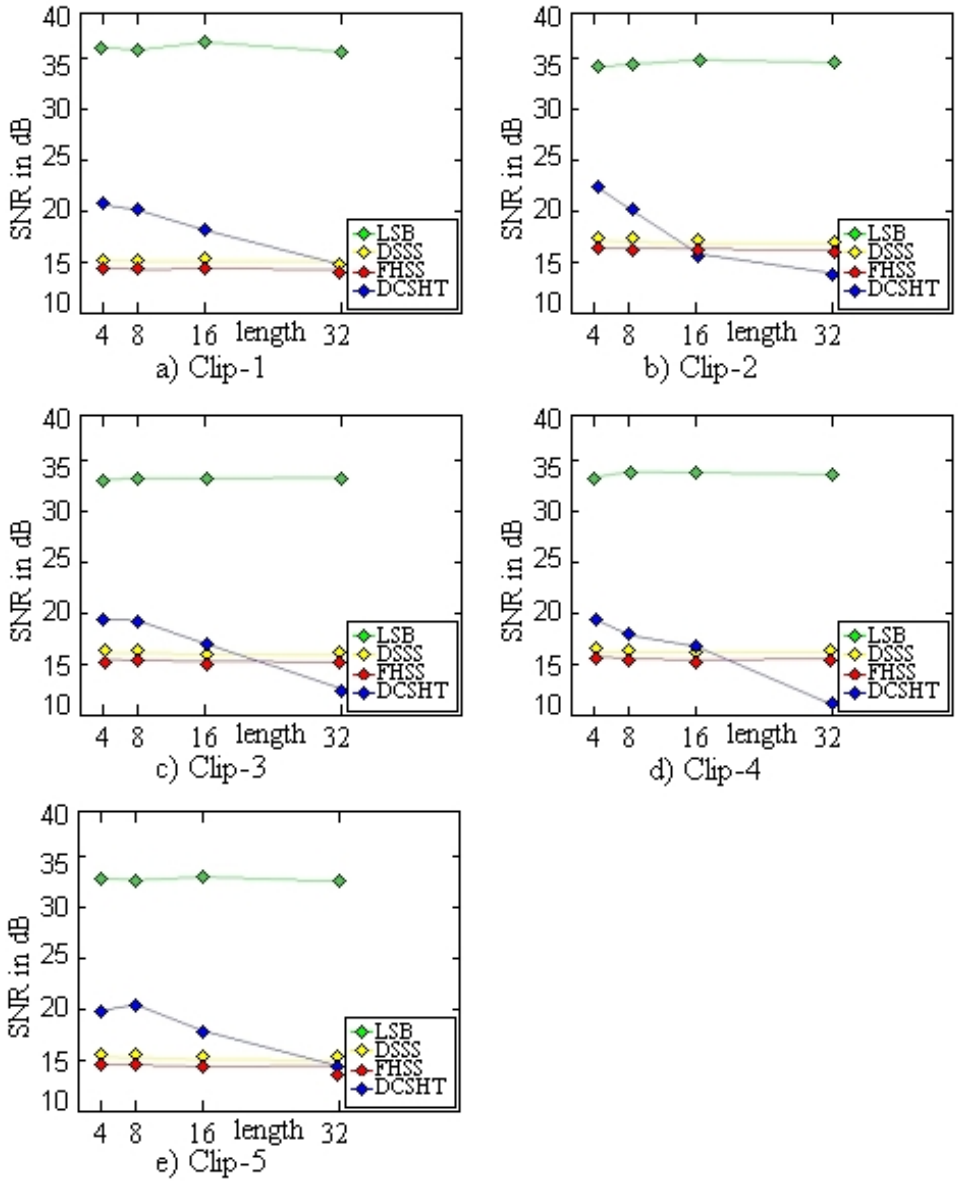


Figure IV.6. SNR in dB Values for Each Clips in Format-1 Before Encoding.

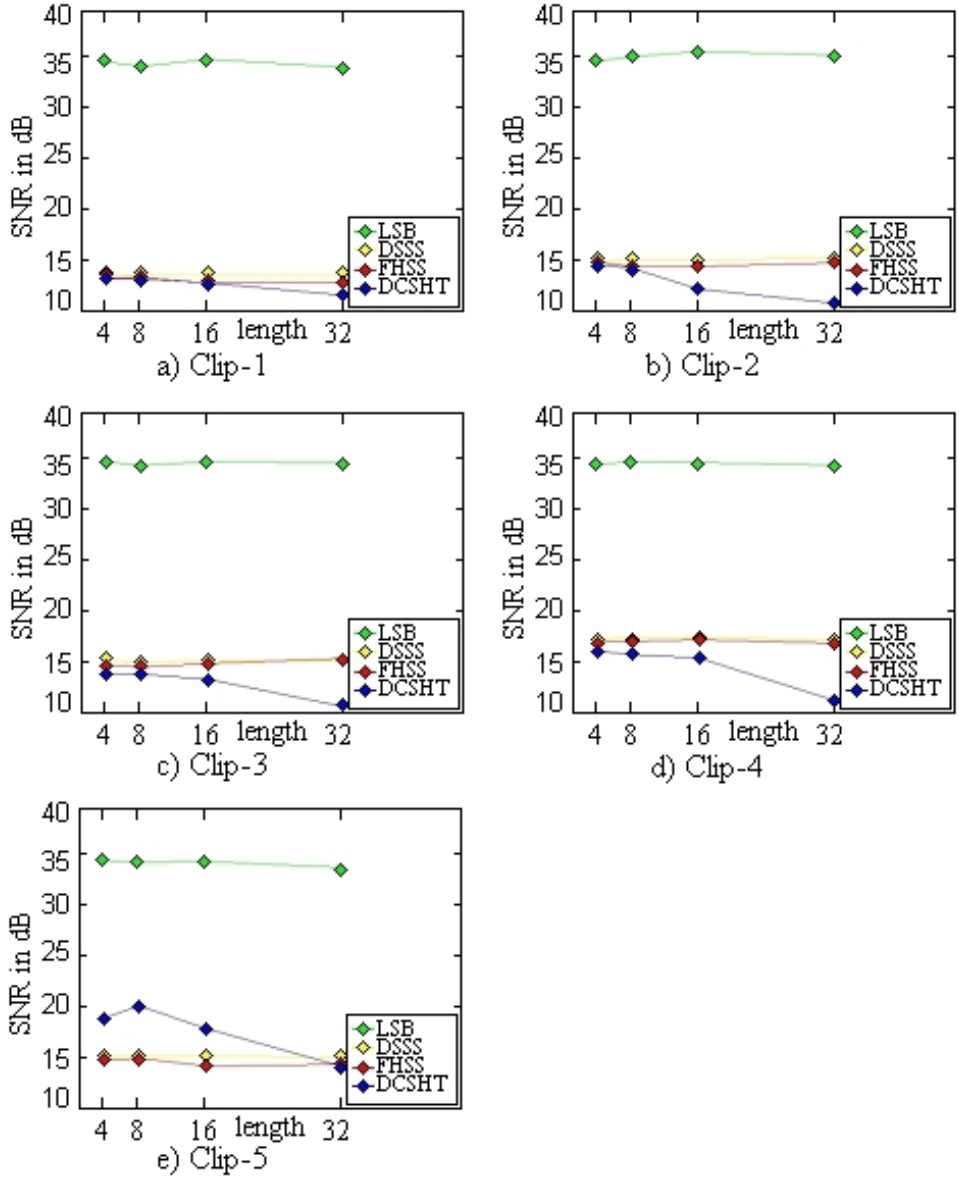


Figure IV.7. SNR in dB Values for Each Clips in Format-2 Before Encoding.

IV.2.2. Comparison of Audio Qualities After Encoding Process over Watermarked RTP Audio Content

Experiment 1, experiment 2, experiment 3 and experiment 4 uses encoding-1, encoding-2, encoding-3, encoding-4 correspondingly in RTP.

Comparison of SNR in dB values after encoding process of all four experiments are demonstrated in Figure IV.8., Figure IV.9, Figure IV.10 and Figure IV.11.

As Shown in Figure IV.8, LSB offers better SNR in dB performance than others. Order of SNR in dB performances are not affected by encoding process. However, SNR in dB performances of DC-SHIFT algorithm dramatically decreased with length of embedded WM data. SNR in dB performances are decreased by ~%44.7, ~%0.2, ~%0.1, ~%8.6 correspondingly LSB, DSSS, FHSS and DC-SHIFT. Decrease ratios show that audio quality of audio contents watermarked by LSB algorithm is much more negatively affected with encoding-1(G.711.1 A-LAW 16000 Hz, 128kbps).

As shown in Figure IV.9, Figure IV.10, Figure IV.11, LSB offers better SNR in dB performance than others also. SNR in dB performances decrease rations in Experiment 2 is ~%49.9, ~%1.2, ~%6.8, ~%0.1 correspondingly LSB, DSSS, FHSS and DC-SHIFT. SNR in dB performances decrease rations in Experiment 3 is ~%52.2, ~%3.9, ~%8.4, ~%0.1 correspondingly LSB, DSSS, FHSS and DC-SHIFT. SNR in dB performances decrease rations in Experiment 4 is ~%75.66, ~%49.2, ~%47.9, ~%47.1 correspondingly LSB, DSSS, FHSS and DC-SHIFT.

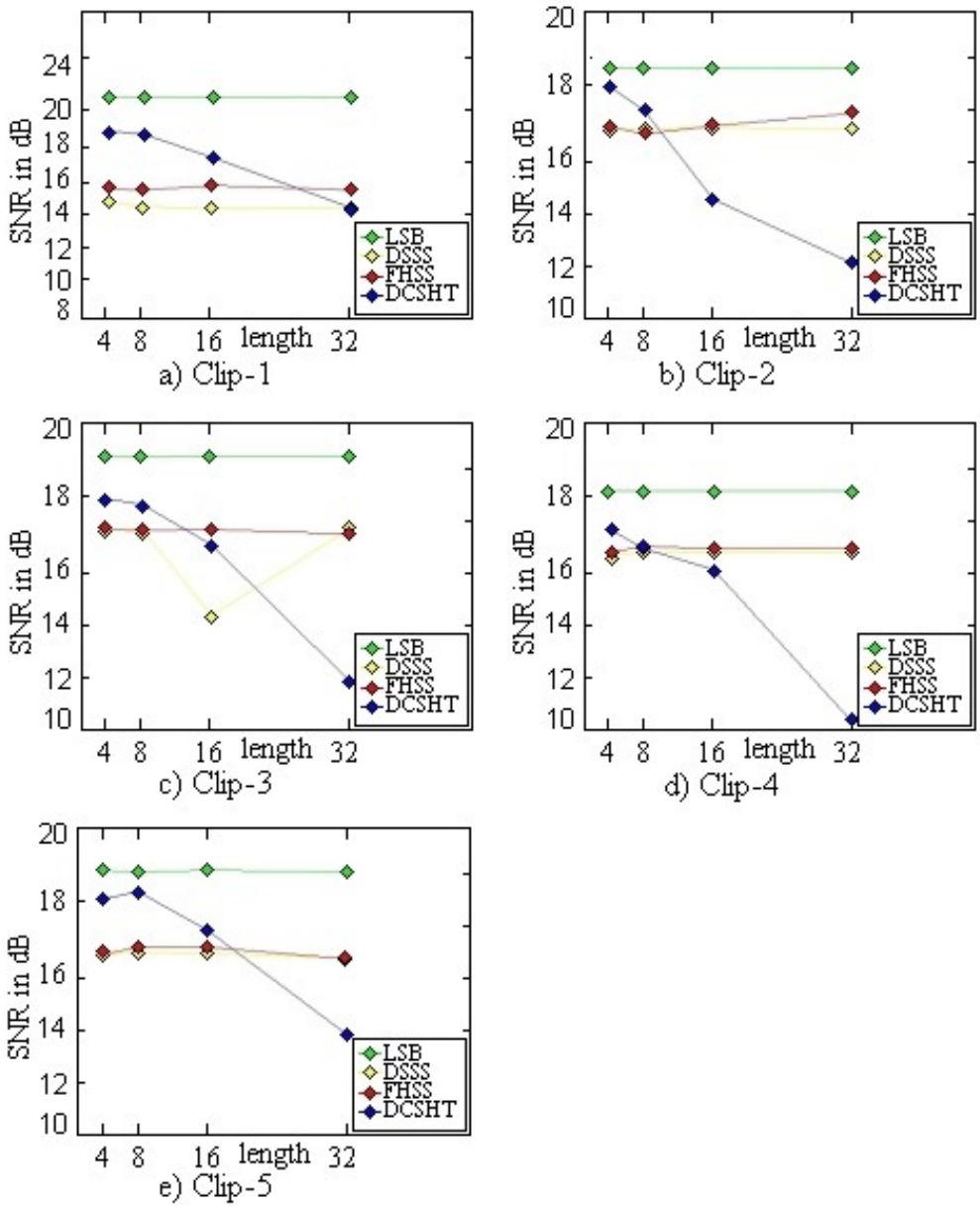


Figure IV.8. SNR in dB Values for Each Clips in Format-1 After Encoding using Encoding-1 in Experiment-1

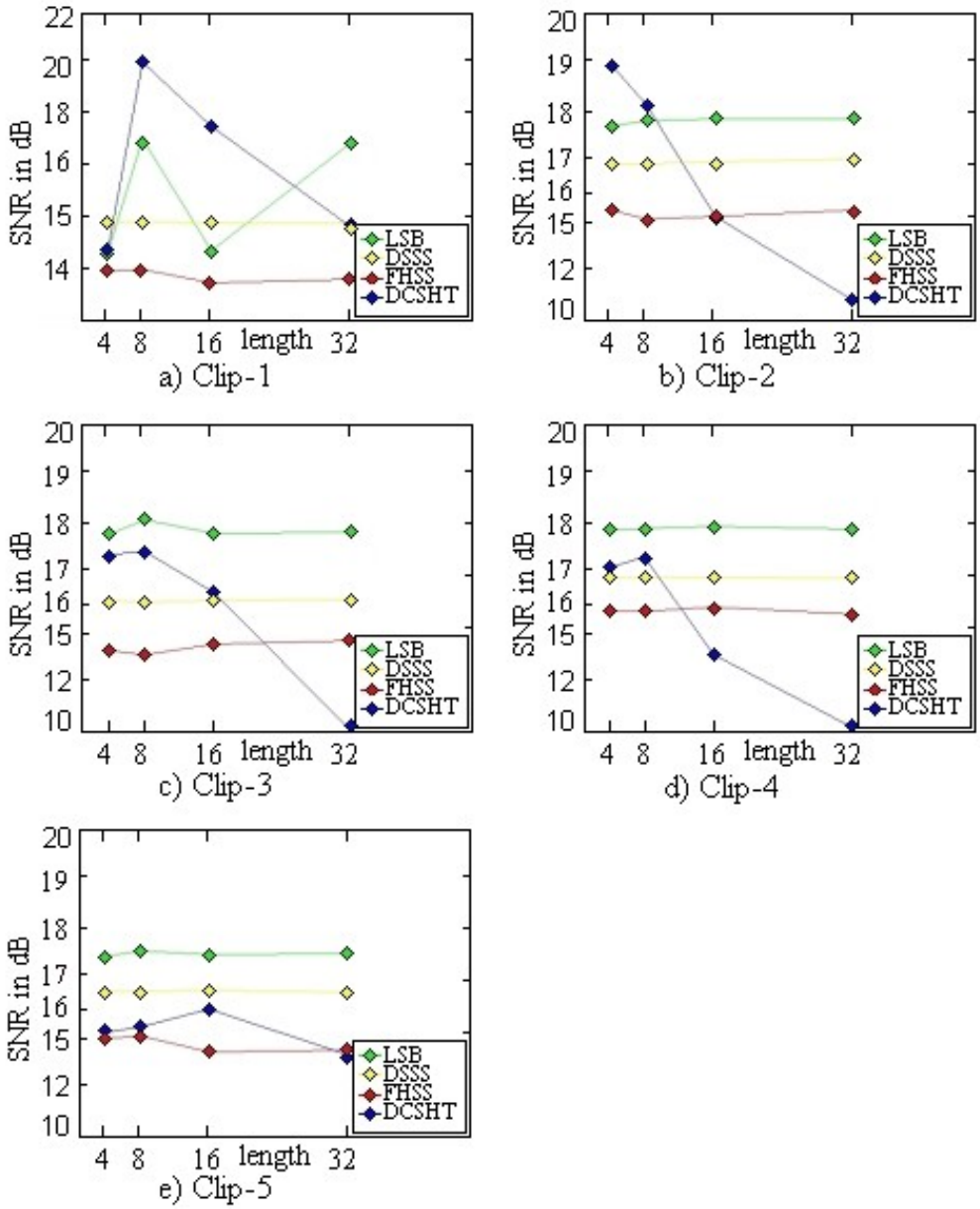


Figure IV.9. SNR in dB Values for Each Clips in Format-2 After Encoding using Encoding-2 in Experiment-2.

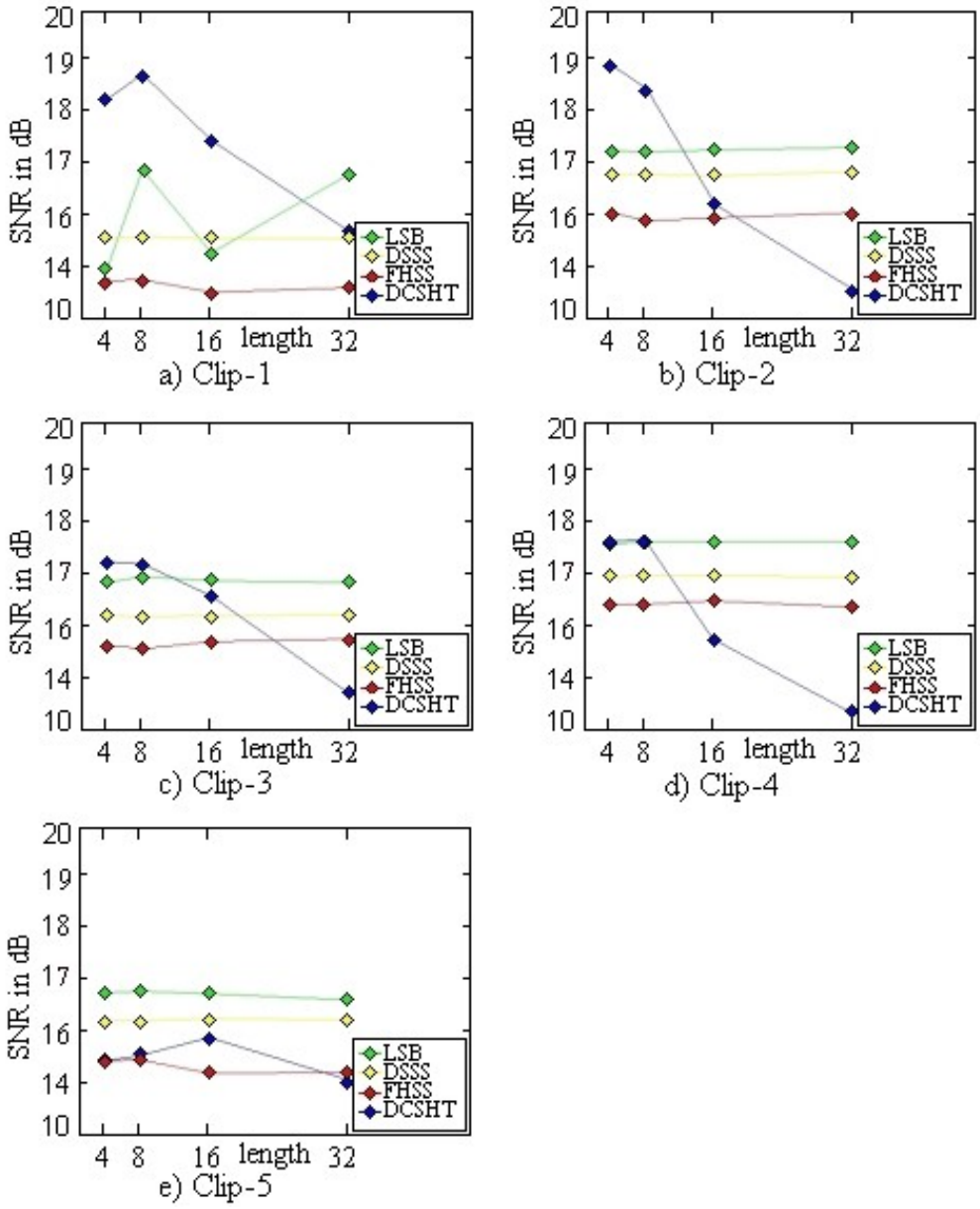
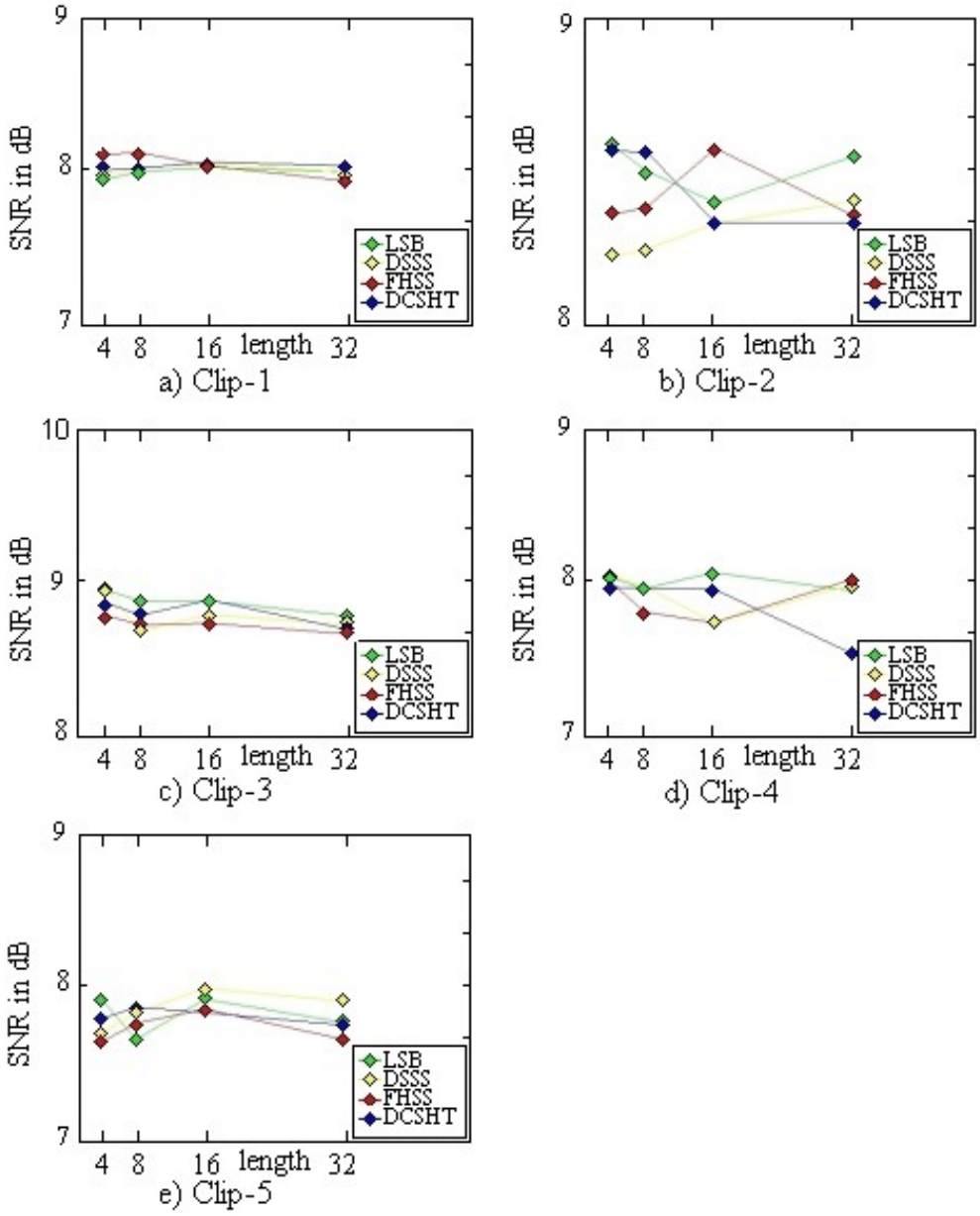


Figure IV.10. SNR in dB Values for Each Clips in Format-2 After Encoding using Encoding-3 in Experiment-3



.**Figure IV.11.** SNR in dB Values for Each Clips in Format-2 After Encoding using Encoding-4 in Experiment-4

Comparison of those decrease ratios are in following Figure IV.12. According to Figure IV.12, in experiment 4 that means GSM 6.10 8000 Hz encoding dramatically decreased audio quality for each WM algorithms. Audio quality order for LSB, FHSS and DSSS is same, which is a-law 16000 Hz, a-law 8000 Hz, u-law 8000 Hz and GSM 6.10 8000 Hz ordered from high quality to low quality.

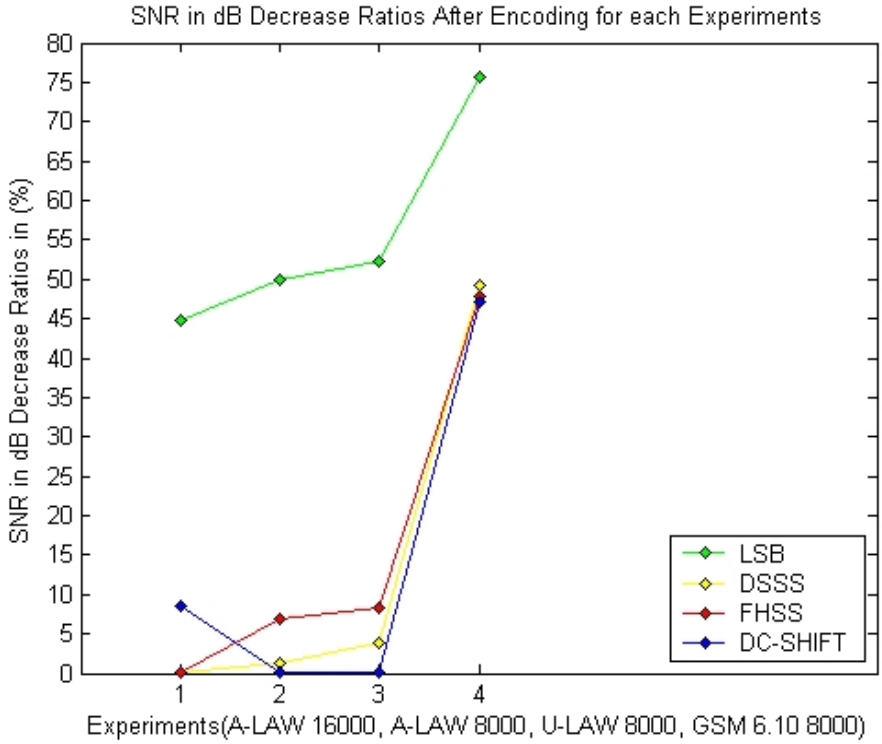


Figure IV.12. SNR in dB Decrease Ratios After Encoding for Each Experiment.

Table IV.2 is the tabular view of graph shown in Figure IV.12. This table shows that audio contents those are watermarked with LSB algorithm or audio contents those are encoded with GSM encoding have a terrible effect to the perceived speech quality.

Table IV.2. SNR in dB Decrease Ratios After Encoding for Each Experiment.

Decrease Ratios After Encoding	LSB	DSSS	FHSS	DC-SHIFT
Experiment-1	~%44.7	~%0.2	~%0.1	~%8.6
Experiment-2	~%49.9	~%1.2	~%6.8	~%0.1
Experiment-3	~%52.2	~%3.9	~%8.4	~%0.1
Experiment-4	~%75.66	~%49.2	~%47.9	~%47.1

IV.2.3. Comparison of Embed/Extract Times of Watermarking Algorithms

Following Table IV.3. shows embed and extract time durations of Watermark algorithms for each experiments. The results are marked with red squares, which total duration is above human tolerance of audio delay limit(~200ms) in RTI(real-time intolerant applications).

As shown in Table IV.3, environment of experiment-1 is not suitable for WM-Enabled VoIP communications. LSB algorithm adds less extra time to communication than other three algorithms in all four experiment environments. DSSS and FHSS has extra delay times in all four experiment environments.

Table IV.3. Embed/Extract Time Durations in Each Experiments

Emded/Extarct Times	Experiment – 1	Experiment – 2	Experiment – 3	Experiment – 4
LSB	0,04 / 0,04	0,03 / 0,02	0,03 / 0,02	0,03 / 0,04
DSSS	0,39 / 0,16	0,08 / 0,02	0,08 / 0,02	0,08 / 0,02
FHSS	0,14 / 0,12	0,09 / 0,04	0,09 / 0,04	0,09 / 0,02
DC-SHIFT	0,12 / 0,09	0,05 / 0,06	0,05 / 0,06	0,05 / 0,13

IV.2.4. Comparison of Capacities of Watermarking Algorithms

WM-1, WM-2, WM-3 and WM-4 watermarks in each experiments are extracted both before encoding and after encoding. Before encoding, all four watermarks are extracted successfully; on the other hand some are lost after encoding.

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

As show in Table IV.4, Environment simulated in experiment-4 is useless for this WM-Enabled VoIP communications. LSB algorithm is clip dependent, so it is useless also.

Also, according to Table IV.4, FHSS and DSSS are resistant to encoding in experiment-1, experiment-2 and experiment-3. FHSS and DSSS hold more capacity in a-law encoding environments(experiment-1 and experiment-2) than u-law encoding(experiment-3). Experiment-3 for DSSS algorithm is resulted with a little bit clip dependent result, it makes DSSS algorithm unreliable for u-law encoding.

Table IV.4. Capacities of Watermarking Algorithms

Capacities for LSB	Experiment – 1	Experiment – 2	Experiment – 3	Experiment – 4
Clip-1(~1.9sn)	8	4	4	0
Clip-2(~1.14 sn)	0	4	4	0
Clip-3(~1.87 sn)	8	0	0	0
Clip-4(~1.48 sn)	8	4	0	0
Clip-5(~1.71 sn)	16	0	0	0

Capacities for DSSS	Experiment – 1	Experiment – 2	Experiment – 3	Experiment – 4
Clip-1(~1.9sn)	8	4	4	0
Clip-2(~1.14 sn)	4	4	0	0
Clip-3(~1.87 sn)	16	8	8	0
Clip-4(~1.48 sn)	4	8	8	0
Clip-5(~1.71 sn)	8	8	4	0

Capacities for FHSS	Experiment – 1	Experiment – 2	Experiment – 3	Experiment – 4
Clip-1(~1.9sn)	8	4	4	0
Clip-2(~1.14 sn)	8	4	4	0
Clip-3(~1.87 sn)	8	8	8	0
Clip-4(~1.48 sn)	8	4	4	0
Clip-5(~1.71 sn)	8	4	4	0

Capacities for DC-SHFT	Experiment – 1	Experiment – 2	Experiment – 3	Experiment – 4
Clip-1(~1.9sn)	4	0	0	0
Clip-2(~1.14 sn)	4	0	0	0
Clip-3(~1.87 sn)	0	0	0	0
Clip-4(~1.48 sn)	0	0	0	0
Clip-5(~1.71 sn)	4	0	0	0

CHAPTER V

CONCLUDING REMARKS AND RECOMMENDATION

V.1. CONCLUSION

In this thesis we defined WM-Enabled VoIP mechanism which may be utilized for source origin authentication. We also implemented several audio watermarking techniques to demonstrate applicability of such a system. Additionally, implemented audio watermarking techniques were compared in terms of SNR, evaluation times, capacity, complexity and robustness. Finally, experimental results show that LSB is simplest algorithm to implement and offers better SNR (in dB) performance than others but it is fragile against a-law encoding and it is clip-dependent. Although evaluation times of DSSS and FHSS are high, those are more resistant to a-law encoding. FHSS hold much more capacity despite of a-law encoding. SNR in dB performances are much more affected after encoding in experiment-1 than other experiments and evaluation times are higher in experiment-1 also.

As a conclusion, FHSS and DSSS algorithms could be used in WM-Enabled VoIP mechanisms which may be utilized for source origin authentication, in which source origins are represented by 4-bit source origin indicators(ID-keys).

16 different source origin indicator could be defined with 4-bit length watermark data. 4-bit source origin indicator is successfully embedded into ~2sn audio content using DSSS or FHSS algorithms, which means authentication process needs 2sn conversation between calling parties. To define more users in the system, initial conversation time for authentication should be 4sn for 8-bit length source origin indicators. So that 256 different users could be defined in the system.

REFERENCES

[1] [Mazurczyk-2007] Mazurczyk, W.; Kotulski, Z.: "Adaptive VoIP with Audio Watermarking for Improved Call Quality and Security", *Journal of Information Assurance and Security 2*, **(2007)** 226-234.

[2] [Yuan-2005] Yuan, S.: "Digital Watermarking-Based Authentication Techniques For Real-Time Multimedia Communication", Vom Fachbereich Informatik der Technischen Universität, Liaoning, China, **(2005)**.

[3] [Mazurczyk-2008] Mazurczyk, W.; Szczypiorski, K.: "Covert Channels in SIP for VoIP Signalling," Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Telecommunications, Warsaw, Poland. **(2008)**.

[4] [Khanvilkar-2008] Khanvilkar, S.; Bashir F.; Schonfeld, D.; Khokhar A.: "Multimedia Networks and Communication", University of Illinois at Chicago, United States, **(2008)**.

[5] [ref-H323] ITU-T Recommendation H.323: “Packet-based multimedia communications systems”, http://www.itu.int/rec/T-REC-H.323/en/.

[6] [ref-RFC3550] Schulzrinne, H.; Casner, S.; Frederick, R.; Jacobson, V.: “RTP: A Transport Protocol for Real-Time Applications”, RFC 3550, IETF, July **(2003)**, http://www.ietf.org/rfc/rfc3550.txt.

[7] [ref-RFC4961] Wing, D.: “RTCP :Symmetric RTP / RTP Control Protocol”, RFC 4961, IETF, July **(2007)**, http://www.ietf.org/rfc/rfc4961.txt.

[8] [ref-RFC2327] Handley, M.; Jacobson, V.: “SDP: Session Description Protocol”, RFC 2327, IETF, April **(1998)**, http://www.ietf.org/rfc/rfc2327.txt.

[9] [ref-RFC3209] Awduche, D.; Berger, L.; Gan, D.; Li, T.; Srinivasan, V.; Swallow, G.: “RSVP-TE: Extensions to RSVP for LSP Tunnels”, RFC 3209, IETF, December **(2001)**, http://www.ietf.org/rfc/rfc3209.txt.

[10] [ref-RFC3261] Rosenberg, J.; Schulzrinne, H.; Camarillo, G.; Johnston, A.; Peterson, J.; Sparks, R.: “SIP: Session Initiation Protocol”, RFC 3261, IETF, June **(2002)**, http://www.ietf.org/rfc/rfc3261.txt.

[11] [Evans-2005] Donald L,; Phillip J. Bond , Shashi Phoha , Security Considerations for Voice Over IP Systems, NIST Special Publication 800-58. Gaithersburg, MD 20899-8930, January, **(2005)** 41-87.

[12] [Kirovski-2001] Kirovski, D. and Malvar, H: "Robust Spread-Spectrum Audio Watermarking", ICASSP, IEEE **(2001)**.

[13] [Mazurczyk-2006] Mazurczyk, W.; Kotulski, Z.: "New VoIP Traffic Security Scheme with Digital Watermarking", Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Telecommunications, Poland, **(2006)**.

[14] [Uludag-2001] Uludag, U.; Arslan, L.: "Audio watermarking using DC-level shifting", *Project Report*, Bogazici University, **(2001)** http://busim.ee.boun.edu.tr/~speech/publications/audio_watermarking/uu_la_audio_wm2001.pdf.

[15] [Schawartz-2001] Schewartz, M.: "FHSS vs. DSSS," Feb, **(2001)** 1-16.

[16] [Thanuja-2008] Thanuja, T.; Dr. Nagaraj, R.: "Schemes for Evaluating Signal Processing Properties of Audio Watermarking", *IJCSNS International Journal of Computer Science and Network Security*, VOL.8 No.7, July **(2008)**.

[17] [Cvejic-2004] Cvejic, N.: "Algorithms for Audio Watermarking and Steganography," ISBN 951-42-7384-2, University of Oulu, Finland, **(2004)** 19-20